Chemically modified CRISPR-Cas9 enables targeting of individual G-quadruplex and i-motif structures, revealing ligand-dependent transcriptional perturbation.

Sabrina Pia Nuccio[1,2], Enrico Cadoni[1,3], Roxani Nikoloudaki[1], Silvia Galli[1], An-Jie Ler[1], Claudia Sanchez-Cabanillas,[1,4] Thomas E. Maher [1,2], Ella Fan[1], Dilek Guneri[5], Gem Flint[1], Minghui Zhu[1], Ling Sum Liu[1], Christopher R. Fullenkamp[7], Zoë Waller[5], Luca Magnani[4,6], John S. Schneekloth, Jr[7], Marco Di Antonio[1,2,8] *

1 Imperial College London, Department of Chemistry, Molecular Sciences Research Hub, 82 Wood Lane, London, W12 0BZ, UK

2 Imperial College London, Institute of Chemical Biology, Molecular Sciences Research Hub, 82 Wood Lane, London W12 0BZ, UK

3 Department of Organic and Macromolecular Chemistry - Organic and Biomimetic Chemistry Research Group (OBCR), Faculty of Sciences, Campus Sterre, Krijgslaan 281, Building S4, 9000 Gent, Belgium

4 The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, 123 Old Brompton Road, SW7 3RP, London, UK

5 UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX, UK

6 Department of Oncology and Haemato-Oncology, Università degli Studi di Milano, Milan, Italy

7 Chemical Biology Laboratory, National Cancer Institute, Frederick, MD 21702, USA

8 The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

*To whom correspondence should be addressed. Email: m.di-antonio@imperial.ac.uk (MDA); Tel: +44 (0)20 7594 5866

Abstract

The development of selective ligands to target DNA G-quadruplexes (G4s) and i-motifs (iMs) has revealed their relevance in transcriptional regulation. However, most of these ligands are unable to target individual G4s or iMs in the genome, severely limiting their scope. Herein, we describe a new Approach to Target Exact Nucleic Acid alternative structures (ATENA) that relies on the chemical conjugation of established G4 and iM ligands to a catalytically inactive Cas9 protein (dCas9), enabling their individual targeting in living cells. ATENA demonstrated that the selective targeting of the G4 present in the oncogene *c-MYC* leads to the suppression of transcripts regulated exclusively by one of its promoters (P1). Conversely, targeting the *c-MYC* iMs on the opposite strand leads to the selective increase of P1-driven transcripts. ATENA revealed that G4-mediated transcriptional responses are highly ligand-specific, with different ligands eliciting markedly different effects at the same G4-site. We further demonstrated that the basal expression levels of the gene targeted can be used to predict the transcriptional impact associated with G4-stabilization. Our study provides an innovative platform to investigate G4- and iM-biology with high precision and unveils the therapeutic relevance of individual DNA structures with unprecedented selectivity.

Introduction

G-quadruplex (G4) structures can promptly form within single-stranded DNA sequences rich in guanines by generating stacks of G-quartets held together by Hoogsteen hydrogen-bonding that are further stabilized by coordination of potassium ions (Fig.1a) [1]. The ability of G-rich sequences to form G4s under physiological conditions has been known for decades [2,3]. Yet, the biological relevance of these structures has been heavily disputed until recently. Over the past decade, the development of orthogonal approaches to detect and map G4s has provided robust evidence to support their formation in living cells. These include immuno-fluorescence [4], live-cell imaging [5,6], and genome-wide mapping strategies [7-10].

While G4-formation in cells has been extensively validated, its active contribution in regulating biological processes is yet to be fully demonstrated. Given the high enrichment of G4s at gene promoters measured experimentally *in vitro* and in cells [9,11], it has long been speculated that these structures may play an essential role in regulating gene expression [12]. Indeed, many G4-selective ligands have been developed to date, revealing that targeting G4s within gene promoters is generally associated with transcriptional suppression [13]. Since key oncogenes like *c-MYC, KRAS,* and *BCL-2* bear a G4 in their promoter, applying G4 ligands for cancer intervention has been extensively investigated [14]. However, the human genome displays more than 700,000 experimentally detected G4-structures *in vitro* [11,15] and around ~10,000 G4s detected in chromatin using genomics strategies [7-10,16], making single-G4 targeting by small molecules extremely challenging. Indeed, most ligands recognize G4s by establishing π-π end-stacking with G-tetrads, a promiscuous interaction that hampers these molecules from displaying any significant inter-G4 selectivity[17]. For example, evidence supports that the transcriptional suppression of *c-MYC* observed upon treatment with certain G4-ligands is an indirect consequence of global G4 stabilization rather than a specific response regulated exclusively by the G4 in the *c-MYC* promoter [18].

Additionally, there are experimental discrepancies between what could be the endogenous biological function of G4s and what is instead observed when ligands bind these structures. A series of

independent genomic studies consistently indicated that G4-formation is strongly associated with active gene expression [12,19], contrasting the transcriptional repression mostly observed upon G4-ligand treatment. Similarly, recent studies have leveraged gene-editing techniques to demonstrate how the selective deletion of the G-rich sequence in the *c-MYC* promoter responsible for the G4-structure formation on this promoter (MYC-G4) is associated with loss of transcriptional activity [20], further pointing to an active role of G4s in stimulating transcription. To justify the discrepancies between endogenous G4-function versus the responses observed by G4-targeting with small-molecules, it has been postulated that ligand binding can prevent key transcriptional factors from recognizing G4s, leading to transcriptional repression[21]. However, it remains challenging to discern whether transcriptional suppression at specific genes is caused by protein displacement at a given G4 site or is a response triggered by global G4-stabilization, as reported for *c-MYC*[18]. It is also often supposed that structurally different G4-ligands should elicit similar responses when targeting the same G4s, with the assumption that G4 binding proteins respond identically to a ligand-bound G4 irrespective of the ligand used. Overall, this highlights the urgent need for tools that can provide inter-G4 selectivity to widely used G4 ligands to fully underpin G4-biology and harness the therapeutic potential of these DNA secondary structures.

Similarly, i-motifs (iMs) structures have been recently mapped through the human genome[22], visualized in cells[23], and have been heavily linked with transcriptional regulation[22]. Although several ligands have been shown to bind the i-motif-forming sequences in *c-MYC*[23], the exact biological responses elicited by targeting iMs still remain elusive, owing to their lack of inter-iMs selectivity that confounds phenotypes arising from global versus site-specific iM targeting.

Therefore, it is evident that the absence of inter-G4 or inter-iMs selectivity displayed by most of the available ligands strongly hampers their use as therapeutic agents and their application as reliable tools to investigate the biology of these DNA structures. Hence, the development of ligands displaying preferential binding toward a specific G4 or iM structure (i.e., *c-MYC*) has represented a

108 longstanding quest in this area of research. To this end, Schneekloth and co-workers recently designed
109 a small-molecule ligand (DC-34) that displays preferential binding to the G4 in the *c-MYC*
110 promoter[24]. Treatment of multiple myeloma cell lines with DC-34 revealed suppression of *c-MYC*
111 transcription with minimal perturbation of other G4-bearing oncogenes, such as *KRAS* and *BCL-2*[24].
112 Similarly, our group and others have shown that short peptides and oligonucleotides can also display
113 binding selectivity towards specific G4s[25-28]. Nevertheless, these remain isolated examples that require
114 ligand optimization to achieve binding selectivity against any given DNA structure of interest and are
115 not suitable for high-throughput screening of G4 or i-motif function at scale. Furthermore, the highly
116 diverse chemical and structural nature of ligands targeting G4s and iMs might represent an additional
117 confounding factor. For example, different G4-ligands might elicit different biological responses
118 when binding to the same G4s, leading to experimental observations that are dependent on the type of
119 ligand used, rather than reflecting the endogenous biological function of the targeted G4.
120
121 To overcome these limitations, we have developed ATENA (Approach to Target Exact Nucleic Acid
122 alternative structures), a CRISPR-Cas9-based platform that enables selective localization of different
123 G4 or iM ligands in the proximity of a given DNA structure of interest (Fig.1b). To achieve this, we
124 have exploited HaloTag technologies to selectively install modified G4 or iM binding ligands onto
125 Cas9 protein in living cells[29]. More specifically, we have designed a small library of established G4-
126 ligands (pyridostatin (PDS)[30] and PhenDC3[31]) functionalized with chloroalkane side chains that
127 ensure incorporation into a nuclease-inactive Cas9 protein fused to a HaloTag (dCas9-Halo) [32].
128 Similarly, we decorated one of the iM selective peptides (Pep-RVS) [33], recently developed by the
129 Waller's group, with a chloroalkane side chain to explore selective iM targeting with ATENA. The
130 synthesized analogs of PDS, PhenDC3, and Pep-RVS contain different polyethylene-glycol (PEG)-
131 linkers (n = 0, 2, 4, Fig.1 c,d), with the PEG chain serving to tether the DNA-binding scaffold to the
132 HaloTag-binding chloroalkane. By fine-tuning the PEG linker length, we optimized the spacing
133 between the DNA-binding moiety and the dCas9-Halo fusion to attain ideal G4 engagement with
134 ATENA, using a dedicated FRET-based assay. We leveraged this knowledge to deploy ATENA in
135 cells and achieve selective targeting of either the G4 or the iM present in the promoter of the proto-

136 oncogene *c-MYC*. ATENA-mediated G4-targeting resulted in the reduction of *c-MYC* transcripts
137 generated exclusively by one of the four promoters regulating *c-MYC* expression (P1), irrespective of
138 the ligand used, which is in agreement with recent literature [20]. Additionally, ATENA has shown, for
139 the first time, that selective MYC-G4 targeting does not result in a net reduction of *c-MYC* expression,
140 as increased transcription from the alternative P2 promoter counterbalances the P1-specific
141 downregulation induced by G4 engagement. The P1-dependent downregulation upon G4-targeting
142 was also confirmed by treatment with the MYC-G4 selective ligand DC-34, further validating the
143 ability of ATENA to target individual G4s. Importantly, we have also observed that guiding G4-
144 ligands in the proximity of the *c-MYC* promoter TATA-box can lead to G4-independent transcriptional
145 suppression. This strongly suggests that previous observations obtained with dCas9 decorated with
146 multiple G4-ligands are likely reflecting occupancy of transcriptional regulatory regions rather than
147 genuine G4-binding [34].
148
149 Upon employing ATENA to target selectively the iM present in the promoter of *c-MYC*, using a
150 HaloTag-compatible version of Pep-RVS, we have also observed P1-specific transcriptional
151 perturbation. An increase in P1-expression was associated with iM targeting, opposing the
152 transcriptional inhibition observed by G4 stabilisation, in line with other systems which indicate that
153 iM-binding results in induction of gene expression[35,36] and that iMs and G4 shape opposing effects in
154 cells[37].
155
156 While downregulation of P1-mediated *c-MYC* expression was observed with two distinct ligands,
157 PDS and PhenDC3, ATENA also revealed that other G4s could provide different transcriptional
158 outcomes when targeted by the same compounds. This suggested that the biological response elicited
159 by G4 stabilization might reflect more the structural nature of the G4-ligand complex rather than
160 providing direct insights into the endogenous function of G4s, underscoring the importance of the
161 choice of ligand used in the different reports aimed at unveiling G4-biology. Finally, we showcased
162 the ability of ATENA to infer the biological relevance of cell line-specific G4s, revealing a strong

dependence of the transcriptional perturbation attained upon ligand treatment on the expression level of the gene targeted.

Altogether, our study provides robust evidence supporting ligand and transcriptional-dependent responses to both G4s and iM targeting. ATENA functions as a modular platform to target individual DNA secondary structures in living cells, enabling a precise study of G4- and iM-biology. We anticipate that ATENA will offer unprecedented potential for screening cell and ligand-specific responses to DNA secondary structure targeting in a high-throughput manner, which can be further translated for therapeutic design and development.

189 **Results**

190

191 **Chemical labelling of CRISPR-Cas9 proteins with G4 ligands**

192 Catalytically inactive CRISPR-Cas9 (dCas9) fused to functional proteins has been widely used in
193 biology to attain site-selective perturbation of gene expression [38]. This strategy takes advantage of the
194 selectivity provided by dCas9 bound to a short-guiding RNA (sgRNA) in recognizing a specific
195 genomic site by base-pairing, which can be used to recruit an effector protein (i.e., a transcription
196 factor or an epigenetic enzyme) at the targeted site [39]. A similar strategy has been recently devised to
197 decorate dCas9 with G4-ligands using non-covalent biotin-streptavidin recognition[34]. Irreversible
198 chemical functionalization of dCas9 proteins has been previously achieved using commercially
199 available chloroalkane-modified fluorophores to label a dCas9-Halo fusion protein in living cells [39].
200 Therefore, we hypothesized that generating modified G4-ligands functionalized with chloroalkane
201 moieties could have been exploited to decorate with higher control and irreversible covalent chemistry
202 Cas9 proteins with G4 ligands under physiological conditions. To achieve this, we designed analogues
203 of the widely established G4 ligands PDS based on the previously described PyPDS scaffold [5]. Unlike
204 PDS, PyPDS presents a single primary amine within its structure, which can be selectively
205 functionalized with chloroalkane sidechains by peptide coupling (Fig. 1c) [5]. After successful synthesis
206 of PyPDS following previously established methods[5], we functionalized the primary amine of the
207 molecule with chloroalkane side chains of different lengths, enabling systematic investigation of the
208 ideal distance between the G4-binding scaffold PyPDS and the HaloTag protein to achieve optimal
209 G4-engagement. Specifically, we have used linkers containing different PEG repeats (n = 0, 2, 4) to
210 vary the distance between PyPDS and the chloroalkane (Cl-PDS$_n$, Fig. 1c, SI-1).
211 To avoid limiting the use of ATENA to a single G4-ligand, we also functionalized with chloroalkane
212 moieties another widely characterized G4 ligand called PhenDC3 using the same chemical strategy
213 (Cl-PhenDC3$_n$) [31]. Unlike PDS, PhenDC3 has a cationic side chain and is structurally bulkier,
214 displaying a phenanthroline core wider than the pyridine one present in the PDS scaffold (Fig. 1c, d).
215 While PDS and PhenDC3 have been extensively validated as selective G4-ligands, the structural

differences between these two ligands might lead to distinct biological responses when targeting G4s in cells, which cannot be assessed with current methods. Therefore, we decided to systematically compare these two ligands with ATENA when recruited to a single-G4 site. To achieve this, we have synthesized a previously reported PhenDC3 analogue displaying an exocyclic primary amine [40], which can be functionalized by peptide coupling using the same synthetic strategy described for PyPDS to afford chloroalkane-modified PhenDC3 analogs that are compatible with HaloTag conjugation (Fig. 1d, SI-1).

In vitro validation of covalent conjugation of G4-ligands to dCas9-Halo

With both PyPDS and PhenDC3 analogues in hand, we initially assessed whether functionalization with the chloroalkane side chains could affect the G4-binding properties of these molecules. To test this, we have subjected all the ligands to FRET or Circular Dichroism (CD) melting to evaluate their ability to stabilize G4-structures after chloroalkane functionalization. All analogues tested displayed good G4-stabilization, providing an increase in melting temperature (ΔT_m >10 ºC at 4 μM) against four distinct G4 structures tested (c-MYC, hTelo, BCL-2 and c-KIT2; Extended Data Tables 1, 2, 3 and 4), with ΔT_m values comparable to the unfunctionalized ligands. This suggested that the addition of the chloroalkane side chains had a negligible effect on the G4-stabilization properties of both PhenDC3 and PyPDS.

Having confirmed that both chloroalkane-functionalized PyPDS and PhenDC3 analogues retained good G4-binding recognition properties, we next investigated if these molecules could be covalently engaged to dCas9-Halo and performed a competition assay *in vitro*. To this end, we expressed and purified the dCas9-Halo protein (see methods) and incubated it for 45 minutes with increasing concentrations of Cl-PDS$_n$ and Cl-PhenDC3$_n$. This was followed by incubation with an excess (5 μM) of the commercially available HaloTag®TAMRA (Cl-TAMRA) ligand to label dCas9-Halo with the TAMRA fluorophore. Since labeling of HaloTag is a covalent irreversible process [29], we reasoned that initial exposure of the dCas9-Halo protein to the chloroalkane functionalized G4-ligands would prevent subsequent incorporation of the fluorescent Cl-TAMRA, leading to a dose-dependent

243 reduction of TAMRA incorporation (Fig. 1e). Indeed, all tested analogues induced a robust dose-
244 dependent decrease of the dCas9-Halo TAMRA signal (Fig. 1f and Extended Data Fig. 1a), indicating
245 high efficiency in labeling dCas9-Halo irrespectively of the G4-ligand (i.e., Cl-PDS$_n$ or Cl-PhenDC3$_n$)
246 or the PEG-linker used to connect the chloroalkane to the G4-binding scaffold. To quantify labeling
247 efficiency, we measured fluorescence intensity for each line of the gel and generated dose–response
248 curves to extract the concentration of ligand required to attain 50% labeling of the dCas9-Halo *in vitro*
249 (IV-CP$_{50}$). As depicted in Extended Data Fig. 1b, c, all the analogues displayed very low IV-CP$_{50}$
250 values (≤ 2 μM), demonstrating a strong ability to functionalize dCas9-Halo *in vitro*. This data
251 supports the use of chloroalkane-modified G4 ligands to covalently label dCas9-Halo.
252
253 ***In vitro* optimization of sgRNAs and PEG-linker to attain G4-engagement**
254 We next asked if dCas9-Halo functionalized with G4-ligands could be used to target individual G4s.
255 To achieve this, we investigated the ideal distance between the dCas9 binding site and the targeted G4
256 to achieve optimal engagement of the G4 ligands with the targeted structure by systematically varying
257 both the short guiding RNA (sgRNA) sequences used and the PEG-linker connecting the Halo
258 reactive moiety (chloroalkane) to the G4-binding scaffold tested. To quantify G4-engagement, we
259 designed a dually fluorescent-labelled DNA template containing an established G4-forming sequence
260 (c-KIT2-G4) at its 3' end to monitor G4-stabilization through FRET (Fig. 1g). We then designed two
261 sgRNAs to orient the dCas9-Halo complex towards the G4 structure sitting at either 18 (NT-
262 sgRNA$_{FRET-18}$) or 42 (NT-sgRNA$_{FRET-42}$) base pairs from the targeted G4 and whose Protospacer
263 Adjacent Motifs (PAM) are located on the non-template strand bearing the G4 (NT, Fig. 1g). This is
264 based on previous structural studies indicating that the C-terminal domain of the Cas9 protein, where
265 the Halo protein is situated, will point towards the 3' end of the PAM sequence [41]. Moreover, we have
266 used a standard non-targeting RNA control (NTC) to rule out any potential unspecific binding that is
267 not strictly mediated by sgRNA-driven proximity. We then used Cl-PDS$_n$ molecules as prototype G4-
268 ligand to assess the extent of G4 targeting by ATENA under different conditions, by measuring
269 changes in FRET when targeting the oligo construct with the dCas9-Halo complex in the presence or
270 the absence of Cl-PDS$_n$. We failed to detect any significant changes ($p>0.05$) in the FRET signal for

