



**HAL**  
open science

## G-quadruplex forming motifs in the promoter region of the B-MYB proto-oncogene

André Miranda, Anne Cucchiarini, Cyril Esnault, Jean-Christophe Andrau, Paula Oliveira, Jean-Louis Mergny, Carla Cruz

► **To cite this version:**

André Miranda, Anne Cucchiarini, Cyril Esnault, Jean-Christophe Andrau, Paula Oliveira, et al.. G-quadruplex forming motifs in the promoter region of the B-MYB proto-oncogene. *International Journal of Biological Macromolecules*, 2024, 270, pp.132244. 10.1016/j.ijbiomac.2024.132244 . hal-04753347

**HAL Id: hal-04753347**

**<https://hal.science/hal-04753347v1>**

Submitted on 25 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1 **G-quadruplex forming motifs in the promoter region of** 2 **the B-MYB proto-oncogene**

3  
4 André Miranda <sup>a,d</sup>, Anne Cucchiarini <sup>b</sup>, Cyril Esnault <sup>c</sup>, Jean-Christophe Andrau <sup>c</sup>, Paula  
5 A. Oliveira <sup>d</sup>, Jean-Louis Mergny <sup>b\*</sup>, Carla Cruz <sup>a, e\*</sup>

6  
7 <sup>a</sup> CICS-UBI - Health Sciences Research Centre, University of Beira Interior, Covilhã,  
8 Portugal;

9 <sup>b</sup> Laboratoire d'Optique et Biosciences, École Polytechnique, CNRS, INSERM, Institut  
10 Polytechnique de Paris, 91128 Palaiseau, France;

11 <sup>c</sup> Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS-UMR  
12 5535, Montpellier, France;

13 <sup>d</sup> Centre for Research and Technology of Agro-Environmental and Biological Sciences  
14 (CITAB), Inov4Agro, University of Trás-os-Montes and Alto Douro (UTAD), Quinta de  
15 Prados, 5000-801 Vila Real, Portugal;

16 <sup>e</sup> Departamento de Química, Universidade da Beira Interior, Rua Marquês de Ávila e  
17 Bolama, 6201-001 Covilhã, Portugal.

18  
19 \* Authors to whom correspondence may be addressed. [jean-louis.mergny@inserm.fr](mailto:jean-louis.mergny@inserm.fr)  
20 (JLM) [carlacruz@fcsaude.ubi.pt](mailto:carlacruz@fcsaude.ubi.pt) (CC)

29 **ABSTRACT**

30 To combat cancer, a comprehensive understanding of the molecular mechanisms and  
31 behaviors involved in carcinogenesis is crucial, as tumorigenesis is a complex process  
32 influenced by various genetic events and disease hallmarks. The B-MYB gene encodes  
33 a transcription factor involved in cell cycle regulation, survival, and differentiation in  
34 normal cells. B-MYB can be transformed into an oncogene through mutations, and  
35 abnormal expression of B-MYB has been identified in various cancers, including lung  
36 cancer, and is associated with poor prognosis. Targeting this oncogene is a promising  
37 approach for anti-cancer drug design. B-MYB has been deemed undruggable in previous  
38 reports, necessitating the search for novel therapeutic options. In this study, we found  
39 that the B-MYB gene promoter contains several G/C rich motifs compatible with G-  
40 quadruplex (G4) formation. We investigated and validated the existence of G4 structures  
41 in the promoter region of B-MYB, first *in vitro* using a combination of bioinformatics,  
42 biophysical, and biochemical methods, then *in cell* with the recently developed *G4access*  
43 method.

44

## 45 1. INTRODUCTION

46 Cancer is widely recognized as a significant global public health issue [1]. In order to  
47 reduce the millions of annual deaths caused by this disease, it is crucial to identify new  
48 targets and explore synergistic therapeutic combinations. Gaining a profound  
49 understanding of the molecular mechanisms and behaviors involved in carcinogenesis  
50 is imperative, as cancer progression is a complex process that necessitates a series of  
51 genetic events and the acquisition of disease hallmarks [2]. The proto-oncogenes  
52 regulate a number of biological processes in healthy cells, acting as growth factors,  
53 transducers of cellular signals and nuclear transcription factors [3]. However, some  
54 mutations (*e.g.*, point mutations, gene amplification and chromosomal translocation or  
55 transposition events) can turn a proto-oncogene into an oncogene, which plays a critical  
56 role in cancer progression. These mutations can be caused by external stimuli, such as  
57 radiation, chemical compounds or viruses that alter the normal cell behavior, leading to  
58 uncontrolled overexpression of proteins resulting in the acquisition of cancer hallmarks,  
59 such as fast cell division, inhibition of cell differentiation and cell death [4]. For these  
60 reasons, oncogenes are considered as good molecular targets for anti-cancer drug  
61 design.

62 The B-MYB (V-Myb avian myeloblastosis viral oncogene homolog-like 2), also known  
63 as MYBL2, is a proto-oncogene located on chromosome 20, locus 20q13.12. This gene  
64 encodes a 21 kDa transcription factor involved in the regulation of the expression of  
65 multiple genes affecting cell cycle progression, cell survival and cell differentiation [5–8].  
66 The knockout of B-MYB results in an early embryonic lethal phenotype, denoting the  
67 importance of the proto-oncogene in proliferating cells [7,9]. During the malignant  
68 transformation of tumor cells, these properties are dysregulated by the B-MYB oncogene  
69 that mediates these alterations [10]. Nevertheless, some reports reported that this  
70 transcription factor is currently undruggable [11,12]. To overcome this obstacle, it is  
71 imperative to consider novel therapeutic options, for example by examining the genetic  
72 components responsible for encoding these transcription factors.

73 In recent studies, B-MYB expression was shown to be strictly correlated with the  
74 occurrence and development of malignant tumors in various organs and tissues [13].  
75 The aberrant expression of the B-MYB (denoting the oncogenic role) leads to gene  
76 alterations causing tumorigenesis and sustained tumor growth [14]. B-MYB deregulation  
77 in cancer may be due to oncogene amplification, deregulation of the DREAM complex  
78 assembly or, at the post-transcriptional level, by microRNAs [7,15]. Abnormal  
79 overexpression is commonly observed in several cancers such as prostate cancer,

80 melanoma, liver cancer, renal clear cell carcinoma and lung cancer (LC) [16–21] and is  
81 a predictor of poor prognosis [14,22–25]. In LC, B-MYB is up-regulated and promotes  
82 cell growth and motility [26] and its overexpression enhances transcription and  
83 deregulates gene expression [27]. B-MYB can induce activation of extracellular signal-  
84 regulated kinases (ERK) and phosphorylated-protein kinase B (Akt) signalling pathways,  
85 involved in tumorigenesis, revealing a potential tumor-role of this oncogene in LC [26].  
86 MYBL2 has also been reported to participate in different aspects of cell apoptosis and  
87 survival by regulating protein interactions [21].

88 Bioinformatic analysis of the human genome showed that promoter regions of several  
89 oncogenes are G/C rich with local G/C skewness. These characteristics facilitate the  
90 formation of non-canonical structures called G-quadruplexes (G4) [28,29]. G4s are  
91 formed by the self-association of four guanine bases in a planar arrangement via  
92 Hoogsteen bonds. As G-rich sequences are significantly enriched around transcription  
93 start sites (TSS), G4s may have a gene regulatory function [28]. In addition, such G4s  
94 can be considered as novel anticancer targets [30] as they provide a potential recognition  
95 motif for small molecules, given that G4 stabilization has been reported to suppress  
96 oncogene transcription and inactivate downstream pathways involved in tumorigenesis  
97 resulting in the reduction of cancer cell proliferation [31]. Recently, Papp *et al.*  
98 hypothesized, using a computational model, that the promoter regions of genes that  
99 encode transcription factors (TFs) are crucial for the regulatory signals for transcription  
100 processes in cells, as over 90% of protein-coding genes contain potential non canonical  
101 G4 in their promoter or gene body [32].

102 Thus, considering its ubiquitous role in cancer biology, downregulating B-MYB  
103 constitutes an attractive strategy for cancer therapy. In this context, the main aim of this  
104 work is to identify and confirm G4 structures in the promoter region of the B-MYB  
105 oncogene as a target for anticancer therapy. For this validation, several bioinformatics,  
106 biophysical and biochemical approaches were carried out, and these *in vitro* results were  
107 compared with data provided by *in cell* analysis of G4 formation, using the recently  
108 developed G4Access method [33].

109

110

111

112

113

## 114 2. MATERIALS AND METHODS

115

### 116 2.1. Oligonucleotides and reagents

117 All sequences (**Table S1**) were purchased lyophilized from Eurofins (Nantes, France) or  
118 Eurogentec (Liège, Belgium) with double-HPLC grade purification. Oligonucleotides  
119 were resuspended in a proper volume of ultrapure water ( $18.2 \Omega \text{ cm}^{-1}$  resistivity), until  
120 reaching a final concentration of 100  $\mu\text{M}$ , according to the manufacturer's indication, and  
121 stored at  $-20 \text{ }^\circ\text{C}$ . Oligonucleotide concentration was checked using the absorbance at  
122 260 nm (NanoDrop™ 2000c, Thermo Scientific™, Massachusetts, USA) and the molar  
123 extinction coefficients ( $\epsilon$ ) provided by the manufacturers.

124 Biophysical experiments were performed in potassium-containing buffers keeping total  
125 ionic strength constant. These buffers always included 10 mM lithium cacodylate  
126 adjusted to pH 7.2 and. Their compositions were adjusted by adding potassium and/or  
127 lithium chloride to obtain buffers called K1 (1 mM KCl, 99 mM LiCl), K10 (10 mM KCl, 90  
128 mM LiCl), K20 (20 mM KCl, 80 mM LiCl) and K100 (100 mM KCl). For the annealing  
129 process, the samples were heated at  $95 \text{ }^\circ\text{C}$  for 5 min and slowly cooled down to room  
130 temperature (inside the heating block until reach  $25^\circ\text{C}$ ).

131 All reagents and commercially available G4 ligands (PhenDC3, ThT and NMM; **Table**  
132 **S1**) were obtained from Sigma-Aldrich (St. Louis, MO, USA).

133

### 134 2.2. Bioinformatic analysis

135 The B-Myb promoter sequence was found using the UCSC Genome Browser database  
136 (<http://genome.ucsc.edu/index.html>; Genomics Institute, University of California Santa  
137 Cruz, USA) to access 2000 nucleotides upstream of the TSS. The sequence was then  
138 analyzed using the G4Hunter v.2.0 Web tool  
139 (<https://bioinformatics.cruk.cam.ac.uk/G4Hunter/>). The software was set for a window  
140 size of 25 nucleotides and a threshold score  $\geq 1.2$  (G4H score). This value appears as a  
141 reasonable compromise to minimize the number of false negatives and false positives.  
142 Higher scores correspond to sequences more likely to form stable G4s (high guanine  
143 content) [33]. To design control sequences (mutants) we employed the G4Killer Web tool  
144 (<https://bioinformatics.ibp.cz/#/analyse/g4-killer>).

145 In addition, we performed a multiple sequences alignment of B-MYB promoter regions  
146 of several primate species. Sequences were taken from Ensembl Genome Browser

147 (<http://www.ensembl.org/index.html>]; EMBL-EBI, Wellcome Genome Campus, Hinxton,  
148 Cambridgeshire, UK; [34]) and subsequently aligned using Clustal Omega webtool  
149 (<https://www.ebi.ac.uk/Tools/msa/clustalo/>; EMBL-EBI, Wellcome Genome Campus,  
150 Hinxton, Cambridgeshire, UK). The promoter region alignment and conservation were  
151 visualized using Jalview v2.11.2.6 software [35] and were used to generate the LOGO  
152 consensus plot of the G4 prone sequence region in aligned sequences using the  
153 WebLogo 3 tool [36]. The analyses of B-MYB expression and epigenetic profiles were  
154 carried out using the GEPIA database [37] and the Factorbook database  
155 (<https://www.factorbook.org/>), respectively.

156 Additionally, we performed a prediction of the transcription factor binding sites thanks to  
157 the PROMO software applying default parameters ([https://algggen.lsi.upc.es/cgi-  
158 bin/promo\\_v3/promo/promoinit.cgi?dirDB=TF\\_8.3](https://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3)) and analyzed the gene expression  
159 profiling of proto-oncogenes [37].

160

## 161 **2.3. Förster resonance energy transfer (FRET) experiments**

162

### 163 **2.3.1. FRET-melting competition (FRET-MC)**

164 FRET-MC assay was performed as described [38]. Experiments were carried out in a  
165 CFX Connect™ Real-Time PCR instrument (Bio-Rad, CA, USA), equipped with a FAM  
166 filter ( $\lambda_{ex} = 492 \text{ nm}$ ;  $\lambda_{em} = 516 \text{ nm}$ ), using 96-well plates and a working volume of 25  $\mu\text{L}$ .  
167 Each well contained 0.2  $\mu\text{M}$  of F21T fluorescent oligonucleotide in the presence or  
168 absence of 0.4  $\mu\text{M}$  G4 ligand (PhenDC3) and 3  $\mu\text{M}$  competitors. All solutions were  
169 dissolved in K10 buffer, and each condition was tested on three separate plates.

170 Equipment was parametrized to record FAM emission and make a stepwise increase of  
171 0.5 °C every 1 min from 25 °C to 95 °C. The obtained curves were normalized and fitted,  
172 applying a Boltzmann model, using OriginPro2021 software (OriginLab, Massachusetts,  
173 USA) to determine the melting temperature ( $T_m$ ).  $\Delta T_m$  is calculated by the difference in  
174  $T_m$  of F21T in the presence of a competitor and the absence of PhenDC3.

