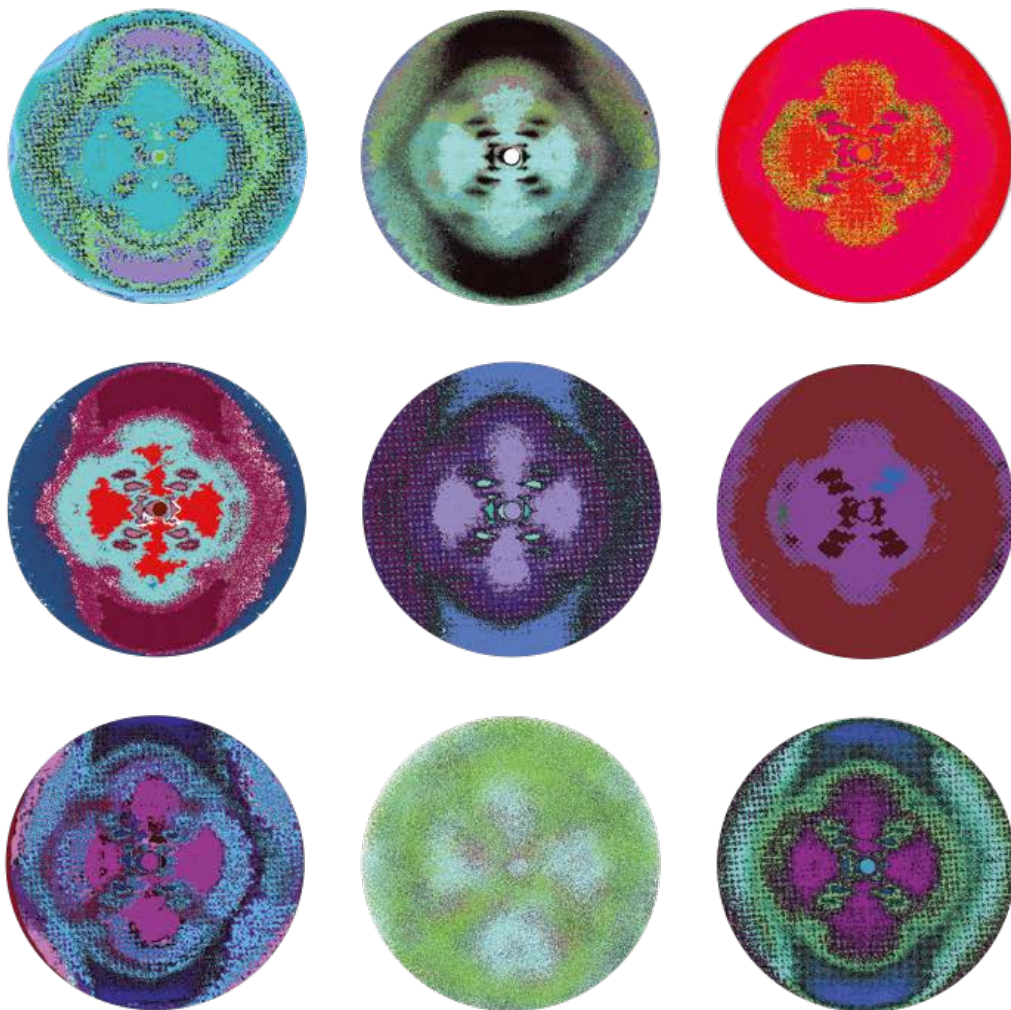# Estudio de "G-quadruplexes" del ADN en cáncer colorrectal y su uso como diana terapéutica

Victoria Sánchez Martín

Tesis Doctoral con Mención Internacional

Dirigida por José A. García Salcedo y Marta Cuadros Celorrio

Programa de Doctorado en Bioquímica y Biología Molecular

"La verdadera ciencia enseña, por encima de todo,

a dudar y a ser ignorante"

Miguel de Unamuno

"Un país sin investigación,

es un país sin desarrollo"

Margarita Salas

Quiero agradecer y dedicar mi Tesis a mi mamá y a mi papá, Ana y Antonio, porque gracias a vosotros estoy aquí y soy como soy. Me habéis dado la vida, apoyo incondicional y todo lo que he necesitado. Y aunque no sepáis qué es un doctorado o qué logro es importante en la carrera científica, sé que estáis muy orgullosos de mí, siempre. Os quiero muchísimo y os tengo en mente en cada paso que doy.

También quiero agradecer a mis hermanos, Antonio, Carlos y Leandro, y a mis cuñadas Inma, Susana y Tere, por estar siempre ahí a pesar de la distancia. Y a mis cuatro sobrinas, Andrea, Aida, Julia y Ana. Ellas son la alegría de mis días y me han hecho ver cómo de grande puede ser el amor de una tita.

Al resto de familia, mis abuelos que siguen cuidándome desde arriba, a mis tíos y a mis primos. Y es imposible no acordarme de mis churris, mi Blanqui, y mi Danna esté donde esté.

A mi love Miguel Ángel, por quererme bien, motivarme, contagiarme su optimismo y conectarme con la realidad. Tú eres quien ha vivido de primera mano esta tesis, con la montaña rusa emocional que ha conllevado. Ya lo sabes, te quiero mucho. Ojalá sigamos alcanzando metas en el futuro juntos.

A mi director de tesis José Antonio, por darme la oportunidad de unirme al grupo tras una pésima entrevista, por confiar en mis ideas y darme libertad para ponerlas en práctica, por permitirme vivir la experiencia de Bradford, Kosice, Cambridge y, aunque el covid lo impidiera, Alabama. He aprendido mucho y me he hecho fuerte con los obstáculos del camino.

A mi directora de tesis Marta, mi mamá científica, por enseñarme de todo en la vida desde hace casi 10 años. Y cuando digo de todo, es de todo: técnicas experimentales, últimas tendencias en moda, buenos restaurantes o incluso gestión emocional. Gracias por apoyarme, contar conmigo y buscar siempre lo mejor para mí. Eres mi referente de mujer en ciencia. Quiero ser como tú.

A mi tercer director de tesis Miguel porque aunque la burocracia haya impedido ese reconocimiento, siempre ha estado ahí el primero para enseñarme, supervisarme y hacer la bioinformática. Tus charlas son muy enriquecedoras y me hacen pensar. Pero aún tengo que meditar más.

A mis compañeros del Laboratorio 22, porque formamos un equipazo, diverso, pero equipazo. A mis alpaquitas Inma, Mati, Ana y Virginia, junto a mi pastor preferido, Ángel. Gracias por convertiros en compañeros de experimentos, de alegrías y desdichas, de cervezas y vinos, de contaminaciones y limpiezas, de desayunos y comidas, de viajes y gym. Todo pasa por algo. Pero esto ya es para siempre. Os quiero mucho.

A Manu y Javi, antiguos del Laboratorio 22. Manu cuando estabas tú había otro nivel de felicidad. Javi gracias a ti empecé aquí.

A Sonia, Michi, Carmen y Andrea, que durante sus TFMs han trabajado conmigo en el maravilloso mundo G4 y hemos aprendido juntos.

A las personas tan increíbles que he conocido en mis estancias, sobre todo a Nick, Maria y Emily (lovely Babraham), y Mangesh y Patri (Kosice).

A mis amigos-familia de Genyo, es imposible nombraros a todos, pero gracias por estos increíbles años, por los festejos varios y por estar dispuestos a ayudarme y echarme una mano siempre.

A Alberto, gracias además por visitarme en territorio y colchón eslovaco.

A mis amigos de siempre y para siempre, en especial Laura, Rosa, Maite y José. Gracias por estar ahí y ser tan buenos amigos. Después de la tesis, recuperaremos el tiempo perdido, no hay duda.

A mi amiga y compañera Mari Carmen. Siempre vamos a la par y siempre juntas, para lo bueno y para lo menos bueno. Contigo no hay distancia. Gracias por escucharme, apoyarme y enseñarme. Eres perfecta.

A muchos otros amigos que, en poco tiempo, se han convertido en grandes apoyos. En especial, a Silvia, Anna y mis BQ favoritos.

A Dios, a mi Virgen de los Dolores y a Fray Leopoldo. Siempre con fe.

INDEX

ABBREVIATIONS

A	Adenine
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
ATCC	American type culture collection
BLISS	Breaks labeling in situ and sequencing
Bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CCLE	Cancer cell line encyclopedia
CD	Circular dichroism
cDNA	Complementary deoxyribonucleic acid
CDR	Complementarity determining region
CH	Constant domain of heavy chain of immunoglobulin
ChIP	Chromatin immunoprecipitation
CL	Constant domain of light chain of immunoglobulin
CMS	Consensus molecular subtypes of colorectal cancer
CPP	Cell-penetrating peptide
CRC	Colorectal cancer
DAPI	4',6-diamidino-2-phenylindole
DC_{50}	Required concentration to induce a 50% fluorescence decrease
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSB	Double-strand break
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme link immunosorbent assay
ER	Endoplasmic reticulum
Fab	Fragment antigen biding region of immunoglobulin
FBL	Fibrillarin
Fc	Fragment crystallizable region of immunoglobulin
FID	Fluorescent intercalator displacement assay
FR	Framework region
G	Guanine
G0	Resting phase of cell cycle

G1	Growth 1 phase of cell cycle
G2	Growth 2 phase of cell cycle
G4	Guanine-quadruplex
GA	Gallic acid
GLUT	Glucose transporter
H	Heavy chain of immunoglobulin
HRP	Horseradish peroxidase
HTYR	Hydroxytyrosol
IC_{50}	Half-maximal inhibitory concentration
IL	Interleukin
IPTG	Isopropyl-β-D-thiogalactoside
L	Light chain of immunoglobulin
LLME	Leucine-leucine methyl-ester hydrobromide
M	Mitosis phase of cell cycle
MAPK	Mitogen-activated protein kinase
MSI	Microsatellite instability
NCL	Nucleolin
NDI	Naphthalene diimide
NMR	Nuclear magnetic resonance
NP40	Octylphenoxypolyethoxyethanol
NSG	NOD scid gamma mice
NT	Non-treated (vehicle-treated)
NTP	Nucleotide triphosphate
OG	8-oxo-7,8-dihydroguanine
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFA	Paraformaldehyde
PI	Propidium iodide
PIC	Piceid or protease inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
POLR1A	Ribonucleic acid polymerase I subunit A
Pol I	Ribonucleic acid polymerase I
Pol II	Ribonucleic acid polymerase II