271 both sgRNAs tested (NT-sgRNA$_{FRET-18}$ vs NT-sgRNA$_{FRET-42}$) when using Cl-PDS$_0$ (Extended Data Fig.
272 1d). Considering that Cl-PDS$_0$ can efficiently bind to dCas9-Halo (Fig. 1f, IV-CP$_{50}$ = 1.6 μM), the
273 lack of G4-stabilization displayed by this molecule indicates that the linker connecting the PDS-
274 scaffold to the dCas9-Halo is inadequately short for engaging with the G4-structure. Nevertheless, a
275 trend showing higher changes in FRET efficiency when using sgRNAs closer to the G4 (NT-
276 sgRNA$_{FRET-18}$ vs NT-sgRNA$_{FRET-42}$) suggests that placing the dCas9 closer to the G4 facilitates ligand
277 engagement (Extended data Fig. 1d). Indeed, when decorating dCas9-Halo with a PDS analogue with
278 a longer PEG-linker (Cl-PDS$_2$), a significant increase ($p<0.05$) in FRET-signal could be measured
279 when using NT-sgRNA$_{FRET-18}$ (ΔFRET = 0.42), which is indicative of G4-engagement. However, when
280 using NT-sgRNA$_{FRET-42}$ we failed to measure an increase in FRET signal, confirming that placing
281 dCas9-PDS closer to the G4 provides better ligand engagement. To further investigate ideal conditions
282 to obtain G4-targeting with ATENA, we also designed T-sgRNA$_{FRET-21}$ and T-sgRNA$_{FRET-41}$ that sit at
283 21 and 41 base pairs from the G4 but whose PAM is located on the template strand (T, Fig. 1g) to
284 investigate the effect of the dCas9-Halo orientation on G4 targeting (Fig. 1g, h). Consistent with our
285 observations indicating that closer placement of dCas9-Halo to the G4 is linked with better ligand
286 engagement, we detected a significant ($p<0.05$) increase in FRET signal when using T-sgRNA$_{FRET-21}$
287 (ΔFRET = 0.33, Fig. 1h) that was abrogated when using T-sgRNA$_{FRET-41}$ (Fig. 1h, SI). This further
288 indicated that optimal G4-targeting by ATENA is achieved by using sgRNAs closer to the G4
289 regardless of the orientation imposed by the sgRNAs used (Fig.1h and SI Fig. 1g). When further
290 increasing the PEG-linker using Cl-PDS$_4$, we failed to observe any significant ($p<0.05$) increase in
291 FRET efficiency with both NT-sgRNA$_{FRET-18}$ and NT-sgRNA$_{FRET-42}$, suggesting that using PEG-linkers
292 that are excessively long is detrimental to G4-engagement, possibly due to high entropic penalty
293 associated with ligand-recognition (Extended Data Fig. 1e). Altogether, our study demonstrated that
294 dCas9-driven G4-ligand engagement is both PEG-linker and sgRNA dependent, with optimal G4-
295 engagement achieved when using a PEG2 linker and sgRNAs placing the dCas9-Halo complex as
296 close as possible (depending on the PAM availability) to the targeted G4.
297

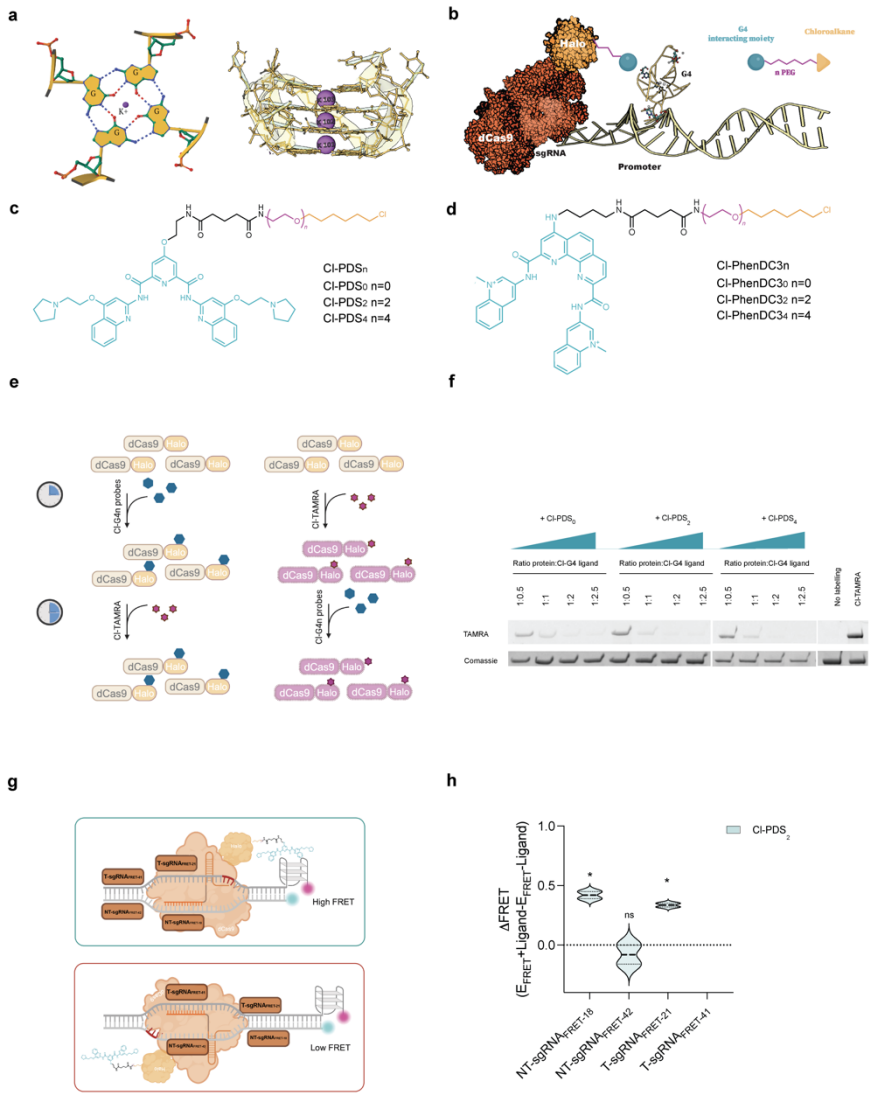


Figure 1: Chemical labeling of dCas9 for selective G4-targeting. **a**, Schematic representation of a G-Tetrad *via* Hoogsteen base pairing with central cation (K^+) (modified from PDB file:6W9P) (Left) and G-Tetrad stacking to form G4-structure (modified from PDB file:6W9P using Protein Imager [42]). (Right). **b**, Schematic overview of ATENA: dCas9-Halo fusion protein functionalized with chloroalkane-modified G4-ligands enables single G4-targeting through sgRNA guidance. **c,** Chemical structure of chloroalkane-modified PyPDS, where n indicates the different PEG linker lengths (Cl-PDS$_n$). **d,** Chemical structure of chloroalkane-modified PhenDC3 (Cl-PhenDC3$_n$), where n indicates the different PEG linker lengths (Cl-PhenDC3$_n$). **e,** Illustration of the competition assay workflow to test the binding ability of Cl-PDS$_n$ probes to dCas9-Halo purified recombinant protein. **f,** SDS-Page gel of the Cl-PDS$_n$ competition assay shows each sample's fluorescent level acquired in the TAMRA

309 channel (542 nm) and the corresponding protein level (Coomassie staining): (n=2). **g,** Schematic
310 representation of the dually labeled FRET oligos containing c-KIT2-G4 forming sequence bound by
311 dCas9-Halo labeled with Cl-PDS$_n$ probes to study G4 stabilization with respect to the PEG-linker
312 length and sgRNA positioning. **h,** ΔFRET efficiency of the decorated dCas9-PDS (with Cl-PDS$_2$)
313 complex targeting c-KIT2-G4. The values indicated were extrapolated from the band intensity
314 measured in the Cy3 and Cy5 channels (Typhoon FLA 9500). The signals in both channels were
315 normalized for the background and the sgRNA NTC control. These normalized fluorescence values
316 were then used to calculate the ΔFRET efficiency for each sgRNA: FRET-Efficiency (E)+ligand $_{sgRNAx}$
317 – FRET-Efficiency (E)-ligand $_{sgRNAx}$ (n=2). Data presented are the mean of n = number of
318 independent experiments. Statistical significance was calculated using a two-tailed t-test in GraphPad
319 Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.
320
321 **Chloroalkane-modified ligands can label dCas9-Halo efficiently in cells.**
322 Having optimized conditions to achieve G4-engagement *in vitro* with ATENA - using Cl-PDS$_n$ and c-
323 KIT2 G4s as a model system - we next investigated whether ATENA can be used to stabilize
324 individual G4 structures in living cells. The use of cell lines that constitutively express dCas9-Halo is
325 essential to ensure consistent cellular levels of the protein across different experiments, avoiding bias
326 introduced by the significantly variable levels of protein expression typically with transient
327 transfection used in previous reports. [34] To this end, we integrated dCas9-Halo into the genome of the
328 breast cancer cell line (MCF7) using standard lentiviral integration approaches (see methods). We
329 selected MCF7 cells in light of the highly diverse transcriptional response previously reported upon
330 treatment with PDS [43], which we wanted to investigate further with ATENA. Successful integration of
331 dCas9-Halo was confirmed by PCR-based genotyping and Western Blot (Extended Data Fig. 2a, b).
332 Next, we evaluated the ability of the chloroalkane functionalized ligands to bind dCas9-Halo in cells.
333 Using an established chloroalkane penetration assay (CAPA), we compared each ligand's relative
334 potency to label covalently dCas9-Halo under physiological conditions [44]. During CAPA, cells are
335 initially exposed to increasing concentrations of the chloroalkane-modified G4-ligands, before

336 incubation with a Halo-reactive Oregon Green fluorophore (Cl-OG), which reacts with any Halo-tag
337 binding site that has been left unoccupied by the previous exposure to the chloroalkane G4-ligands
338 (Fig. 2a, i-ii). The efficiency of G4-ligand incorporation can be, therefore, measured as an inverse
339 function of the Oregon Green fluorescence emission, as successful G4-ligand incorporation to Halo
340 prevents subsequent fluorophore functionalization (Fig. 2a, iii-iv). To quantify this numerically, we
341 calculated the half-maximal chloroalkane penetration value (CP_{50}), which is the ligand concentration
342 required to label 50% of the available dCas9–Halo molecules and can be used as a direct readout of
343 target occupancy[44]. As displayed in Fig. 2a (right), treatment of MCF7 cells expressing dCas9-Halo
344 with both Cl-PDS_0 and Cl-PDS_4 revealed a modest dose-dependent reduction of Oregon Green
345 emission, providing CP_{50} values of 15.9 μM and 5.4 μM, respectively (Extended Data Fig. 2c).
346 Conversely, Cl-PDS_2 could label ~90% dCas9-Halo at a concentration as low as 0.25 μM (CP_{50}
347 0.012 μM, Extended Data Fig. 2c), saturating at 2.5 μM (Fig. 2a), suggesting that the cellular
348 permeability and bioavailability of Cl-PDS_2 were particularly suitable for its application in ATENA.
349 Given that the PEG2 linker also led to best G4-engagement *in vitro* (Fig. 1h), we decided to assess the
350 compatibility with ATENA of a different G4-ligand (PhenDC3) bearing a PEG2 linker (Cl-$PhenDC3_2$)
351 through CAPA. Gratifyingly, Cl-$PhenDC3_2$ labelled very efficiently dCas9-Halo in cells, yielding a
352 CP_{50} value of 1.7 μM (Extended Data Fig. 2d). Cl-$PhenDC3_4$ showed a similar trend to Cl-PDS_4,
353 indicating that PEG4 functionalized ligands could not be employed in ATENA.
354

355 **Selective targeting of MYC-G4 through ATENA reveals no changes in *c-MYC* expression**

356 After having identified conditions to decorate dCas9-Halo with G4-ligands in cells, we next set out to
357 investigate transcriptional responses associated with individual G4-targeting using ATENA. We
358 started our study by considering the G4 present in the promoter of the *c-MYC* proto-oncogene (MYC-
359 G4), as this is one of the most studied and described G4s in the literature. Many studies have linked
360 the targeting of MYC-G4 with ligands to transcriptional suppression of *c-MYC* [12]. To assess the extent
361 of transcriptional perturbation mediated exclusively by G4 stabilization, we designed a panel of
362 sgRNAs to direct ATENA at MYC-G4.

363 Specifically, we designed sgRNAs targeting either the non-template strand (NT) bearing the G4-
364 structure at its 3' end or the opposite strand at its 5' end (T), Fig. 2b. Based on our biophysical
365 investigation, we reasoned that placing sgRNAs close enough to the G4 would have ensured G4-
366 stabilization by G4-ligands tethered to dCas9-Halo (Fig. 1h). Considering PAM sequences available
367 for dCas9-Halo binding at either G4 ends, we designed $sgRNA_{MYC-19}$ and $sgRNA_{MYC-67}$ that would
368 place the protein complex on the T strand, respectively 19 and 67 base pairs away from the MYC-G4.
369 Similarly, we generated $sgRNA_{MYC+22}$ and $sgRNA_{MYC+58}$, targeting the MYC-G4 from its 5' end at a
370 distance of 22 and 58 base pairs on the NT strand, respectively. To further investigate the optimal
371 distance to achieve G4 stabilization in a cellular context, which might differ from our simple
372 biophysical model, we have also designed $sgRNA_{MYC-89}$ and $sgRNA_{MYC-93}$, along with $sgRNA_{MYC+75}$,
373 $sgRNA_{MYC+77,}$ and $sgRNA_{MYC+119}$, also targeting the MYC-G4 at its 3' and 5' end, respectively, but at a
374 further distance from the targeted structure (Fig. 2b). After cloning sequences encoding the various
375 sgRNAs into a vector for mammal expression (see methods), we have transfected MCF7 cells stably
376 expressing dCas9-Halo with individual sgRNAs, followed by treatment with either Cl-PDS_2 or mock
377 (DMSO) for 48 hours. We then measured changes in *c-MYC* expression using RT-qPCR, normalizing
378 the expression level against individual samples transfected with the respective sgRNAs and mock-
379 exposed. This enabled us to consider any change in gene expression potentially triggered by the
380 positioning of the dCas9-Halo complex on the targeted site and, therefore, control for any
381 transcriptional perturbation imposed by dCas9 not functionalized with ligands.
382 Unexpectedly, we observed that sgRNAs placing ATENA within a ~50 bp window of the MYC-G4 on
383 the T strand ($sgRNA_{MYC-19}$, $sgRNA_{MYC-67}$) in conjunction with Cl-PDS_2 treatment led to negligible
384 effects on the global expression of *c-MYC* (Fig. 2c), contrasting our biophysical predictions (Fig. 1h).
385 Conversely when placing the complex on the NT strand with $sgRNA_{MYC+22}$ and $sgRNA_{MYC+58}$, we
386 observed a modest decrease of *c-MYC* expression to 0.8-fold the mock (~20% reduction), which could
387 be indicative of G4-engagement of PDS mediated by ATENA (Fig. 2c). To investigate this further, we
388 analyzed the transcriptional effects elicited by placing the ATENA further away from the MYC-G4
389 using $sgRNA_{MYC+75}$, $sgRNA_{MYC+77,}$ and $sgRNA_{MYC+119}$, which should lead to abrogation of G4-
390 engagement in a distance-dependent fashion, as observed in our biophysical measurements. To our

391 surprise, we observed the opposite with *c-MYC* expression being drastically reduced the further away
392 the complex was from the G4 (Fig. 2c), which cannot be consistent with a G4-mediated effect. By
393 closer inspection of the promoter annotation, we noticed that $sgRNA_{MYC+58}$ to $sgRNA_{MYC+119}$
394 overlapped with the P1 promoter and TATA-box sequences. Since $sgRNA_{MYC+75}$ and $sgRNA_{MYC+77}$ are
395 complementary to the region identifying a TATA box, we reasoned that ATENA occupies the TATA-
396 box region and hinders transcription initiation, resulting in the observed *c-MYC* suppression. This
397 effect is exacerbated when using $sgRNA_{MYC+119}$ that targets the area next to the Transcriptional Starting
398 Site (TSS) of the P1 promoter (4 bp downstream), further indicating a G4-independent transcriptional
399 suppression.
400 Collectively, our results indicated that the reduction in *c-MYC* transcript levels is driven by ATENA's
401 interference with the TATA-box region, rather than by ligand-induced stabilization of the MYC-G4,
402 suggesting that previous observations obtained with the equivalent of our $sgRNA_{MYC+58}$ are likely
403 affected by this[34]. To test this hypothesis further, we replaced the G4-stabilizer (Cl-PDS_2) with a
404 fluorophore (Cl-OG) and monitored *c-MYC* expression while using the sgRNA previously reported to
405 provide the strongest downregulation ($sgRNA_{MYC+58}$) [34]. Under these conditions, we observed similar
406 transcriptional downregulation of *c-MYC* compared to what was measured when treating with Cl-
407 PDS_2 (Fig. 2d), demonstrating that anchoring the dCas9 complex in the proximity of key
408 transcriptional regions of the *c-MYC* promoter prevents any reliable evaluation of G4-mediated
409 transcriptional effects. The interference of ATENA with *c-MYC* expression when placed in proximity
410 to key promoter sites is consistent with what has been reported for CRISPRi studies[45,46] and needs to
411 be carefully considered when using dCas9-based strategies to target G4s[34].
412
413 **ATENA confirms P1-dependent transcriptional expression associated with MYC-G4**
414 It has been shown that multiple promoters globally contribute to regulating *c-MYC* expression [47,48].
415 Therefore, we decided to examine how G4-targeting affects *c-MYC* expression regulated by specific
416 promoters by analyzing transcripts originating from the two main ones: P1 and P2.
417 Indeed, it has been recently shown that genetic deletion of the MYC-G4 is associated with selective
418 suppression of transcription from the P1 promoter and only a modest reduction of overall *c-MYC*