175 To classify and quantify the competition effect of competitors, the S Factor was  
176 calculated following **Equation (1)**.

$$177 \quad S \text{ factor} = \frac{\Delta T_m \text{ of F21T with competitors}}{\Delta T_m \text{ of F21T alone}} \quad \text{Eq. (1)}$$

178

179 Strong competitors give a S factor value around 0 while S remains close to 1 for  
180 ineffective competitors.

181

### 182 **2.3.2. Isothermal-FRET (iso-FRET) assay**

183 The isothermal competition experiment was performed in 96-well plates as reported [39].  
184 Each well contains a final concentration of 200 nM 37Q, 5  $\mu$ M competitor, 1  $\mu$ M  
185 PhenDC3, and 20 nM F22 in a total volume of 25  $\mu$ L. For this assay, the samples were  
186 not pre-denatured and diluted in K20 buffer. Briefly, we mixed 5  $\mu$ L of competitor (2.5  
187  $\mu$ M) with 5  $\mu$ L of 37Q (1  $\mu$ M) and 10  $\mu$ L of PhenDC3 (2.5  $\mu$ M). The blend was then mixed  
188 and incubated for 5 minutes. After this time, 5  $\mu$ L of F22 (100 nM) and 5  $\mu$ L of K20 buffer  
189 were added. Additionally, controls were also performed with F22 in the presence or  
190 absence of 37Q and with G4 and non-G4 forming sequences. Before reading, plates  
191 were kept in an incubator at 37  $^{\circ}$ C and the fluorescence signal was measured using a  
192 TECAN Infinite M1000 Pro plate reader (Salzburg, Austria).

193 We calculated the F factor using the OriginPro2021 software (OriginLab, Massachusetts,  
194 USA) with **Equation (2)** and fluorescence intensities were plotted. The F value  
195 parameter is calculated from the fluorescence intensities (FI); it reflects the extent to  
196 which a competitor affects F22-37Q hybridization in the presence of PhenDC3.

$$197 \quad F \text{ value} = \frac{FI_{\text{Competitors}} - FI_{\text{F22} + 37\text{Q duplex}}}{FI_{\text{F22}} - FI_{\text{F22} + 37\text{Q duplex}}} \quad \text{Eq. (2)}$$

198

199 The global alignment analysis function, based on Needleman-Wunsch algorithm, of  
200 NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; National Center for  
201 Biotechnology Information, Maryland, USA) was used to calculate the CF factor  
202 (**Equation (3)**) and confirm possible complementarity among competitors and F22  
203 (**Table S1**). Analysis was performed using the following parameters: NUC4.4 scoring  
204 matrix, gap and extension penalty of 10.0 and 0.5, respectively.

$$205 \quad \text{CF factor} = \frac{\text{Numbers of base pairs expected in (F22+Competitor duplex)}}{\text{Length F22}} \quad \text{Eq. (3)}$$

206

### 207 **2.4. Fluorescence light-up probe assays**

208 For these assays, the fluorescence signal of probes was measured using a TECAN  
209 Infinite M1000 Pro plate reader (Salzburg, Austria). ThT (Thioflavin T) and NMM (N-  
210 methyl mesoporphyrin IX) were excited at 420 nm and 380 nm and fluorescence signals



211 were recorded at 490 nm and 610 nm, respectively. Before reading, the plate was shaken  
212 for 5 min and incubated for 10 min at room temperature. Each well contained 52  $\mu$ l of  
213 K100 buffer, 40  $\mu$ l of oligonucleotide (final concentration of 3  $\mu$ M) and 8  $\mu$ l of probe (final  
214 concentration of 2  $\mu$ M). The ligand had be the last component to be added to the mixture.  
215 Experiments were repeated in three different 96-well black plates using positive or  
216 negative controls using different sequences (G4s with different topologies and duplexes).

217

## 218 **Circular dichroism spectroscopy**

219 Using a Jasco J-1500 CD spectropolarimeter (Jasco, Japan), equipped with a multi-  
220 holder and Peltier-type temperature controller, we recorded the CD spectra between 210  
221 – 350 nm (scanning rate of 100 nm/min, 0.5 nm bandwidth, 1 s integration time over 4  
222 averaged accumulations). Samples were typically annealed in K100 buffer (or other  
223 working buffers when specified) and scanned using high-precision quartz cuvettes (ref.  
224 115B-10-40 Hellma Analytics, Germany; path-length 10 mm  $\times$  2 mm). Final sample  
225 concentrations were adjusted for each oligonucleotide to reach an absorbance value  
226 close to 0.8, following the Lambert-Beer equation and considering  $\epsilon$ . Obtained data was  
227 smoothed and converted to molar dichroic absorption ( $\Delta\epsilon$ ), through **Equation (4)**:

$$228 \quad \Delta\epsilon = \frac{\theta}{32980 \times c \times l} \quad \text{Eq. (4)}$$

229

230 where  $\theta$  is ellipticity in millidegrees,  $c$  is the oligonucleotide concentration ( $\text{mol L}^{-1}$ ) and  $l$   
231 is the path of the length in cm. Additionally, the conformation index  $r$  was calculated  
232 following **Equation 5**, as reported by Chen *et al.* [40]:

$$233 \quad r = \frac{CD_{265}}{|CD_{265}| + CD_{290}} \quad \text{Eq. (5)}$$

234

235 where  $CD_{265}$  and  $CD_{290}$  are the molar dichroic absorption of peaks. Parallel, hybrid and  
236 antiparallel conformation have  $1 > r \geq 0.5$ ,  $0.5 > r > 0$  and  $r < 0$ , respectively.

237 In CD-melting experiments, we monitored the ellipticities at 240, 260 and 290 nm as a  
238 function of temperature, in the range of 20 to 100  $^{\circ}\text{C}$ , with a rate of  $1^{\circ}\text{C}/\text{min}$ . Using the  
239 OriginPro2021 program, data were normalized into fraction folded ( $f$ ) plots, according to  
240 **Equation (6)**, and fitted to a Boltzmann distribution to determine the  $T_m$  value.

241

$$242 \quad f = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \quad \text{Eq. (6)}$$

243

244 **UV-Vis spectroscopy**

245 All UV-Vis experiments were performed using a Cary 300 double-beam  
246 spectrophotometer (Agilent Technologies, France) equipped with a thermostable 6-cell  
247 holder and a high-performance peltier temperature controller (Agilent Technologies,  
248 France). Absorbance spectra were recorded in 1 cm pathlength quartz cuvettes (115B-  
249 QS, Hellma Analytics, Germany) between 220 and 340 nm with a scan rate of 300  
250 nm/min, 0.5 nm data intervals, 0.05 s of integration time and automatic baseline  
251 correction. For all experiments, the oligonucleotides were used at a 3  $\mu\text{M}$  concentration.

252 In **thermal difference spectra (TDS)** oligonucleotide solutions were annealed in K100  
253 buffer and we recorded the spectra at 20°C (folded state). Then the temperature was  
254 increased to 95°C and a second spectrum was recorded after 15 min of equilibration  
255 (unfolded state). TDS corresponds to the arithmetic difference between the spectra in  
256 folded and unfolded conditions (20°C and 95°C, respectively) [41,42]. The resultant  
257 spectra were normalized relative to the maximum absorbance and TDS factors were  
258 calculated [43,44].

259 For **isothermal difference spectra (IDS)** oligonucleotide solutions were annealed in 10  
260 mM LiCaCo pH 7.2 buffer as previously described. Firstly, we acquired a first spectrum  
261 without potassium cation at 25°C (unfolded), and then a second after adding this cation  
262 (presumably folded). For this last step, 1 M KCl was added directly into the cuvette to  
263 obtain a final concentration of 100 mM and equilibrated for 30 min before recording.  
264 Spectra were corrected for dilution. IDS corresponds to the arithmetic difference between  
265 the spectra in the unfolded and folded states (in the absence and presence of cation,  
266 respectively).

267 In **the UV-melting** assay, we recorded the variations of absorbance at 295 nm as a  
268 function of temperature. Oligonucleotides were prepared in K10 buffer and first heated  
269 to 95 °C for 15 min. The absorbance was recorded at distinct wavelengths using a cycle  
270 composed of cooling down to 10 °C at a rate of 0.2 °C/min. followed by heating to 95 °C  
271 at the same rate [45]. Melting temperatures were determined. Assuming a two-step  
272 melting process, thermodynamic parameters were calculated (**Equation (7)**) [46,47].

273 
$$\Delta G_{fold}^{\circ} = -RT \ln(K_A) = \Delta H_{fold}^{\circ} - T\Delta S_{fold}^{\circ} \quad \text{Eq. (7)}$$

274 Using the  $\Delta G_{fold}^{\circ}$  vs T plot was extracted from the slope and interception the values of  
275 enthalpy  $\Delta H^{\circ}$  and entropy  $\Delta S^{\circ}$  of folding.

276

## 277 **Nuclear Magnetic Resonance (NMR) spectroscopy**

278 <sup>1</sup>H NMR spectra were recorded using a 600 MHz Bruker Avance III spectrometer (Bruker  
279 Corporation, Massachusetts, USA), operating at a <sup>1</sup>H Larmor frequency of 600.10 MHz,  
280 equipped with an inverse detection triple-resonance z-gradient QCI cryoprobe. We  
281 applied the *zgesgp* water suppression pulse sequence using the next parameters: 32 K  
282 data points, relaxation delay of 2 s and 1024 scans for a spectral width of 12.019 Hz  
283 centered at the water resonance.

284 All samples (50 μM of oligonucleotide) were acquired at 25°C in LiCaCo buffer  
285 supplemented with 10% (v/v) D<sub>2</sub>O (Eurisotop, France). In salt titration experiments,  
286 increasing amounts of KCl 1M were directly added into the 3 mm tube and performed  
287 the annealing. Additionally, we studied the topological conversion analyzing the spectral  
288 changes with temperature (20 °C and 37 °C).

289 Data were processed and analyzed using Bruker TopSpin4.1. software and chemical  
290 shifts (δ) were measured in ppm.

291

## 292 **Size-Exclusion Chromatography (SEC)**

293 SEC experiments were conducted with an AKTA FPLC (GE Healthcare Life Sciences,  
294 USA) equipped with a ThermoAcclaim SEC-300 column (4.6 × 300 mm; 5μm hydrophilic  
295 polymethacrylate resin spherical particles, 300 Å pore size; ref. 079723; Thermo  
296 Scientific™, Massachusetts, USA) at 20 °C according previously reported [48].  
297 Equilibration was performed before the first injection (with at least three column volumes,  
298 i.e. 15 ml). Oligonucleotides were annealed in elution buffer (Tris-HCl 50 mM pH 7.5  
299 supplemented with 100 mM KCl) and eluted at 0.150 ml/min rate. Eluted species were  
300 monitored by the absorbance at 260 nm and then chromatograms were processed  
301 (normalized to 0-1) and plotted in OriginPro2021 using the relative elution volume  $V_e/V_0$   
302 ( $V_e$  is the elution volume and  $V_0$  is the dead volume).

303

## 304 **Native Polyacrylamide gel electrophoresis**

305 Native polyacrylamide gel electrophoresis (PAGE) was used to check the ability of  
306 sequences to adopt a folded structure. For this experiment, a 20% polyacrylamide gel  
307 supplemented with 100 mM KCl was cast and pre-run for 30 min before loading.  
308 Oligonucleotides were first diluted in water at 2 μM and heated for 5 min at 95°C, followed  
309 by the addition of buffer (20 mM LiCaco and 200 mM of KCl). Subsequently, samples

310 were heated at 90°C for 5 min and slowly cooled down to room temperature. G4 forming  
311 sequences and mutated sequences were mixed with sucrose (Sigma-Aldrich, MO, USA)  
312 until reaching a final concentration of 15% sucrose, 20 mM LiCaco, 100 mM KCl and 1  
313 µM DNA. As a control, single-stranded oligothymidylate markers, 9 to 57 nucleotide-long  
314 (dT<sub>9</sub>, dT<sub>15</sub>, dT<sub>22</sub>, dT<sub>30</sub>, dT<sub>45</sub> and dT<sub>57</sub>) were also loaded on the gel.

315 The loaded gel was run in a vertical electrophoretic cell Mini-Protean II (BioRad,  
316 Hercules, CA, USA) at a temperature of 4°C (110 V, 210 min). After running, the gel was  
317 stained with 2.5 µM of NMM for 30 minutes under continuous and gentle agitation and  
318 visualized in ChemiDoc XRS+ (BioRad, Hercules, CA, USA). Lastly, the gel was stained  
319 with 1X SYBR Gold during 20 min before visualization in ChemiDoc XRS+. Each gel was  
320 briefly washed with water before reading.

321

## 322 **G4access**

323 The G4access was applied according to the protocol recently described by Esnault *et al.*  
324 [33]. Briefly, K562 (human myelogenous leukaemia), Raji (human Burkitt's lymphoma)  
325 and HaCaT (spontaneously immortalized human keratinocyte) cell lines were cultured  
326 in RPMI 1640 or DMEM medium supplemented with 10% heat-inactivated fetal bovine  
327 serum, 100 units/mL penicillin, and 100 µg/mL streptomycin at 37°C in a humidified 5%  
328 CO<sub>2</sub> incubator.

329 Prior to MNase digestion, the nuclei were purified and refined using a sucrose cushion.  
330 The nuclei were resuspended in MNase digestion buffer. Prewarmed MNase reaction  
331 buffer supplemented with MNase (Merck, 10107921001) was then added. Samples were  
332 MNase digested for 10 minutes at 37°C, stopped on ice and by addition of EDTA and  
333 SDS lysis buffer. Before DNA purification, the samples were lysed for two hours at 37 °C  
334 using 5 µL of RNase A (ThermoFisher, EN0531) and for two hours at 56 °C using 8 µL  
335 of proteinase K (Euromedex, 09-0911). The samples were cleaned up using the  
336 QIAquick PCR Purification Kit (QIAGEN, 28106) and assessed using agarose gels and  
337 the Bioanalyzer 2100, in order to guarantee the quality of the MNase digestions. Size  
338 selection was applied using a 4–20% polyacrylamide Novex™ TBE gel (ThermoFisher,  
339 EC6225BOX) in order to select fragments smaller than 100 bp. Segments ranging from  
340 0 to 100 bp were extracted from the gel, eluted in water, and purified using an Agilent  
341 5185–5990 0.22 µm spin filter. DNA libraries were prepared from the size-selected  
342 fragments from MNase digestions that contain approximately 30% of mono-  
343 nucleosomes. Using a starting material of 50 ng, paired-end libraries were prepared  
344 using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs,

345 E7645S). Libraries were sequenced using paired 75 bp reads (Raji), paired 50-30 bp  
346 reads (K562 cells), or paired 76 bp reads (HaCaT cells) using an Illumina NextSeq-500  
347 Sequencer. Datasets were aligned with Bowtie2 and analysed using PASHA (version  
348 0.99.21) R (version 3.3.1) pipeline. Aligned reads were elongated in silico to create  
349 wiggle files.