PP242	Torkinib
qPCR	Quantitative polymerase chain reaction
rDNA	Ribosomal deoxyribonucleic acid
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RSV	Resveratrol
RT	Room temperature
RT-PCR	Reverse transcription–polymerase chain reaction
S	Synthesis phase of cell cycle
ScFv	Single-chain variable fragment antibody
SD	Standard deviation
Seq	Sequencing
SDS	Sodium dodecyl sulfate
T	Thymine
ThT	Thioflavin T
T-PBS	Phosphate-buffered saline with Tween 20
TYR	Tyrosol
T3S	Type III secretion system of bacteria
UMI	Unique molecular identifier
UV	Ultraviolet
VH	Variable domain of heavy chain of immunoglobulin
VHH	Nanobody
Vis	Visible
VL	Variable domain of light chain of immunoglobulin
5'ETS	5′ external transcribed spacer
γH2AX	Gamma H2A histone family member X

ABSTRACT/

RESUMEN

DNA guanine quadruplexes (G4s) are non-canonical structures formed through self-recognition of four guanines into stacked tetrads. G4s are highly prevalent at regulatory genomic regions. Considerable evidence has linked G4 formation with key biological processes ranging from telomere maintenance and transcription to genome instability and cancer. In particular, colorectal cancer (CRC) is the third most diagnosed cancer and constitutes the second leading cause of cancer death worldwide. New treatment options for CRC are required.

To mimic the progression of CRC, we established a cellular model including non-tumoral, primary tumor and metastatic stages. Despite helicases involved in unwinding of G4s were overexpressed in CRC, stabilization of G4s and induction of DNA damage increased along CRC progression, both at G0/G1 and S phases. We identified a link between the presence of G4s and the accumulation of double-strand breaks in their vicinity. The folding status of G4s played a role in the abnormal gene expression of CRC-relevant genes such as *CMYC*. The G4 harbored in its promoter region did not contain any mutation. Several well-known G4 ligands induced cytotoxicity, but lacked selectivity for tumoral cells.

In addition, we screened in CRC the antitumoral activity of several naphthalene-diimides (NDIs), a class of G4 ligands. We identified the leading compound T5 with a potent and selective inhibition of tumoral cell growth by high-affinity binding to G4s present in ribosomal DNA, thereby impairing RNA polymerase I (Pol I) elongation. Consequently, T5 induced a rapid inhibition of Pol I transcription, nucleolus disruption, proteasome-dependent Pol I catalytic subunit A degradation, and autophagy. Moreover, we attributed the higher selectivity of carbohydrate-conjugated T5 for tumoral cells to its preferential uptake through the overexpressed glucose transporter 1. We succinctly demonstrated that T5 could be explored as a therapeutic agent in a patient cohort with CRC.

Furthermore, we screened in CRC the antitumoral potential of several natural phenolic compounds. We selected gallic acid (GA) as candidate in terms of potency and selectivity. We reported on the role of

GA as a G4 ligand explaining several of its antitumoral effects, including the transcriptional inhibition of ribosomal and *CMYC* genes. In addition, GA shared some effects with other established G4 ligands such as cell cycle arrest, nucleolar stress, and induction of DNA damage and autophagy. We further confirmed antitumoral and G4-stabilizing properties of GA in vivo using a xenograft model of CRC. Moreover, we concisely demonstrated that treatment with GA could be exploitable in a patient cohort with CRC.

Finally, we generated a library of nanobodies targeting G4s through in vitro immunization. Nanobodies derive from heavy chain camelid antibodies by recombinant DNA technology. Although nanobodies displayed cross-reactivity with different G4 structures, specific *CMYC* G4-targeting nanobodies were enriched after phage-display biopanning.

Overall, in this PhD thesis, we have analyzed G4 involvement in colorectal carcinogenesis, and we have investigated the therapeutic potential of several G4 ligands in CRC. We have disclosed a mode of action for NDIs that involves ribosomal G4s targeting, and that GA affects gene expression by interaction with G4s both in vitro and in vivo.

Los cuartetos de guanina (G4) del ADN son estructuras no canónicas formadas mediante el autorreconocimiento de cuatro guaninas en tétradas que se apilan entre sí. Los G4s son muy frecuentes en regiones reguladoras del genoma. En numerosos estudios, se ha relacionado la formación de G4s con procesos biológicos clave que abarcan desde el mantenimiento de los telómeros y la transcripción hasta la inestabilidad genómica y el cáncer. En concreto, el cáncer colorrectal (CRC) es el tercer tipo de cáncer más diagnosticado y constituye la segunda causa de muerte por cáncer en todo el mundo. Se requieren nuevas opciones de tratamiento para el CRC.

Para simular la progresión del CRC, establecimos un modelo celular incluyendo la etapa no tumoral, de tumor primario y metastásica. A pesar de que las helicasas implicadas en desenrollar G4s se sobreexpresaron en CRC, la estabilización de los G4s y la inducción del daño en el ADN aumentaron a lo largo de la progresión del CRC tanto en la fase G0/G1 como S. Identificamos una asociación entre la presencia de G4s y la acumulación de roturas de doble cadena del DNA en su vecindad. El estado de plegamiento de los G4s desempeñó un papel en la expresión anormal de genes relevantes en CRC como *CMYC*. El G4 que se encuentra en su región promotora no estaba mutado. Varios ligandos de G4s conocidos provocaron citotoxicidad, careciendo de selectividad para células tumorales.

Por una parte, examinamos la actividad antitumoral en CRC de varias diimidas de naftaleno (NDIs), un tipo de ligandos de G4s. Identificamos el compuesto T5 que inhibió de manera potente y selectiva el crecimiento celular tumoral mediante la unión a G4s presentes en el DNA ribosómico con gran afinidad, afectando así a la elongación por la RNA polimerasa I (Pol I). En consecuencia, T5 provocó una inhibición rápida de la transcripción por Pol I, desorganización del nucleolo, degradación vía proteasoma de la subunidad A catalítica de Pol I y autofagia. Asimismo, atribuimos una mayor selectividad de T5 para las células tumorales a que, al tratarse de un compuesto conjugado con carbohidrato, se captaría preferentemente por células tumorales que

sobreexpresan el transportador de glucosa 1. Asimismo, demostramos sucintamente en una cohorte de pacientes con CRC que T5 podría ser explorado como agente terapéutico.

Por otra parte, analizamos el potencial antitumoral en CRC de varios compuestos fenólicos naturales. Seleccionamos el ácido gálico (GA) como el mejor candidato en términos de potencia y selectividad. Demostramos el papel del GA como ligando de G4s, lo que explica varios de sus efectos antitumorales, incluida la inhibición transcripcional de los genes ribosomales y de *CMYC*. Además, el GA compartió varios efectos con otros ligandos de G4s conocidos como son la detención del ciclo celular, estrés nucleolar, inducción del daño en el DNA y de autofagia. Utilizando un modelo de xenógrafo de CRC, confirmamos que el GA poseía efecto antitumoral y estabilizador de G4s in vivo. También demostramos concisamente que el tratamiento con GA podría explotarse en una cohorte de pacientes con CRC.

Finalmente, generamos una batería de nanoanticuerpos dirigidos a G4s mediante inmunización in vitro. Los nanoanticuerpos se producen a partir de anticuerpos de cadena pesada de camélidos por ingeniería genética. Aunque estos nanoanticuerpos presentaron reactividad cruzada con diferentes estructuras de G4, los nanoanticuerpos dirigidos específicamente al G4 de *CMYC* se enriquecieron a través de "*biopanning*" con disposición en fagos.

En general, en la presente tesis doctoral, hemos analizado la implicación de los G4s en la carcinogénesis colorrectal y hemos investigado el potencial terapéutico de varios ligandos de G4s en CRC. Hemos desvelado un mecanismo de acción para los NDIs que involucra su acción sobre los G4s del DNA ribosómico y que el GA afecta la expresión génica interaccionando con G4s tanto in vitro como in vivo.

INTRODUCTION

1. DNA G-quadruplexes

1.1. Discovery

In the 1950s, Francis Crick and James Watson pieced together the puzzle of the DNA molecule using Rosalind Franklin's X-ray crystallography data, and determined the three-dimensional double helix structure of DNA (1). However, nucleic acids also can fold into alternative secondary structures. More than 40 years before the elucidation of the DNA double helix, it was first reported that guanosine monophosphate and its analogues, but no other nucleosides, form gels at high concentration in aqueous solution (2). This unique property of guanine derivatives was largely overlooked until the 1960s when the G-tetrad (also known as G-quartet) structure, derived from X-ray fiber diffraction studies, was postulated to be the basis for the aggregation into such gelatinous substance (3). In comparison to well-known Watson-Crick base pairing in DNA, G-quartets are formed when four guanine bases are arranged in a planar tetrameric square via Hoogsteen base pairing (Figure 1A). When several G-quartets are proximally located, they can stack upon each other by means of π-π interactions, to form a three-dimensional structure, called a G-quadruplex (G4) (Figure 1B). Stacks of G-quartets are stabilized by monovalent or divalent cations centrally coordinated to O6 of the guanines (4). To note, G4s could be also formed in RNA molecules, but RNA G4s are out of the scope of the present thesis.