419 expression, which is instead represented by the combined transcriptional output of both the P1 and P2
420 promoters [20]. When using ATENA with Cl-PDS$_2$ and monitoring P1-driven *c-MYC* expression, we
421 observed a strikingly distance-dependent suppression of P1-mediated expression with sgRNA$_{MYC-19}$
422 and sgRNA$_{MYC-67}$. In particular, when using sgRNA$_{MYC-19}$ we observed an 85% reduction of P1-
423 mediated *c-MYC* expression (0.15-fold), whereas use of sgRNA$_{MYC-67}$ led to a 72% reduction of the *c-*
424 *MYC* expression (0.28-fold), Fig. 2e. Importantly, no significant changes in expression were detected
425 when using ATENA in conjunction with sgRNA$_{MYC-89}$ and sgRNA$_{MYC-93}$ that place the Cl-PDS$_2$
426 excessively distant from the targeted G4, which agrees with a G4-dependent transcriptional
427 suppression (Fig. 2e). We then asked why the reduction in P1-driven expression observed with
428 ATENA does not result in an overall decrease in *c-MYC* transcription. To achieve this, we also
429 measured changes in transcription originating from the P2 promoter using promoter-specific qPCR
430 primers (see methods). Notably, we observed an increase in P2-selective expression that suggests a
431 compensatory mechanism activated by the cells in response to the G4-induced reduction of P1
432 transcription (Extended Data Fig. 3a), justifying the absence of statistically significant changes in
433 global *c-MYC* expression detected when using primers amplifying regions common to both P1- and
434 P2-derived transcripts.
435
436 Targeting the G4 from its 5' end with sgRNA$_{MYC+22}$ led to a 75% reduction of P1-driven *c-MYC*
437 expression (0.25-fold) Fig. 2e. We also detected strong P1-dependent transcriptional repression when
438 using sgRNA$_{MYC+58}$ and sgRNA$_{MYC+75}$, which might be affected by G4-independent transcriptional
439 perturbation that we already observed when using these sgRNAs (Fig. 2e). Indeed, P1-mediated *c-*
440 *MYC* expression was entirely abrogated when the ATENA was directed at sites overlapping close to
441 the P1-TSS with sgRNA$_{MYC+119}$ (Fig. 2e), which is consistent with G4-independent transcriptional
442 inhibition. These observations further confirmed that using ATENA on the 5' end of the MYC-G4
443 cannot reliably detect changes in gene expression that G4-targeting strictly mediates, and careful
444 consideration of the promoter regulatory elements is needed.
445

446 Next, we sought to confirm that the observed changes in P1-driven *c-MYC* expression following
447 treatment with $sgRNA_{MYC-19}$ and $sgRNA_{MYC-67}$ result specifically from ATENA-mediated targeting of
448 the MYC-G4, rather than non-specific ligand interactions with other G4 structures. To address the
449 potential for ligand-mediated off-target effects, we monitored the expression of *KRAS*, a gene known
450 to contain a stable G4 structure in its promoter region. As shown in Extended Data Fig. 3b, directing
451 ATENA specifically to the MYC-G4 did not alter *KRAS* expression, supporting the selectivity of
452 ATENA-mediated G4 targeting. In contrast, free PyPDS treatment significantly lowered *KRAS*
453 expression (0.20-fold), leading to an 86% transcriptional suppression (Extended Data Fig. 3c), which
454 validates the ability of ATENA to confer G4-ligand selectivity towards individual G4s, whilst
455 minimizing off-target effects.
456
457 **P1-dependent transcriptional inhibition is linked with protein displacement from MYC-G4**
458 We further assessed the ability of ATENA to mediate selective MYC-G4 targeting by investigating
459 perturbation in protein binding at MYC-G4 upon ligand stabilization. It has been proposed that
460 ligands bound to G4s can displace key transcription factors and regulatory proteins, leading to the
461 observed transcription suppression [21,49]. Therefore, we reasoned that if ATENA was correctly
462 positioned to enable ligand-G4 interaction, we should have observed reduced protein accessibility to
463 the G4. To measure this, we used the G4-selective antibody BG4 [4] and performed CUT&Tag [7]
464 coupled with qPCR to compare the efficiency of BG4 in enriching for MYC-G4, targeted by ATENA,
465 against 3 independent validated G4-sites (MAZ, RPA3, RBBP4) that should not be affected as they
466 are not targeted by ATENA. This enabled us to assess relative protein accessibility at these individual
467 G4 sites under different conditions, as previously described [20]. As displayed in Fig. 2f, when treating
468 cells with $Cl\text{-}PDS_2$ using $sgRNA_{MYC-19}$, we observed a consistent reduction of BG4 signal that was not
469 detected for the Non-Targeting Control (NTC), irrespective of the reference G4 used. This result
470 strongly suggests that ATENA can be used to guide $Cl\text{-}PDS_2$ selectively to the MYC-G4 structure,
471 resulting in a decrease in binding of the BG4 antibody to MYC-G4 due to the binding competition of
472 the ligand that leads to displacement of the antibody from the G4. This supports a model in which

473 ligand-mediated G4 stabilization suppresses P1-driven *c-MYC* transcription by outcompeting binding
474 of transcriptional effectors at the MYC-G4[21].
475
476 **P1-dependent transcriptional suppression is recapitulated with PhenDC$_3$**
477 To further validate our findings, we used ATENA to deploy a different G4 ligand - PhenDC3 [31] - to
478 stabilize the MYC-G4. The PEG2-functionalized analogue, Cl-PhenDC3$_2$, also efficiently labeled
479 dCas9–Halo in cells, as confirmed by CAPA (Extended Data Fig. 2d). When using ATENA with Cl-
480 PhenDC3$_2$ and sgRNA$_{MYC-19}$, we observed a significant inhibition of P1-mediated *c-MYC*
481 transcription (0.09-fold), leading to a 91% reduction of expression (Fig. 2g), which is even greater
482 than what observed with Cl-PDS$_2$, likely reflecting the stronger G4 stabilization capacity of Cl-
483 PhenDC3$_2$, as indicated by biophysical CD measurements (Extended Data Table 4). This result
484 suggests that ATENA can also be leveraged to compare the relative potency and biological impact of
485 different G4 ligands when deployed at the same genomic target.
486 Collectively, our data indicate that ATENA can successfully target MYC-G4 in a ligand-independent
487 fashion, leading to a detectable G4-engagement and corresponding P1-specific *c-MYC* suppression,
488 validating recent findings generated by genetic deletion of the sequence responsible for the MYC-G4
489 folding[20].
490

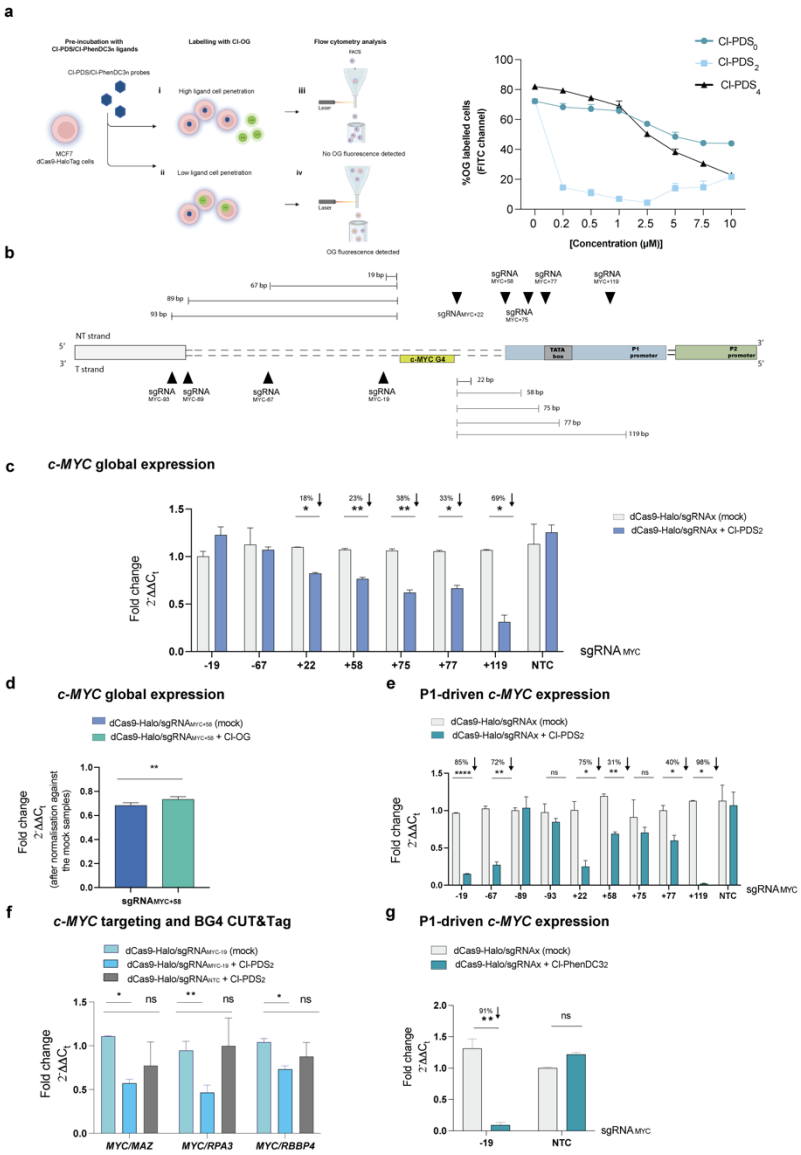


Figure 2: ATENA enables selective targeting of a G4 in the *c-MYC* promoter. a, (left) Schematic representation of the CAPA assay; (**i**) high probe cell penetration/reactivity results in Cl-OG inability to bind the HaloTag, conversely (**ii**) low probe cell penetration/reactivity allows Cl-OG to bind the HaloTag, resulting respectively in (**iii**) no Oregon Green fluorescence detected *via* Flow cytometry or (**iv**) presence of Oregon Green fluorescence signal in the FITC channel. (right) CAPA assay curves of Oregon Green fluorescence signal obtained upon treatment of MCF7 cells stably expressing dCas9-Halo with Cl-PDS$_n$ followed by fluorophore incubation (n=2). Flow cytometry was performed using Attune NxT Flow cytometer (FITC channel), and data were then analyzed with FlowJo software. **b**, Schematic illustration of the *c-MYC* promoter with the annotated G4 (MYC-G4), sgRNA targeting region (black triangles), and their relative distance in bp from the MYC-G4. **c**, RT-qPCR for *c-MYC*

502 expression in MCF7 cells stably expressing dCas9-Halo transfected with the indicated sgRNAs and
503 incubated for 48h in the presence of (2.5 μM) Cl-PDS_2 or DMSO (mock). The expression values are
504 represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) and normalized for the
505 housekeeping gene *GAPDH*. n=3, biological replicates, each with three technical replicates. **d,** RT-
506 qPCR for *c-MYC* expression in MCF7 cells stably expressing dCas9-Halo transfected with
507 $sgRNA_{MYC+58}$ and incubated for 24h in the presence of either Cl-PDS_2 (2.5 μM) or Cl-OG (5 μM). The
508 expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated)
509 and normalized for the housekeeping gene *GAPDH*. n=2, biological replicates, each with two
510 technical replicates. **e,** RT-qPCR for P1-driven *c-MYC* expression in MCF7 cells stably expressing
511 dCas9-Halo transfected with the indicated sgRNAs and incubated for 48h in the presence of (2.5 μM)
512 Cl-PDS_2. The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock
513 (DMSO-treated) and normalized for the housekeeping gene *GAPDH*. n=3, biological replicates, each
514 with three technical replicates. **f,** BG4 CUT&Tag-qPCR for MCF7 cells stably expressing dCas9-Halo
515 transfected with either $sgRNA_{MYC-19}$ or sgRNA NTC and treated with DMSO (mock) or (2.5 μM) Cl-
516 PDS_2. BG4 accessibility was analyzed for *c-MYC* and normalized to three G4s in control gene sites
517 (*RPA3*, *MAZ*, *RBBP4*). n=2, biological replicates, each with three technical replicates for BG4 and one
518 for the negative (no BG4 treatment). **g,** RT-qPCR for P1-dependent *c-MYC* expression in MCF7 cells
519 stably expressing dCas9-Halo transfected with $sgRNA_{MYC-19}$ or sgRNA NTC and incubated for 48h in
520 the presence of (2.5 μM) Cl-$PhenDC3_2$. The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$)
521 with respect to the mock (DMSO-treated) and normalized for the housekeeping gene *GAPDH*. n=3,
522 biological replicates, each with two technical replicates. Data presented are the mean of n = number of
523 independent biological samples. Statistical significance was calculated using a Welch-corrected two-
524 tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.
525

526 **The MYC-G4 selective molecule DC-34 validates ATENA.**

527 Encouraged by the P1-driven *c-MYC* suppression observed upon targeting MYC-G4 with ATENA, we
528 explored whether a similar phenotype could be elicited when using ligands that display some inter-G4
529 selectivity. To this end, we leveraged the MYC-G4 selective ligand DC-34 (Fig. 3a), which shows a

530 strong binding affinity for the MYC-G4 with high selectivity over other G4 structures such as *KRAS*
531 and *c-KIT* [24]. We treated MCF7 cells with increasing concentrations of DC-34 for 48 hours before
532 measuring changes in *c-MYC* expression by RT-qPCR. As observed with ATENA, treatment with DC-
533 34 caused negligible dose-dependent changes in global *c-MYC* expression (Fig. 3b), supporting the
534 notion that selective targeting of MYC-G4 does not impact the overall expression of *c-MYC* in MCF7
535 cells. Conversely, when measuring P1-mediated transcription, DC-34 revealed a strong dose-
536 dependent suppression that plateaued at 74% reduction when treating with 10 μM of the ligand, (0.26-
537 fold), Fig. 3c. This indicated that using an inter-G4 selective ligand for selective MYC-G4 targeting
538 led to observations comparable to those obtained using ATENA, further corroborating the validity of
539 our platform for single G4-targeting.
540 Next, we assessed the selectivity of DC-34 for MYC-G4 relative to another G4-containing promoter,
541 as we have done for ATENA. Our observations indicated that ATENA did not lead to detectable
542 changes in *KRAS* expression when guided to MYC-G4 (Extended Data Fig. 3b). In contrast, treatment
543 with DC-34 resulted in a dose-dependent reduction in *KRAS* expression (~30%, Fig. 3d), which was
544 less pronounced than what was observed for *c-MYC* under the same conditions, indicating that MYC-
545 G4 is the primary target of DC-34, but residual off-target binding to other G4s may occur at high
546 concentrations. This likely reflects the structural similarity shared between different G4s, which
547 substantially complicates inter-G4 selective targeting with small-molecule ligands.
548 To further confirm that DC-34 downregulates *c-MYC* through direct G4 binding, we also performed
549 the BG4 CUT&Tag qPCR upon ligand treatment that we optimized for ATENA to assess protein
550 occupancy upon treatment. As shown in Fig. 3e, DC-34 treatment significantly reduced the BG4
551 signal at MYC-G4, which is consistent with the decreased protein accessibility induced by MYC-G4
552 selective targeting with ATENA. This further supports the use of BG4 CUT&Tag qPCR as an indirect
553 measure of ligand-mediated G4 stabilization associated with transcriptional suppression.
554
555 **Transcriptome-wide comparison of ATENA with DC-34**
556 To further assess the inter-G4 selectivity provided by ATENA and DC-34, we analyzed transcriptome-
557 wide gene-expression changes through mRNA-seq. Specifically, we generated mRNA-Seq datasets

558 for MCF7 cells treated with either DC-34 or ATENA (Cl-PDS$_2$ in conjunction with sgRNA$_{MYC-19}$) and
559 compared those to transcriptome-wide changes induced by the generic G4-ligand PyPDS. We
560 hypothesized that the inter-G4 selectivity provided to Cl-PDS$_2$ by ATENA should be reflected by a
561 substantially lower number of differentially expressed genes (DEGs) when compared to free PyPDS.
562 Indeed, treatment of MCF7 cells with PyPDS (2.5 μM) for 6 hours altered the expression of 2,228
563 genes (319 downregulated, 1907 upregulated, FDR<0.05, Extended Data Fig. 4). In contrast,
564 treatment with the MYC-G4 selective ligand DC-34 (7.5 μM, 6h) affected only the expression of 120
565 genes (39 downregulated, 81 upregulated; FDR<0.05, Fig. 3f), indicating that the enhanced inter-G4
566 selectivity of DC-34 is translated into a lower number of genes being differentially expressed when
567 compared to PyPDS. Additionally, mRNA-Seq confirmed that DC-34 did not significantly alter *c-*
568 *MYC* expression globally, as already observed in our RT-qPCR data, further indicating the selective
569 downregulation of P1 transcripts when targeting MYC-G4, an essential consideration when devising
570 G4-based therapies aiming at suppressing *c-MYC* expression. Given the substantial reduction in
571 DEGs observed with DC-34 relative to a global G4-stabilizer like PyPDS, we hypothesized that
572 conjugation of Cl-PDS$_2$ to dCas9 in ATENA should also reduce the extent of transcriptional
573 perturbations. To investigate this, we have transfected MCF7 cells with sgRNA$_{MYC-19}$ and treated
574 them with Cl-PDS$_2$ (6 h) before performing mRNA-Seq.
575 Initially, we performed a standard differential expression analysis comparing ATENA (sgRNA$_{-MYC-19}$
576 or sgRNA NTC + Cl-PDS$_2$) to their DMSO-treated controls using DESeq2. We filtered the resulting
577 DEGs for statistical significance (adjusted p-value < 0.05) and magnitude of change
578 (|log2FoldChange| ≥ 1), Table S9.
579 However, to identify genes uniquely responsive to MYC-targeted ligand recruitment, we then
580 excluded all genes that were also differentially expressed in the sgRNA NTC + Cl-PDS$_2$ *vs* DMSO
581 comparison, using the same filtering criteria. This filtering allowed us to isolate transcriptional effects
582 that were specific to the combination of the sgRNA$_{MYC-19}$ and the ligand, rather than shared responses
583 with sgRNA NTC that reflect more unspecific effects of the dCas9 platform (Fig. 3g). This analysis
584 identified only 43 DEGs, underscoring the exceptional specificity conferred to Cl-PDS$_2$ when
585 conjugated to dCas9 as opposed to the free PyPDS ligand.

To evaluate if ATENA, DC-34, and free PyPDS shared any of the transcriptional changes elicited, we compared DEGs observed in: DC-34 vs ATENA (NTC-filtered), PyPDS vs ATENA (NTC-filtered), and PyPDS vs DC-34, while also testing for enrichment in *c-MYC*-related pathways[50].