350

## 351 RESULTS AND DISCUSSION

352 Proto-oncogenes are essential for normal cell homeostasis; however, their conversion  
353 into oncogenes plays a pivotal role in cancer. One example is the B-MYB oncogene that  
354 encodes a transcription factor abnormally overexpressed in cancer and associated with  
355 the occurrence and development of malignant tumors. Due to the undruggable profile of  
356 the B-MYB transcription factor, it is necessary to develop new and efficient therapeutic  
357 approaches.

358 One proposed approach is to attack upstream pathways, *i.e.*, before the expression of  
359 the transcription factor, namely before transcription. For this, we can take advantage of  
360 the sequence bias found in oncogene promoters near their TSS, that potentially allows  
361 the formation of G4. This quadruplex can act as a binding site for small molecules,  
362 leading to the repression of transcription and consequently the inhibition of downstream  
363 pathways. Multiple examples in the literature testify to the importance of G4 in promoter  
364 regions for the regulation of transcription processes in cells.

365

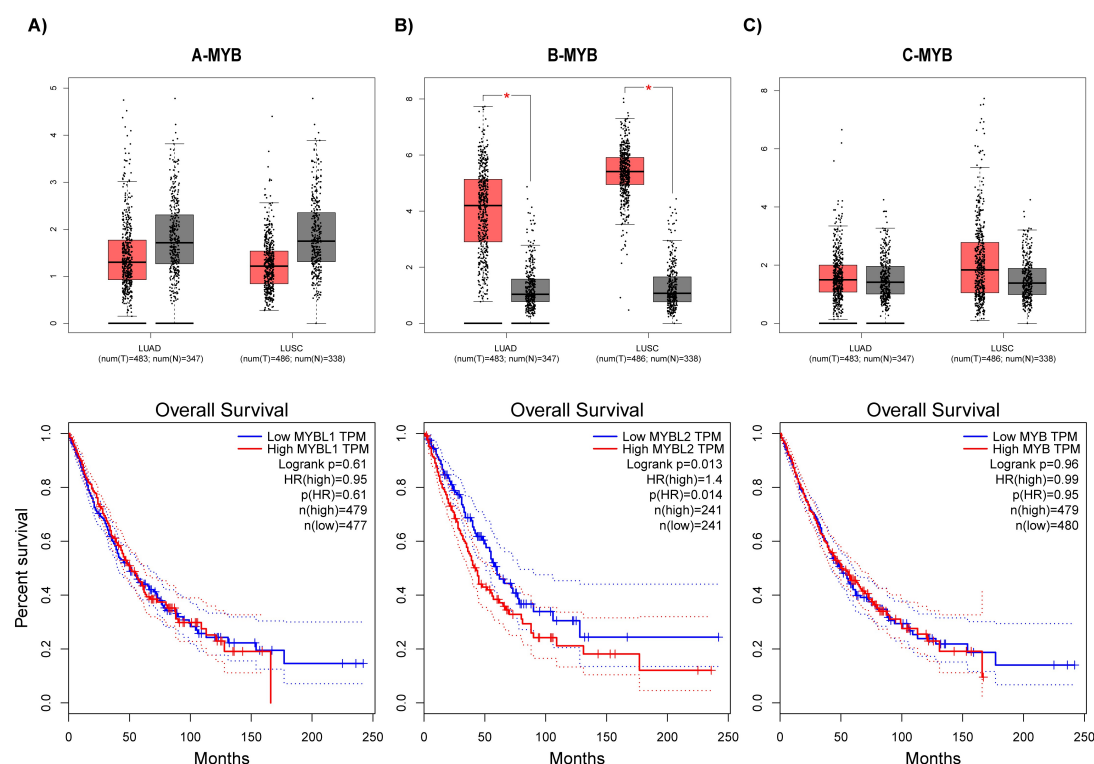
### 366 Bioinformatic analysis

367 The B-MYB gene (ID 4605) is located on locus 20q13.12 of chromosome 20.  
368 Interestingly, this is the chromosome with the higher ratio of G4-positive genes *per*  
369 chromosome [32]. We started by analyzing the human promoter region of the B-MYB  
370 proto-oncogene to evaluate its ability to form G4 structures. This analysis was performed  
371 with the G4 Hunter algorithm that considers G-richness and G-skewness of a given  
372 sequence and provides a score (G4H score) as an output [29]. This scoring is calculated  
373 taking into account the nucleotide nature of the sequence and the running length of G.  
374 More stable quadruplexes are correlated with higher G4H scores [33]. The evaluation  
375 was carried out using standard parameters namely, a window size of 25 nucleotides and  
376 a threshold score of 1.2, which is considered a reasonable compromise to identify G4-  
377 prone motifs *in vitro* and maximize accuracy (*i.e.*, minimize the fraction of false negatives  
378 and false positives) [29].

379 The predicted G4 sequences with their G4H score and location are summarized in  
380 **Figure 1** and **Table S2**. The algorithm revealed 4 sequences (B-MYB 26, B-MYB 25R,  
381 B-MYB 26RA and B-MYB 43) among which B-MYB 43 shows the highest score (G4H  
382 score = 2.14) while the other motifs give G4H scores near the threshold value. In the  
383 case of B-MYB 43R, due to its high score and length, shorter sequence derivatives, B-  
384 MYB 25RC and B-MYB 22R, were designed. Both sequences exhibit high G4H scores



394 Continuing with bioinformatics analysis, we observe that B-MYB is highly expressed in  
 395 lung cancer (**Figure 2A**), as consistently described in the literature [7,14,17,26,27],  
 396 emphasizing the relevance of this oncogene. Comparing the relative gene expression of  
 397 B-MYB with other transcription factors of the same family, namely A-MYB and C-MYB,  
 398 a clear overexpression was evidenced for lung cancer. In addition, elevated levels of B-  
 399 MYB have been associated with advanced tumor stages and poor prognosis [5,50,51].  
 400 Thus, overexpression of B-MYB may contribute to the progression of lung cancer by  
 401 promoting the survival and growth of cancer cells.



402

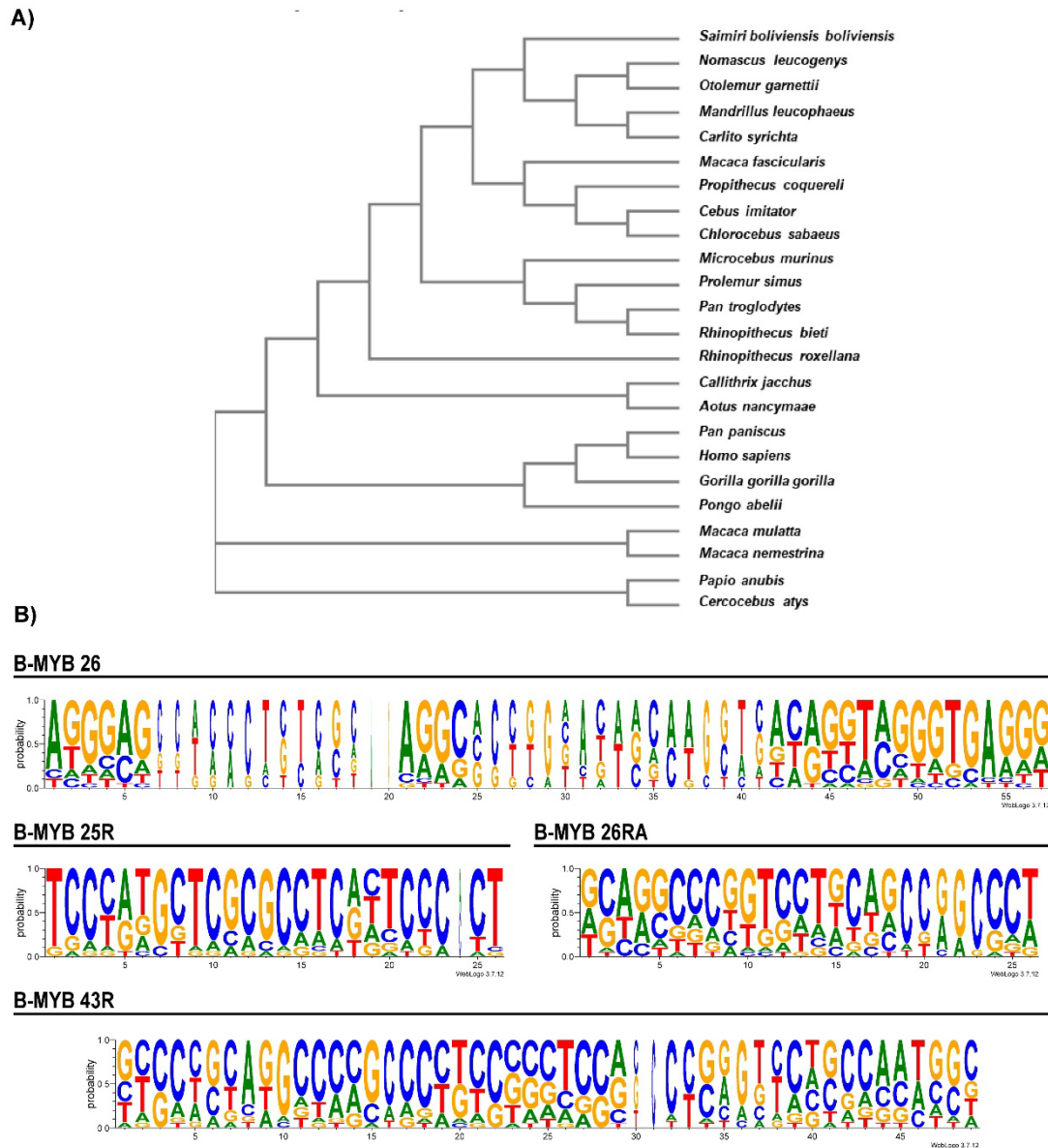
403 **Figure 2.** Bioinformatic analysis of expression of MYB genes in lung cancer. Relative expression  
 404 and overall survival in MYB gene family obtained thanks to the GEPIA database, in lung cancer  
 405 diseases (LUAD – lung adenocarcinoma; LUSC - Lung squamous cell carcinoma) and compared  
 406 with healthy individuals. Comparison among **A)** A-MYB **B)** B-MYB and **C)** C-MYB proto-  
 407 oncogenes.

408

409 Additionally, the analysis of epigenetic profiles can provide an additional clue on the  
 410 biological relevance of this gene. Histone changes are useful epigenetic markers of the  
 411 chromatin condition linked to either repression or activation of genes. Usually, actively  
 412 transcribed genes are marked by trimethylated H3K4 (H3K4me3) and acetylated H3K27  
 413 (H3K27ac) in TSS [52]. Analyzing the epigenetic data, taken from the FactorBook



414 database (<https://www.factorbook.org/>), we verified an enrichment in H3K4me3,  
 415 H3K4me2 and H3K27ac marks, as demonstrated by the peak intensity density  
 416 distribution around its TSS (**Figure S1**). Thus, we observed a correlation between the  
 417 higher expression of B-MYB in cancer patients, when compared to normal tissues, and  
 418 the epigenetic profile of histones that reveal an activation mechanism of the gene.



419

420 **Figure 3.** Bioinformatic analysis of the promoter region of the B-MYB gene. **A)** Phylogenetic tree  
 421 of B-MYB promoter region between Human and primates. **B)** Logo representation of putative G4  
 422 sequences after multiple alignment of primate sequences.

423

424 After the identification of putative G4 sequences in the B-MYB promoter and expression  
425 analysis, we performed a multiple sequence alignment of the promoter region of the B-  
426 MYB gene against the sequencing data reported for primates (**List S1-S24**). The  
427 promoter sequences were downloaded from the Ensembl Genome database and aligned  
428 using the Clustal Omega webtool. The phylogenetic tree evidenced a close relationship  
429 between *Homo sapiens* and *Pan paniscus* followed by *Gorilla gorilla gorilla* and *Pongo*  
430 *abellii* (**Figure 3A**) in light of functional, structural, and evolutionary relationships [53,54].  
431 The data also showed good conservation of the promoter, including in the region of the  
432 identified putative G4s (**Figure S3-S4**). Sequence logos quickly reveal important aspects  
433 of this alignment, as depicted in **Figure 3B**. They also offer a deeper and more accurate  
434 representation of sequence similarity than consensus sequences and the height of each  
435 letter is proportional to the observed frequency [36]. The logos are only considering the  
436 template sequence of the promoter. Comparing all the studied sequences, we observed  
437 that B-MYB 25R is the most conserved along 24 primate's species, followed by B-MYB  
438 26RA and B-MYB 43R.