1.2. Topological diversity

G4s arise in guanine-rich sequences. The potential G4 motif is commonly described as G_X-N_{1-7}-G_X-N_{1-7}-G_X-N_{1-7}-G_X, where x is 3-6 and N corresponds to any nucleotide (A, G, T or C) (Figure 1C). Although G4s are related to each other in primary sequence, they in fact comprise a diverse family of structures that can fold into various topologies (Figure 1D). Their conformational properties are influenced by several parameters, including the nature of the central cation, the relative direction of the strands (parallel or antiparallel), the nature and the length of the sequence

connecting the strands (i.e., the loops), the number of separate strands associating together (from intramolecular/monomolecular to intermolecular), and by the number of stacking G-quartets (5). Depending on the nucleic acid strand orientation, G4s are mainly classified as parallel, anti-parallel or they can be mixed. A parallel G4 is formed when all four strands are oriented in the same direction, whilst antiparallel G4s are formed when at least one strand runs in an orientation opposite to the other strands. In general, G4s arise intermolecularly from two (bimolecular) or four (tetramolecular) different strands. However, trimolecular G4 structures, commonly believed to be thermodynamically unstable, have been also reported (6). Interestingly, a G4 structure can recruit a guanine derivative from the environment to fulfill a vacancy (7). In addition, G4-flanking nucleotide fragments may significantly affect the structural polymorphisms of G4s (8). Altogether, these facts increase the topological complexity of G4s. However, the extent to which distinct topologies influence G4 formation remains unknown.
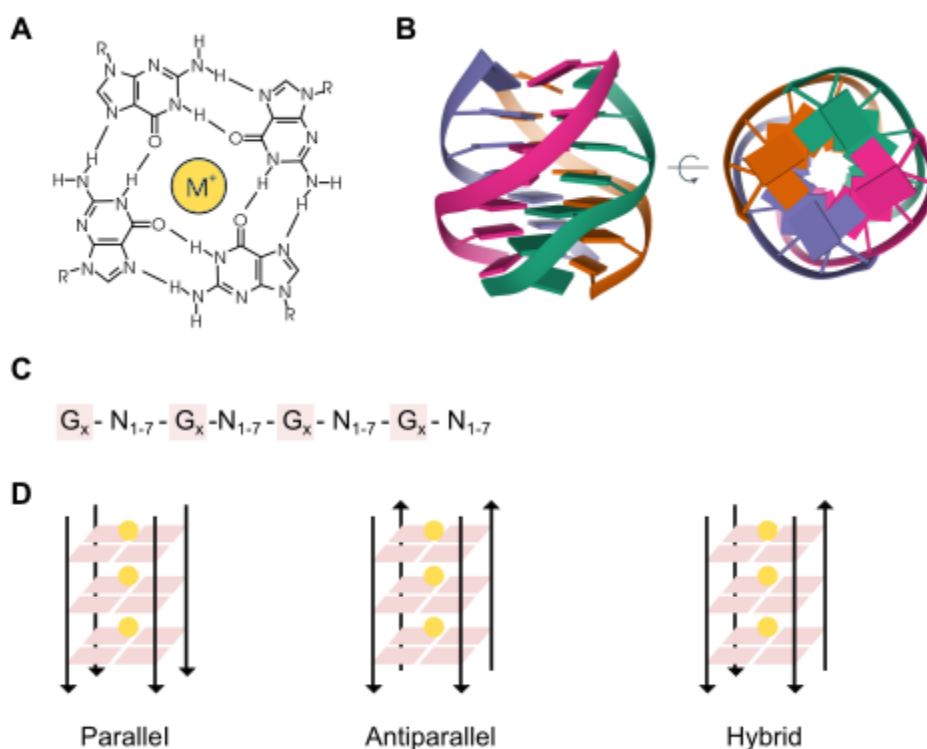
Figure 1. The structure and topologies of G4s. (A) Chemical structure of a guanine tetrad which is stabilized by Hoogsten base-pairing and a central cation (M^+). (B) Nuclear

magnetic resonance solution structure of a tetrameric parallel G4 from different rotation angles (Protein Data Bank: 139D). (C) Consensus sequence for the putative G4. x denotes the number of guanines in the tract (3-6). (D) Schematic representation of some G4 topologies. Image made with Biorender (https://biorender.com).

1.3. Detection

G4s have been identified through a combination of computational analyses of sequence and experiments aiming to detect G4s both in purified nucleic acids and in a cellular context, using chemical, molecular biology, and imaging methods.

Upon sequencing of the human genome, simple bioinformatic algorithms were used to search for the consensus sequence, and identified putative G4 sequences with the potential to form G4s in the human genome and others (9). Although early computational predictions were unable to account for structural variants including longer loops, mismatches, vacancies or the importance of flanking regions, more recent tools have accommodated them (10).

The putative G4s derived from bioinformatic predictions need to be further characterized in vitro using structural and biophysical techniques such as nuclear magnetic resonance (NMR), circular dichroism (CD), and ultraviolet (UV) spectroscopy on G4-forming oligonucleotides. Since the resonance frequency of the iminos in Watson-Crick base pairing at 12-15 ppm is clearly different from the iminos in G4s (10-12 ppm), NMR is able to distinguish G4s from helix structures in vitro (11). Apart from NMR, CD is a popular method to evidence G4 formation because the G4 structure gives rise to specific patterns of hyperchromism and hypochromism (12). In particular, parallel G4 shows a positive peak near 260 nm whereas antiparallel G4 shows a positive peak at 290 nm and a negative peak at 260 nm. Hybrid G4 structures show a positive peak at 290 nm, a positive hump at 270 nm and a negative peak at 235 nm (13). Moreover, UV measurements rely on a unique G4 hypochromic signature at 295 nm, along with an absorbance peak for nucleic acids at 260 nm (14). These

methods are used to monitor G4 stability by determination of the G4 melting temperature in G4-stabilizing or -destabilizing buffer conditions.

Additional approaches have been developed to detect G4 structures, and complement computational prediction and biophysical experiments. G4s in DNA can stall a DNA polymerase and it is the basis for DNA G4-sequencing. Comparison of polymerase pause sites in G4-stabilizing conditions (for example, in the presence of K^+ ions) versus in conditions that do not stabilize G4s (for example, in the presence of Li^+ ions) enables the detection of the 5' end of G4s in vitro when adapted into a genome-wide polymerase-stop assay followed by high-throughput sequencing (15) (Figure 2A). DNA G4s have been also identified using chemical methods by exploiting the different reactivity of nucleobases following the formation of G4 structures. For instance, in $KMnO_4$-S1 nuclease footprinting, only single-stranded DNA is digested by S1 nuclease, and subsequent computational analyses of the nuclease footprints provide a snapshot of DNA G4s (16) (Figure 2B). Another method of chemical mapping is based on the relative protection from methylation of guanines involved in G4s. The Hoogsteen hydrogen interactions between these guanines protect them from methylation by dimethyl sulfate and subsequent cleavage by piperidine, and thus, the location of G4s can be deduced (17) (Figure 2C).

Within cells, G4s have been visualized thanks to the development of specific antibodies (Figure 2D). Use of the single-chain variable fragment antibody (scFv) Sty49 revealed G4 formation at telomeres of ciliates (18). In addition, the scFv antibody BG4 revealed G4s in DNA of human cells (19). In the following years, alternative G4-specific antibodies like IgG 1H6 (20) and scFv D1 (21) corroborated these observations. Aside from antibodies, synthetic small molecules (fluorescently tagged or those with intrinsic fluorescence) that recognize and stabilize G4s have also been used to probe cellular G4 structures in live cells. Such is the case of derivatives of pyridostatin (22) and PhenDC3 (23) conjugated to fluorescent probes, or the fluorescent probe DAOTA-M2 (24).

DNA G4s have been mapped in the chromatin of human cells using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) using BG4 antibody (25) (Figure 2E). In an alternative approach, G4s can be inferred using antibodies against well-known G4-binding proteins. For instance, the helicases α-thalassemia mental retardation X-linked protein (ATRX) (26), xeroderma pigmentosum group B (XPB) and group D (XPD) (27) have been mapped to G4 motifs in human chromatin using ChIP-seq (Figure 2F).
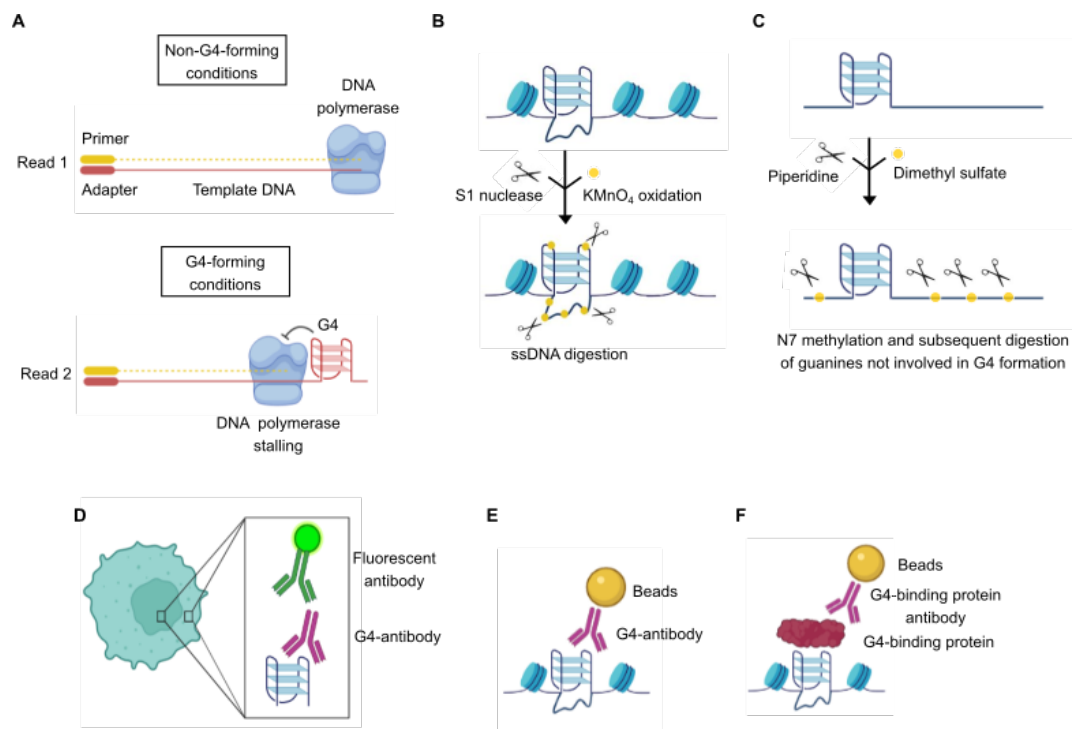
Figure 2. Approaches to detect and map G4s. (A) Mapping DNA G4s in the genome by DNA polymerase stalling followed by high-throughput sequencing (G4-seq). A library of fragmented genomic DNA is sequenced twice, first in non-G4-forming conditions (Read 1) to provide a reference, and then in G4-stabilizing conditions (Read 2) to determine the positions of G4-dependent DNA polymerase stalling. (B) DNA G4s are mapped by potassium permanganate ($KMnO_4$) single-strand nuclease (S1 nuclease) footprinting. $KMnO_4$ selectively oxidizes and traps single-stranded DNA (ssDNA), thereby allowing its digestion by S1 nuclease. Subsequent computational analyses infer the formation of DNA G4s based on the nuclease footprints. (C) DNA G4s are mapped by dimethyl sulfate (DMS) methylation and subsequent piperidine digestion. DMS selectively methylates N7 of guanines not involved in G4 formation and guides their digestion by piperidine. Subsequent computational analyses infer the formation of DNA G4s based on the piperidine footprints. (D) Visualization of DNA G4s by immunofluorescence in the nucleus using G4 antibodies together with fluorescently conjugated secondary or tertiary

antibodies. Alternatively, fluorescence-labeled G4-targeting ligands can be used. (E) Mapping of DNA G4s using chromatin immunoprecipitation and next-generation sequencing (ChIP-seq) with G4-specific antibodies. (F) DNA G4s are indirectly inferred by mapping the location of G4-binding proteins using ChIP-seq. Image made with Biorender (https://biorender.com).