When comparing DC-34 *vs* ATENA (NTC-filtered), and free PyPDS *vs* ATENA (NTC-filtered), we found an overlap of a few genes that were not enriched in any pathway, suggesting that these approaches yield largely orthogonal transcriptomic profiles, consistent with their distinct mechanisms of delivery and engagement (Tables S10-11). Finally, when comparing PyPDS vs DC-34, we observed 24 shared DEGs (Table S12). However, these DEGs failed to enrich for any known pathways or MYC-related functions, indicating that even structurally distinct G4 ligands with overlapping target preferences can elicit unique transcriptional responses, likely due to differences in binding affinity, cellular uptake, and selectivity.

Altogether, these analyses indicated that ATENA, DC-34, and PyPDS can elicit distinct transcriptional responses, with negligible overlap and no shared enrichment for *c-MYC*-related genes. This reinforces the conclusion that ATENA - by spatially confining ligand activity to a single G4 at the *c-MYC* promoter - induces highly selective gene expression changes, contrasting with the broader, less discriminating effects of freely diffusing G4-ligands. Additionally, our mRNA-Seq analysis further confirmed that MYC-G4 targeting is not associated with significant *c-MYC* downregulation in MCF7 cells, and caution is needed when designing MYC-based therapeutics based on G4-targeting.

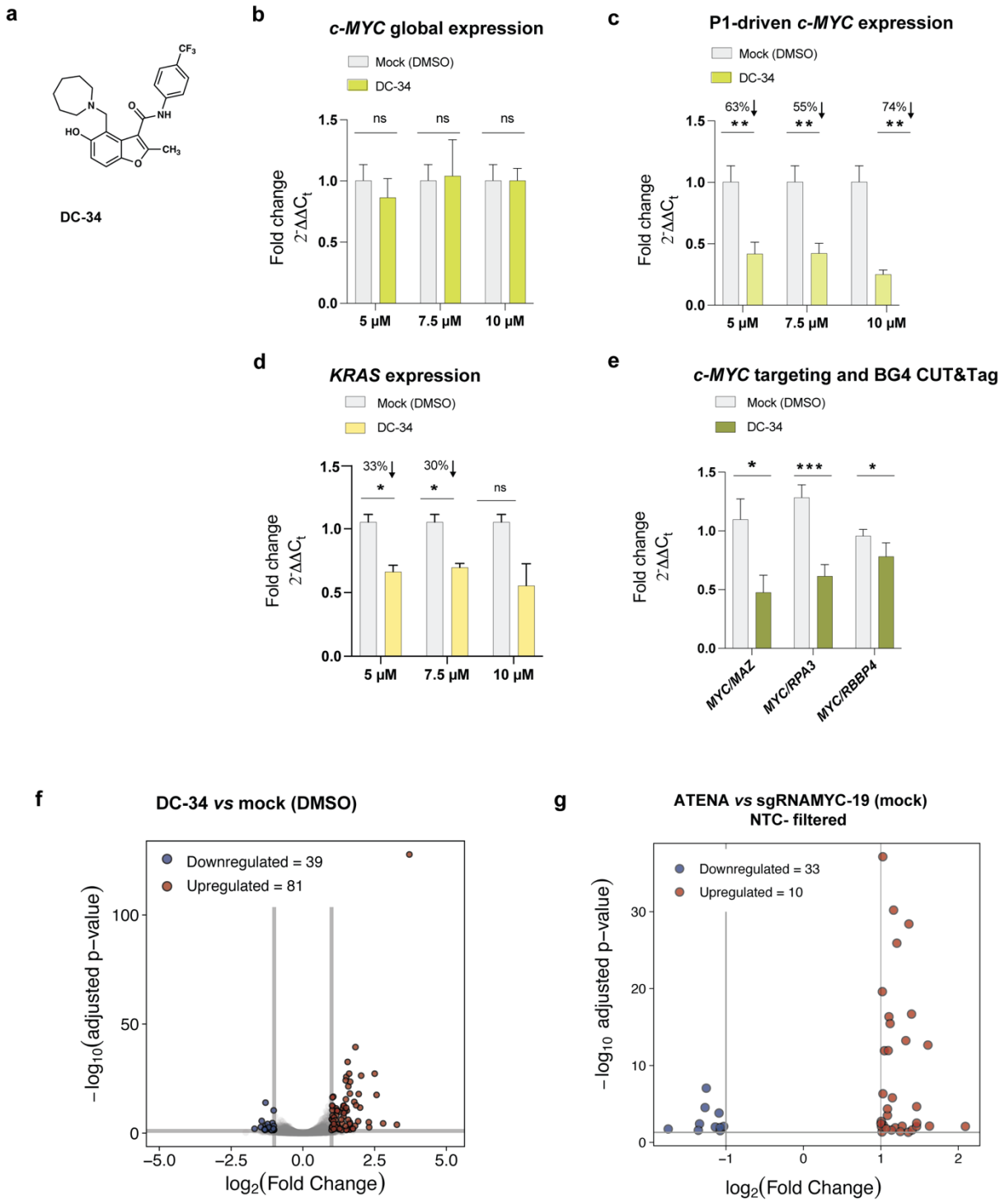


Figure 3: The selective G4-MYC ligand DC-34 matches ATENA. a, Chemical structure of DC-34. **b,** RT-qPCR for *c-MYC* expression in MCF7 cells treated with increasing concentration of DC-34 for 48h. Values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) and normalized for the housekeeping gene *GAPDH*. n=3, biological replicates, each of which included two technical replicates. **c**, RT-qPCR for P1-driven *c-MYC* expression in MCF7 cells treated with

617 different concentrations of DC-34 for 48h. Values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect
618 to the mock (DMSO-treated) and normalized for the housekeeping gene *GAPDH*. n=3, biological
619 replicates, each of which included two technical replicates. **d**, Evaluation of potential DC-34 off-
620 targets by analyzing the *KRAS* expression using the same samples reported in c. **e,** BG4 CUT&Tag-
621 qPCR for MCF7 cells treated with DMSO (mock) or (7.5 μM) DC-34. BG4 accessibility was analyzed
622 for *c-MYC* and normalized to three G4s in control gene sites (*RPA3*, *MAZ*, *RBBP4*). n=2, biological
623 replicates, including three technical replicates for BG4 and one for the negative (no BG4 treatment). **f,**
624 Volcano plot of DEGs in MCF7 cells treated with DC-34 *vs* cells treated with DMSO (mock). Gray =
625 non-significant genes (FDR ≥ 0.05 or $|\log_2 FC| < 1$); red = up-regulated DEGs (FDR < 0.05, $\log_2 FC \geq$
626 +1); blue = down-regulated DEGs (FDR < 0.05, $\log_2 FC \leq -1$).**g,** Volcano plot of DEGs in MCF7 cells
627 transfected with $sgRNA_{MYC\text{-}19}$ treated with Cl-PDS_2 *vs* mock, after sgRNA NTC filtering; Red = up-
628 regulated DEGs (FDR < 0.05, $\log_2 FC \geq +1$); blue = down-regulated DEGs (FDR < 0.05, $\log_2 FC \leq$
629 -1). Plot includes only genes passing FDR < 0.05 and $|\log_2 FC| \geq 1$. The data presented are the mean
630 of n = number of independent biological samples. Statistical significance was calculated using a
631 Welch-corrected two-tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, ***
632 ≤0.001, **** ≤0.0001.
633
634 **Targeting of the *c-MYC* i-motif with ATENA is associated with transcriptional stimulation.**
635 After validating the suitability of ATENA for the selective targeting of individual G4-structures in the
636 genome - exemplified by the MYC-G4 - we next explored whether this platform could be adapted to
637 interrogate other DNA secondary structures, such as i-motifs (iMs). Similarly to G4s, iMs are
638 stabilized by Hoogsteen hydrogen bonding. However, they form in cytosine-rich regions of the
639 genome, typically complementary to G-rich G4-forming sequences[51]. The formation of iMs in cells
640 has been recently validated in living cells by both immunofluorescence[23] and genome-wide
641 mapping[22]. Like G4s, iMs have been implicated in transcriptional regulation[35], although their
642 mechanistic roles remain less well characterized compared to G4s.

643 To assess the potential of individual iMs to modulate gene expression, we decided to use ATENA in
644 combination with selective iM ligands. Conveniently, the *c-MYC* promoter also bears an iM structure
645 complementary to the G-rich sequence forming the MYC-G4[52], which has been reported to modulate
646 *c-MYC* expression when targeted with selective iM ligands[53-55]. We therefore reasoned that the same
647 sgRNAs optimized for selective targeting of MYC-G4 could be repurposed to localize ATENA to the
648 MYC-iM, enabling selective iM targeting when decorated with an appropriate ligand. To this end, we
649 utilized a recently developed class of short peptides from the Waller group, which show high affinity
650 and specificity for iMs over G4s[33]. Among those, we selected the RVS peptide (pep-RVS) for
651 synthetic ease and potential amenability for modification. To enable compatibility with ATENA, we
652 chemically modified pep-RVS with chloroalkene side chains to ensure covalent attachment onto the
653 Halo-Tag (SI-2 and SI-7), generating Cl-pep-RVS$_n$ analogs with different PEG-linkers (n). We first
654 confirmed that functionalization with chloroalkene did not affect the ability of pep-RVS to bind iMs
655 *via* UV titrations, confirming that it bound the c-MYC i-motif structure with a K_d of 0.35 ± 0.12 μM
656 compared to >33 μM for G4 (Fig. SI-18). We next assessed the ability of Cl-pep-RVS$_n$ analogs to
657 covalently bind to the dCas9-Halo in living cells using CAPA. As shown in Extended Data Fig. 5a, all
658 Cl-pep-RVS$_n$ analogs displayed good cellular permeability and effectively labelled dCas9-Halo at low
659 μM concentrations. However, Cl-pep-RVS$_4$ performed best in CAPA, providing a CP_{50} value of 2.3
660 μM, and was selected for further application in ATENA.
661 To evaluate whether targeting the MYC-iM affects *c-MYC* transcription, we employed ATENA with
662 sgRNA$_{MYC-19}$ and treated the cells with Cl-pep-RVS$_4$ for 48h, using the same conditions previously
663 optimized for G4 targeting. Notably, ATENA-mediated iM-targeting led to a significant increase of
664 P1-driven (~2-fold increase) *c-MYC* transcription (Fig. 4b), in contrast to the transcriptional
665 repression observed upon MYC-G4 targeting. Interestingly, this P1-mediated upregulation was
666 accompanied by a decrease in P2-driven transcription (Fig. 4b), resulting in minimal net changes in
667 global *c-MYC* expression - mirroring the outcome observed when targeting the G4 motif.
668 Overall, these results underscore the modularity of ATENA in targeting distinct DNA secondary
669 structures within the same genomic locus simply by varying the conjugated ligand. By exploiting the
670 high iM selectivity of pep-RVS, we show that selective iM targeting is associated with P1-dependent

transcriptional activation of *c-MYC* - a functional effect opposite to that seen with G4 engagement. This highlights the capacity of ATENA to disentangle the complex regulatory roles of overlapping secondary structures in gene promoters.

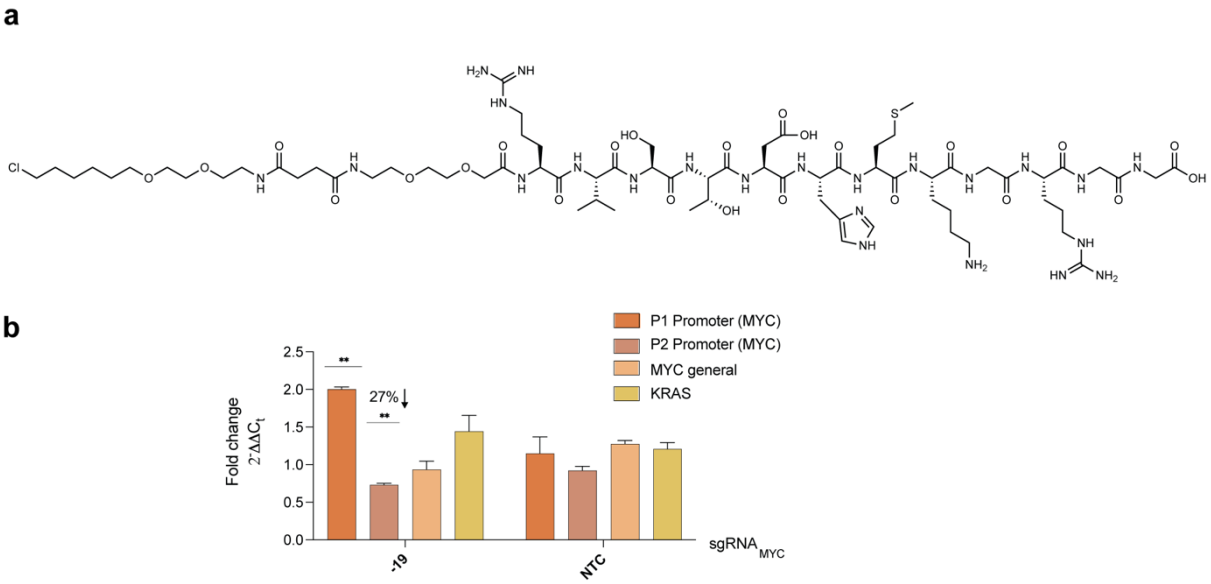


Figure 4: ATENA and *c-MYC* iM targeting. a, Chemical structure of chloroalkane-modified pep-RVS with a 2-PEG linker (Cl-pep-RVS$_2$). **b,** RT-qPCR of the indicated genes in MCF7 cells stably expressing dCas9-Halo transfected with sgRNA$_{MYC-19}$ or sgRNA NTC and incubated for 48h in the presence of 10 µM of Cl-pep-RVS2 or DMSO (mock). The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) transfected samples and after normalization for the housekeeping gene (GAPDH). n=2, biological replicates, each with three technical replicates. The data presented are the mean of n = number of independent biological samples. Statistical significance was calculated using a Welch-corrected two-tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.

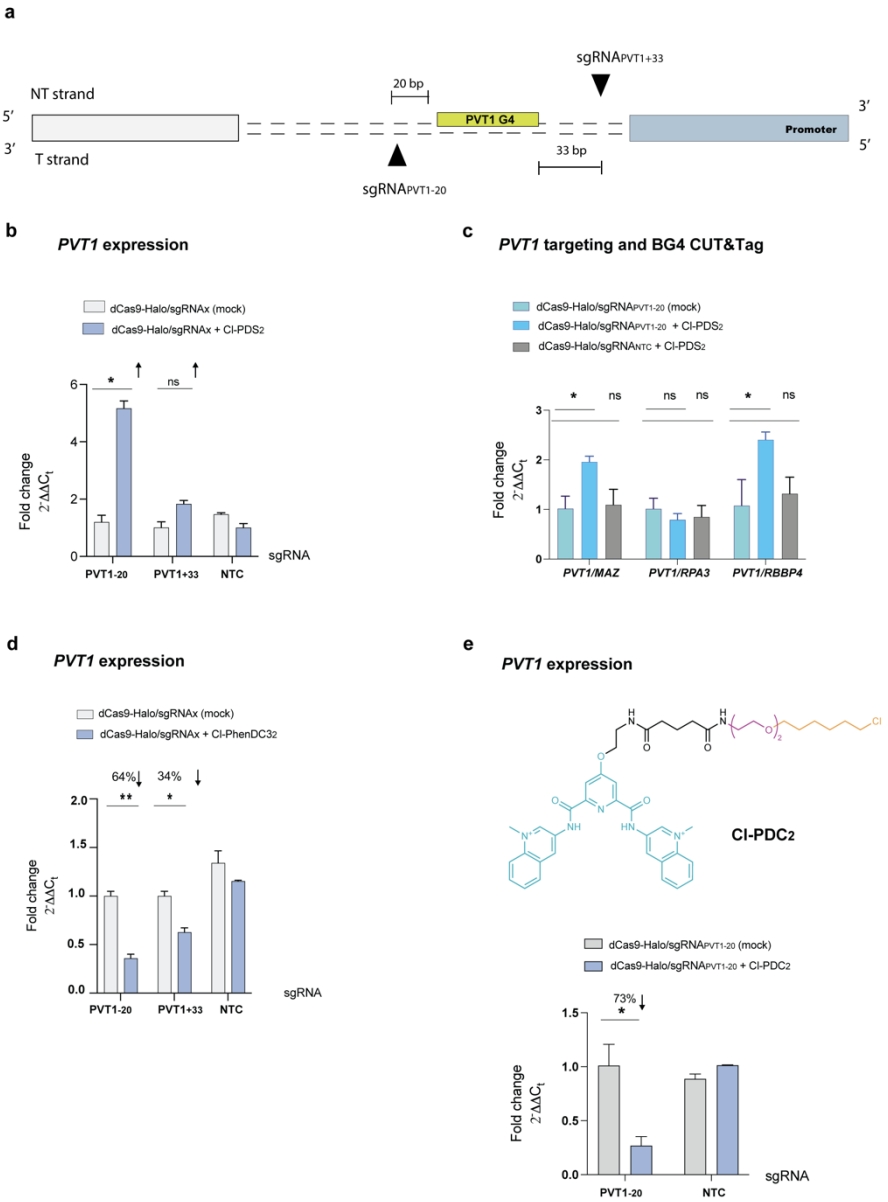
PDS can act as a molecular glue of specific G4-protein interactions.