439 Transcription factor binding sites often overlap or are located near putative G4s [55,56]  
440 and functionally active quadruplex motifs are often associated with one or more  
441 transcription factors [56]. Also, other evidences demonstrate that the transcription  
442 regulation by G4 can be influenced by the recruitment of transcription factors [55,57–60].  
443 Therefore, G4 structures can be considered as novel epigenetic regulators of the  
444 transcription machinery, working in concert with their associated transcription factors to  
445 generate cell-specific transcriptional programs [57]. Taking this into consideration, we  
446 performed a prediction of the transcription factor binding sites, thanks to the PROMO  
447 software ([https://alggen.lsi.upc.es/cgi-  
448 bin/promo\\_v3/promo/promoinit.cgi?dirDB=TF\\_8.3](https://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3)) [37]. The program predicted the  
449 existence of 71 distinct transcription factors binding sites in the analyzed promoter region  
450 of 2000 nucleotides. We observed an increased density of binding sites close to the TSS.  
451 Considering only the regions capable of forming G4 structures, we identified possible  
452 transcription-binding sites (**Figure S4**), also detailed for each G4 region in **Table S3**.  
453 From all predicted transcription factors, we noted the presence of the Sp1 transcription  
454 factor, which is extensively described as the most significantly represented in G4-folded  
455 sequences. Sp1 is a zinc finger transcription factor that binds specifically to GC-rich  
456 motifs, such as the consensus sequence 5'-GGGCGGGG-3', commonly found in the  
457 promoter regions of a wide range of genes [56]. This consensus it is also observed in  
458 the B-MYB 43R sequence. It functions as a transcriptional activator or repressor,  
459 depending on the context and the target gene. This effect is dependent on the behaviour

460 among them, namely if they exert (i) cooperative regulation (facilitate the formation or  
461 stabilization of G4, thereby enhancing transcriptional activation) or if they have (ii)  
462 antagonist interaction (the presence of G4 may hinder the recruitment of Sp1, thereby  
463 modulating gene expression negatively) [61].

464

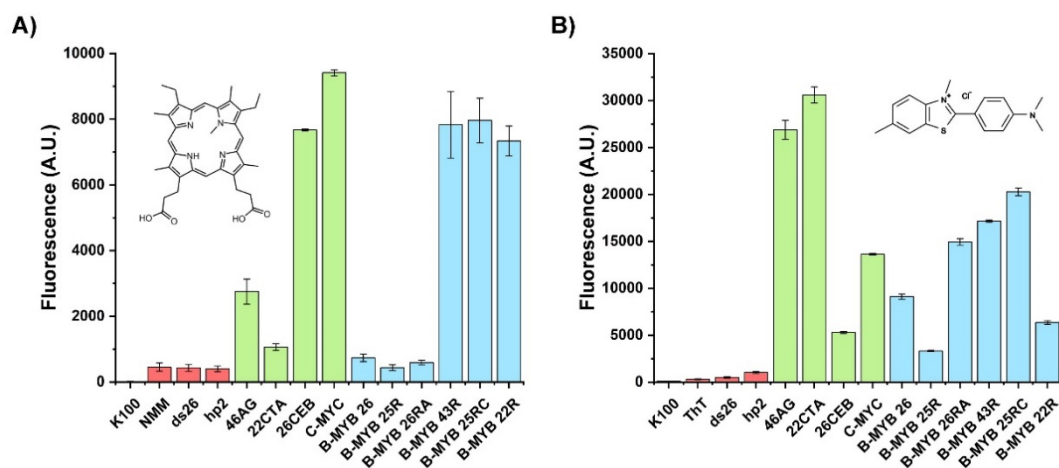
465

## 466 Experimental evidence of G4 forming motifs

### 467 High throughput G4-specific probes

468 After the identification of putative G4 forming motifs in B-MYB proto-oncogene and  
469 analysis of conservation among primates, we wanted to experimentally confirm G4  
470 formation *in vitro* using a combination of techniques [62].

471 A biophysical examination was carried out combining distinct methods to validate G4  
472 formation. We first analyzed the fluorescence emission of G4-specific probes [41],  
473 namely N-methyl mesoporphyrin IX (NMM) [63] and Thioflavin T (ThT) [64] (**Table S2**).  
474 Upon binding to G4s, these dyes emit a strong fluorescence signal. The results are  
475 depicted in **Figure 4**, in which positive (e.g., 46AG, 26CEB, C-MYC) and negative (e.g.,  
476 ds26, hp2) controls were also added for comparison.



477

478 **Figure 4.** Light-up fluorescence experiments using **A)** NMM and **B)** ThT G4 probes. The chemical  
479 structures of the molecules are represented in each panel. For both assays, an increase in  
480 fluorescence happens in the presence of a sequence adopting a G4 fold. Negative controls are  
481 colored in red, positive controls in green and the tested sequences in blue.

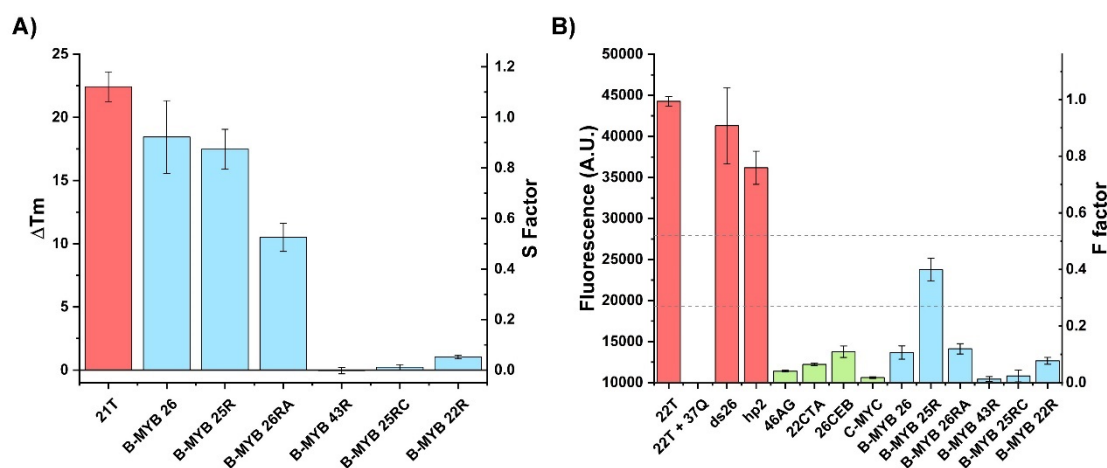
482

483 The results show, in a general way, distinct behaviours for NMM and ThT which can be  
484 explained by the nature and properties of these two ligands.

485 The NMM results (**Figure 4A**) evidenced, in a first look, an important fluorescence only  
486 in the presence of the B-MYB 43R sequence and its designed derivatives (B-MYB 25RC  
487 and B-MYB22R). This result not only provides evidence for G4 formation but also gives  
488 clues on topology, as NMM prefers parallel G4 and tend to have higher fluorescence  
489 levels in the presence of parallel-stranded structures [63].

490 In contrast, the ThT results (**Figure 4B**) tell a different story. The fluorescence profiles  
491 vary among G4 structures, even in the positive control group. The signal emission is  
492 dependent on the binding mode of organic molecules to the oligonucleotide and, in some  
493 cases, can lead to positive and false negative results [64]. From the obtained data, we  
494 can infer the ability of all sequences identified in the B-MYB promoter to form a G4,  
495 except for B-MYB 25R, which shows a low fluorescence level compared to all positive  
496 controls.

497 Proceeding with biophysical evaluation, and continuing with high throughput  
498 approaches, we then performed FRET-based experiments, namely employing the  
499 recently described FRET-MC [38] and iso-FRET [39] methods. Both methods allow the  
500 multiplex testing of sequences in parallel and results are presented in **Figure 5**.



501

502 **Figure 5.** Evidence of formation of G4 structures by B-MYB sequences in promoter region using  
503 FRET experiments. **A)** FRET-MC assays show the formation of G4 in B-MYB 43R and derivative  
504 sequences (B-MYB 25RC and B-MYB 22R), also corroborated by S factor. **B)** iso-FRET results  
505 of sequences under study and controls performed at 37°C. Negative controls are colored in red,  
506 positive controls in green and the tested sequences in blue.

507

508 The FRET-MC assay is based on the ability of the tested sequence to act as a competitor  
509 for the well-known PhenDC3 G4 ligand [65]. If the competitor is able to form a G4, it will  
510 trap the compound, leading to a noticeable decrease in ligand-induced stabilization ( $\Delta T_m$ )  
511 of the dual-labeled telomeric G4-forming probe (F21T) [38].

512 As demonstrated by **Figure 5A**, the only sequences demonstrating efficient competition  
513 were B-MYB 43R, B-MYB 25RC and B-MYB 22R, as shown by a strong reduction in  $\Delta T_m$   
514 and an S Factor value around 0. For the other sequences, the same effects were not  
515 observed. Nevertheless, we cannot categorically exclude G4 formation, as it is important  
516 to notice that the main limitation of the FRET-MC method is its inability to detect unstable  
517 quadruplexes, having a low thermal stability. In that case, at the temperature where F21T  
518 begins to melt, these unstable G4s are “seen” as single strands, as previously reported  
519 for other well-known low stability G4s [38].

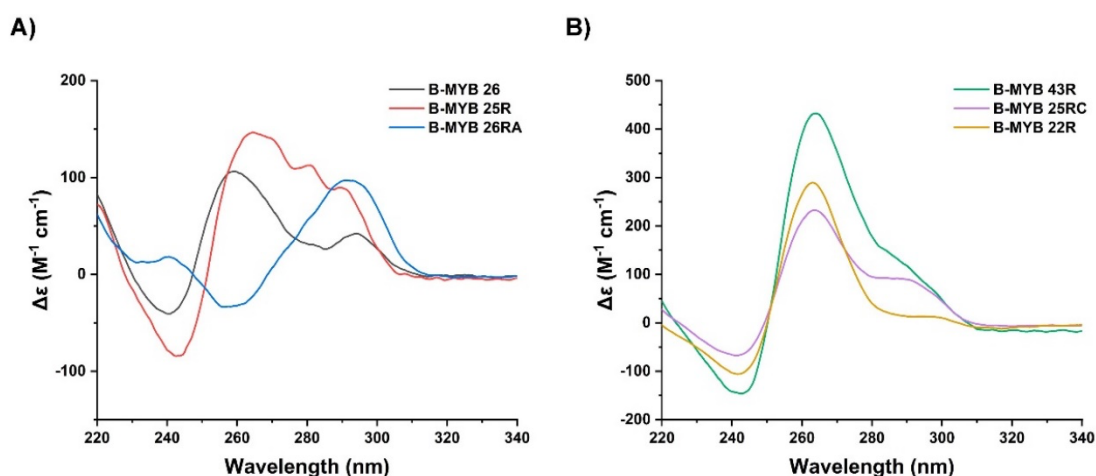
520 To overcome this limitation, a new method was developed employing an isothermal  
521 version of the competition assay called iso-FRET [39]. In the same way, G4-forming  
522 competitors act as decoys and capture the PhenDC3 ligand, preventing it from binding  
523 to the 37Q probe. This in turn leads to the hybridization to the FAM-labeled short  
524 complementary C-rich strand (F22), resulting in fluorescence quenching. On the other  
525 hand, if the competitor does not form a G4, it cannot act as a trap, and PhenDC3 remains  
526 bound to 37Q, preventing the formation of the F22 + 37Q complex. In addition to the  
527 studied sequences, negative control and positive controls were used for comparison. A  
528 potential caveat of this assay is a partial complementarity of competitors with F22, which  
529 was checked by calculating the CF Factor (**Table S1**). All the competitors showed values  
530 lower than 0.68, meaning that the assay could be used [39].

531 All iso-FRET experiments were performed at physiological temperature (37°C). The iso-  
532 FRET results (**Figure 5B**) show that all sequences from the B-MYB promoter form a G4,  
533 except B-MYB 25R, as evidenced by fluorescence intensity and F Factor values. The  
534 standard F Factor values are defined according to the temperature that assay was  
535 performed [39]. F values below 0.27 are characteristic of a G4 competitor, above 0.54  
536 correspond to non-G4 forming samples while values in between correspond to unknown  
537 structures. One more time, we found no clear evidence of G4 formation for the B-MYB  
538 25R sequence, in agreement with previous results.

539 The secondary structure of these sequences was then checked by CD spectroscopy.  
540 This technique has the potential to discriminate among G4 topologies as different G4  
541 structures exhibit various CD spectra [66,67]. Parallel G4 conformations exhibit a positive  
542 band at 260–265 nm and a negative band at 240–245 nm, while antiparallel G4

543 structures display a positive band around 290–295 nm and a weaker negative band at  
544 260–265 nm. Finally, hybrid topologies present a positive band at 290 nm, a negative  
545 band at 240 nm and a shoulder around 270 nm.

546 To access the formation and G4 conformation the putative G4 sequences were annealed  
547 in a K100 buffer and spectra were acquired. The CD spectra for all sequences in the  
548 study are presented in **Figure 6**. As a general tendency, the B-MYB G4 sequences  
549 showed the ability to assemble into a G4 structure with distinct conformations (parallel  
550 and antiparallel), except for B-MYB 25R (described in details below).



551

552 **Figure 6.** Circular dichroism spectra of the G-rich sequences identified in the B-MYB promoter  
553 region. The spectra were acquired in K100 buffer at 20°C.

554

555 Analyzing each sequence in detail, the sequence B-MYB 26 demonstrated a mix of  
556 conformations including parallel (predominant) and antiparallel topologies as shown by  
557 the characteristic CD bands. Additionally, the effect of the nature of the cation was also  
558 evaluated by replacing 100 mM  $K^+$  by 100 mM of  $Na^+$ . Results demonstrated the  
559 maintenance of spectral shape associated with a hypochromism of bands at 240 nm and  
560 290 nm (**Figure S5A**).

561 In the case of the B-MYB 25R sequence, the spectrum was atypical, with some bands  
562 matching common features of G4 but also an unexpected band at 280 nm. To decipher  
563 this, experiments with different cations were carried out (**Figure S5B**). Results  
564 demonstrated no changes in the shape and significant alterations in CD signal,  
565 supporting that, despite the presence of bands suggesting G4 formation, the shape of  
566 the CD spectrum rather suggests the formation of Watson-Crick complexes, as observed  
567 for G/C rich heterogeneous DNA (e.g., 71 % with *Sarcina Lutea* DNA or 42% of Calf

568 Thymus DNA) or G/C rich oligonucleotides with G block at the 5' end, e.g., GGGGCCCC  
 569 or GGCGCC [68]. This observation illustrates the potential limitation of CD analysis,  
 570 arguing for combining several methods [62].