1.4. Evolutionary conservation and distribution

G4 structures largely remained as laboratory curiosities and unexplored until it was discovered that the ends of human chromosomes, the telomeres, are composed of tandem repeats of guanine-rich DNA sequences (28). From then on, several reports noted that G4s are highly prevalent in the human genome (29). In addition, G4s have been identified in multiple viral (30), prokaryotic (31), and other eukaryotic genomes (32). The evolutionary context of G4 distribution in the genome has been extensively studied. It has been reported that G4s are not randomly distributed in the genome, but their distribution depends on the specie (33). Within the human genome, G4s are mainly clustered in key regulatory sites including 40% of gene promoters, telomeres, as well as in gene bodies (34). In particular, G4s are primarily accumulated at regulatory, nucleosome-depleted regions, and promoters of actively transcribed genes (35). These results strongly support their involvement in transcriptional regulation not only in humans, but also across the evolutionary trees.

1.5. Dynamics/Control

In particular, in the human genome, the first computational prediction algorithms indicated that up to 375,000 G4s could be formed simultaneously (36). Years later, high-throughput G4-sequencing (G4-seq) analysis in purified single-stranded human DNA using G4-stabilizing ligands revealed ~700,000 G4s in our genome, and ~450,000 G4s from them were not previously predicted by bioinformatics, including non-canonical G4s (15). More recently, mapping of G4s in chromatin by ChIP-seq with the anti-G4 antibody, BG4, only retrieved ~1,000-10,000 G4s

(25). Endogenous G4s have also been visualized in cells thanks to the antibody BG4, and were stabilized by a small-molecule ligand such as pyridostatin (19).

Interestingly, the number of G4s within cells accounted for ~1% of those G4s identified by direct G4-seq, possibly owing to chromatin-associated and other proteins that control the formation of these DNA structures (37). In fact, not all sequences with G4 potential form these structures in a cellular context. It is increasingly clear that G4s fold under certain conditions such as during specific stages of the cell cycle, under specific stress stimuli, or in a cell type-specific manner. G4 stability is affected by numerous factors in a cellular context. First, G4s are stabilized by centrally located cations which may be physiologically relevant, like K^+ which is the most abundant metal ion in mammalian cells (38). Second, G4 formation is favored by the induction of negative torsional stress behind RNA and DNA polymerases (39), and by molecular crowding (40). Third, during replication, transcription, and DNA damage repair, nucleic acids adopt a single stranded conformation, and Watson-Crick base-pairing is disrupted, thereby favoring the G4 formation (41). Moreover, the formation of DNA:RNA hybrids from the hybridization of the nascent RNA with the template DNA during transcription (known as R-loops) may contribute to G4 formation on the displaced DNA strand (42). In contrast, the relationship with other alternative structures may be more complex. For example, i-Motifs are formed in cytosine-rich DNA by stacking intercalated and hemi-protonated cytosine base pairs (C+:C), and their formation seems to be mutually exclusive by steric hindrance with the formation of G4s (43). To add another level of complexity in the control of G4 formation, there are numerous proteins that can bind and resolve them in a cellular context (44). Therefore, G4 formation is dynamic in live cells and largely dependent on interaction with other molecules, protein factors or ligands. These interactors shift the equilibrium of unfolded guanine-rich sequences towards fully folded G4s or vice versa (stabilization versus destabilization).

1.6. Functions

Overall, the existence of G4s over a long evolutionary period suggests their relevance for numerous biological processes. G4s can be considered as physical obstacles that impede the movement of the replicational and transcriptional machinery and must be overcome. Alternatively, G4s can recruit protein factors which influence replication and transcription in a positive manner. In this way, G4s also regulate genome processes like recombination or telomere homeostasis (45).

Apart from functioning in the neighboring regions, the role of G4 structures expands beyond the local context into additional long-range mechanisms of epigenetics. Thus, it is important to consider the G4 not as an isolated entity within a specific genomic location, but rather as a structure that exists as part of an interconnected network of other biomolecules within living cells (46).

- **Telomere homeostasis**

Telomeres are nucleoprotein structures located at chromosome ends that maintain genome integrity (47). Specifically, telomeres consist of recurrent TTAGGG-containing sequences, and constitute the regions of eukaryotic genomes with the highest concentration of G4s (48). There, G4 formation controls access to telomeres of human telomerase, hTERT, the non-coding RNA-reverse transcriptase complex that extends 3' ends of chromosomes to prevent telomere shortening. It has been long assumed that G4s can sequester the 3' end of the telomere and prevent it from being extended by hTERT (49) (Figure 3A). However, hTERT is able to unwind and extend G4s in vitro, and this ability is conserved among evolutionarily distant species (50). In addition, the *hTERT* core promoter forms an unusually long G4 that disables all critical binding sites for SP1 transcription factor, thus dramatically downregulating *hTERT* expression, and exerting telomere shortening (51). Therefore, through different mechanisms, G4s actively participate in telomere homeostasis.

- **Replication and DNA damage**

G4s can induce replication stress by obstructing the progression of DNA replication forks and causing replication-fork collapse, which generates DNA double-strand breaks leading to genome instability (52) (Figure 3B). In fact, bioinfomatic analyses of large datasets associate G4s with breakpoints that accompany somatic copy-number alterations and gene amplifications (53). G4 structures increase the probability of recurrent mutations and are important determinants of mutagenesis (54). In this sense, G4 stabilization with small molecules exceptionally induces DNA damage (55). To counteract, helicases protect the genome by unfolding G4s that can cause DNA damage. For this reason, cells compromised in their ability to process G4s (due to loss of G4 helicase activity) are particularly sensitive to G4-stabilizing ligands (56). Therefore, G4s induce genome instability that may impact on gene expression.

- **Transcription**

The initial finding that more than 40% of human genes contain putative G4s in their promoter regions suggests a role for G4s in regulating gene transcription (34). Moreover, a high number of studies have identified transcriptional changes in putative G4-containing genes upon treatment with G4-ligands (57). Such correlations are consistent with the existence of a G4-transcription link. In fact, G4 motifs function in two nuclear and interconnected processes, transcriptional termination and activation, to affect transcriptional output.

Several observations support the view that putative G4s act as terminator sequences that cause RNA polymerase II (Pol II) transcription to pause (58). The underlying mechanism involves the formation of DNA:RNA hybrids known as R-loops. These R-loops are favored during transcription as the newly synthesized RNA can base pair with the complementary template DNA strand to form a DNA:RNA hybrid (59). When sequences involved in R-loop formation contain two or more neighboring guanines, they can potentially fold into an intermolecular DNA:RNA G4. Interestingly, the formation of this intermolecular G4 only

requires as few as two tandem G-tracks on the nascent RNA transcript, and a non-template DNA strand, instead of four or more G-tracks required for the formation of canonical intramolecular G4s. These hybrid G4s have been implicated in transcription termination by a mechanism that relies on the helicase Senataxin (SETX) and R-loops (60). SETX resolves R-loop structures and allows access of the 5' to 3' exonuclease Xrn2 at 3' cleavage poly(A) sites. This action affords 3' cleavage product degradation and finally, Pol II termination as consequence (61) (Figure 3C). In addition, a substantial inhibitory effect on transcription elongation was observed when a G4 motif was present on the template strand acting as a roadblock, which is consistent with impairment of Pol II progression (62) (Figure 3D).

Based on these works, there is a general notion suggesting that G4s have a negative effect on transcription. In contrast, G4 formation in the non-template strand is associated with enhanced transcription (63). Genes with a greater number of G4s on the non-template strand up to 500 base pairs downstream of the transcription start site are associated with higher than average steady-state transcription levels and Pol II occupancy. It is explained because G4s on the non-template strand maintain the DNA in an open state, and thus, aid transcription reinitiation (64). Moreover, G4s on the non-template strand stabilize R-loops formation, giving rise to transcriptional activation (63) (Figure 3E). In accordance, G4s are prevalently found at gene promoters of transcriptionally active genes, acting as transcriptional enhancers rather than repressors (65).

Moreover, G4s display a high affinity to interact with multiple transcriptional-regulatory proteins, including transcription factors as SP1 (66), MAZ (67), and YY1 (68), resulting in altered transcription (Figure 3F). The presence of G4s at loop boundaries increases the stability of DNA loops, and in turn facilitates long-distance DNA interactions such as distal enhancer-promoter interaction (69) (Figure 3G). In addition, it has been found that G4 sequences that are split over long distances may come together to assemble into a full G4, and are significantly enriched within super-enhancer regions (70).

Within this framework, it is evident that the specific G4 positions (for example, in template or non-template strand), and the chromatin context (for example, the presence of different regulatory proteins and the formation of long-distance loops) may contribute to different regulatory mechanisms in transcription. Interestingly, low and very high abundance of putative G4s correlates with specific functional classes of genes, suggesting that different gene transcription regulatory mechanisms exist for different gene clases (71).