After establishing the reliability of ATENA to accurately measure biological responses mediated by individual DNA secondary structures, using *c-MYC* as a case study, we next questioned if this platform could be expanded to study transcriptional responses that are uniquely associated with specific ligands. For instance, previous reports indicated that PDS treatment of MCF7 cells led to

690 significant upregulation, rather than repression, of the long non-coding RNA *PVT1*- an observation we
691 confirmed also for PyPDS by RT-qPCR and mRNA-seq (Extended data Fig. 3a) [43]. Therefore, we
692 sought to leverage ATENA to determine if PDS-mediated *PVT1* upregulation was a direct response to
693 the specific stabilization of the G4 in its promoter or rather an indirect effect caused by global G4
694 stabilization. Previous CUT&Tag experiments performed in MCF7 cells identified a clear G4-peak in
695 the *PVT1* promoter, [56] which we have used to design sgRNAs for ATENA-based targeting of the
696 PVT1-G4. Considering available PAM sequences, we designed two sgRNAs targeting the PVT1-G4
697 at either 20 base pairs from its 5' end ($sgRNA_{PVT1-20}$) or 33 base pairs from its 3' end ($sgRNA_{PVT1+33}$)
698 and that display no overlap with any annotated regulatory site (Fig. 5a). Upon transfection with either
699 $sgRNA_{PVT1-20}$ or $sgRNA_{PVT1+33}$ and treatment with Cl-PDS_2, we observed a boosted *PVT1* expression
700 of ~4-fold, consistent with observations reported using free PDS (Fig. 5b). G4-ligands typically
701 compete with regulatory proteins for G4-binding, which leads to transcriptional suppression, as we
702 recapitulated measuring BG4 occupancy upon MYC-G4 targeting (Fig. 2f). Therefore, we questioned
703 whether PDS could instead act as a molecular glue when binding to the PVT1-G4, leading to
704 enhanced protein binding to the PVT1-G4 and, thus, justifying the observed transcriptional increase
705 upon PDS treatment. We performed BG4 CUT&Tag qPCR on PVT1-G4 target by ATENA and in
706 conjunction with Cl-PDS_2 and observed a rise in BG4 occupancy at *PVT1* of ~2-fold (Fig. 5c). This
707 contrasts the transcriptional response elicited by the same ligand when targeting a different G4 (MYC-
708 G4, Fig. 2f) and points to the role of PyPDS as a molecular glue for protein-G4 interactions within the
709 PVT1-G4. Moreover, this observation suggests that the previously described increase in *PVT1*
710 expression elicited by PyPDS treatment in MCF7 cells reflects the genuine response of the ligand
711 targeting the PVT1-G4 and cannot be ascribed to a secondary response associated with global G4-
712 targeting.
713
714 We next asked whether the increase in *PVT1* expression measured with PyPDS was limited to this
715 molecule or if a more general response could be observed with any G4-ligand. MCF7 treatment with
716 free PhenDC3 led to the suppression rather than the enhancement of *PVT1* expression, suggesting that

717 different G4-ligands might elicit different responses when targeting an identical G4 (Extended Data
718 Fig. 5b).
719 To further investigate this, we targeted selectively the PVT1-G4 using ATENA in conjunction with
720 Cl-PhenDC3$_2$ and sgRNA$_{PVT1-20}$ or sgRNA$_{PVT1+33}$. Under these conditions, we observed a suppression
721 of *PVT1* transcription, 64% and 34%, respectively, (expression level of the sample was 0.36-fold and
722 0.66-fold the mock respectively), (Fig. 5d), which is consistent with free PhenDC3 treatment and
723 opposite to what was observed with PyPDS (Extended Data Fig. 5c). This suggested that G4-ligands
724 can elicit a different response when bound to the same G4, possibly reflecting binding modalities that
725 can either increase or prevent protein accessibility, which can be indirectly measured by BG4
726 occupancy (Extended Data Fig. 5d). This is an essential factor to consider when using different G4-
727 ligands to infer the biology associated with these secondary structures, as it is often assumed that G4
728 ligands will all behave the same.
729
730 To further investigate these ligand-specific observations, we have also synthesized a chloroalkene
731 analog of a third widely used G4-ligand: Pyrido Dicarboxamide (PDC, Fig. 5e) [57]. We functionalized
732 the PDC scaffold with a PEG2 chloroalkane side chain (Cl-PDC$_2$) to mimic the PyPDS and PhenDC3
733 analogues used in ATENA (Fig. 4e, SI-1). After validating the labelling efficiency of Cl-PDC$_2$ through
734 CAPA (Fig. SI-5), we used it in conjunction with sgRNA$_{-20}$ to target the *PVT1* promoter and
735 investigate associated transcriptional responses. Similarly, to what was observed with PhenDC3, PDC
736 lowered *PVT1* expression (0.27-fold), leading to a 73% reduction (Fig. 5e), suggesting that PDC
737 interacts with the PVT1-G4 in a manner reminiscent of PhenDC3 and causes protein displacement
738 from the G4. Structurally, the PDC scaffold is indeed similar to PhenDC3, displaying methylated
739 nitrogens on the quinolines that are facing opposite orientation compared to PDS (Fig. 5e). Moreover,
740 both PhenDC3 and PDC lack the amino-side chains present in PyPDS, further highlighting the
741 structural similarity between these two scaffolds, which might recapitulate the similar response
742 observed. Collectively, our results indicated that the transcriptional responses elicited by ligands at
743 individual G4s depend heavily on the structural nature of the ligand and on their binding modality,
744 which might lead to protein displacement at the G4-site or might act as a molecular glue enhancing

745 G4-protein interactions. This strongly suggests that transcriptional changes observed upon G4-ligand
746 treatment should be interpreted as ligand-specific outcomes, reflecting the response to a specific
747 ligand at a specific G4 structure rather than the endogenous function of the DNA structure.



748

749 **Figure 5: ATENA unveils a ligand-dependent transcriptional response of the *lncPVT1*.**
750 **a**, Schematic overview of the *PVT1* promoter containing annotation of the predicted G4, sgRNA
751 targeting region (black triangles) and their relative distance in bp from the G4 forming sequence **b,**
752 RT-qPCR for *PVT1* expression in MCF7 cells stably expressing dCas9-Halo, transfected with either
753 $sgRNA_{PVT1-20}$, $sgRNA_{PVT1+33}$ or sgRNA NTC and treated with (2.5 µM) Cl-PDS_2 for 48h after

754 transfection. The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock
755 (DMSO-treated) and normalized for the housekeeping gene *GAPDH*; n=3, biological replicates, each
756 of which included two technical replicates. **c,** BG4 CUT&Tag-qPCR for MCF7 cells stably expressing
757 dCas9-Halo transfected with either $sgRNA_{PVT1-20}$ or sgRNA NTC and treated with DMSO (mock) or
758 (2.5 μM) Cl-PDS_2. BG4 accessibility was analyzed for *PVT1* and normalized to three G4s in control
759 gene sites (*RPA3*, *MAZ*, *RBBP4*). n=2, biological replicates, including three technical replicates for
760 BG4 and one for the negative (no BG4 treatment). **d**, RT-qPCR for *PVT1* expression in MCF7 cells
761 stably expressing dCas9-Halo, transfected with either $sgRNA_{PVT1-20}$, $sgRNA_{PVT1+33}$ or sgRNA NTC
762 and treated with (2.5 μM) Cl-$PhenDC3_2$ for 48h after transfection. Values are represented as fold
763 change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) and normalized for the housekeeping gene
764 *GAPDH*. n=3, biological replicates, each of which includes two technical replicates. **e,** (Top)
765 Chemical structure of chloroalkane-modified PDC (Cl-PDC_2) with the PEG linker length of two. **e**,
766 (bottom) RT-qPCR for *PVT1* expression in MCF7 cells stably expressing dCas9-Halo, transfected
767 with either $sgRNA_{PVT1-20}$ or sgRNA NTC and treated with (2.5 μM) Cl-PDC_2 for 48h after
768 transfection. Values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated)
769 and normalized for the housekeeping gene *GAPDH*. n=3, biological replicates, each of which includes
770 two technical replicates. Statistical significance was calculated using a Welch-corrected two-tailed t-
771 test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.
772

773 **Targeting cell-specific G4s with ATENA reveals transcription-dependent response to ligands.**
774 CUT&Tag and other chromatin-compatible G4-mapping methods, such as BG4-ChIP and Chem-Map,
775 have shown that genomic G4 distribution is cell-specific and predominantly located at promoters of
776 highly expressed genes [7,9,10,58]. Therefore, we decided to investigate biological responses attained
777 when directing a ligand towards previously unexplored MCF7-specific G4s. Specifically, we aimed to
778 determine whether the biological relevance of individual G4s and their response to ligand binding
779 correlate with the expression levels of the associated genes. We leveraged the existing dataset on G4-
780 distribution in MCF7 cells previously obtained using CUT&Tag[56]. This dataset identified a G4-peak
781 in the promoter of the highly expressed *HMGN1* gene - encoding for a non-histone chromosomal

protein able to interact with nucleosomes and regulate chromatin structure [59,60] - as unique to MCF7 cells compared to other cell lines[56]. We designed $sgRNA_{HMGN1-22}$ and $sgRNA_{HMGN1+34}$ to target the HMGN1-G4 at 22 and 34 base pairs, respectively, at its 3' and 5' ends, ensuring no overlap with known functional regions (Fig. 6a). After transfecting MCF7 cells expressing dCas9-Halo with $sgRNA_{HMGN1-22}$ and $sgRNA_{HMGN1+34}$, we incubated them with Cl-PDS_2 for 48 hours, as per the optimized ATENA protocol. Under these conditions, we measured a striking 99% reduction of *HMGN1* expression (0.01-fold) when targeting its G4 at the closest distance of 22 base pairs with $sgRNA_{HMGN-22}$ (Fig. 5b). *HMGN1* downregulation was partially attenuated when placing ATENA further away from the G4 with $sgRNA_{HMGN+34}$, consistent with the distance-dependent ligand engagement observed for other G4s (Fig. 6b). To place the observed *HMGN1* down-regulation in a biologically meaningful context, we examined the role of this gene in breast-cancer dormancy - an epigenetic-driven, non-replicative state, from which cancer cells "awaken", causing cancer relapse and resistance to therapy. In dormant MCF7 cells (estrogen-deprived), we inspected the epigenetic changes, chromatin accessibility, and transcriptional profile at the *HMGN1* promoter. Dormant cells exhibited loss of the active histone mark H3K4me3, gain of the repressive histone mark H3K27me3, and a corresponding drop in *HMGN1* expression; these epigenetic and transcriptomic changes were partially reversed upon cell-cycle re-entry - "awakening"[61] (Extended data Fig. 6a, 6b). Remarkably, ATENA-mediated stabilization of the HMGN1-G4 reproduced this repressive transcriptional state (Fig. 6b). Therefore, these convergent observations suggest that the HMGN1-G4 could act as an epigenetic switch: stabilizing the structure could indeed reinforce the *HMGN1* repressive state characteristic of dormancy and might be investigated as a strategy to maintain residual tumor cells in a dormant state.

Altogether, these findings indicated that targeting a cell-specific G4 located in the promoter of a highly transcribed gene can profoundly suppress gene expression, suggesting that maintaining G4-homeostasis at the promoter of highly transcribed genes is key to sustaining elevated expression levels and making these G4s particularly sensitive to ligands. Therefore, the varied transcriptional response observed upon G4-ligand treatment might reflect the relative relevance of different G4s in sustaining gene expression in the specific cell line studied.

To corroborate this hypothesis, we used ATENA to target a G4 present in the promoter region of a gene expressed at very low levels in MCF7 cells: *IL17RA*. Indeed, BG4 CUT&Tag performed in MCF7 cells revealed a distinct G4-peak in the promoter region of *IL17RA* [56], a gene that is only marginally expressed in this cell line. *IL17RA* encodes for the Interleukin 17 Receptor A, a proinflammatory cytokine secreted by activated T-lymphocytes and, therefore, not essential for breast cancer cell homeostasis. We generated $sgRNA_{IL17RA-20}$ to target the IL17RA-G4 at 20 base pairs from its 5' end and within a region that does not overlap with other regulatory elements of this promoter. Following transfection with $sgRNA_{IL17RA-20}$ or $sgRNA_{NTC}$ and incubation with Cl-PDS_2, we failed to detect any measurable changes in *IL17RA* expression levels (Fig. 5c). This observation indicates that targeting a G4 in a promoter of a transcriptionally inactive gene is not associated with gene-expression perturbation, linking tightly the functional relevance of G4s to the transcriptional levels of the genes associated. Considering that both G4s in the *HMGN1* and *IL17RA* promoters are equally detected in MCF7 by CUT&Tag [56] and targeted with similar sgRNA designs (within PAM sequence constraints), our findings suggest that the transcriptional levels linked to the targeted G4 can be used to anticipate the extent of transcriptional perturbation associated with ligand treatment. This model also explains the relatively modest transcriptional changes observed by the selective targeting of *c-MYC* and *PVT1* G4s, which are only moderately expressed in MCF7 cells. Increasing evidence in the literature suggests a model whereby G4s act as epigenetic factors to mark highly transcribed genes[12]. Our data support this model by showcasing how the extent of gene suppression/activation elicited by G4 ligands is linked to basal transcriptional levels, underscoring the relevance of maintaining G4 homeostasis in preserving transcriptional profiles characteristic of specific cell lines.

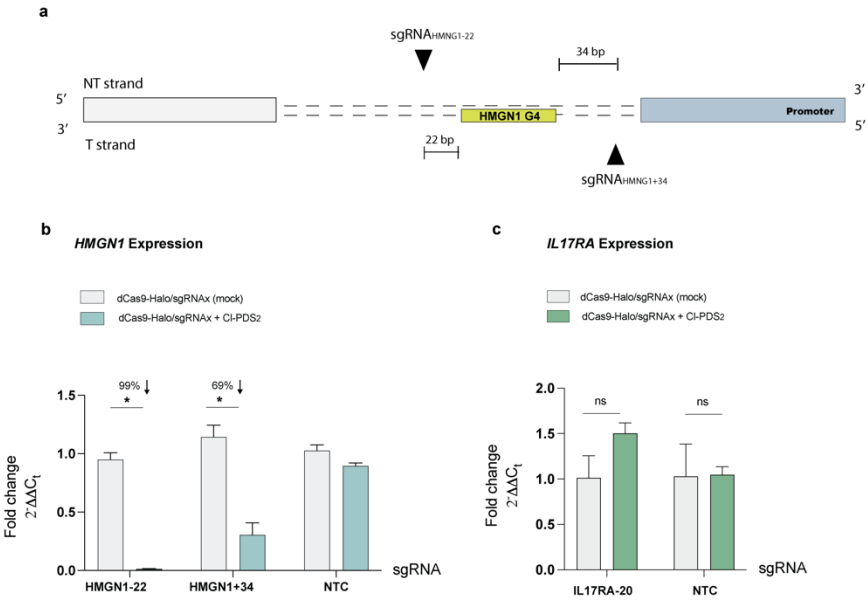


Figure 6: Targeting of *de novo* G4s with ATENA uncovers a transcriptionally dependent functional response. A schematic overview of the *HMGN1* promoter contains annotation of the predicted G4, two sgRNA designed to target HMGN1-G4 (black triangles), and their relative distance in bp from the G4-forming sequence. **b,** RT-qPCR for *HMGN1* expression in MCF7 cells stably expressing dCas9-Halo, transfected with either $sgRNA_{HMGN1-22}$, $sgRNA_{HMGN1+34}$ or sgRNA NTC and treated with (2.5 µM) Cl-PDS_2 for 48h after transfection. The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) and normalized for the housekeeping gene *GAPDH*; n=2, biological replicates, each of which includes two technical replicates. **c**, RT-qPCR for *IL17RA* expression in MCF7 cells stably expressing dCas9-Halo, transfected with either $sgRNA_{IL17RA-20}$ or sgRNA NTC and treated with (2.5 µM) Cl-PDS_2 for 48h after transfection. The expression values are represented as fold change ($2^{-\Delta\Delta Ct}$) with respect to the mock (DMSO-treated) and normalized for the housekeeping gene *GAPDH*; n=2, biological replicates, each of which includes two technical replicates. Data presented are the mean of n = number of independent biological samples. Statistical significance was calculated using a Welch-corrected two-tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.

851 **Discussion**

852

853 There is now substantial evidence to support that G4 structures form within endogenous chromatin
854 and that their formation is intimately linked to transcriptional activity [12]. For instance, G4s have been
855 detected in the promoter regions of key oncogenes in cancer cells, and recent studies have
856 demonstrated that they constitute critical structural features required to sustain high transcriptional
857 rate, such as in the case of c-*MYC* [20]. Over the past decades, the development of selective G4 ligands
858 has revealed that stabilizing these structures often results in suppression of oncogene expression [13].
859 Consequently, G4s have represented a therapeutically appealing target for decades. However, using
860 G4-ligands for clinical application has not yet gained traction, reflecting two intrinsic limitations.
861 Firstly, the recognition mechanism leveraged by most G4 ligands relies on end-stacking interactions,
862 which lack the selectivity to discriminate amongst different G4s - while effective at distinguishing
863 G4s from duplex DNA. Given that the prevalence of G4s in the genome is highly cell-type specific,
864 the lack of inter-G4 selectivity displayed by G4 ligands results in widespread transcriptional
865 perturbations and inconsistent phenotypes across cell models used. Secondly, the binding affinity of
866 individual ligands to different G4s varies broadly, making it difficult to pinpoint which G4s are
867 functionally responsible for the biological responses elicited by any given ligand.