571 The B-MYB 26RA was the only one that displayed a CD spectra indicative of an  
 572 antiparallel topology. Salt changes (replacing 100 mM of KCl by 100 mM NaCl) lead to  
 573 an interconversion from the antiparallel to a hybrid topology (**Figure S5C**). We also  
 574 observed a topology conversion from antiparallel to hybrid topology in response to  
 575 temperature increases in the K100 buffer (**Figure S6**).

576 Finally, B-MYB 43R and its derivative sequences (B-MYB 25RC and B-MYB 22R) all  
 577 displayed a CD spectra compatible with a parallel topology (**Figure 6**) and a K<sup>+</sup>  
 578 dependency, as denoted by a hyperchromism with increasing cation concentration  
 579 (**Figure S7**).

580 All of these conclusions were supported by the quantitative conformation *r* index and are  
 581 summarized in **Table 1** [40,69]. This parameter is calculated as detailed in the  
 582 methodology section and each *r* value is associated with a specific topology, namely  $1 >$   
 583  $r \geq 0.5$  for parallel,  $0.5 > r > 0$  for hybrid and  $r < 0$  for antiparallel conformations.

584 Using CD-melting experiments the thermal stability of oligonucleotides was evaluated by  
 585 monitoring the characteristic G4 bands (**Figure S8**). For all the sequences the  $T_m$  values  
 586 are listed in **Table 1** and **Table S4**. While some  $T_m$  were close to the physiological  
 587 temperature, three sequences, all derived from the same region just upstream of the  
 588 TSS, exhibited strikingly higher  $T_m$  values, above 80°C. This contrasts with the  $T_m$  of the  
 589 other sequences, between 43.2 and 58.7°C. We observed that the B-MYB 22R sequence  
 590 presents the highest  $T_m$  (87.9°C).

591

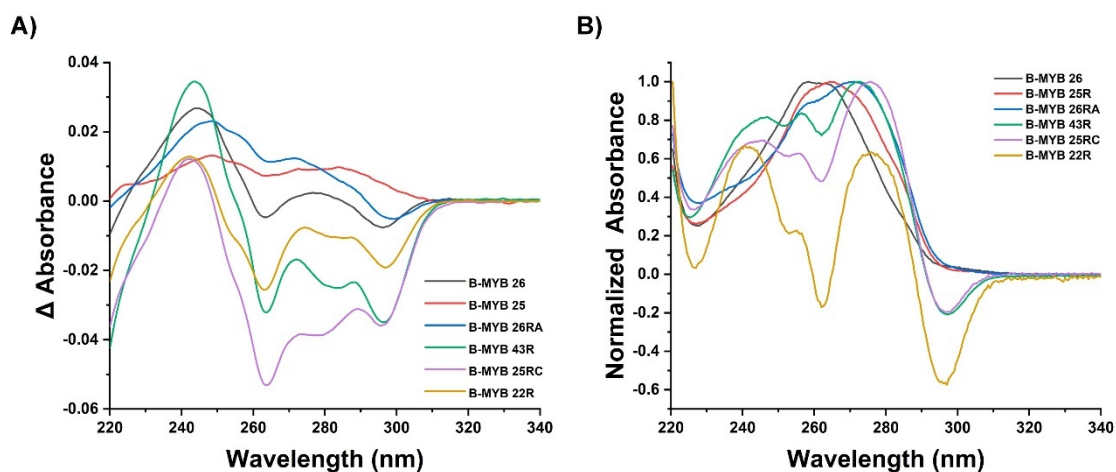
**Table 1.** Values of *r* index and respective topologies of studied sequences, calculated throughout CD data, and melting temperature ( $T_m$ ) in K100 buffer acquired by CD-melting

|                   | <i>r</i> index | Topology     | $T_m$ (°C) at 260 nm |
|-------------------|----------------|--------------|----------------------|
| <b>B-MYB 26</b>   | 0.72           | Parallel     | 43.2                 |
| <b>B-MYB 25R</b>  | 0.62           | Parallel     | 58.7                 |
| <b>B-MYB 26RA</b> | -0.25          | Antiparallel | 50.6                 |
| <b>B-MYB 43R</b>  | 0.78           | Parallel     | 83.7                 |
| <b>B-MYB 25RC</b> | 0.71           | Parallel     | 78.1                 |
| <b>B-MYB 22R</b>  | 0.95           | Parallel     | 87.9                 |

592



593 We then carried out isothermal and thermal difference spectra measurements (IDS and  
 594 TDS, respectively) to confirm the ability of identified sequences to form a G4 structure.  
 595 Both methods are based on recording the spectra of unfolded and folded forms of the  
 596 same sample and performing the arithmetic difference between the two [42]. The  
 597 approaches differ in the variable responsible for the folding/unfolding process, namely  
 598 temperature or presence of salt. Thus, the IDS method does not involve temperature  
 599 variation (isothermal process), only salt addition as a driving force to induce folding,  
 600 which can replicate the absorbance characteristics more accurately [44,45] and give  
 601 distinct indicators of various structural conformations of DNA [29]. The IDS spectra are  
 602 depicted in **Figure 7A**. The IDS spectra are indicative of G4 formation, as denoted by  
 603 the hypochromism at 295 nm and 260 nm and positive bands at 240 nm, for all  
 604 sequences with the exception of B-MYB 25.



605

606 **Figure 7.** IDS and TDS spectra of B-MYB G-rich sequences. Both **A)** IDS spectra and **B)** TDS  
 607 spectra were acquired in the K10 buffer. The spectra were obtained by calculating the arithmetic  
 608 difference between the spectra in folded and unfolded states.

609

610 In the case of TDS, the spectral shape captures the nuances of base stacking  
 611 interactions that are specific to each kind of nucleic acid structure [42,43]. Given the high  
 612 thermal stabilization denoted by CD-melting the spectra were recorded in K10 buffer to  
 613 ensure unfolding at high temperature. The results are presented in **Figure 7B** and  
 614 evidence G4 formation for B-MYB 43R, B-MYB 25RC and B-MYB 22R. This conclusion  
 615 is supported by the bands at 245 nm, 273 and 295 nm that match the postulated  
 616 characteristic values for G4 structures [42]. The intense hypochromic band at 295 nm  
 617 (typical of the G4 topologies) is explained by the tetrads stacking in the folding and  $n \rightarrow \pi^*$   
 618 and  $\pi \rightarrow \pi^*$  transitions of bases [70]. As cation concentration plays an important role in



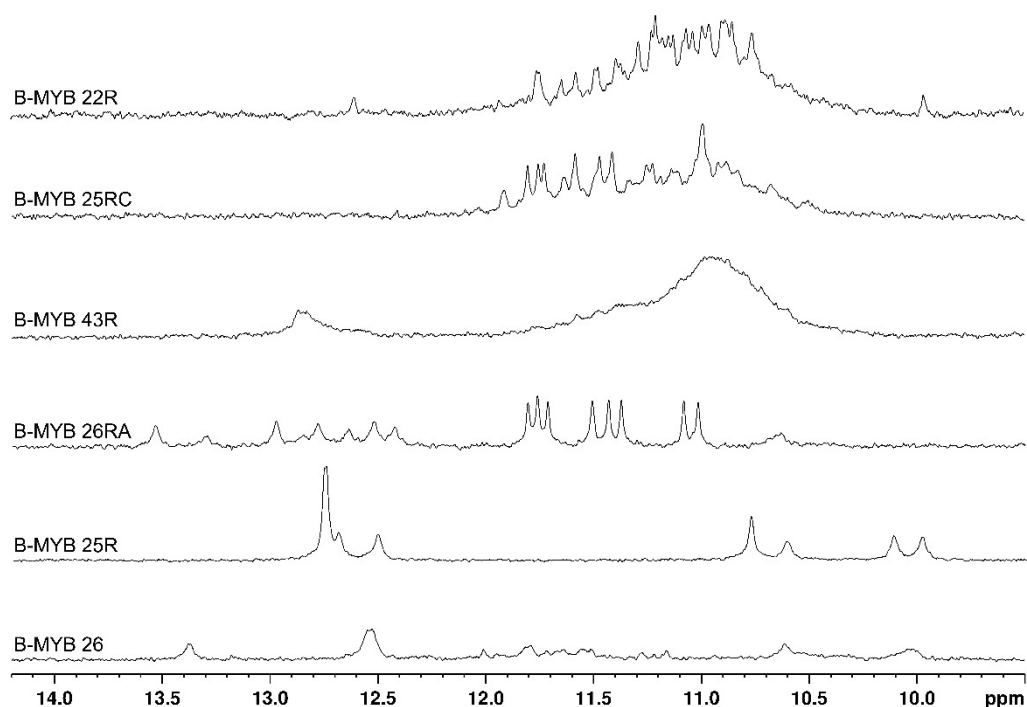
619 the folding and stability of G4 structures, we hypothesized that the absence of  
620 characteristic bands in the other sequences (B-MYB 26, B-MYB 25R and B-MYB 26RA)  
621 could be explained by the low potassium concentration chosen. We therefore repeated  
622 the assay in the K100 buffer and the corresponding TDS spectra (**Figure S9**) exhibited  
623 a faint negative band at 295 nm, possibly indicative of G4 formation.

624 Additionally, we carried out UV-melting experiments at 295 nm and then performed a  
625 van't Hoff analysis to calculate the thermodynamic parameters [46,47]. We could only  
626 obtain exploitable melting profiles allowing van't Hoff analyses for sequences B-MYB 26,  
627 B-MYB 43R, B-MYB 25RC and B-MYB 22R (**Figure S10**). Assuming a two-state process  
628 and an intramolecular equilibrium, as suggested by the absence of a significant  
629 hysteresis between cooling and heating curves, we performed a thermodynamics  
630 analysis. The  $T_m$  values and thermodynamics parameters as listed in **Tables S5-S6**.  
631 Overall, the thermodynamic data suggest a spontaneous (at low temperatures) and  
632 strongly enthalpy-driven reaction for G4 formation.

633 Finally, in order to once again validate G4 formation, NMR assays were carried out. This  
634 technique stands out for its atomic resolution and its ability to discriminate the type of  
635 bonds between nucleic acids. The  $^1\text{H}$  NMR experiments began with the KCl titrations to  
636 evaluate the effect of cation concentration on the G4 structure. For each titration point,  
637 a proper amount of KCl was directly added to the tube and the sample annealed as  
638 previously described. The spectra of the last titration point (100 mM of KCl) are depicted  
639 in **Figure 8**. One can observe the presence of two types of imino protons resonances: i)  
640 10-12 ppm characteristic from Hoogsteen pairing and ii) 12-14 ppm typical from Watson-  
641 Crick bonds for all sequences, except B-MYB 25RC. In addition, the  $^1\text{H}$  NMR spectra  
642 show more or less well-resolved spectra with defined peaks and enable us to deduce  
643 that single conformations are present, except for B-MYB 43R that shows a broad  
644 spectrum, perhaps as a result of its long length. Aiming to understand the feasibility of  
645 structure formation at physiological temperature and access the effect of temperature in  
646 secondary structure, more spectra were acquired at 37°C (**Figure S11**). Comparing the  
647 spectra at 20°C and 37°C allowed us to observe the maintenance of spectral shape,  
648 namely in the guanine imino proton region, for all sequences except B-MYB 26, B-MYB  
649 25R and B-MYB 26RA that exhibited a temperature-dependent behavior.

650

651



652

653 **Figure 8.**  $^1\text{H}$  NMR spectroscopy of B-MYB sequences in the study. The spectra were acquired in  
 654 K100 buffer at  $20^\circ\text{C}$ . Note the presence of two distinct sets of imino protons: 10-12 ppm  
 655 characteristic of Hoogsteen bonds formation and 12-14 ppm typical for Watson-Crick pairing.

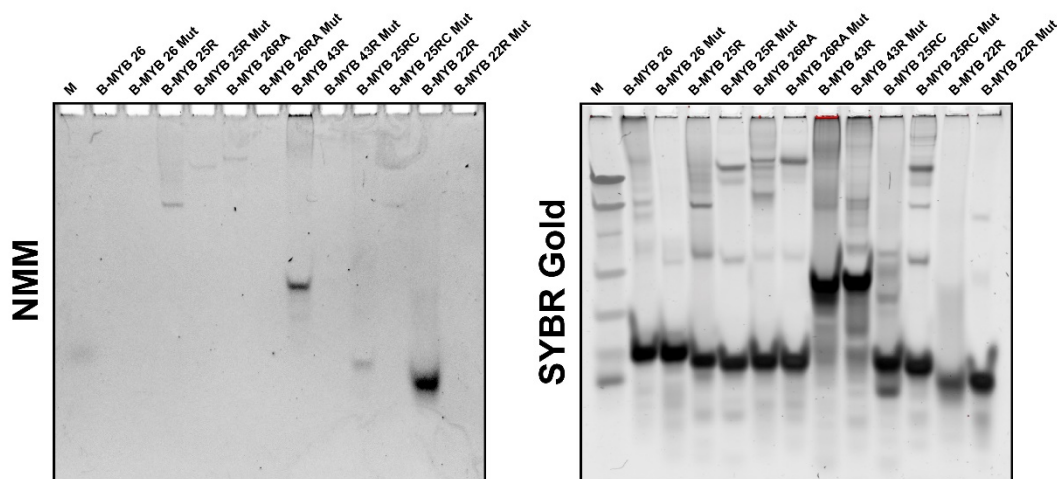
656

657 Analyzing in detail the  $\text{K}^+$  titrations of B-MYB sequences (**Figures S12-S17**) we observed  
 658 a general induction of G4 and the appearance of WC bonds signal upon salt addition,  
 659 except for B-MYB 25R sequence that did not show any difference in imino signal along  
 660 the titration (**Figure S13**). For this sequence, the spectral shape is maintained and gives  
 661 us information about a robust fold independent of salt concentration or nature as  
 662 confirmed by CD results (**Figure S5B**). This intriguing result can also support the  
 663 presence of G/C-rich heterogeneous DNA, which is capable of establishing Hoogsteen  
 664 bonds but not able to form a stable G4. Further experiments will be needed to decipher  
 665 the structural behavior of the structure in response to the type of cation and  
 666 concentration.