- **Epigenetics**

DNA methyltransferases, which catalyze the formation of 5-methylcytosine, have a biophysical preference to bind G4 DNA over double-stranded DNA (72). In particular, DNA (cytosine-5)-methyltransferase 1 (DNMT1), is sequestered at G4 sites to inhibit methylation of proximal CpG island promoters (73) (Figure 3H). In addition, it is known the G4-dependent recruitment of different epigenetic complexes to regulate gene expression. For example, the recruitment of RE1-silencing transcription factor - lysine-specific histone demethylase 1A (REST-LSD1) repressor complex is mediated by G4s in non-metastatic 2 (*NME2*) gene. REST-LSD1 removes the gene-activating monomethylation H3K4me1 and dimethylation H3K4me2, inducing gene repression (74) (Figure 3I). In contrast, G4 might guide the recruitment of chromatin remodeling complexes like BRD3 to favor transcription at G4 sites (75). Moreover, G4s epigenetically control gene expression by promoting DNA oxidation and repair. Among the four DNA bases, guanine has the lowest redox potential, and is the most frequent site to be oxidized by reactive oxygen species, generating 8-oxo-7,8-dihydroguanine (OG) DNA (76). OG formation in gene promoters containing G4-forming sequences can stimulate DNA repair mechanisms (77), recruit different transcription factors (78), and act as on-off switches for transcription (79) (Figure 3J). The underlying mechanism relies on the recruitment of mediators in the base excision repair (e.g., APE1 or PARP1) to OG DNA by G4s, which in turn tune the activity of transcriptional regulators (80). Altogether these works confirm the relationship between G4s and the installation of

epigenetic marks, underlining the potential of G4s in molding the epigenetic landscape.
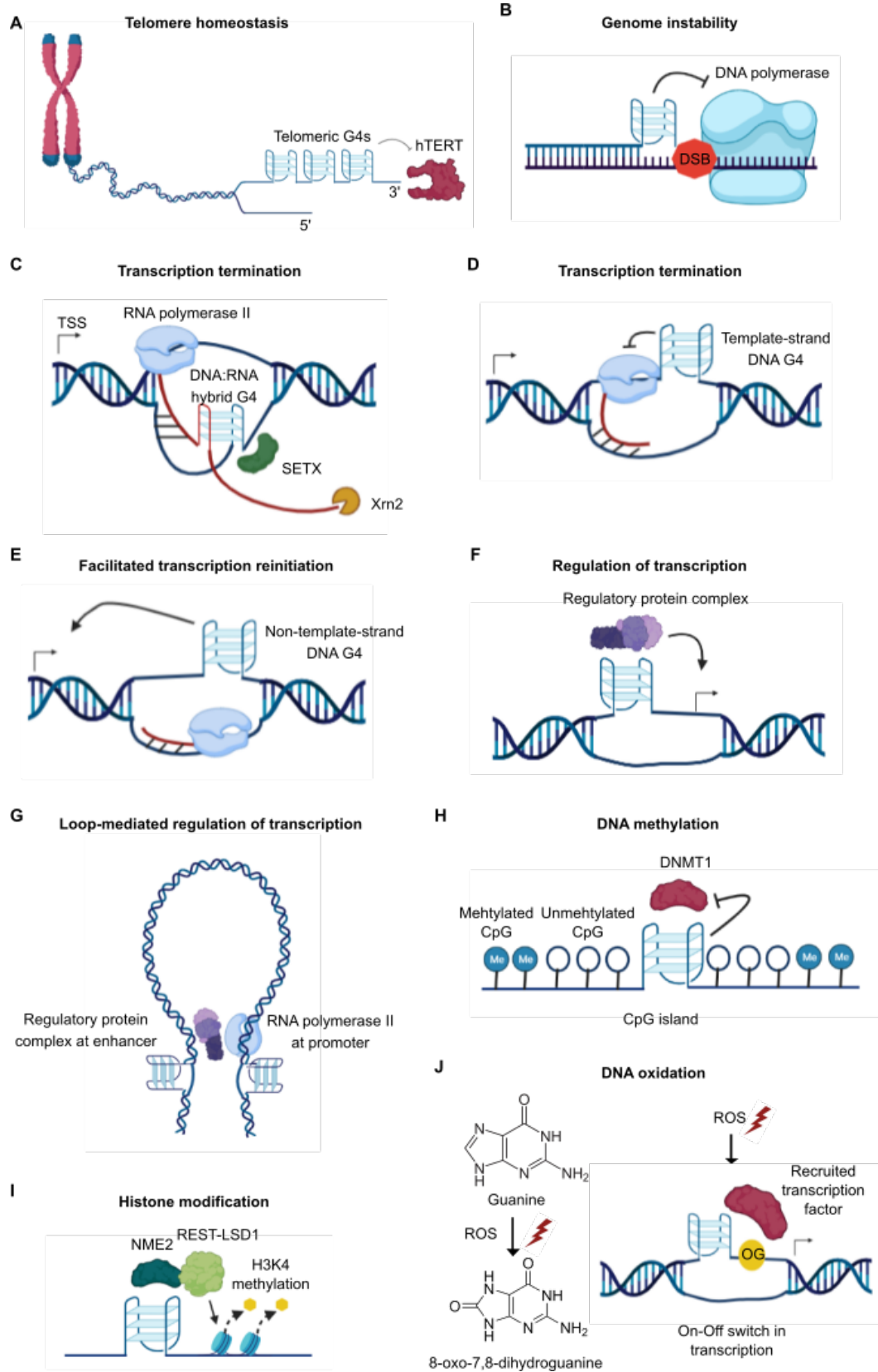
Figure 3. Functions of G4s. (A) Telomeric G4s interfere with human telomerase (hTERT), mediating telomere homeostasis. (B) G4s act as roadblocks to DNA polymerase obstructing its progression and inducing DNA double-strand breaks (DSBs). (C) Formation of DNA:RNA hybrid G4s between the non-template DNA and the nascent RNA can lead to premature transcription termination. This process implies cooperation between helicase Senataxin (SETX) and exoribonuclease Xrn2. (D) During transcription elongation, the separation of DNA strands may result in the formation of G4s on the template strand that can block the progression of RNA polymerase II. (E) G4s on the non-template strand may facilitate transcription reinitiation. (F) G4s could bind or displace transcription factors, resulting in altered transcription. (G) G4s enrichment at loop boundaries possibly act as stabilizers of enhancer-promoter loops, resulting in altered transcription. (H) G4s at CpG islands inhibit DNA (cytosine-5)-methyltransferase 1 (DNMT1), contributing to hypomethylation at CpG islands. (I) G4s via non-metastatic 2 (NME2), recruit the RE1-silencing transcription factor (REST)-lysine-specific histone demethylase 1A (LSD1) repressor complex to remove the gene-activating methylation of histone H3 Lys4 (H3K4). (J) Guanine is frequently oxidized by oxidative stress (ROS) generating 8-oxo-7,8-dihydroguanine (OG). OG formation in gene promoters containing a G4 stimulates the recruitment of transcription factors with consequent on-off switches in transcription. Image made with Biorender (https://biorender.com).

1.7. Involvement in cancer

Owing to the wide role of G4 structural elements in multiple molecular, cellular, and physiological levels, G4s quickly became potential targets in the context of diseases (81). In particular, G4s have been widely associated with carcinogenesis. There is an increase of G4s in the chromatin of cancer tissues (82) and cancer cell lines (19), in comparison with non-tumoral tissues or cells, respectively.

In this regard, G4s display key cancer-related functions. The first validated DNA G4 was found in the promoter of the oncogene *CMYC*, functioning as a roadblock to transcriptionally downregulate *CMYC* expression (57). Subsequent analyses have assessed G4 structures in an endless list of genes. Six vital cellular and microenvironmental processes are considerably dysregulated during oncogenic transformation and malignancy. These processes include sustained proliferative signaling, evasion of growth suppressors, resistance to cell death, angiogenesis,

replicative immortality, and metastasis (83). When examining each one of these six hallmarks of cancer, critical genes with a G4 structure in the core or proximal promoter are found, and new ones are being continually identified (84) (Figure 4). Therefore, G4s are linked to the control of the expression of several oncogenes and tumor suppressors. Moreover, G4s play a key role in telomere biology as detailed above, and lengthening of telomeres is a frequently activated mechanism in cancer to sustain limitless replication (85). In addition, G4s obstruct the progression of DNA replication forks inducing DNA damage, which is characteristic of many tumors. Furthermore, DNA mutations can lead to genomic instability, and there is a notable association of G4s with tumoral gene amplification (86). Therefore, the cancer-related functions of G4s offer an alternative therapeutic approach in cancer.
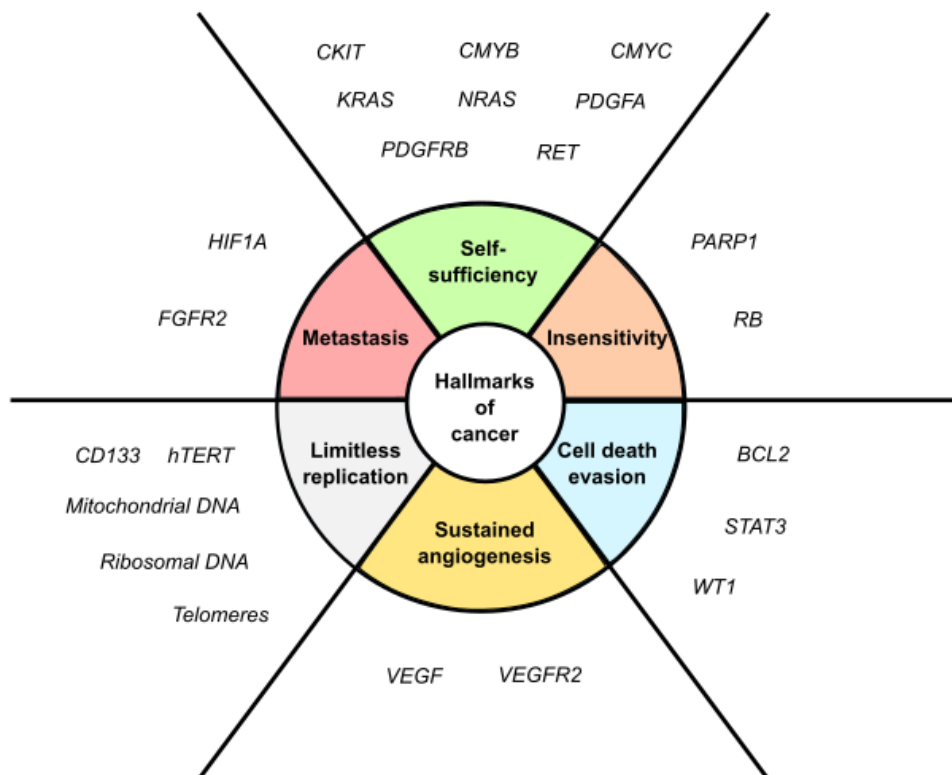
Figure 4. Genes containing G4s at their promoters are associated with cancer hallmarks. DNA G4s are present in multiple cancer-relevant genes playing key regulatory functions. To facilitate visualization, each gene is only associated with one feature, even though they might be involved in several cancer hallmarks. Image modified from Sanchez-Martin V *et al.* Int. J. Mol. Sci. 2020, 21, 8900.