868

869 To address these limitations, tools enabling single-G4 targeting have been considered essential to
870 unravel the fundamental biology regulated by specific G4s and to validate their therapeutic potential.
871 Although several locus-directed and ligand-based strategies for single-G4 targeting have been
872 reported[34,62-65], intrinsic limitations associated with these methods summarized in Table S13,
873 motivated us to develop ATENA- a CRISPR-guided platform in which catalytically inactive dCas9 is
874 chemically functionalized with G4 ligands.
875 This system allows the positioning of a ligand in proximity to a specific G4 of interest using a short
876 guiding RNA. We demonstrated that this approach enables transcriptional modulation attributable to
877 G4 engagement at a single genomic locus

878 We initially optimized conditions for single-G4 targeting *in vitro*, before applying ATENA in cells to
879 investigate the transcriptional role of a G4-structure located in the promoter region of the proto-
880 oncogene *c-MYC*. While independent studies have previously reported *c-MYC* suppression upon
881 treatment with G4-ligands[53], it remains unclear whether this effect is exclusively mediated by
882 engagement of ligands with the MYC-G4 or from broader transcriptomic changes induced by global
883 G4-stabilisation. Indeed, Hurley and co-workers – who initially proposed that *c-MYC* downregulation
884 was exclusively attributed to the targeting of MYC-G4[53] - challenged the previous proposed model
885 suggesting that transcriptional suppression was more likely a response to global G4-stabilization[18].
886 Using ATENA, we observed that G4-mediated *c-MYC* suppression in MCF7 cells is minimal and
887 limited to P1-controlled transcription. These findings are not only in line with what has been shown
888 by recent genomic studies[20] but were also fully recapitulated when using the MYC-G4 selective small
889 molecule DC-34 [24]. This strongly indicates that the biological response obtained by selective MYC-
890 G4 targeting is limited to P1-mediated transcription, regardless of the targeting approach.
891
892 Importantly, we have noted that the MYC-G4 lies in proximity of key regulatory regions of the
893 promoter, including the P1 promoter and the TATA-box sequence. Therefore, using CRISPR-based
894 tools for selective MYC-G4 targeting can easily lead to misleading results when using sgRNAs
895 targeting those regions and placing small molecules near these key regulatory elements. For instance,
896 a recent study reported global *c-MYC* downregulation when using either dCas9-Nucleolin fusion or
897 dCas9 poly-labeled with ten G4-ligands in tandem[34]. However, this effect was observed when
898 targeting the same region as our $sgRNA_{MYC+58}$, which places the dCas9 protein 9-bp apart from the
899 TATA-box and we demonstrated to cause unspecific *c-MYC* downregulation (Fig. 2d). These findings
900 indicated that placing ligands near core promoter elements can lead to transcriptional changes
901 unrelated to G4-stabilization and that careful design of sgRNAs is required when using dCas9-based
902 tools to avoid false positives. Additionally, our data suggested that the use of multiple ligands on a
903 single dCas9 protein to induce transcriptional perturbation is likely leading to false positives by
904 inducing local overcrowding at promoters [34]. Similarly, using multiple sgRNAs on a single target can

905 perturb the homeostasis of regulatory elements in a ligand-dependent manner, without necessarily
906 reflecting G4-specific effects, and should thus be avoided [34].
907
908 Using the selective G4-antibody BG4, we confirmed that the P1-specific *c-MYC* downregulation
909 induced by ATENA is accompanied by reduced protein accessibility at the MYC-G4 site, supporting
910 the notion that G4 engagement underlies the observed transcriptional suppression by hampering G4-
911 protein interactions. Notably, a similar effect was observed with DC-34, further suggesting that
912 reduced BG4 binding, rather than the previously reported enhancement[34], is to be expected when the
913 ligand successfully engages the G4 within the *c-MYC* promoter. This also provides evidence that
914 selective G4 targeting at the *c-MYC* promoter can be leveraged therapeutically to suppress its
915 expression by interfering with protein-G4 interactions. However, this effect is limited to *c-MYC*
916 transcription mediated by the P1 promoter and, therefore, heavily dependent on the cellular system
917 investigated [24]. For example, the substantial *c-MYC* downregulation, observed upon treatment with
918 DC-34 in multiple myeloma cells[24], is consistent with the high levels of P1-driven *c-MYC* expression
919 characteristic of these cancers [,67], while it is ineffective in other cells that are less reliant on the P1
920 promoter for *c-MYC* expression (i.e., MCF7).
921
922 ATENA further enabled us to investigate the transcriptional response elicited when targeting the iM
923 structure, also present within the *c-MYC* promoter. By decorating ATENA with an iM-selective
924 peptide recently developed by the Waller's group (RSV)[33], we also observed transcriptional changes
925 limited to the P1-promoter. However, when targeting the iM, we measured an increase of P1-driven
926 transcription, rather than a reduction, suggesting that distinct DNA secondary structures might
927 differently affect the expression at the same promoter. Moreover, this demonstrated the modularity of
928 ATENA, which can be easily adapted to target any DNA structure of interest using the same design
929 principles.
930
931 ATENA also enabled us to explore ligand-dependent variation in responses when targeting the same
932 G4-structure. It is well established that G4-ligands can cause either transcriptional activation or

933 repression depending on the G4-associated promoter, and this variability has often been attributed to
934 context-dependent biological roles of the specific G4 investigated. However, this model contradicts
935 genomic studies that strongly indicate a global association of G4-formation with transcriptional
936 activation rather than a context-dependent function of these structures. Using ATENA in combination
937 with CUT&Tag, we could demonstrate that the variation in gene expression responses at specific G4
938 sites stems not from inherent differences in G4 function, but rather from how structurally different
939 ligands affect protein-G4 interactions.
940 We tested this by characterizing the different transcriptional responses observed upon targeting the G4
941 in the promoter of the long non-coding RNA *PVT1* when using two established G4 ligands: PDS and
942 PhenDC3. The previously observed upregulation in *PVT1* transcription following PDS treatment[43]
943 was recapitulated with ATENA. This transcriptional increase was accompanied by increased BG4-
944 binding at this G4-site, as quantified by BG4 CUT&Tag qPCR, suggesting that PDS binding might
945 enhance local protein accessibility, acting as a molecular glue and, therefore, stimulate transcription.
946 In contrast, PhenDC3 treatment resulted in transcriptional suppression - consistent with the more
947 commonly reported effects of G4 ligands- demonstrating that structurally distinct ligands can drive
948 divergent transcriptional responses at the same G4 site.
949 Notably, these ligand-dependent responses were also observed when using the free - not bound to
950 dCas9 - PyPDS or PhenDC3, indicating that the lack of inter-G4 selectivity does not necessarily
951 prevent these molecules from providing meaningful information on the specific G4 site. Taken
952 together, these findings suggest that G4-ligands - while useful tools for perturbing G4-homeostasis
953 and investigating the consequent biological responses - cannot be used to directly infer the native
954 biological roles of G4s,but should instead be used to gain insights into the responses triggered by their
955 binding to these structures. Moreover, our data strongly indicate that it is not safe to assume that
956 different G4-ligands will lead to similar biological responses, which is a common assumption often
957 reported in the literature.
958
959 Finally, we utilized ATENA to target uncharacterized G4s previously identified in MCF7 cells with
960 CUT&Tag but never investigated their biological functions. Given that G4s are typically found at

961 active promoters, we reasoned that a gene's transcriptional status can strongly influence its
962 responsiveness to G4-ligand targeting. We took advantage of ATENA to direct PyPDS to a G4-
963 promoter of either a highly expressed gene (*HMGN1*) or a low-expressed one (*IL17RA*). Single
964 targeting of the G4 located in the *HMGN1* promoter - unique to MCF7 cells and not detected in other
965 cancer cell lines[56] - led to the higher transcriptional suppression measured in MCF7 (a 99%
966 reduction), confirming a strong link between transcriptional activity and ligand sensitivity that
967 indicates a substantial contribution of the HMGN1-G4 in sustaining elevated expression levels of
968 *HMGN1* in breast cancer cells.
969 We recently observed that PDS treatment in chemo-resistant ovarian cancer cells resensitizes them to
970 chemotherapy, due to the enrichment of G4s in highly transcriptionally active genes that are essential
971 to establish a chemo-resistant state, such as *WNT*[68]. Given the relevance of epigenetic changes in
972 establishing dormancy in breast cancer[61] and our observation linking expression levels of *HMGN1*
973 with dormancy/awakening transition (Extended Data Fig. 5), targeting of the HMGN1-G4 with
974 ligands might offer scope for the development of a therapeutic strategy to prevent awakening of
975 dormant cells and cancer relapse.
976 Altogether, these results support a model in which the biological significance of G4 structures within
977 the genome is contingent on the epigenetic landscape and the gene transcriptional status in a given
978 cell line.
979
980 In conclusion, ATENA provides a robust method for targeting single G4s within the genome of living
981 cells using distinct G4-ligands. This technology allowed us to dissect the diverse biological responses
982 elicited by different ligands at the same G4 site, as well as the effects of targeting various G4s with
983 the same ligand with high precision. We found that the chemical nature of the ligand used can perturb
984 the local protein-binding homeostasis of promoter G4s differently, leading to either increased protein
985 accessibility and transcriptional activation or decreased accessibility and transcriptional suppression.
986 These effects can be monitored directly using BG4 CUT&Tag qPCR, providing an assessment of
987 protein accessibility at the targeted G4 as a response to ligand binding. We demonstrated that some

988 ligands could act as molecular glue of G4-protein interactions or as displacers, depending on the
989 specific G4-targeted. This can be valuable for developing therapeutic agents based on G4-targeting,
990 which can be tailored chemically to either diminish or amplify transcription based on their binding
991 modalities. Moreover, ATENA enabled us to disentangle local effects triggered by individual G4-
992 targeting from broader responses driven by global G4-stabilization obtained when using canonical
993 G4-ligands. Interestingly, our data suggest that G4-ligands - despite their lack of inter-G4 selectivity -
994 often recapitulate the transcriptional changes observed with selective targeting of G4s and can be used
995 to infer the therapeutic potential of individual structures. Indeed, the response to G4-ligand treatment
996 is mainly shaped by the intrinsic level of transcription of the targeted promoter, making the ligand
997 responses highly epigenetic and cell-dependent.
998
999 We envisage that further development of ATENA can be leveraged to generate genome-wide
1000 screening of G4-functional response to various ligands in different cell lines. Additionally, this
1001 technology can be used to generate Halo-functional ligands and screen their ability to perturb protein
1002 accessibility at a given G4 site. We also anticipate that - given the minimal perturbation of ligands
1003 required for Halo-functionalization (chloroalkane) - ATENA is potentially a promising platform for
1004 gene-selective localization of other DNA-interacting therapeutics, including Topoisomerase inhibitors
1005 or cross-linking agents.
1006
1007
1008
1009
1010
1011
1012
1013 **Methods**

1014 Comprehensive synthetic protocols and purification methods for Cl-PDS$_n$, Cl-PhenDC3$_n$, and Cl-pep-
1015 RVS$_n$ ligands are provided in the Supplementary Information (SI1).

1016 **FRET-melting assay**

1017 Before each experiment, a solution containing 20 µM DNA ordered from Integrated DNA
1018 Technologies (IDT) (Table S1), labeled at 5'- with FAM and at 3'- with TAMRA, was freshly prepared
1019 in 10 mM lithium cacodylate buffer, pH 7.4 supplemented with 20 mM KCl. For the non-G4-forming
1020 sequence dsDNA26 (26 mer), the solution was supplemented with 150 mM KCl. The DNA solution
1021 was placed on an Eppendorf Thermomixer, annealed by heating at 95°C for 10 minutes, and then
1022 cooled at a rate of 0.5°C/minutes. The DNA was aliquoted into a 96-well RT-PCR plate (Thermo
1023 Fisher Scientific) and supplemented with increasing equivalent amounts of each ligand, resulting in a
1024 final oligonucleotide concentration of 0.4 µM. Fluorescence readings were performed on an Agilent
1025 Stratagene MX3000P, in the range of 25-95°C, in technical duplicate. The data points obtained were
1026 analyzed with GraphPad Prism 10, and the melting temperature was extrapolated from a Boltzmann
1027 sigmoidal function. Statistical significance was calculated using a two-tailed t-test in GraphPad Prism.
1028 p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.
1029

1030 **CD-melting assay**

1031 Circular Dichroism (CD) melting experiments were conducted on an AVIV Biomedical Inc. (Lakewood,
1032 NJ, USA) Model 410 Circular Dichroism Spectrometer. Before each experiment, c-MYC-Pu22 (Table
1033 S1) was annealed at a concentration of 2 µM in lithium phosphate buffer (pH 7.4, 20 mM phosphate)
1034 supplemented with 1 mM KCl. When required, the ligand was added to the annealed solution and
1035 equilibrated for 1h before initiating the melting process.
1036 Each experiment was performed in triplicate using a 10 mm optical path length quartz cuvette, with a 2
1037 mm bandwidth. The temperature was ramped from 30°C to 95°C, with scans recorded every 5°C after
1038 a 2 min equilibration time. CD signals were measured at 264 nm (corresponding to the parallel G-
1039 quadruplex maximum).

1040 The CD values (mdeg) at 264 nm were plotted as a function of temperature, normalized, and fit to a
1041 Boltzmann sigmoidal equation using GraphPad Prism 10.0. This analysis yielded the melting
1042 temperature (V50) of the G4 structure in the presence of increasing ligand equivalents. Statistical
1043 significance was calculated using a two-tailed t-test in GraphPad Prism. p-value: ns > 0.05, * ≤0.05,
1044 ** ≤0.01, *** ≤0.001, **** ≤0.0001.
1045

1046 **UV-binding assay**

1047 Oligonucleotides c-MYC-C52 and c-MYC-G52, purified by reverse-phase HPLC, were purchased from
1048 Eurogentec (Table S1). The oligonucleotide stock solutions were dissolved in MilliQ water and diluted
1049 to a final concentration of 250 µM in stock buffer (10 mM sodium cacodylate, pH 6.6) and annealed by
1050 heating at 95 °C for 5 min and cooling to room temperature overnight. The peptide was dissolved to
1051 10 mM in DMSO, and 1 mM in DMSO was used in the experiment. UV-based binding curves were
1052 determined using the wavelength at 350 nm for 10 µM pep-RVS or 10 µM Cl-pep-RVS_4 in stock buffer.
1053 DNA (250 µM in stock buffer) was added in increments up to 50 µM final concentration.

1054 **dCas9-Halo purification**

1055 The expression vector for dCas9-Halo (Addgene #72269) was introduced into BL21 cells (D3, NEB-
1056 C2527H) by heat shock. Bacteria were then plated in LB agar plates and selected with Ampicillin
1057 (100 µg/ml). One bacteria colony was picked and grown overnight in a 50 ml flask at 37°C. Overnight
1058 precultures were diluted 1:100 into the main culture (LB media, Ampicillin 100 µg/ml) and incubated
1059 at 37°C, 200 rpm until an OD600 of 0.6. The expression of the recombinant dCas9-Halo was induced
1060 by using 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). Upon induction, the temperature
1061 was reduced to 16°C for dCas9-Halo expression. BL21-dCas9-Halo cells were harvested after
1062 overnight incubation at 16°C by centrifugation (3500 g, 20 minutes, 4°C). Cells were lysed in lysis
1063 buffer (50 mM sodium phosphate, 300 mM NaCl, pH 7.2, 20% glycerol, 0.1% Tween-20, 1 mM
1064 TCEP, 1 mg/ml lysozyme, 1 mM PMSF, 1 mM $MgCl_2$, and 5U/g of culture Benzonase). Cell lysates
1065 were then homogenized, and cell debris was removed by centrifugation (16000xg, 15 minutes, 4°C).
1066 The soluble fraction of the lysate was filtered and then loaded onto a cobalt column, and the bound

1067 protein was eluted with 150 mM imidazole. After imidazole removal and concentration by using a
1068 50,000-MWCO centrifugal filter (Millipore Amicon), fractions were loaded into a 1-ml HiTrap SP XL
1069 (GE Healthcare), ÄKTA pure 25 for cationic exchange chromatography, and elution was performed
1070 with a gradient from 0% to 70% of Buffer (50 mM HEPES pH 7.5, 1 M KCl, 1mM TCEP). Purest
1071 fractions were pooled and then used for gel filtration chromatography. Gel filtration was performed
1072 using a Superdex 200 10/300 GL column, and isocratic elution was performed in a buffer containing
1073 50 mM HEPES (pH 7.5), 150 mM KCl, and 1 mM TCEP. Protein aliquots in storing buffer (50 mM
1074 HEPES, pH 7.5, 150 mM KCl, 20% glycerol) were snap-frozen and stored at -80°C.

1075 **BG4 purification**

1076 BL21 *E. coli* competent cells (D3, NEB-C2527H) were transformed with a pSANG10-3F-BG4
1077 plasmid (Addgene #55756). Single colonies were inoculated in 2xYT medium supplemented with 2%
1078 glucose and 50 µg/ml kanamycin and incubated at 30 °C at 250 rpm overnight. The starting culture
1079 was diluted 1:500 into ZYM-5052 autoinduction medium prepared following the previously described
1080 method[69] supplemented with 50 µg/ml kanamycin. The culture was then incubated at 37 °C for 6
1081 hours at 250 rpm, followed by an overnight incubation at 25 °C at 280 rpm. The culture was
1082 centrifuged at 4,000 x *g*, 4 °C for 30 minutes, and the collected pellet lysed in 80 ml ice-cold TES
1083 buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 20% sucrose) and incubated on ice for 10
1084 minutes. This was followed by 15 minutes of incubation in ice-cold TES (diluted 1:5), supplemented
1085 with 15 U/mL benzonase, 2 mM MgSO4, and a protease inhibitor cocktail. The lysate was centrifuged
1086 at 16,000 x g for 20 minutes. The supernatant was collected and incubated with Nickel resin (Sigma,
1087 P6611), previously equilibrated in PBS, at 4 °C for 1 hour in rotation. The complex was washed three
1088 times with wash buffer (PBS, 100 mM NaCl, 10 mM imidazole). Then, BG4 was eluted in 4 mL of
1089 elution buffer (PBS, pH 8.0, 250 mM imidazole) and dialyzed against PBS at 4 °C using GeBaflex
1090 tubes (Generon). Purified BG4 was snap-frozen in liquid nitrogen and stored at -80 °C.
1091

1092 **dCas9-binding competition assay**

1093

1094 dCas9-Halo purified protein (4 µM) was incubated with several dilutions of Cl-PDS$_n$ or Cl-PhenDC3$_n$
1095 probes or with commercially available Halo-TAMRA for the positive control (Promega, G8252) for
1096 30 minutes at room temperature in PBS. Pretreated dCas9-Halo proteins were then incubated with 5
1097 µM Halo-TAMRA, or Cl-PDS$_n$ for the positive control, and samples were loaded in an SDS-PAGE
1098 gel (8% polyacrylamide) and run at 180V for 45 minutes. Gels were then imaged using a Typhoon
1099 FLA 9500 (GE) with the TAMRA filter set (excitation at 542 nm) and then stained with Coomassie.
1100 Images of Coomassie-stained gels were then acquired with ImageQuant LAS 4000 (Cytiva).
1101 Each band intensity was then quantified using Image Studio software and normalized for the
1102 corresponding Coomassie signal, expressed as a percentage (with 100% labeling corresponding to the
1103 positive control). IV-CP$_{50}$ values were determined using nonlinear regression (dose-response
1104 inhibition curves with constrained fitting) in GraphPad Prism with n = (number of independent
1105 experiments) = 2.
1106
1107 **In vitro transcription for sgRNA synthesis**
1108
1109 Oligonucleotides containing sgRNA sequences (Table S2) were ordered from Integrated DNA
1110 Technologies (IDT), and a PCR was set up to generate the corresponding duplex following the
1111 previously described method[70]. Following the manufacturer's instructions, the DNA product was
1112 transcribed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB, E2040S). The resulting
1113 sgRNA was purified using the RNeasy Mini kit (74104) and stored at -80°C.
1114
1115 **dCas9-Halo_FRET assay**
1116
1117 Before the experiment, Cy5-3' end-labeled and Cy3-5' end-labeled oligos (Table S1) were annealed in
1118 10 mM Tris HCl pH 7.5, 100 mM KCl at a final concentration of 1 µM (95 °C for 10 minutes and left
1119 for overnight cooling down at room temperature). dCas9-Halo purified protein at a final concentration
1120 of 4 µM was incubated with Cl-PDS$_n$ probes in a 1:2.5 ratio for 45 minutes in binding buffer (20 mM
1121 HEPES, pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 1 mM TCEP, 10% glycerol). The pretreated dCas9-Halo

1122 was then incubated with 120 pmol of purified sgRNA for 20 minutes at room temperature, followed
1123 by incubation with 200 nM of pre-annealed oligos at 37°C for 1h. The complexes were then loaded
1124 into an 8% polyacrylamide native gel and ran for 120 minutes at 4°C in 1x TBE supplemented with 2
1125 mM MgCl2. Gels were then imaged using a Typhoon FLA 9500 (GE) with the Cy3 and Cy5 filter
1126 sets. Analysis of the fluorescent band intensity (I) and relative FRET efficiency (E) was performed
1127 using a Python code. Particularly, FRET efficiency was calculated as: $E=I_{Cy5}/(I_{Cy3}+I_{Cy5})$. The signals
1128 in both channels were normalized for the background and the sgRNA NTC control, and for every
1129 sgRNA, the ΔFRET is defined as: (E+ligand) - (E−ligand). Statistical significance was calculated
1130 using a two-tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, ****
1131 ≤0.0001 with n = (number of independent experiments) = 2.
1132
1133
1134 **Cloning of lentiviral dCas9-Halo construct**
1135
1136 The lentiviral dCas9-Halo plasmid was cloned using Gibson assembly. A lentiviral backbone
1137 containing already dCas9 (Addgene #61425) was digested using BamHI and BsrGI (New England
1138 Biolabs, R3575S, R3136S) and then assembled with the HaloTag sequence amplified (Table.S3) from
1139 pET302-6His-dCas9-Halo (Addgene #72269) in a 1:3 molar ratio of backbone: insert using HiFi DNA
1140 Assembly Mix (E2621S, New England Bioscience) following manufacturing protocols. Post
1141 incubation, assembled products were diluted with water, and 5 µL of the product was transformed by
1142 heat shock into 10-beta competent cells (New England Biolabs, C3019I). Cells were then plated on
1143 agarose plates (supplemented with ampicillin 100 µg/mL) for overnight outgrowth at 37 °C. Single
1144 clones were picked and grown in 5 mL of Amp LB media overnight at 37 °C while shaking. Plasmid
1145 DNA was purified from cells using the Promega PureYield plasmid miniprep system (Promega,
1146 A1223) and sequenced using the Genewiz service.
1147 The positive plasmid was further modified by substituting the blasticidine resistance gene with the
1148 mCherry coding sequence using Gibson assembly after digestion with BsrGI and EcoRI enzymes
1149 (Table S3). Positive clones were then sequenced using the Genewiz service.