667 For B-MYB 26RA, well-resolved Hoogsteen imino peaks only appear at high salt  
 668 concentrations and WC imino signals are hugely altered, suggesting the pivotal role of  
 669 cations for the structure (**Figure S14**). A similar stance is observed for B-MYB 43R  
 670 (**Figure S15**).

671 After the spectroscopic evaluation of G4 formation in the promoter region of B-MYB  
672 proto-oncogene, we proceeded to the assessment of molecularity of G4 sequences  
673 throughout non-denaturing polyacrylamide gel electrophoresis (PAGE) and size-  
674 exclusion chromatography (SEC).

675 PAGE analysis was performed thanks to post-staining with a G4-specific light-up probe,  
676 NMM, followed by a general probe, SYBR Gold. During electrophoresis, G4 mobility is  
677 affected by the number of strands and topology, and it is possible to have direct evidence  
678 of G4 by the fluorescent probes, which bind and illuminate G4 bands [62]. For this  
679 experiment, mutated sequences (**Table S1**) were designed using the G4Killer algorithm  
680 (<https://bioinformatics.ibp.cz/#/analyse/g4-killer>) and were run parallelly to the B-MYB  
681 oligonucleotides in the study. The gel stained with an NMM light-up probe (**Figure 9**)  
682 evidenced a single band and a G4 formation only in sequence B-MYB 43R and  
683 derivatives B-MYB 25RC and B-MYB 22R, in agreement with previous data (NMM light-  
684 up fluorescence studies). In the case of B-MYB 26RA, which forms a G4 with an  
685 antiparallel topology, we did not observe any NMM fluorescent band. This phenomenon  
686 can be explained by the preferential binding of NMM to parallel topologies, as previously  
687 detailed in fluorescence experiments. Mutated sequences did not lead to NMM  
688 fluorescent bands, as expected. SYBR gold was then used to stain and prove the  
689 presence of oligos in the gel (**Figure 8**).



690

691 **Figure 9.** Detection of G4 by NMM or SYBR Gold staining after non-denaturing polyacrylamide  
692 gel electrophoresis (PAGE). Samples were placed onto a 20% non-denaturing acrylamide gel  
693 supplemented with 10 mM KCl after being prepared in a 50 mM Tris/HCl, pH 7.5, and 100 mM  
694 KCl solution. The gel was run at 4°C.

695

696 While several independent methods allowed to demonstrate G4 formation, independent  
 697 techniques are required to determine the molecularity of these structures, such as SEC.  
 698 Compared to non-denaturing gel electrophoresis (PAGE), SEC allows us to evaluate the  
 699 hydrodynamic volume of oligonucleotides, under their native settings and uses the  
 700 absorbance signal to calculate the relative amount of each structure [62]. The obtained  
 701 chromatograms are presented in **Figure S19**. From these results, we found that inter  
 702 molecular complexes (dimers or tetramers) were observed for most oligonucleotides,  
 703 with the exception of B-MYB 26 and B-MYB 26RA which are eluted as monomers.

704 To easily understand all the biophysical information, all the data were summarized in  
 705 **Table 2**.

706

**Table 2.** Summary of the results obtained with different biophysical methods. -: no evidence for G4 formation; +: results in agreement with G4 formation; ++: results strongly supporting G4 formation; ?: unclear. For SEC (size exclusion chromatography), we indicate the molecularity (intra- or inter-molecular). All acronyms are defined in the experimental section.

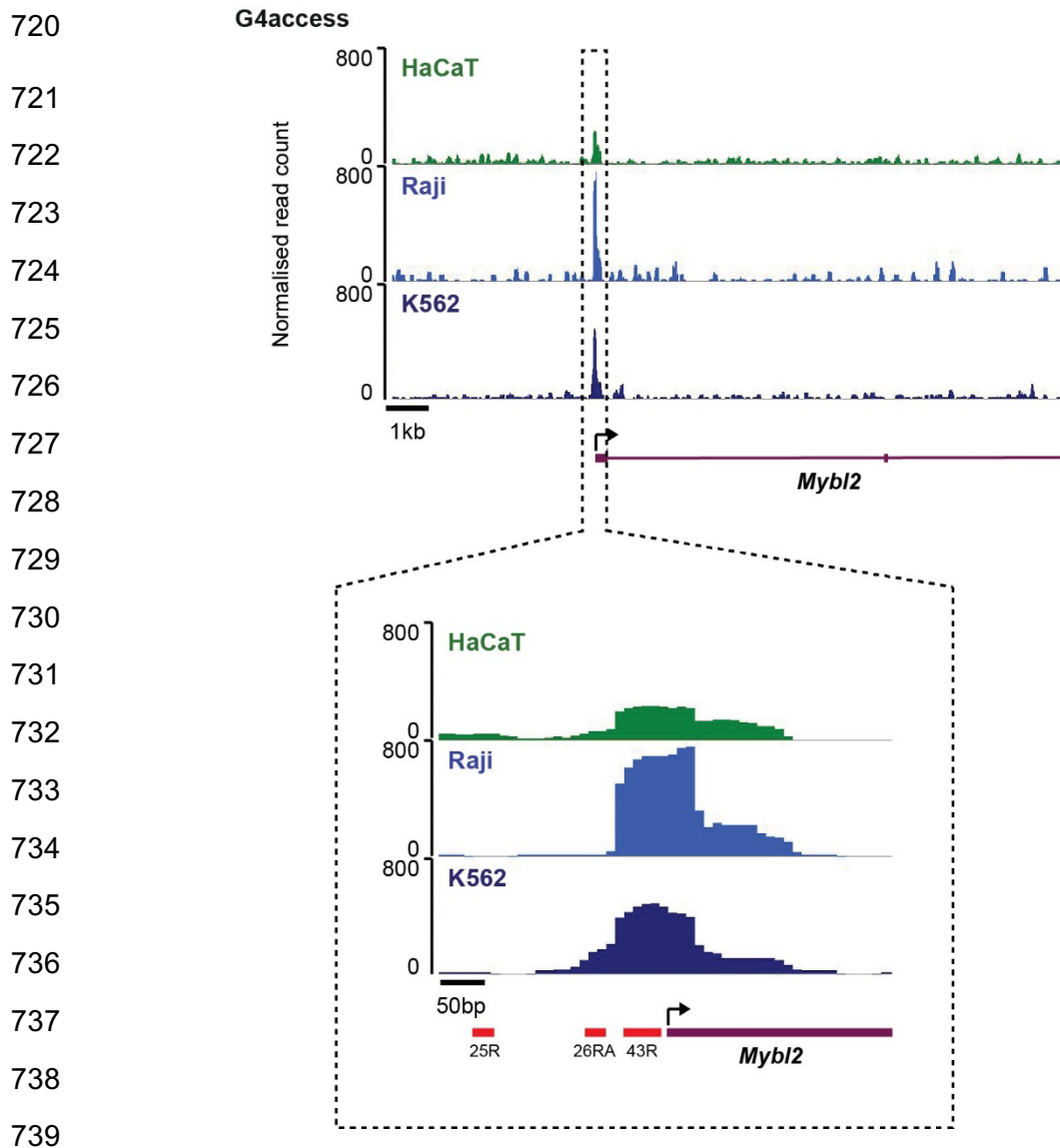
| <b>Sequence</b>   | <b>NMM</b> | <b>ThT</b> | <b>FRET-<br/>MC</b> | <b>Iso-<br/>FRET</b> | <b>CD</b> | <b>IDS</b> | <b>TDS</b> | <b>NMR</b> | <b>PAGE</b> | <b>SEC</b> | <b>Conclusion</b> |
|-------------------|------------|------------|---------------------|----------------------|-----------|------------|------------|------------|-------------|------------|-------------------|
| <i>B-MYB 26</i>   | -          | +          | -                   | ++                   | +         | +          | +          | +          | -           | intra      | G4                |
| <i>B-MYB 25R</i>  | -          | ?          | -                   | ?                    | ?         | ?          | -          | +          | -           | inter      | ?                 |
| <i>B-MYB 26RA</i> | -          | ++         | ?                   | ++                   | ++        | +          | +          | +          | -           | intra      | G4                |
| <i>B-MYB 43R</i>  | ++         | ++         | ++                  | ++                   | ++        | ++         | ++         | +          | +           | intra      | G4                |
| <i>B-MYB 25RC</i> | ++         | ++         | ++                  | ++                   | ++        | ++         | ++         | +          | +           | mix        | G4                |
| <i>B-MYB 22R</i>  | ++         | ?          | ++                  | ++                   | ++        | ++         | ++         | +          | +           | inter      | G4                |

707

708

### 709 **G4access evaluation**

710 After the bioinformatic identification of putative G4 sequences and our biophysical  
 711 evaluation, it was also important to evaluate the *in-cell* relevance and formation of  
 712 structures. For this, we employed the recently developed methodology called G4access  
 713 [33] that is capable of isolating and sequencing G4s associated with open chromatin via  
 714 nuclease digestion in an antibody- and crosslinking-independent manner. G4access  
 715 allows the study of G4s and cellular dynamics and link this with chromatin opening,  
 716 transcription, and their antagonism to DNA methylation. Taking all of these into  
 717 consideration we took advantage of G4access data performed in three different cell lines  
 718 (HaCat, Raji, K562) and we investigated the formation and conservation of G4s at the B-  
 719 MYB promoter region and around the TSS.



740 **Figure 10.** Assessment of G4 formation in living cells using G4access. G4access was performed  
741 in HaCaT, Raji and K562 cells. A genome browser view of G4access signal at the MYBL2 locus  
742 is shown (upper panel) and a zoom-in view depicts the signal at the B-MYB 25R, B-MYB 26RA  
743 and B-MYB 43R predicted G4 (lower panel). Only the peak at B-MYB 43R was called by MACS2  
744 as significant with a  $p$ -value  $< 10^{-5}$ .

745

746 The results, presented in **Figure 10**, evidenced one reproducible and highly significant  
747 enriched G4 signal located 44 nucleotides upstream of the MYBL2 TSS, which  
748 corresponds to one of the quadruplex motifs predicted by G4Hunter (B-MYB 43R  
749 sequence; with a G4Hunter score above 2.0). Furthermore, no significant G4access  
750 peaks were matching the other predicted G4 sequences, which had a significantly lower  
751 G4Hunter score, associated with a much lower thermal stability. Thus, *in cell* G4  
752 formation was evidenced for the sequence with the highest  $T_m$  and highest G4H score.

753 For the other sequences, which form a G4 at 37 °C *in vitro*, no noticeable G4access peak  
754 was found. This may indicate that the actual G4Hunter threshold for G4 formation is  
755 significantly higher in cells than *in vitro* with isolated oligonucleotides. The chromatin  
756 context and presence of the complementary sequence allowing stable duplex formation  
757 can prevent G4 formation for moderate G4H scores, at least in the absence of a G4  
758 ligand. Esnault *et al.* observed that G4s with low-to-medium *in silico* prediction scores  
759 were preferentially stabilized by the PDS ligand. This would argue that the actual  
760 threshold to be used for *in-cell* study may depend on the presence of G4 stabilizing  
761 factors altering the duplex-quadruplex equilibrium, potentially interfering with the  
762 hybridization of a complementary sequence.

763 Overall, taking the *in-cell* results in consideration, we can hypothesize that while the  
764 optimal threshold for predicting G4 in ssDNA (such as for an aptamer) may be 1.2, as  
765 confirmed experimentally with the sequences analyzed *in vitro* here, *in cell* applications  
766 in a chromatin environment may require a higher threshold.

767 Given the previously described relation between G4access *signals at* promoters and  
768 open chromatin, we investigated nucleosome density and positioning, RNA Polymerase  
769 (Pol) II recruitment around the G4access peaks at the B-MYB promoter region using  
770 published data sets. Clearly, in this area, the G4access signal that relates to the B-MYB  
771 43R G4 sequence correlates to nucleosome exclusion and strong Pol II recruitment  
772 (**Figure S20**), suggesting that this specific G4 could directly contribute to chromatin  
773 opening and transcription activation at this locus. This result is in agreement with the  
774 literature [33,71–73]. We propose that this G4 is formed prior transcription, while the act  
775 of transcription likely reinforces this G4 through the action of negative supercoiling as  
776 previously described at the global scale [33].

777 Regarding the methylation profile, the G4access peak close to TSS and surrounding  
778 areas are essentially unmethylated (**Figure S20**) and define a methylation valley. This  
779 data suggests that G4 formation in the B-MYB promoter antagonizes methylation (and  
780 vice-versa) and associates with nucleosome repositioning on DNA.

781 Thanks to this information we gathered regarding the *in cell* formation of the identified  
782 G4 sequences in the promoter region of B-MYB oncogene, our study opens up a new  
783 perspective for the further functional study and evaluation of the biological role of this  
784 oncogene in cancer. We provide a strong demonstration of G4 formation *in vitro* thanks  
785 to the combination of an unprecedented number of independent methods. Even then,  
786 some ambiguity remains for one motif, B-MYB 25R. In any case, while more than one  
787 motif was shown to adopt a G4 fold, a single peak was found with G4Access, arguing

788 that only the most proximal G-rich sequence adopts a G4 conformation in cells. Future  
789 work will be needed to understand the contribution of G4 structures on B-MYB  
790 transcription.