1.8. Targeting G4s in cancer therapy

New insights are put into targeting G4s for the treatment of cancer (87). G4 ligands have been explored with a view to interfere with gene expression and telomere lengthening (Figure 5A). Ligands are chemical compounds that specifically bind to and stabilize the structure of G4s. In general, G4 ligands possess an aromatic surface for π-π interactions with G4s, a positive charge or basic groups to selectively bind to the loops or grooves of the G4, and a steric bulk to prevent intercalation with double-stranded DNA (88). To date, an arsenal of around 1,000 small molecules that target G4s has been reported, and the majority of them have emerged in recent years (89).

The distinct molecular features of the different G4s enable structure-selective recognition by small molecules (90). Initial efforts were focused on targeting G4 structures at telomeres with the view to inhibiting telomere extension by hTERT in cancer cells (91). Subsequently, several studies addressed the modulation of individual cancer genes by targeting their G4s (92). However, the prevalence of G4s in many cancer-related genes suggests that collectively targeting multiple G4s, thus altering the expression of many such genes, would be a feasible strategy (93). Furthermore, the capacity of G4 ligands to generate synthetic lethality in tumor cells provides another potential G4-based therapeutic avenue. Synthetic lethality is classically defined as the process by which the inactivation of an individual gene has little effect on cell viability, but the loss of function of two or more genes simultaneously leads to cell death (94). For instance, G4 ligands enhance killing of *BRCA1/2*-deficient cancer cells by exploiting their deficiency in DNA repair (95). Moreover, G4 ligands can be also used in combination with other agents to synergize in their cytotoxic activities (96).

Many G4 ligands have characteristic cores that can be chemically modified, rendering various analogues whose therapeutic activity in cancer is being investigated. During the last decades, extensive research efforts identified naphthalene diimides (NDIs) (Figure 5B) as favored chemotypes

for G4 binding due to their high target affinity and potential for chemical variability (97). NDIs were originally reported to bind to telomeric G4s resulting in telomerase inhibition (98). However, after easily tunable synthesis, NDIs target G4s formed in several oncogene promoters such as *CKIT* (99) and *BCL2* (100), interfering with their oncogenic signaling pathways. In addition, another NDI derivative binds to G4s in *MDM2* oncogene, which is a master regulator of TP53, reducing *MDM2* transcription. Such an approach is being explored to defeat all tumors in which the restoration of wild-type TP53 is sought (101). Various NDIs, such as MM41 (102) and CM03 (93) have demonstrated promising results in cancer therapeutics in vivo.

Interestingly, some G4 ligands have even entered clinical trials. CX3543 (Figure 5C), also named as quarfloxin, is a fluoroquinolone that was originally designed to target the G4 found in the *CMYC* promoter (103). Further studies demonstrated that CX3543 also interacts with a G4 found in ribosomal DNA, and disrupts the binding between these G4s and nucleolin complexes in the nucleolus, thereby inhibiting ribosome biogenesis (104). Although CX3543 passed phase II trials as a candidate therapeutic agent against several tumors, phase III trials were not completed because of its high binding to albumin. Another fluoroquinolone, CX5461 (Figure 5D), was found to selectively inhibit ribosomal RNA synthesis by reducing the binding affinity of the SL1 pre-initiation complex and RNA polymerase I complex to the ribosomal DNA promoter (105). Very recently, it was shown that SL1 recruitment to ribosomal DNA is performed in a G4-dependent manner, and CX5461 traps such G4 structures, interfering with SL1 DNA binding activity (106). Similar to CX3543, CX5461 selectively binds and stabilizes a broad spectrum of G4 structures, including those harbored in *CMYC*, *CKIT*, and telomeres (95). Notably, CX5461 is currently in phase I clinical trials, constituting the most advanced G4 ligand in the clinics at the moment.
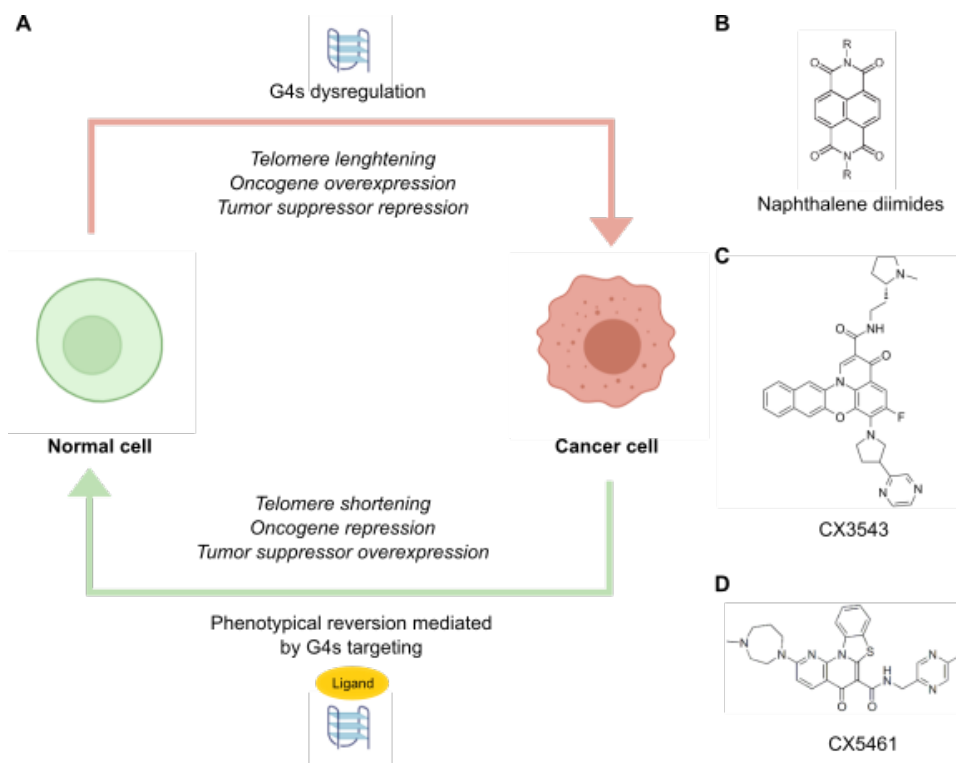
Figure 5. G4s as targets in cancer therapy. (A) Schematic representation of antitumoral effects mediated by G4 ligands for cancer therapy. (B-D) Chemical structure of several G4 ligands including naphthalene diimides (B), CX3543 (C), and CX5461 (D). Image modified from Sanchez-Martin V *et al*. Cancers 2021, 13, 3156.

2. Colorectal cancer

According to the Centre for Disease Control and Prevention, colorectal cancer (CRC) is cancer that occurs in the colon or rectum. The colon is the large intestine or large bowel. The rectum is the passageway that connects the colon to the anus.

2.1. Epidemiology

CRC is the third most commonly diagnosed cancer and the second leading cause of cancer death in the world. In numbers, more than 1.9 million new CRC cases and 935,000 deaths were estimated to occur in 2020 in the globe, representing about 10% of cancer cases and deaths. Incidence rates are approximately 4-fold higher in developed countries compared with developing countries, but there is less variation in the

mortality rates because of higher fatality in developing countries (107). Its incidence is steadily rising in developed nations apparently due to lifestyle factors (108).

2.2. Risk factors

In epidemiological studies, male sex and increasing age have consistently shown strong associations with CRC incidence. Both hereditary and environmental risk factors play a role in the development of CRC (Figure 6).

Figure 6. List of modifiable and non-modifiable risk factors for CRC. Factors displayed in red increase the risk of CRC, whilst factors in green may prevent CRC. NSAIDs: non-steroidal anti-inflammatory drugs. Image made with Biorender (https://biorender.com) and inspired from Dekker E *et al.* Lancet 2019, 394, 1467–1480.

Positive family history seems to have a part, with varying risk depending on number and degree of affected relatives and age of diagnosis (109). Estimation for heritability of CRC range from 12% to 35% (110). However, most factors causing heritability are still elusive and subject to further study (111). A subgroup of approximately 5-7% of the patients with CRC is affected by hereditary CRC syndromes (112). These hereditary syndromes are subdivided as polyposis and non-polyposis (Lynch syndrome) syndromes. The polyposis syndromes are more easily recognised as the physician is alerted by the number of polyps. This syndrome is caused by mutations in the adenomatous polyposis coli (*APC*) gene, which controls the activity of the WNT signalling pathway (113). In turn, Lynch syndrome is frequently missed as those patients have

few adenomas that morphologically resemble sporadic lesions. Therefore, molecular analyses are required to diagnose Lynch syndrome, which is characterized by dysfunction of the DNA mismatch repair system. Such alteration lead to the accumulation of DNA mutations, which occur, in particular, in microsatellite DNA fragments with repetitive nucleotide sequences (microsatellite instability, MSI) (114). Apart from hereditary CRC syndromes, patients with long-standing inflammatory bowel disease (115) or with a previous history of adenomas (116) are at an increased risk, and require adequate surveillance.

Among the main, largely modifiable, environmental factors are exercise for prevention (117), smoking (118) and excessive alcohol intake (119) as risk factors, and diet (120). Respect to diet, whilst consumption of red and processed meats increases the risk of CRC, diets enriched on fish, fruits, vegetables, and fibers are proposed to reduce the risk of disease onset and progression (121). Other risk factors exist, including menopausal hormone therapy (122) and use of statins (123). In contrast, regular aspirin and non-steroidal anti-inflammatory drugs intake have been associated with reduced CRC risk (124).