1150

1151 **Cloning of sgRNA in pLG1 plasmid**

1152

1153 The pLG1 backbone (Addgene #109003) was digested with BstXI and BlpI (FastDigest,
1154 ThermoFisher) for 1 hour at 37°C. Oligos containing the sgRNA sequence (Table S4) were ordered
1155 from IDT and annealed in water at 10 µM in a thermocycler (37°C, 30 minutes- 95°C, 5 minutes and
1156 ramp down 5 degrees/minute to 25°C). The annealed oligos were diluted (1:50) and then assembled
1157 with the digested plasmid in a 1:2 molar ratio of backbone: insert using HiFi DNA Assembly Mix
1158 (E2621S, New England Bioscience) following the manufacturer's protocols. Post-incubation, the
1159 assembled products were diluted 1:2 in water, and 5 µL of the product was transformed by heat shock
1160 into 10-beta competent cells (New England Biolabs, C3019I). Cells were then plated on agarose
1161 plates (supplemented with 100 µg/mL ampicillin) for overnight outgrowth at 37 °C. Single clones
1162 were picked and grown overnight in 5 mL of Amp LB medium at 37 °C in a rotating shaker. Plasmid
1163 DNA was purified from cells using the Promega PureYield plasmid miniprep system (Promega,
1164 A1223) and sequenced using the Genewiz service.

1165

1166 **Mammalian cell culture**

1167

1168 HEK293FT cells were cultured in DMEM GlutaMAX™ (Thermo Fisher Scientific, 10566016)
1169 supplemented with 1x penicillin-streptomycin (Thermo Fisher Scientific, 15140122) and 10%
1170 (vol/vol) fetal bovine serum (FBS) and maintained at 37 °C and 5% CO2.
1171 MCF7 cells were cultured in DMEM GlutaMAX™ (Thermo Fisher Scientific, 10566016)
1172 supplemented with 1x β-Estradiol (Sigma, 50-28-2), 1x penicillin-streptomycin (Thermo Fisher
1173 Scientific, 15140122), and 10% (vol/vol) fetal bovine serum (FBS) and maintained at 37 °C and 5%
1174 CO2.

1175

1176 **Lentivirus production and transduction**

1177 For lentiviral production, HEK293T cells were seeded at $3.8x10^6$ in a 10 cm tissue culture plate the
1178 day before transfection. 8.4 µg of the envelope plasmids pCMV-VSV-G (Addgene #8454) and 6.4 ug
1179 of packaging plasmid R8.74 (Addgene #22036) were co-transfected along with 2.1 µg of the target
1180 plasmid (dCas9-Halo-T2A-mCherry) using polyethyleneimine (PEI-Linear, MW 25000) in a µg
1181 DNA: ug PEI ratio of 1:3. Viral supernatant was harvested after 48h and 72h, and before usage, it was
1182 filtered using a 0.45 µm filter unit. MCF7 cells were plated on 6-well plates the day before infection.
1183 They were infected with lentiviruses in DMEM 10% FBS, β-Estradiol, in the presence of polybrene
1184 with a final concentration of 10 µg/mL. Half of the media was changed the day after, and cells were
1185 grown for one week and examined by flow cytometry (Attune NxT, ThermoFisher) to confirm
1186 successful transduction (YL2-Channel). After genotyping (Table S3), cells were single-cell sorted
1187 using FACS BD FACS Diva 9.0.1 (Fig.SI-3).

1188

1189 **Western blot**

1190 MCF7 stably expressing dCas9-Halo cells were harvested and processed using 1x RIPA buffer
1191 (Merck,20-188) supplemented with 1x Protein inhibitor cocktail. The total protein concentration of
1192 the cleared lysate was then measured by Bradford assay (Pierce™ Bradford Plus Protein Assay Kits,
1193 Thermo Fisher Scientific, 23236) at 595 nm using a ClarioStar plate reader. A total of 20/25 µg
1194 protein was then loaded into an 8% polyacrylamide SDS-PAGE gel and transferred to a pre-assembled
1195 PVDF membrane (Trans-Blot Turbo Mini 0.2 µm PVDF Transfer Packs, Bio-Rad) using a semi-dry
1196 method (Bio-Rad). The membrane was blocked (5% milk in TBS-T) and incubated with primary
1197 antibodies (1:1000 Cas9 antibody- 14697T, 1:1000 GAPDH antibody- 2118T, Cell Signaling
1198 Technology) overnight at 4 °C. The membrane was then washed in TBS-T and incubated with
1199 secondary antibodies (1:10,000) goat anti-rabbit/mouse HRP-Advansta, R-05071-500, R-05072-500)
1200 for 1h at room temperature. After three washing steps of the membrane in TBS-T, 0.5 ml of HRP
1201 substrate (Merck, WBLUC0100) was added to the top of the membrane, and the excess was removed.
1202 The signal was developed using Image Quant LAS 4000 (Cytiva) with the following settings:
1203 chemiluminescence for the WB signal and Cy5 for visualizing the protein marker.

1204

1205 **CAPA assay**

1206 MCF7 cells stably expressing dCas9-Halo were plated the day before at 30x10^3 cells/ well in 96-
1207 well plates pre-coated with Poly-D-Lysine (Thermo Fisher Scientific, A3890401). Cells were
1208 incubated for 2h in the presence of serial dilutions of Cl-PDS$_n$, Cl-PhenDC3$_n$,or Cl-pep-RVS$_n$ probes
1209 ranging from 0.2µM to 10 µM (5% CO2, 37 °C). Cells were then washed two times with media (every
1210 washing step included 10 minutes of incubation, 5% CO2, 37 °C) and incubated with HaloTag®
1211 Oregon Green® Ligand (Promega, G2801) for 45 minutes, followed by two washing steps (10
1212 minutes,5% CO2, 37 °C). Lastly, cells were washed with DPBS, treated with 0.25% trypsin, and
1213 resuspended in FACS buffer (5% BSA in DPBS) before undergoing flow cytometry analysis (Attune
1214 Nxt). Data analysis was performed in FlowJo 10.9.0. Mean fluorescence values from two biological
1215 replicates (each with three technical replicates) were normalized to the positive-control signal and
1216 expressed as percent labeling. Following this, CP_{50} values were obtained by nonlinear regression
1217 (dose–response inhibition curves with constrained fitting) in GraphPad Prism (n = 2).

1218

1219 **Transfection and incubation with G4 ligands**

1220 MCF7 cells stably expressing dCas9-Halo were plated the day before at 25x10^3 cells/ well in 96-
1221 well plates pre-coated with Poly-D-Lysine (Thermo Fisher Scientific, A3890401). According to the
1222 manufacturer's protocol, 50 ng of guide expressing plasmid was transfected using Lipofectamine 3000
1223 (Thermo Fisher Scientific, L3000001). The day after, cells were incubated with either DMSO (0.5%)
1224 or Cl-PDS$_{n,}$Cl-PhenDC3$_{n,}$ or Cl-pep-RVS$_n$ probes and incubated for 48h.

1225 **RNA isolation**

1226 RNA was harvested 72 hours post-transfection. Cells were washed with 100 µL of 1X DPBS
1227 (Gibco, 14190144) and incubated with RLT buffer from RNeasy mini kits (Qiagen, 74104)
1228 according to the manufacturer's instructions. Following the manufacturer's instructions, the eluted
1229 RNA was reverse-transcribed into cDNA using the RevertAid cDNA Synthesis Kit (K1621, Thermo
1230 Fisher Scientific).

1231

1232 **Quantitative Real-Time PCR (RT-qPCR)**

1233 RT-qPCR reactions were prepared as follows: cDNA was mixed with 500 nM of forward and reverse
1234 primer (Table S5) and 5µL of Fast SYBR Green Master Mix (Applied Biosystems: 4385612) in a final
1235 volume of 10 µL. RT-qPCR was performed on a 96-well plate format using an Agilent Stratagene
1236 Mx3000P machine with the following program: 95 °C for 20 seconds, 40× (95 °C for 3 s, 60 °C for 30
1237 s), 95 °C for 15 s, 60 °C for 1 minute, and 95 °C for 15 s (melting curve). The C_t values obtained were
1238 used to calculate the relative fold-change in gene expression using the $\Delta\Delta C_t$ method and normalized
1239 to the housekeeping gene *GAPDH* in both treated and untreated samples. Statistical significance was
1240 calculated using a Welch-corrected two-tailed t-test in GraphPad Prism; p-value: ns > 0.05, * ≤0.05,
1241 ** ≤0.01, *** ≤0.001, **** ≤0.0001.

1242

1243 **BG4 CUT&Tag**

1244 Cells were plated in a 6-well plate format one day before transfection, following the previously
1245 described transfection protocol. 24h after transfection, cells were incubated with Cl-PDS_n probes for
1246 6h and then harvested, counted, and inspected for viability. To prepare nuclei, 200k cells per reaction
1247 (plus 10% excess) were centrifuged for 3 minutes at 600×*g* at room temperature and then resuspended
1248 in nuclei extraction buffer (20 mM HEPES–KOH, pH 7.5, 10 mM KCl, 0.1% Triton X-100, 20 mM
1249 Glycerol, 0.5 mM Spermidine, 1X Roche cOmplete Mini EDTA-free Protease Inhibitor (Roche,
1250 11836170001) and incubated for 10 minutes on ice. Nuclei were spun at 600 x g for 3 min at 4 °C and
1251 then gently resuspended in 200 µL/per reaction cold Nuclei Extraction Buffer. A 2 µL aliquot was
1252 taken and mixed with 2 µL trypan blue to examine nuclei integrity under the microscope. In the
1253 meantime, ConA beads (BioMag®Plus Concanavalin A, BP531) were prewashed and equilibrated in
1254 binding buffer (20 mM HEPES–KOH, pH 7.5, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). Nuclei for
1255 each condition that needed to be tested were mixed with 10 µL of beads and incubated for 10 minutes
1256 at RT while shaking. After checking the binding of the beads to the ConA beads under the microscope
1257 (2 µL sample + 2 µL trypan blue), the samples were washed two times with 300 µL of 1%BSA
1258 antibody buffer (20 mM HEPES–KOH, pH 7.5, 150 mM KCl,0.5 mM Spermidine, 1X Roche

1259 cOmplete Mini EDTA-free Protease Inhibitor, 5% Digitonin, 2mM EDTA, and 1%BSA). The tubes
1260 were placed on the magnetic rack, and the liquid was withdrawn. Samples were then resuspended in
1261 50µL/ reaction of 1%BSA antibody buffer and incubated for 1h at RT while shaking. The samples (3
1262 technical replicates for each condition) were incubated with 216 nM of an in-house prepared BG4
1263 antibody and 0.5 µL of H3K27me3 (Cell Signaling Technology,9733S) antibody for the positive
1264 control (1 replicate for each condition) and left overnight at 4°C while shaking.

The day after, samples were checked to ensure that they were bound to the beads and then washed two times with Dig-wash buffer (20 mM HEPES–KOH, pH 7.5, 150 mM KCl,0.5 mM Spermidine, 1X Roche cOmplete Mini EDTA-free Protease Inhibitor, 5% Digitonin). Samples containing BG4 and samples for the negative control were resuspended in 50 µL of Dig-wash buffer containing 2 µL of anti-flag antibody (DYKDDDDK Tag Antibody, 2368S) and incubated for 1h at RT while shaking. Then, the BG4 and negative samples were washed three times in 100 µL Dig wash buffer, resuspended in 50 µL Dig wash buffer supplemented with 0.5 µL of tertiary antibody (anti-rabbit, Abcam, ABIN101961) and incubated for 1h at RT while shaking. Samples were then washed three times with 100µL Dig-wash buffer, resuspend in 50 µL Dig-300 buffer (20 mM HEPES–KOH, pH 7.5, 300 mM KCl, 0.5 mM Spermidine, 0.01% Digitonin, 1X Roche cOmplete Mini EDTA-free Protease Inhibitor) containing pA-Tn5 (CUTANA™ pAG-Tn5 for CUT&Tag, 15-1017) adapter complex (1:20) and incubated for 1h at RT while shaking. Samples were then washed three times in Dig-300 buffer, resuspended in 100 µL of Tagmentation buffer (Dig-300 buffer supplemented with 10 mM $MgCl_2$) and incubated for 1h at 37°C. Samples were then, washed twice with TAPS buffer (10 mM TAPS buffer- J63268.AE, Thermo Scientific Chemicals, 0.2 mM EDTA), mixed by vortexing in 100 µL of Protenaise K buffer (0.5mg/mL Proteinase K, Thermo Fisher Scientific EO0492/EO0491, 0.5% SDS in 10mM Tris-HCl buffer pH 8.0) and incubate 1 hr at 55 ºC. Samples were then purified with Zymo DNA Clean & Concentrator-5 (D4013) and samples were then, amplified using NEBNext HiFi 2x PCR Master mix (NEB, M0541) with the following set up (Initial 72 °C for 5 minutes, 20X cycle (98 °C for 30 sec, 98 °C for 10 sec, 63 °C for 10 sec) and 72°C for 1 minutes). Samples were then purified using Ampure XP beads (A63880) according to the

manufacturer's instructions. Quality control of the sample was performed using TapeStation ScreenTape HSD1000 (5067-5584) and TapeStation ScreenTape HSD1000 reagents (5067-5585) according to the manufacturer's instructions on a TapeStation (TapeStation Agilent 4150).

1265

1266 **BG4 CUT&Tag-qPCR**

1267 CUT&Tag libraries were diluted 1:10, and 2 µL of the diluted CUT&Tag library was mixed with 1
1268 µM of primer mix (Table S6) and 5 µL of SYBR Green PCR Master Mix (Applied Biosystems:
1269 4385612) according to the following protocol 20 seconds at 95 °C, 39 cycles of 10 seconds at 95 °C,
1270 30 seconds at 60 °C, and 10 seconds at 72 °C, and finally heating to 90 °C. The C_t values obtained
1271 were used to calculate the relative fold-change in gene expression using the $\Delta\Delta C_t$ method and assess
1272 the relative fold change at G4 target sites when compared against G4-positive regions (MAZ, RBBP4,
1273 RPA3). Statistical significance was calculated using a Welch-corrected two-tailed t-test in GraphPad
1274 Prism; p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.

1275

1276 **Histone modifications CUT&Tag**

1277 The CUT&Tag assay was performed using the EpiCypher CUTANA Direct-to-PCR CUT&Tag
1278 Protocol v1.7, which included all the mentioned materials and buffer recipes with minor
1279 modifications. Briefly, cryopreserved nuclei were thawed quickly and immobilized to Concanavalin A
1280 (ConA) Conjugated Paramagnetic Beads (Epycypher, cat. 21-1401). Bead-bound nuclei were spiked
1281 with 2 µL of spike-in SNAP-CUTANA™ K-MetStat Panel (1:50, Epicypher, cat. 19-1002) and
1282 immediately resuspended and incubated in Primary Antibody diluted in Antibody buffer at the
1283 manufacturer's CUT&Tag recommended dilutions, at 4°C by nutation overnight. The next day, the
1284 primary antibody was removed, and nuclei were incubated with either anti-Rabbit or anti-Mouse
1285 Secondary Antibodies at Room Temperature (RT) in Digitonin 150 buffer for 1h with nutation. Nuclei
1286 were washed twice with wash 150 buffer and incubated with CUTANA pAG-Tn5 enzyme (Epicypher,
1287 cat. 15-1017) diluted in Digitonin 300 buffer for 1h at RT with nutation and thoroughly rewashed
1288 twice. Nuclei were resuspended in Tagmentation Buffer containing MgCl2 and incubated 1 h at 37°C.