791

## 792 CONCLUSION

793 In conclusion, the identification of G4 structure in the B-MYB gene opens new  
794 perspectives to target this oncogene. Interestingly, while most predicted sequences were  
795 confirmed *in vitro*, only the most stable one, both in terms of G4Hunter score and thermal  
796 stability, was shown to be significantly formed in cells. This motif is also the one closest  
797 to the TSS, only -44 nucleotides away, therefore highly likely to have an impact on the  
798 transcription of this gene. Further studies will be required to **understand the functional**  
799 **role of this G4 in B-MYB transcription. This work may later open up new ways for the**  
800 **appropriate design of** G4 ligands to target this region, based on high-resolution structure  
801 determination of this quadruplex motif.

802

## 803 SUPPLEMENTARY DATA

804 Supplementary Data are available online and include a list of oligonucleotides and  
805 ligands used in the experiments, epigenetic profile of B-MYB gene considering the  
806 modifications in histones, multiple alignments of promoter regions B-MYB 26 and B-MYB  
807 25R sequences, prediction of transcription factor binding sites in the region of putative  
808 G4 sequences in promoter of B-MYB, CD spectra of B-MYB 43R, B-MYB 25RC and B-  
809 MYB 22R, CD-melting curves of B-MYB 26, B-MYB 25R, B-MYB 26RA, B-MYB 43R, B-  
810 MYB 25RC and B-MYB 22R, TDS spectra of oligonucleotides studied, UV-melting  
811 curves, Van't Hoff equations of B-MYB G4-forming motifs and thermodynamic  
812 parameters extracted from UV-melting curves, <sup>1</sup>H NMR spectra of B-MYB sequences,  
813 SEC chromatography results and correlation of G4access peaks with nucleosome,  
814 promoter methylation and Pol II features in K562 cell line.

815

## 816 ACKNOWLEDGMENTS

817 We thank Lionel Guittat (LOB) for helpful discussions.

818

## 819 FUNDING

820 A. M. acknowledges a doctoral fellowship grant from FCT – Foundation for Science and  
821 Technology ref. 2021.04785.BD.

822 Thanks are due to FCT/MCT for the financial support to CICS-UBI ref. UIDB/00709/2020  
823 and UIDP/00709/2020 research unit, PPBI-Portuguese Platform of Bioluminescence research  
824 unit (POCI-01-0145-FEDER-022122), and to the Portuguese NMR Network



825 (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where  
826 applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and  
827 PIDDAC.

828 C.C. acknowledges the grants from project PAPILOMA ref. CENTRO-01-0145-FEDER-  
829 181235, NRC-LPCC Bolsa Dr Rocha Alves 2022, and Instruct-ERIC Pilot R&D  
830 application ID 2473.

831 J.L.M., C.E. and J.C.A. benefited from support from Inserm, CNRS, INCa G4 Access  
832 and ANR (ANR-20-CE12-0023) grants.

833 P.A.O. acknowledges European Investment Funds by FEDER/COMPETE/POCI-  
834 Operational Competitiveness and Internationalization Program, and National Funds by  
835 FCT-Portuguese Foundation for Science and Technology, under the project  
836 UIDB/04033/2020 (<https://doi.org/10.54499/UIDB/04033/2020>).

837 J.C.A and C.E acknowledges support from ANR (grants ANR-18-CE12-0020 and ANR-  
838 20-CE12-0023), FRM (grant AJE20130728183) and INCA PLbIO (grant 2020-117).

839 Conflict of interest statement. None declared.

840

841 **REFERENCES**

- 842 1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2022. *CA*  
843 *Cancer J Clin* **2022**, *72*, 7–33, doi:10.3322/caac.21708.
- 844 2. Dhanasekaran, R.; Deutzmann, A.; Mahauad-Fernandez, W.D.; Hansen, A.S.;  
845 Gouw, A.M.; Felsher, D.W. The MYC Oncogene — the Grand Orchestrator of  
846 Cancer Growth and Immune Evasion. *Nat Rev Clin Oncol* **2022**, *19*, 23–36,  
847 doi:10.1038/s41571-021-00549-2.
- 848 3. Kontomanolis, E.N.; Koutras, A.; Syllaios, A.; Schizas, D.; Mastoraki, A.; Garpis,  
849 N.; Diakosavvas, M.; Angelou, K.; Tsatsaris, G.; Pagkalos, A.; et al. Role of  
850 Oncogenes and Tumor-Suppressor Genes in Carcinogenesis: A Review.  
851 *Anticancer Res* **2020**, *40*, 6009–6015, doi:10.21873/ANTICANRES.14622.
- 852 4. Mechanisms of Oncogene Activation - Holland-Frei Cancer Medicine - NCBI  
853 Bookshelf Available online: <https://www.ncbi.nlm.nih.gov/books/NBK12538/>  
854 (accessed on 17 July 2022).
- 855 5. JIN, Y.; QI, G.; CHEN, G.; WANG, C.; FAN, X. Association between B-Myb Proto-  
856 Oncogene and the Development of Malignant Tumors (Review). *Oncol Lett* **2021**,  
857 *21*, 1–8, doi:10.3892/ol.2021.12427.
- 858 6. Sala, A. B-MYB, a Transcription Factor Implicated in Regulating Cell Cycle,  
859 Apoptosis and Cancer. *Eur J Cancer* **2005**, *41*, 2479–2484,  
860 doi:10.1016/j.ejca.2005.08.004.
- 861 7. Musa, J.; Aynaud, M.-M.; Mirabeau, O.; Delattre, O.; Grünewald, T.G. MYBL2 (B-  
862 Myb): A Central Regulator of Cell Proliferation, Cell Survival and Differentiation  
863 Involved in Tumorigenesis. *Cell Death Dis* **2017**, *8*, e2895–e2895,  
864 doi:10.1038/cddis.2017.244.
- 865 8. Liu, M.; Du, Q.; Mao, G.; Dai, N.; Zhang, F. MYB Proto-Oncogene like 2 Promotes  
866 Hepatocellular Carcinoma Growth and Glycolysis via Binding to the *Optic Atrophy*  
867 *3* Promoter and Activating Its Expression. *Bioengineered* **2022**, *13*, 5344–5356,  
868 doi:10.1080/21655979.2021.2017630.
- 869 9. Tanaka, Y.; Patestos, N.P.; Maekawa, T.; Ishii, S. B-Myb Is Required for Inner  
870 Cell Mass Formation at an Early Stage of Development. *Journal of Biological*  
871 *Chemistry* **1999**, *274*, 28067–28070, doi:10.1074/jbc.274.40.28067.

- 872 10. Jiao, M.; Zhang, F.; Teng, W.; Zhou, C. MYBL2 Is a Novel Independent Prognostic  
873 Biomarker and Correlated with Immune Infiltrates in Prostate Cancer. *Int J Gen*  
874 *Med* **2022**, Volume 15, 3003–3030, doi:10.2147/IJGM.S351638.
- 875 11. Du, K.; Sun, S.; Jiang, T.; Liu, T.; Zuo, X.; Xia, X.; Liu, X.; Wang, Y.; Bu, Y. E2F2  
876 Promotes Lung Adenocarcinoma Progression through B-Myb- and FOXM1-  
877 Facilitated Core Transcription Regulatory Circuitry. *Int J Biol Sci* **2022**, 18, 4151–  
878 4170, doi:10.7150/ijbs.72386.
- 879 12. Henley, M.J.; Koehler, A.N. Advances in Targeting ‘Undruggable’ Transcription  
880 Factors with Small Molecules. *Nat Rev Drug Discov* **2021**, 20, 669–688,  
881 doi:10.1038/s41573-021-00199-0.
- 882 13. Wei, M.; Yang, R.; Ye, M.; Zhan, Y.; Liu, B.; Meng, L.; Xie, L.; Du, M.; Wang, J.;  
883 Gao, R.; et al. MYBL2 Accelerates Epithelial-Mesenchymal Transition and  
884 Hepatoblastoma Metastasis via the Smad/SNAI1 Pathway. *Am J Cancer Res*  
885 **2022**, 12, 1960.
- 886 14. Chen, X.; Lu, Y.; Yu, H.; Du, K.; Zhang, Y.; Nan, Y.; Huang, Q. Pan-Cancer  
887 Analysis Indicates That MYBL2 Is Associated with the Prognosis and  
888 Immunotherapy of Multiple Cancers as an Oncogene. *Cell Cycle* **2021**, 20, 2291–  
889 2308, doi:10.1080/15384101.2021.1982494.
- 890 15. Deng, S.; Calin, G.A.; Croce, C.M.; Coukos, G.; Zhang, L. Mechanisms of  
891 MicroRNA Deregulation in Human Cancer. *Cell Cycle* **2008**, 7, 2643–2646,  
892 doi:10.4161/cc.7.17.6597.
- 893 16. Guan, Z.; Cheng, W.; Huang, D.; Wei, A. High MYBL2 Expression and  
894 Transcription Regulatory Activity Is Associated with Poor Overall Survival in  
895 Patients with Hepatocellular Carcinoma. *Curr Res Transl Med* **2018**, 66, 27–32,  
896 doi:10.1016/J.RETRAM.2017.11.002.
- 897 17. Fan, X.; Wang, Y.; Jiang, T.; Cai, W.; Jin, Y.; Niu, Y.; Zhu, H.; Bu, Y. B-Myb  
898 Mediates Proliferation and Migration of Non-Small-Cell Lung Cancer via  
899 Suppressing IGFBP3. *Int J Mol Sci* **2018**, 19, 1479, doi:10.3390/ijms19051479.
- 900 18. Le, L.; Luo, J.; Wu, H.; Chen, L.; Tang, X.; Fu, F. Overexpression of MYBL2  
901 Predicts Poor Prognosis and Promotes Oncogenesis in Endometrial Carcinoma.  
902 *European Journal of Histochemistry* **2021**, 65, 3226, doi:10.4081/ejh.2021.3226.

- 903 19. Sun, S.; Fu, Y.; Lin, J. Upregulation of MYBL2 Independently Predicts a Poorer  
904 Prognosis in Patients with Clear Cell Renal Cell Carcinoma. *Oncol Lett* **2020**, *19*,  
905 2765–2772, doi:10.3892/ol.2020.11408.
- 906 20. Hou, X.; Zhang, Y.; Han, S.; Hou, B. A Novel DNA Methylation 10-CpG Prognostic  
907 Signature of Disease-Free Survival Reveal That MYBL2 Is Associated with High  
908 Risk in Prostate Cancer. *Expert Rev Anticancer Ther* **2020**, *20*, 1107–1119,  
909 doi:10.1080/14737140.2020.1838280.
- 910 21. Zhong, F.; Liu, J.; Gao, C.; Chen, T.; Li, B. Downstream Regulatory Network of  
911 MYBL2 Mediating Its Oncogenic Role in Melanoma. *Front Oncol* **2022**, *12*,  
912 doi:10.3389/fonc.2022.816070.
- 913 22. Long, E.M.; Long, M.A.; Tsirigotis, M.; Gray, D.A. Stimulation of the Murine Uchl1  
914 Gene Promoter by the B-Myb Transcription Factor. *Lung Cancer* **2003**, *42*, 9–21,  
915 doi:10.1016/S0169-5002(03)00279-4.
- 916 23. Guiley, K.Z.; Iness, A.N.; Saini, S.; Tripathi, S.; Lipsick, J.S.; Litovchick, L.; Rubin,  
917 S.M. Structural Mechanism of Myb–MuvB Assembly. *Proceedings of the National*  
918 *Academy of Sciences* **2018**, *115*, 10016–10021, doi:10.1073/pnas.1808136115.
- 919 24. Guan, Z.; Cheng, W.; Huang, D.; Wei, A. High MYBL2 Expression and  
920 Transcription Regulatory Activity Is Associated with Poor Overall Survival in  
921 Patients with Hepatocellular Carcinoma. *Curr Res Transl Med* **2018**, *66*, 27–32,  
922 doi:10.1016/j.retram.2017.11.002.
- 923 25. Li, Q.; Wang, M.; Hu, Y.; Zhao, E.; Li, J.; Ren, L.; Wang, M.; Xu, Y.; Liang, Q.;  
924 Zhang, D.; et al. MYBL2 Disrupts the Hippo-YAP Pathway and Confers Castration  
925 Resistance and Metastatic Potential in Prostate Cancer. *Theranostics* **2021**, *11*,  
926 5794–5812, doi:10.7150/thno.56604.
- 927 26. Jin, Y.; Zhu, H.; Cai, W.; Fan, X.; Wang, Y.; Niu, Y.; Song, F.; Bu, Y. B-Myb Is Up-  
928 Regulated and Promotes Cell Growth and Motility in Non-Small Cell Lung Cancer.  
929 *Int J Mol Sci* **2017**, *18*, 860, doi:10.3390/ijms18060860.
- 930 27. Xiong, Y.-C.; Wang, J.; Cheng, Y.; Zhang, X.-Y.; Ye, X.-Q. Overexpression of  
931 MYBL2 Promotes Proliferation and Migration of Non-Small-Cell Lung Cancer via  
932 Upregulating NCAPH. *Mol Cell Biochem* **2020**, *468*, 185–193,  
933 doi:10.1007/s11010-020-03721-x.