2.3. Prevention

A large proportion of CRC is highly preventable. Thus, primary prevention remains the key strategy to reduce the increasing burden of CRC. A study in nine European countries found that nearly 20% of CRC cases were attributable to not adhering to healthy lifestyle recommendations (125). In addition, implementation of screening tools have the potential to considerably reduce CRC incidence and mortality (126). The long time for development of most CRC allows for detection and removal of precursor lesions in some cases (114). Screening tools include faecal occult blood tests, which are currently the most widely used tests (127), flexible sigmoidoscopy (128), and colonoscopy (129).

2.4. Pathogenesis

A prevailing paradigm is that the cell of origin of most CRCs is a stem cell or stem cell-like cell that resides in the base of the colon crypts (130). In the "classic" CRC formation model, cancer usually emerges from the glandular epithelial cells of the large intestine (a crypt), which acquire genetic and epigenetic alterations that confer on them a selective advantage in a cross-talk manner. With abnormally heightened replication and survival, these hyper-proliferative cells give rise to a neoplastic precursor lesion (a polyp). The polyp then evolves into an early adenoma (<1 cm in size), and progresses to an advanced adenoma (>1 cm in size), before eventually evolving into carcinoma and metastasize. This process occurs over an estimated 10-15 year period, but can progress more rapidly in certain settings (e.g., in patients with Lynch syndrome) (131).

Globally, there are two major distinct precursor lesion pathways: the traditional adenoma-carcinoma pathway (representing up to 80-90% of CRC), and the serrated neoplasia pathway (10-20% of CRC). Both pathways are driven by the accumulation of distinct multiple genetic and epigenetic events in a rather sequential order (Figure 7) (132). Frequently, CRC is initiated by mutations that affect the WNT signalling pathway, and the ensuing neoplastic cells progress upon deregulation of other signalling pathways, including the RAS-RAF-MAPK, PI3K-AKT, TGFb, and TP53 pathways. The most common mutations in CRC include those in *APC*, catenin-b1 (*CTNNB1*), kirsten rat sarcoma virus (*KRAS*), neuroblastoma RAS viral oncogene homolog (*NRAS*), V-Raf murine sarcoma viral oncogene homolog B1 (*BRAF*), SMAD Family Member 4 (*SMAD4*), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit-a (*PI3KCA*), transforming growth factor-b receptor 2 (*TGFBR2*), and tumor protein P53 (*TP53*) (133). Moreover, *CMYC* is overexpressed in approximately 70% of CRC, although such deregulation is not accompanied by amplification or rearrangement of the gene (134). Therefore, additional epigenetic mechanisms commonly cooperate to drive cancer progression. Hypermethylation of CpG islands that silences tumor suppressor genes, and hypomethylation that leads to genomic instability

or oncogene activation, have been associated to colorectal carcinogenesis (135). Some of the mutated or epigenetically altered genes are shared between the two pathways, whereas others are unique. For example, the adenoma-carcinoma pathway is typically initiated by an *APC* mutation, followed by *RAS* activation, and function loss of *TP53*. Conversely, *BRAF* mutations and CpG island methylation phenotype only occur in the serrated pathway (131).
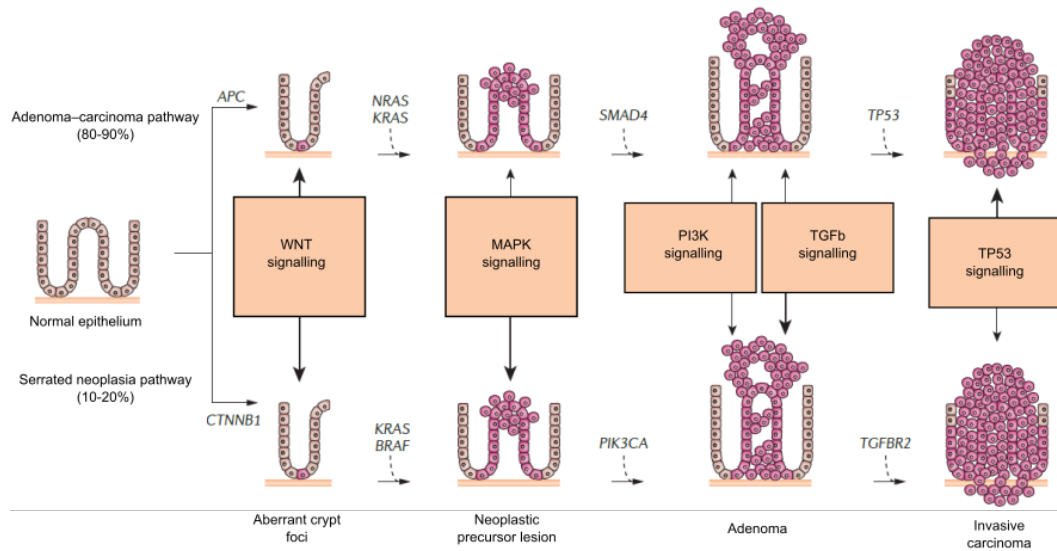
Figure 7. CRC development pathways and their associated genetic alterations. CRC development involves the progression of normal colon epithelial cells to aberrant crypt foci, followed by a neoplastic precursor lesion (polyps) with subsequent progression to early and advanced adenomas, and then carcinoma and metastasis. The "classic" or traditional pathway (top) involves the development of tubular adenomas that can progress to adenocarcinomas. An alternate pathway (bottom) involves serrated polyps, and their progression to serrated CRC. The genes mutated or epigenetically altered are indicated in each sequence; some genes are shared between the two pathways, whereas others are unique. The signalling pathways deregulated during the progression sequence are also shown, with the width of the arrow reflecting the significance of the signalling pathway in tumor formation. Image modified from Kuipers EJ *et al*. Nat. Rev. Dis. Prim. 2015, 1, 1–25.

Importantly, the frequencies of many of these molecular features vary depending on the location of the tumor in the gut. Some studies support a gradual change in frequency of the molecular alterations

("continuum model"), whereas others suggest a more abrupt dichotomy ("proximal versus distal" and "right-sided versus left-sided" models) (136). Both models support the notion that the tumor microenvironment (the gut microbiota and the inflammatory state of adjacent tissue) modulates the way by which these mutations affect cancer formation and disease progression (137). However, the involvement of additional factors such as G4s in colorectal carcinogenesis still remains ellusive.

2.5. Subtypes

CRC is a heterogeneous disease with different subtypes that may be distinguished by their specific clinical and/or molecular features. In 2014, on the basis of gene expression, CRC was classified into four molecular subtypes, referred to as "consensus molecular subtypes" (CMS). CMS1 is characterized by immune activation and tends to occur in the elderly, female, and proximal colon with worse survival after relapse. CMS2 is characterized by canonical features with marked WNT and CMYC pathway activation, and appears to occur in the left-side colon and rectum with superior survival after a relapse. CMS3 exhibits an epithelial signature and metabolic dysregulation. CMS4 is characterized by mesenchymal features, and seems to be diagnosed with advanced stages, showing poorer overall survival. Although CMS classification cannot suggest a therapeutic stratification, it is being explored as a prognostic or predictive marker (138).

2.6. Diagnosis

The diagnosis of CRC either results from an assessment of a patient presenting with symptoms or as a result of population screening. The disease can be associated with a range of symptoms, including blood in stools, change in bowel habits, and abdominal pain. Other symptoms include fatigue, anaemia-related symptoms, such as pale appearance and shortness of breath, and weight loss. However, CRC is largely an asymptomatic disease until it reaches an advanced stage. With the widespread introduction of population screening, many individuals are even diagnosed at the preclinical stage (139).

Colonoscopy is the gold standard for the diagnosis of CRC, both for symptomatic patients and for asymptomatic candidates that derive from population screenings with a positive test. Colonoscopy has a high diagnostic accuracy and assesses the location of the tumor. Importantly, such technique enables simultaneous biopsy sampling, and hence, histological confirmation of the diagnosis. In addition, colonoscopy is the only screening technique that provides both a diagnostic and therapeutic effect (140). Other imaging techniques like computed tomography and magnetic resonance imaging are used as complementary methods (e.g., after incomplete or inadequate colonoscopy) (141).

Molecular detection of biomarkers of CRC by non-invasive assays using proteins, RNA, and DNA in the blood, stool, and urine has been developed. In particular, hypermethylation of the promoter region of *SEPT9,* which belongs to a class of GTPases, is associated with CRC (142). Moreover, mutations of *APC* and *KRAS* have been tested in DNA shed by epithelial cells and isolated from stool samples (143). Other more-discriminating markers such as aberrant NDRG family member 4 (*NDRG4*), bone morphogenetic protein 3 (*BMP3*) methylation, and the presence of β-actin, have been added within a multitarget approach (144).

2.7. Treatment

Advancement in the pathophysiological understanding of CRC has increased the array of treatment options leading to have doubled overall survival for advanced disease to 3 years. However, survival is still best for those with non-metastasised disease (114).

The cornerstone of curative treatment in CRC includes endoscopic resection and surgery for primary disease (145). In the case of liver and lung metastases, local but more aggressive resection is required (146). In particular, rectal cancers enormously benefit from chemoradiotherapy, with a dose of 45-50 gray in 25-28 fractions, and with a fluoropyrimidine as radiation sensitizer (147). In fact, the observation of complete clinical response after neoadjuvant chemoradiotherapy has initiated rectal

preserving treatment approaches (148). In contrast, there is no accepted neoadjuvant treatment for colon cancer.

Different classes of drugs are available for the treatment of CRC (Figure 8A). The adjuvant fluoropyrimidine-based chemotherapy (5-fluorouracil or capecitabine as prodrug) is the essential component of systemic chemotherapy for CRC. 5-fluorouracil is an anti-metabolite that exerts its antitumoral effects mainly through the inhibition of the enzyme thymidylate synthase, leading to disruption of the intracellular deoxynucleotide pools required for DNA replication. Possibly, 5-fluorouracil is also incorporated into RNA and DNA, interfering with their cellular functions. 5-fluorouracil is usually accompanied by leucovorin, as an active metabolite of folic acid, in order to stabilize the bond of 5-fluorouracil to thymidylate synthase, and enhance the activity of the fluoropyrimidine (149). Several landmark studies have established the addition of oxaliplatin to a fluoropyrimidine as the new standard (150). Oxaliplatin is a platinum-based compound that exerts its cytotoxic effect through the formation of platinum-DNA adducts and DNA damage (151). In addition, the combination of irinotecan with 5-fluorouracil significantly prolongs survival compared with 5-fluorouracil alone (152). Irinotecan acts as a potent inhibitor of topoisomerase I, inducing DNA double-strand breaks (153). Therefore, fluoropyrimidines, oxaliplatin, and irinotecan form the chemotherapy backbone in various iterations of two-drug or three-drug regimens.