1289 The reactions were then washed with Post-tagmentation buffer and incubated in SDS Release Buffer
1290 at 58ºC for 1h to quench the tagmentation reaction. Then, SDS Quench Buffer is added to neutralize
1291 SDS, as well as NEBNext High-Fidelity 2 × PCR Master Mix (NEB, cat. M0541L), and 2.5 µM of
1292 universal i5 primer and unique i7 primer sequences previously published in Buenrostro et al. 2015
1293 (IDT technologies). Libraries were amplified at 14 cycles for abundant histone modification samples
1294 (H3K27me3), and 16 cycles for less abundant marks (H3K4me3) and negative control IgG following
1295 the cycling parameters outlined in the EpiCypher® CUTANA™ Direct-to-PCR CUT&Tag Protocol
1296 v1.7. Library clean-up was performed using 1.3x SPRI beads (Beckman Coulter, cat. B23319) to
1297 recover ~75 bp DNA fragments. The beads-DNA were incubated at RT for 5 minutes, followed by
1298 two washes with 85% ethanol. Libraries were eluted in 15 µL of 0.1x TE buffer, quantified using
1299 Qubit™ fluorometer per manufacturer's instructions, and quality controlled for fragment length
1300 enrichment using the Agilent TapeStation Bioanalyzer® with High Sensitivity D1000 reagents. The
1301 following antibodies were used for CUT&Tag: H3K4me3 (Epycypher, cat 13-0041, 1:50),
1302 H3K27me3 (Cell signalling, cat 9733T, 1:50), IgG (Epycypher, cat 13-0042, 1:1000), Anti-Rabbit
1303 Secondary Antibody (Epycypher, cat 13-0047, 1:100) and Anti-Mouse Secondary Antibody
1304 (Epycypher, cat 13-0048, 1:100). Libraries were sequenced at 10 million reads per sample on Paired-
1305 End platform of 150bp per fragment (PE150) by our sequencing provider, Novogene, on a NovaseqX
1306 platform from Illumina.
1307 CUT&Tag data was processed on the Imperial College London HPC using an adapted version of the
1308 pipeline developed by Dr Ye Xheng from the Steven Henikoff lab, available on
1309 https://github.com/clabanillas/cutnTag_processing. This pipeline includes quality control, adapter
1310 trimming, alignment, filtering, and peak calling. Peaks for H3K27me3 were called using SEACR with
1311 no built-in normalisation, against control IgG in stringent mode. Peaks for H3K4me3 were called with
1312 IgG control with MACS2 callpeak command (--keep-dup all --nomodel --shift -75 --extsize 150 -q
1313 0.01).
1314
1315

1316 **ATAC-seq**

1317 scATAC-seq was performed on a Chromium platform (10x Genomics) using Chromium Single Cell
1318 ATAC Reagent Kit' V1 chemistry (manual version CG000168 Rev C) and Nuclei Isolation for Single
1319 Cell ATAC Sequencing (manual version CG000169 Rev B) protocols. Nuclei suspensions were
1320 prepared to get 10,000 nuclei as target nuclei for recovery. Final libraries were loaded onto a Novaseq
1321 6000 platform (Illumina) to obtain 50,000 reads per nucleus with a read length of 2 × 50 bp.
1322 A pseudobulk of single-cell sequencing readout was prepared by concatenating all reads containing
1323 the transposase adaptor sequence barcode per sample. Data processing was done according to the
1324 following repository: https://github.com/harvardinformatics/ATAC-seq. Briefly, raw paired-end
1325 FASTQ files were subjected to quality control using FastQC, and adapter trimming was performed
1326 with FASTP (--detect_adapter_for_pe -l 20). Reads were then aligned to the human genome (T2T-
1327 CHM13v2.0) using Bowtie2 with parameters optimized for ATACseq (--very-sensitive -X 700).
1328 Mitochondrial reads were removed with Samtools. Duplicates were removed using PICARD (picard-
1329 2.27.4-0/picard.jar). Then multimapped reads were removed with samtools (-h -q 30), as well as
1330 unmapped, mate unmapped, not primary alignment, reads failing platform, duplicates (-h -b -F 1804),
1331 and properly paired reads -f 2 were retained. Reads aligned to blacklisted regions were removed.
1332 Coordinates were shifted with deeptools alignmentSieve command (--numberOfProcessors max –
1333 ATACshift). Normalised genome coverage tracks were generated with the BEDtools genomecov
1334 function (--numberOfProcessors max --binSize 10 --normalizeUsing RPGC --effectiveGenomeSize
1335 2786136059) using the effective genome size for T2T. Peaks were called using the MACS2 callpeak
1336 command (-q 0.01 --keep-dup all).

1337

1338 **RNAseq sample preparation and bioinformatic analysis**

1339 For transcriptome-wide gene expression changes in transfected samples, cells were plated the day
1340 before in 6-well plates, and 24h after transfection, they were incubated with either 0.5% DMSO or 2.5
1341 μM Cl-PDS$_2$ for 6 hours. For DC-34/PyPDS-treated samples, cells were plated in a 6-well plate the
1342 day before treatment with 5 μM and 2.5 μM of the compound for 6h. Cells were harvested and

washed with 1x DPBS (300xg, 5 minutes). RNA was then extracted using RNeasy mini kits (Qiagen, 74104) according to the manufacturer's instructions. The quality of the extracted RNA was assessed by TapeStation (High Sensitivity RNA Screen Tape, Agilent) and sent to Novogene service for sequencing. RNA libraries were prepared using Novogene kit for library preparation and then sequenced with Illumina NovaSeq X-plus. CASAVA (version 1.8) was used to perform base calling and Phred score. The quality of the reads was assessed with FastQC (Fastp v.0.23.1). Sequence alignment to the reference genome (GRCh38-NCBI:GCA_000001405.27) was performed using HISAT2. Sorted bam files were generated with (Hisat2 v.2.0.5), and gene counts using FPKM. Raw count data were then imported into DESeq2 (v1.40.2), and differential gene expression analysis was conducted with three comparisons: (1) the generic drug effect in control (NTC) cells treated with the ligand versus mock (DMSO), (2) the effects of the sgRNA ($sgRNA_{MYC-19}$) under mock conditions and data set (2) filtered for NTC. DEGs were identified based on adjusted p-values (FDR < 0.05) and log2 fold-change thresholds (|log2FC| ≥ 1).

Cell proliferation assay

MCF7 cells, wild-type or stably expressing dCas9-Halo were plated one day before the experiment at 24 x 10^3 cells/well in a 96-well plate. Cells were then treated 24h later with serial dilutions (ranging from 0.25 µM to 10 µM) of Cl-PDS_2, Cl-$PhenDC3_2$, Cl-PDC_2, PyPDS, PhenDC3, PDC, DC-34 or Cl-pep-RVS_n and 0.5% DMSO as a negative control. After 48h, cells were incubated with 20 µL/well of CellTiter96®AQ_{ueous}One Solution Reagent for 2 h (5% CO_2, 37 °C), before acquiring the absorbance value at 490 nm using a ClarioStar plate reader. The data were fitted to a sigmoidal dose-normalized response curve with a variable slope. Statistical analysis of curve-fit parameters was performed by independently fitting data from separate biological experiments, followed by comparison of the resulting curve-fit parameters using Extra sum-of-squares F test in GraphPad Prism. p-value: ns > 0.05, * ≤0.05, ** ≤0.01, *** ≤0.001, **** ≤0.0001.

Data availability

All new datasets have been deposited in the gene expression omnibus (GEO): GSE279769.

To review GEO accession GSE279769 Enter token "itwzseukjhwfjon" into the box: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE279769

Acknowledgments

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). S.P.N. acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) (EP/S023518/1). T.E.M acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) (EP/S023518/1). J.S.S. is supported by the Intramural Research Programs of the National Institutes of Health, National Cancer Institute (NCI), Center for Cancer Research, Project BC011585. D.G. was supported by a BBSRC grant (BB/W001616/1). This work was supported by a Leverhulme Trust Research Grant (RPG-2022-184). E.C. was supported by FWO (12B1923N). We are thankful to Prof Luca Magnani for providing MCF7 cells and for helpful discussions. The authors acknowledge Prof Ramon Vilar, Dr Michele Stasi, and Dr Koustav Pal for insightful discussions. The authors also acknowledge Dr Denise Liano, Dr Aisling Minard, Dr Souroprobho Chowdhury and Dr Anna Di Porzio for initial experimental efforts on this project.

Authors Contribution

M.D.A. conceived and supervised the research project. S.P.N. designed the research with M.D.A. and performed all the experiments unless otherwise specified. E.C, L.S.L., R.N., A.J.L., E.F., M.Z., T.E.M. have contributed to synthesizing the G4-ligands. E.C. and A.J.L. perform FRET melting assays. E.C. performed the synthesis and chemical characterization of the pep-RVSn ligands and CD melting experiments. S.G. performed with S.P.N. BG4 CUT&Tag experiments. C.S.C. and L.M. performed and analyzed ATAC-seq and CUT&Tag of histone marks. Z.W. and D.G. advised on i-motif-peptides and D.G. performed binding studies on the pep-RVSn ligands. G.F. helped with chromatin prep. J.S.S. conceived DC-34 and helped edit the manuscript. C.R.F synthesized DC-34 used in this study. M.D.A. wrote the manuscript supported by S.P.N with input from all authors. All authors contributed to critical discussion and data interpretation.

1407 **References**

1408 1 Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular
1409 chemistry. *Angew. Chem. Int. Ed. Engl.* **43**, 668-698 (2004).
1410 2 Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad.*
1411 *Sci. U. S. A.* **48**, 2013-2018 (1962).
1412 3 Parkinson, G. N., Lee, M. P. & Neidle, S. Crystal structure of parallel quadruplexes from human
1413 telomeric DNA. *Nature* **417**, 876-880 (2002).
1414 4 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of
1415 DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182-186 (2013).
1416 5 Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live
1417 cells. *Nat. Chem.* **12**, 832-837 (2020).
1418 6 Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence
1419 lifetime imaging microscopy. *Nat. Commun.* **12**, 162 (2021).
1420 7 Lyu, J., Shao, R., Kwong Yung, P. Y. & Elsasser, S. J. Genome-wide mapping of G-quadruplex
1421 structures with CUT&Tag. *Nucleic Acids Res.* **50**, e13 (2022).
1422 8 Esnault, C. *et al.* G4access identifies G-quadruplexes and their associations with open
1423 chromatin and imprinting control regions. *Nat. Genet.* **55**, 1359-1369 (2023).
1424 9 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat.*
1425 *Genet.* **48**, 1267-1272 (2016).
1426 10 Hansel-Hertsch, R. *et al.* Landscape of G-quadruplex DNA structural regions in breast cancer.
1427 *Nat. Genet.* **52**, 878-883 (2020).
1428 11 Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the
1429 human genome. *Nat. Biotechnol.* **33**, 877-881 (2015).
1430 12 Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex
1431 structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* **49**, 8419-8431
1432 (2021).
1433 13 Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters:
1434 a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10**, 261-275 (2011).
1435 14 Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising
1436 target for cancer therapy. *Mol. Cancer* **20**, 40 (2021).
1437 15 Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple
1438 species. *Nucleic Acids Res.* **47**, 3862-3874 (2019).
1439 16 Lago, S. *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell
1440 type-specific transcriptome. *Nat. Commun.* **12**, 3885 (2021).
1441 17 Figueiredo, J., Mergny, J. L. & Cruz, C. G-quadruplex ligands in cancer therapy: Progress,
1442 challenges, and clinical perspectives. *Life Sci.* **340**, 122481 (2024).

18 Boddupally, P. V. *et al.* Anticancer activity and cellular repression of c-MYC by the G-quadruplex-stabilizing 11-piperazinylquindoline is not dependent on direct targeting of the G-quadruplex in the c-MYC promoter. *J. Med. Chem.* **55**, 6076-6086 (2012).

19 Galli, S., Flint, G., Ruzickova, L. & Di Antonio, M. Genome-wide mapping of G-quadruplex DNA: a step-by-step guide to select the most effective method. *RSC Chem. Biol.* **5**, 426-438 (2024).

20 Esain-Garcia, I. *et al.* G-quadruplex DNA structure is a positive regulator of MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2320240121 (2024).

21 Spiegel, J. *et al.* G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **22**, 117 (2021).

22 Zanin, I. *et al.* Genome-wide mapping of i-motifs reveals their association with transcription regulation in live human cells. *Nucleic Acids Res.* **51**, 8309-8321 (2023).

23 Zeraati, M. *et al.* I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **10**, 631-637 (2018).

24 Calabrese, D. R. *et al.* Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.* **9**, 4229 (2018).

25 Cadoni, E., De Paepe, L., Manicardi, A. & Madder, A. Beyond small molecules: targeting G-quadruplex structures with oligonucleotides and their analogues. *Nucleic Acids Res.* **49**, 6638-6659 (2021).

26 Minard, A. *et al.* A short peptide that preferentially binds c-MYC G-quadruplex DNA. *Chem. Commun. (Camb)* **56**, 8940-8943 (2020).

27 Chowdhury, S., Wang, J., Nuccio, S. P., Mao, H. & Di Antonio, M. Short LNA-modified oligonucleotide probes as efficient disruptors of DNA G-quadruplexes. *Nucleic Acids Res.* **50**, 7247-7259 (2022).

28 Heddi, B., Cheong, V. V., Martadinata, H. & Phan, A. T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9608-9613 (2015).

29 Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86-89 (2003).

30 Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758-15759 (2008).

31 De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P. & Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* **129**, 1856-1857 (2007).

32 Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11870-11875 (2015).

1480 33 Rosonovski, S., Guneri, D., King, J., Morris, C. J. & Waller, Z. A. E. Affinity-selected peptide
1481 ligands specifically bind i-motif DNA and modulate c-Myc gene expression. *bioRxiv*,
1482 2025.2005.2030.656635 (2025). https://doi.org:10.1101/2025.05.30.656635
1483 34 Qin, G. *et al.* Targeting specific DNA G-quadruplexes with CRISPR-guided G-quadruplex-
1484 binding proteins and ligands. *Nat. Cell. Biol.* **26**, 1212-1224 (2024).
1485 35 Kendrick, S. *et al.* The dynamic character of the BCL2 promoter i-motif provides a mechanism
1486 for modulation of gene expression by compounds that bind selectively to the alternative DNA
1487 hairpin structure. *J. Am. Chem. Soc.* **136**, 4161-4171 (2014).
1488 36 Debnath, M. *et al.* Preferential targeting of i-motifs and G-quadruplexes by small molecules.
1489 *Chem. Sci.* **8**, 7448-7456 (2017).
1490 37 King, J. J. *et al.* DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in
1491 Human Cells. *J. Am. Chem. Soc.* **142**, 20600-20604 (2020).
1492 38 Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene
1493 expression. *Nat. Protoc.* **8**, 2180-2196 (2013).
1494 39 McCarty, N. S., Graham, A. E., Studena, L. & Ledesma-Amaro, R. Multiplexed CRISPR
1495 technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281 (2020).
1496 40 Prasad, B. *et al.* A complementary chemical probe approach towards customized studies of G-
1497 quadruplex DNA structures in live cells. *Chem. Sci.* **13**, 2347-2354 (2022).
1498 41 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target
1499 DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).
1500 42 Tomasello, G., Armenia, I. & Molla, G. The Protein Imager: a full-featured online molecular
1501 viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* **36**, 2909-2911
1502 (2020).
1503 43 Lam, E. Y., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable
1504 and detectable in human genomic DNA. *Nat. Commun.* **4**, 1796 (2013).
1505 44 Peraro, L. *et al.* Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *J. Am.*
1506 *Chem. Soc.* **140**, 11360-11369 (2018).
1507 45 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and
1508 Activation. *Cell* **159**, 647-661 (2014).
1509 46 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control
1510 of gene expression. *Cell* **152**, 1173-1183 (2013).
1511 47 Ahmadi, S. E., Rahimi, S., Zarandi, B., Chegeni, R. & Safa, M. MYC: a multipurpose oncogene
1512 with prognostic and therapeutic implications in blood malignancies. *J. Hematol. Oncol.* **14**, 121
1513 (2021).
1514 48 Albert, T. *et al.* The chromatin structure of the dual c-myc promoter P1/P2 is regulated by
1515 separate elements. *J. Biol. Chem.* **276**, 20482-20490 (2001).

49 Zhang, X., Spiegel, J., Martinez Cuesta, S., Adhikari, S. & Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **13**, 626-633 (2021).

50 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417-425 (2015).

51 Pena Martinez, C. D. *et al.* Human genomic DNA is widely interspersed with i-motif structures. *EMBO J.* **43**, 4786-4804 (2024).

52 Simonsson, T., Pribylova, M. & Vorlickova, M. A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.* **278**, 158-166 (2000).

53 Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11593-11598 (2002).

54 Sun, D. & Hurley, L. H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **52**, 2863-2874 (2009).

55 Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III(1). *J. Am. Chem. Soc.* **138**, 14138-14151 (2016).

56 Hui, W. W. I., Simeone, A., Zyner, K. G., Tannahill, D. & Balasubramanian, S. Single-cell mapping of DNA G-quadruplex structures in human cancer cells. *Sci. Rep.* **11**, 23641 (2021).

57 Pennarun, G. *et al.* Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* **24**, 2917-2928 (2005).

58 Yu, Z. *et al.* Chem-map profiles drug binding to chromatin in cells. *Nat. Biotechnol.* **41**, 1265-1271 (2023).

59 Cuddapah, S. *et al.* Genomic profiling of HMGN1 reveals an association with chromatin at regulatory regions. *Mol. Cell Biol.* **31**, 700-709 (2011).

60 Kugler, J. E., Deng, T. & Bustin, M. The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes. *Biochim. Biophys. Acta* **1819**, 652-656 (2012).

61 Rosano, D. *et al.* Long-term Multimodal Recording Reveals Epigenetic Adaptation Routes in Dormant Breast Cancer Cells. *Cancer Discov.* **14**, 866-889 (2024).

62 Berner, A. *et al.* G4-Ligand-Conjugated Oligonucleotides Mediate Selective Binding and Stabilization of Individual G4 DNA Structures. *J. Am. Chem. Soc.* **146**, 6926-6935 (2024).

63 He, Y. D. *et al.* Selective Targeting of Guanine-Vacancy-Bearing G-Quadruplexes by G-Quartet Complementation and Stabilization with a Guanine-Peptide Conjugate. *J. Am. Chem. Soc.* **142**, 11394-11403 (2020).

1553 64 Tan, D. J. Y., Das, P., Winnerdy, F. R., Lim, K. W. & Phan, A. T. Guanine anchoring: a strategy
1554 for specific targeting of a G-quadruplex using short PNA, LNA and DNA molecules. *Chem.*
1555 *Commun. (Camb)* **56**, 5897-5900 (2020).
1556 65 Zhao, H., Lau, H. L., Zhang, K. & Kwok, C. K. Selective recognition of RNA G-quadruplex in
1557 vitro and in cells by L-aptamer-D-oligonucleotide conjugate. *Nucleic Acids Res.* **52**, 13544-
1558 13560 (2024).
1559 66 Felsenstein, K. M. *et al.* Small Molecule Microarrays Enable the Identification of a Selective,
1560 Quadruplex-Binding Inhibitor of MYC Expression. *ACS Chem. Biol.* **11**, 139-148 (2016).
1561 67 Flusberg, D. A. *et al.* Identification of G-Quadruplex-Binding Inhibitors of Myc Expression
1562 through Affinity Selection-Mass Spectrometry. *SLAS Discov.* **24**, 142-157 (2019).
1563 68 Robinson, J. *et al.* G-quadruplex structures regulate long-range transcriptional reprogramming
1564 to promote drug resistance in ovarian cancer cells. *Genome Biol.* **26**, 183 (2025).
1565 69 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein*
1566 *Expr. Purif.* **41**, 207-234 (2005).
1567 70 Lingeman, E., Jeans, C. & Corn, J. E. Production of Purified CasRNPs for Efficacious Genome
1568 Editing. *Curr. Protoc. Mol. Biol.* **120**, 31 10 31-31 10 19 (2017).
1569