- 934 28. Rigo, R.; Palumbo, M.; Sissi, C. G-Quadruplexes in Human Promoters: A  
935 Challenge for Therapeutic Applications. *Biochimica et Biophysica Acta (BBA) -*  
936 *General Subjects* **2017**, *1861*, 1399–1413, doi:10.1016/j.bbagen.2016.12.024.
- 937 29. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity  
938 with G4Hunter. *Nucleic Acids Res* **2016**, *44*, 1746–1759, doi:10.1093/nar/gkw006.
- 939 30. Balasubramanian, S.; Hurley, L.H.; Neidle, S. Targeting G-Quadruplexes in Gene  
940 Promoters: A Novel Anticancer Strategy? *Nat Rev Drug Discov* **2011**, *10*, 261–  
941 275, doi:10.1038/nrd3428.
- 942 31. Siddiqui-Jain, A.; Grand, C.L.; Bearss, D.J.; Hurley, L.H. Direct Evidence for a G-  
943 Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to  
944 Repress c-MYC Transcription. *Proceedings of the National Academy of Sciences*  
945 **2002**, *99*, 11593–11598, doi:10.1073/pnas.182256799.
- 946 32. Papp, C.; Mukundan, V.T.; Jenjaroenpun, P.; Winnerdy, F.R.; Ow, G.S.; Phan,  
947 A.T.; Kuznetsov, V.A. Stable Bulged G-Quadruplexes in the Human Genome:  
948 Identification, Experimental Validation and Functionalization. *Nucleic Acids Res*  
949 **2023**, *51*, 4148–4177, doi:10.1093/nar/gkad252.
- 950 33. Esnault, C.; Magat, T.; Zine El Aabidine, A.; Garcia-Oliver, E.; Cucchiaroni, A.;  
951 Bouchouika, S.; Lleres, D.; Goerke, L.; Luo, Y.; Verga, D.; et al. G4access  
952 Identifies G-Quadruplexes and Their Associations with Open Chromatin and  
953 Imprinting Control Regions. *Nat Genet* **2023**, *55*, 1359–1369,  
954 doi:10.1038/s41588-023-01437-4.
- 955 34. Martin, F.J.; Amode, M.R.; Aneja, A.; Austine-Orimoloye, O.; Azov, A.G.; Barnes,  
956 I.; Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; et al. Ensembl 2023. *Nucleic Acids*  
957 *Res* **2023**, *51*, D933–D941, doi:10.1093/nar/gkac958.
- 958 35. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview  
959 Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench.  
960 *Bioinformatics* **2009**, *25*, 1189–1191, doi:10.1093/bioinformatics/btp033.
- 961 36. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence  
962 Logo Generator. *Genome Res* **2004**, *14*, 1188–1190, doi:10.1101/gr.849004.
- 963 37. Tang, Z.; Li, C.; Kang, B.; Gao, G.; Li, C.; Zhang, Z. GEPIA: A Web Server for  
964 Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic*  
965 *Acids Res* **2017**, *45*, W98–W102, doi:10.1093/nar/gkx247.

- 966 38. Luo, Y.; Granzhan, A.; Verga, D.; Mergny, J. FRET-MC: A Fluorescence Melting  
967 Competition Assay for Studying G4 Structures in Vitro. *Biopolymers* **2021**, *112*,  
968 doi:10.1002/bip.23415.
- 969 39. Luo, Y.; Verga, D.; Mergny, J.-L. Iso-FRET: An Isothermal Competition Assay to  
970 Analyze Quadruplex Formation *in Vitro*. *Nucleic Acids Res* **2022**, *1*, 13–14,  
971 doi:10.1093/nar/gkac465.
- 972 40. Chen, J.; Cheng, M.; Stadlbauer, P.; Šponer, J.; Mergny, J.-L.; Ju, H.; Zhou, J.  
973 Exploring Sequence Space to Design Controllable G-Quadruplex Topology  
974 Switches. *CCS Chemistry* **2022**, *4*, 3036–3050,  
975 doi:10.31635/ccschem.021.202101357.
- 976 41. Cantara, A.; Luo, Y.; Dobrovolná, M.; Bohalova, N.; Fojta, M.; Verga, D.; Guittat,  
977 L.; Cucchiari, A.; Savrimoutou, S.; Häberli, C.; et al. G-Quadruplexes in Helminth  
978 Parasites. *Nucleic Acids Res* **2022**, *50*, 2719–2735, doi:10.1093/nar/gkac129.
- 979 42. Mergny, J.L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J.B. Thermal Difference  
980 Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res* **2005**,  
981 *33*, e138–e138, doi:10.1093/NAR/GNI134.
- 982 43. Karsisiotis, A.I.; Hessari, N.M.; Novellino, E.; Spada, G.P.; Randazzo, A.; Webba  
983 da Silva, M. Topological Characterization of Nucleic Acid G-Quadruplexes by UV  
984 Absorption and Circular Dichroism. *Angewandte Chemie International Edition*  
985 **2011**, *50*, 10645–10648, doi:10.1002/anie.201105193.
- 986 44. Miranda, A.; Santos, T.; Largy, E.; Cruz, C. Locking up the AS1411 Aptamer with  
987 a Flanking Duplex: Towards an Improved Nucleolin-Targeting. *Pharmaceuticals*  
988 **2021**, *14*, 121, doi:10.3390/ph14020121.
- 989 45. Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex  
990 Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double  
991 Switches. *J Am Chem Soc* **2016**, *138*, 2780–2792, doi:10.1021/jacs.5b13130.
- 992 46. Mergny, J.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr Protoc Nucleic Acid*  
993 *Chem* **2009**, *37*, 17.1.1–17.1.15, doi:10.1002/0471142700.nc1701s37.
- 994 47. Mergny, J.-L.; Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides*  
995 **2003**, *13*, 515–537, doi:10.1089/154545703322860825.

- 996 48. Largy, E.; Mergny, J.-L. Shape Matters: Size-Exclusion HPLC for the Study of  
997 Nucleic Acid Structural Polymorphism. *Nucleic Acids Res* **2014**, *42*, e149–e149,  
998 doi:10.1093/nar/gku751.
- 999 49. Guédin, A.; Lin, L.Y.; Armane, S.; Lacroix, L.; Mergny, J.-L.; Thore, S.; Yatsunyk,  
1000 L.A. Quadruplexes in 'Dicty': Crystal Structure of a Four-Quartet G-Quadruplex  
1001 Formed by G-Rich Motif Found in the Dictyostelium Discoideum Genome. *Nucleic*  
1002 *Acids Res* **2018**, *46*, 5297–5307, doi:10.1093/nar/gky290.
- 1003 50. Fan, X.; Wang, Y.; Jiang, T.; Liu, T.; Jin, Y.; Du, K.; Niu, Y.; Zhang, C.; Liu, Z.; Lei,  
1004 Y.; et al. B-Myb Accelerates Colorectal Cancer Progression through Reciprocal  
1005 Feed-Forward Transactivation of E2F2. *Oncogene* **2021**, *40*, 5613–5625,  
1006 doi:10.1038/s41388-021-01961-9.
- 1007 51. Iness, A.N.; Rubinsak, L.; Meas, S.J.; Chaoul, J.; Sayeed, S.; Pillappa, R.; Temkin,  
1008 S.M.; Dozmorov, M.G.; Litovchick, L. Oncogenic B-Myb Is Associated With  
1009 Deregulation of the DREAM-Mediated Cell Cycle Gene Expression Program in  
1010 High Grade Serous Ovarian Carcinoma Clinical Tumor Samples. *Front Oncol*  
1011 **2021**, *11*, 514, doi:10.3389/fonc.2021.637193.
- 1012 52. Kimura, H. Histone Modifications for Human Epigenome Analysis. *Journal of*  
1013 *Human Genetics* **2013**, *58*, 439–445, doi:10.1038/jhg.2013.66.
- 1014 53. Zihlman, A.L.; Bolter, D.R. Body Composition in *Pan Paniscus* Compared with  
1015 *Homo Sapiens* Has Implications for Changes during Human Evolution.  
1016 *Proceedings of the National Academy of Sciences* **2015**, *112*, 7466–7471,  
1017 doi:10.1073/pnas.1505071112.
- 1018 54. Bruce, E.J.; Ayala, F.J. Phylogenetic Relationships Between Man and the Apes:  
1019 Electrophoretic Evidence. *Evolution (N Y)* **1979**, *33*, 1040, doi:10.2307/2407465.
- 1020 55. De Rache, A.; Marquevielle, J.; Bouaziz, S.; Vialet, B.; Andreola, M.-L.; Mergny,  
1021 J.-L.; Amrane, S. Structure of a DNA G-Quadruplex That Modulates SP1 Binding  
1022 Sites Architecture in HIV-1 Promoter. *J Mol Biol* **2023**, 168359,  
1023 doi:10.1016/j.jmb.2023.168359.
- 1024 56. Kumar, P.; Yadav, V.K.; Baral, A.; Kumar, P.; Saha, D.; Chowdhury, S. Zinc-Finger  
1025 Transcription Factors Are Associated with Guanine Quadruplex Motifs in Human,  
1026 Chimpanzee, Mouse and Rat Promoters Genome-Wide. *Nucleic Acids Res* **2011**,  
1027 *39*, 8005–8016, doi:10.1093/nar/gkr536.

- 1028 57. Lago, S.; Nadai, M.; Cernilogar, F.M.; Kazerani, M.; Domínguez Moreno, H.;  
1029 Schotta, G.; Richter, S.N. Promoter G-Quadruplexes and Transcription Factors  
1030 Cooperate to Shape the Cell Type-Specific Transcriptome. *Nat Commun* **2021**,  
1031 *12*, 3885, doi:10.1038/s41467-021-24198-2.
- 1032 58. Uribe, D.J.; Guo, K.; Shin, Y.-J.; Sun, D. Heterogeneous Nuclear  
1033 Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular  
1034 Endothelial Growth Factor Promoter through Interaction with Secondary DNA  
1035 Structures. *Biochemistry* **2011**, *50*, 3796–3806, doi:10.1021/bi101633b.
- 1036 59. Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K.K.; Xodo, L.E. The  
1037 KRAS Promoter Responds to Myc-Associated Zinc Finger and Poly(ADP-Ribose)  
1038 Polymerase 1 Proteins, Which Recognize a Critical Quadruplex-Forming GA-  
1039 Element. *Journal of Biological Chemistry* **2010**, *285*, 22003–22016,  
1040 doi:10.1074/jbc.M110.101923.
- 1041 60. Raiber, E.-A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A Non-  
1042 Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 in  
1043 Vitro. *Nucleic Acids Res* **2012**, *40*, 1499–1508, doi:10.1093/nar/gkr882.
- 1044 61. Balasubramanian, S.; Hurley, L.H.; Neidle, S. Targeting G-Quadruplexes in Gene  
1045 Promoters: A Novel Anticancer Strategy? *Nature Reviews Drug Discovery* **2011**  
1046 *10:4* **2011**, *10*, 261–275, doi:10.1038/nrd3428.
- 1047 62. Luo, Y.; Granzhan, A.; Marquevielle, J.; Cucchiari, A.; Lacroix, L.; Amrane, S.;  
1048 Verga, D.; Mergny, J.-L. Guidelines for G-Quadruplexes: I. In Vitro  
1049 Characterization. *Biochimie* **2023**, *214*, 5–23, doi:10.1016/j.biochi.2022.12.019.
- 1050 63. Sabharwal, N.C.; Savikhin, V.; Turek-Herman, J.R.; Nicoludis, J.M.; Szalai, V.A.;  
1051 Yatsunyk, L.A. N-methylmesoporphyrin IX Fluorescence as a Reporter of Strand  
1052 Orientation in Guanine Quadruplexes. *FEBS J* **2014**, *281*, 1726–1737,  
1053 doi:10.1111/febs.12734.
- 1054 64. Renaud de la Faverie, A.; Guédin, A.; Bedrat, A.; Yatsunyk, L.A.; Mergny, J.-L.  
1055 Thioflavin T as a Fluorescence Light-up Probe for G4 Formation. *Nucleic Acids*  
1056 *Res* **2014**, *42*, e65–e65, doi:10.1093/nar/gku111.
- 1057 65. De Cian, A.; DeLemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D.  
1058 Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J Am*  
1059 *Chem Soc* **2007**, *129*, 1856–1857, doi:10.1021/ja067352b.



- 1060 66. Carvalho, J.; Queiroz, J.A.; Cruz, C. Circular Dichroism of G-Quadruplex: A  
1061 Laboratory Experiment for the Study of Topology and Ligand Binding. *J Chem*  
1062 *Educ* **2017**, *94*, 1547–1551, doi:10.1021/acs.jchemed.7b00160.
- 1063 67. Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.;  
1064 Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57*, 64–  
1065 75, doi:10.1016/j.ymeth.2012.03.011.
- 1066 68. Kypr, J.; Kejnovska, I.; Renciuik, D.; Vorlickova, M. Circular Dichroism and  
1067 Conformational Polymorphism of DNA. *Nucleic Acids Res* **2009**, *37*, 1713–1725,  
1068 doi:10.1093/nar/gkp026.
- 1069 69. Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop  
1070 Permutation Affects the Topology and Stability of G-Quadruplexes. *Nucleic Acids*  
1071 *Res* **2018**, *46*, 9264–9275, doi:10.1093/nar/gky757.
- 1072 70. Mergny, J.L.; Phan, A.T.; Lacroix, L. Following G-Quartet Formation by UV-  
1073 Spectroscopy. *FEBS Lett* **1998**, *435*, 74–78, doi:10.1016/S0014-5793(98)01043-  
1074 6.
- 1075 71. Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V; Marsico, G.; Zyner, K.; Parry, A.;  
1076 Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-Quadruplex Structures  
1077 Mark Human Regulatory Chromatin. *Nat Genet* **2016**, *48*, 1267–1272,  
1078 doi:10.1038/ng.3662.
- 1079 72. Jonkers, I.; Kwak, H.; Lis, J.T. Genome-Wide Dynamics of Pol II Elongation and  
1080 Its Interplay with Promoter Proximal Pausing, Chromatin, and Exons. *Elife* **2014**,  
1081 *3*, doi:10.7554/eLife.02407.
- 1082 73. Xia, Y.; Zheng, K.; He, Y.; Liu, H.; Wen, C.; Hao, Y.; Tan, Z. Transmission of  
1083 Dynamic Supercoiling in Linear and Multi-Way Branched DNAs and Its Regulation  
1084 Revealed by a Fluorescent G-Quadruplex Torsion Sensor. *Nucleic Acids Res*  
1085 **2018**, *46*, 7418–7424, doi:10.1093/nar/gky534.
- 1086