Alongside these combined chemotherapy regimens, targeted agents are used for metastatic CRC treatment. In particular, these include two main groups of drugs: agents targeting the tyrosine kinase EGFR and those targeting the growth factor VEGFA (154). On the one hand, approximately 80% of all CRC express or overexpress EGFR, and it can be blocked by monoclonal antibodies such as cetuximab, which is a recombinant chimeric monoclonal IgG1 antibody, and panitumumab, which is a human EGFR-specific antibody. To note, the *RAS* mutational status of the tumor (*RAS* is mutated in about half of all patients with CRC) must be examined before treatment with EGFR-specific antibodies (155).

On the other hand, VEGFA is an endothelial growth factor which binds to VEGF receptor 1 (VEGFR1) and VEGFR2, and acts as a key effector of tumor angiogenesis. Drugs targeting VEGFA include bevacizumab, which is a humanized anti-VEGF monoclonal antibody, and aflibercept, a recombinant fusion protein that consists of the VEGF-binding portions from the extracellular domains of human VEGFR1 and VEGFR2 fused to the fragment crystallizable (Fc) portion of human IgG1 (156).

In addition, regorafenib (a multi tyrosine-kinase inhibitor) (157) and TAS-102 (combination of anti-mebatolites trifluridine and tipiracil) (158) are newer drugs approved for patients with refractory metastatic CRC who have not responded to upfront systemic therapies. These therapies are commonly known as "salvage therapy drugs".

The aforementioned alternatives for CRC treatment need to be adapted in some tumors with distinctive characteristics. For instance, identification of the *BRAF-V600E* mutant in CRC is important, since outcomes are 2-3 times worse (159). In such cases, combinatorial strategies (BRAF-inhibitors and anti-EGFR antibodies paired with chemotherapy or MEK inhibitors) are now included in consensus guidelines (160). For the 4-5 % of tumors with mismatch repair deficiency or high microsatellite instability, blockade of programmed cell death protein 1 (PD-1) with immunotherapies such as human nivolumab or humanized pembrolizumab is now approved (161). These immune checkpoint inhibitors have shown durable and curative potential. Unfortunately, they do not work for the mismatch repair-proficient CRC, which constitute the fair majority. Therefore, patients with CRC are treated with different drug regimens one after another, depending on patient-related and tumor-related factors (e.g., sidedness, mutations, or mismatch repair status) (Figure 8B).

Despite major advances in treatment, mortality from CRC remains high and 40-50% of patients eventually die (114). In this context, new treatment options are required.
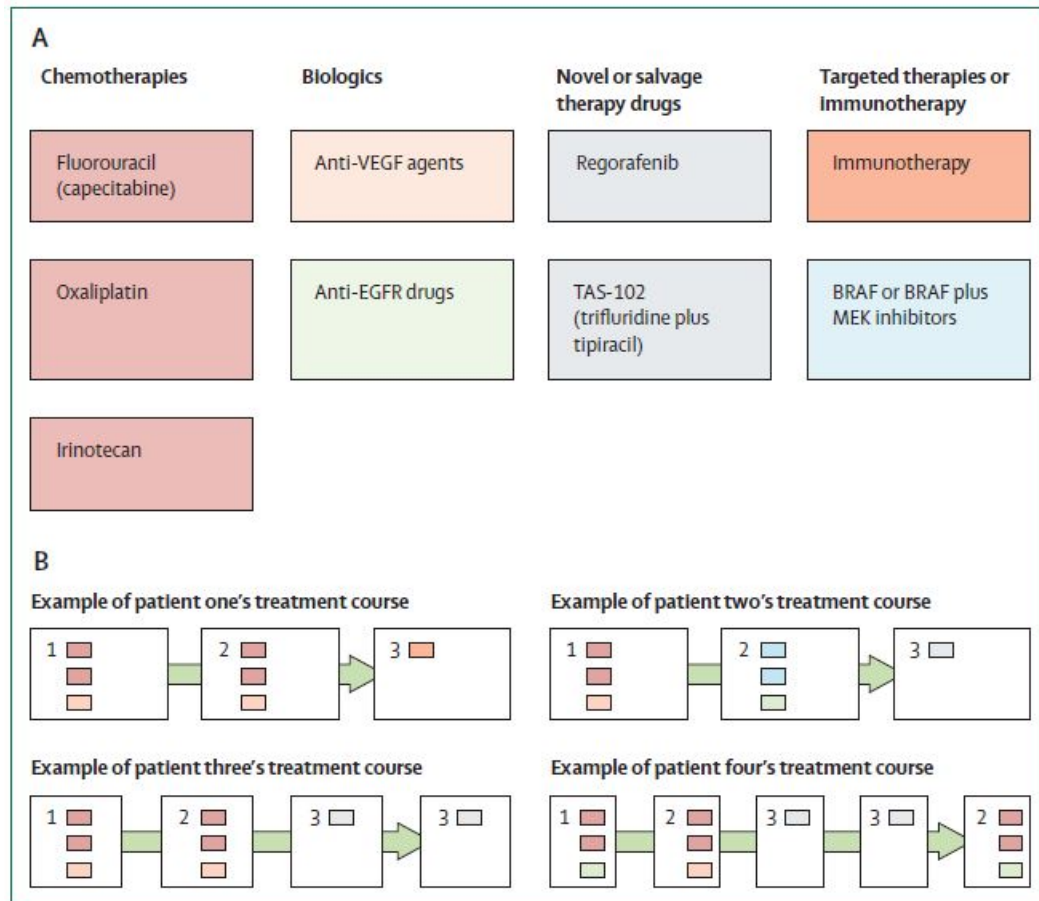
Figure 8. Schematic representation of classes of drugs used for CRC treatment. (A) Different classes of drugs available are often used in combination (e.g., two or three chemotherapy drugs paired with a biologic). (B) Patients with metastatic CRC are often treated with multiple treatment regimens one after another, depending on patient-related and tumor-related factors. The distinct treatments are represented in different colors as in (A) Image obtained from Dekker E *et al*. Lancet 2019, 394, 1467–1480.

2.8. Dietary phenols

Natural bioactive compounds present in food, particularly phenolic compounds, have been reported to possess antitumoral activities that are important not only for prevention, but also for treatment of cancer (162). Additional properties of phenolic compounds include antioxidant and antiinflammatory activities (163). Phenolic compounds are secondary metabolites in plants with a common aromatic ring bearing one or more hydroxyl groups, and more than 8,000 have been identified to date (164).

A range of evidence supports the anticancer properties of phenolic compounds in CRC (165). Their actions comprise scavenging free radicals, induction of enzymes involved in the metabolism of xenobiotics, regulation of gene expression, and modulation of cellular signaling pathways. Specifically, phenolic compounds inhibit the initiation and progression of cancer by modulating genes that regulate key processes such as: (i) oncogenic transformation of normal cells; (ii) growth and development of tumors; and (iii) angiogenesis and metastasis (166). Interestingly, phenolic compounds are accumulated in the cell nucleus rather than in any other organelles (167). However, the exact molecular mechanism underlying many of their actions in CRC are yet to be fully clarified.

3. Nanobodies

The conventional antibodies of all vertebrates are heteromeric molecules composed of two heavy (H) and two light (L) chains, with both chains contributing to two identical antigen-binding sites (Fab). Each target-binding site is composed of one constant (C) and one variable (V) domain from each antibody's H and L-chain, designated as CH, CL, VH and VL respectively (168). In addition, conventional antibodies contain the fragment crystallizable (Fc) region, which is the tail region that mediates interaction with many of the effector functions of the immune system following antigen binding. Nanobodies are synthetically obtained from heavy-chain antibodies present in dromedary (169). This technology was developed after the discovery that camelids possess a fraction of immunoglobulins which lacks the light chain (170). Such heavy-chain antibodies are homodimers where each monomer unit (H-chain) has an antigen-binding fragment reduced to one single variable domain (VHH or nanobody). Interestingly, when a nanobody is expressed as a recombinant protein, it maintains its antigen-binding capacity intact (Figure 9A). The applicability of nanobodies is multiple, similar to conventional antibodies.

Nanobodies can be used both for diagnostic and therapeutic purposes, including cancer (171).

Nanobodies have important advantages over conventional antibodies: i) their small size (15 kDa), which allows access to occluded epitopes inaccessible to conventional antibodies; ii) high affinity and specificity; iii) high stability and solubility at extreme temperatures and pH; iv) they are not immunogenic for humans or animals, v) ease of manipulation and production by recombinant technology, being able to be produced on a large scale in a short period of time and at low cost (171).

The peculiarities of nanobodies are due to a series of differences in the antigen-binding domain (VHH), in relation to the corresponding domain of the heavy chains of conventional antibodies (VH) (Figure 9B). The peptide backbone of this domain in both types of immunoglobulins is very similar. It is composed of four conserved "framework regions" (FR1-FR4) and three hypervariable "complementary determining regions" (CDR1-CDR3). The conserved FR regions are responsible for maintaining the structure and three-dimensional conformation of the antibody, while the hypervariable CDR regions provide the binding sites and antigen specificity. In the VHH domains, the hypervariable CDRs are of greater length than in the conventional VH domains, especially the CDR3 region that provides the basis for antigen specificity. Nanobodies also have a convex conformation that allows them to interact with concave epitopes inaccessible to conventional antibodies. There are also some important substitutions in the conserved FR2 region that provide greater solubility to the VHH domains, and explain the absence of interaction with the light chains. Due to the lack of the constant domain present in conventional immunoglobulins, nanobodies do not elicit complement-mediated or toxic cellular responses (169). Their similarity to human immunoglobulins mean that they are not immunogenic, and their small size makes them suitable for administration in aerosols or orally (172). Furthermore, nanobodies can be humanized by substituting a few amino acids without affecting their affinity (173). Altogether, these intrinsic characteristics of nanobodies

confer them an enormous applicability. Nevertheless, nanobodies have not been used for the treatment of CRC so far.
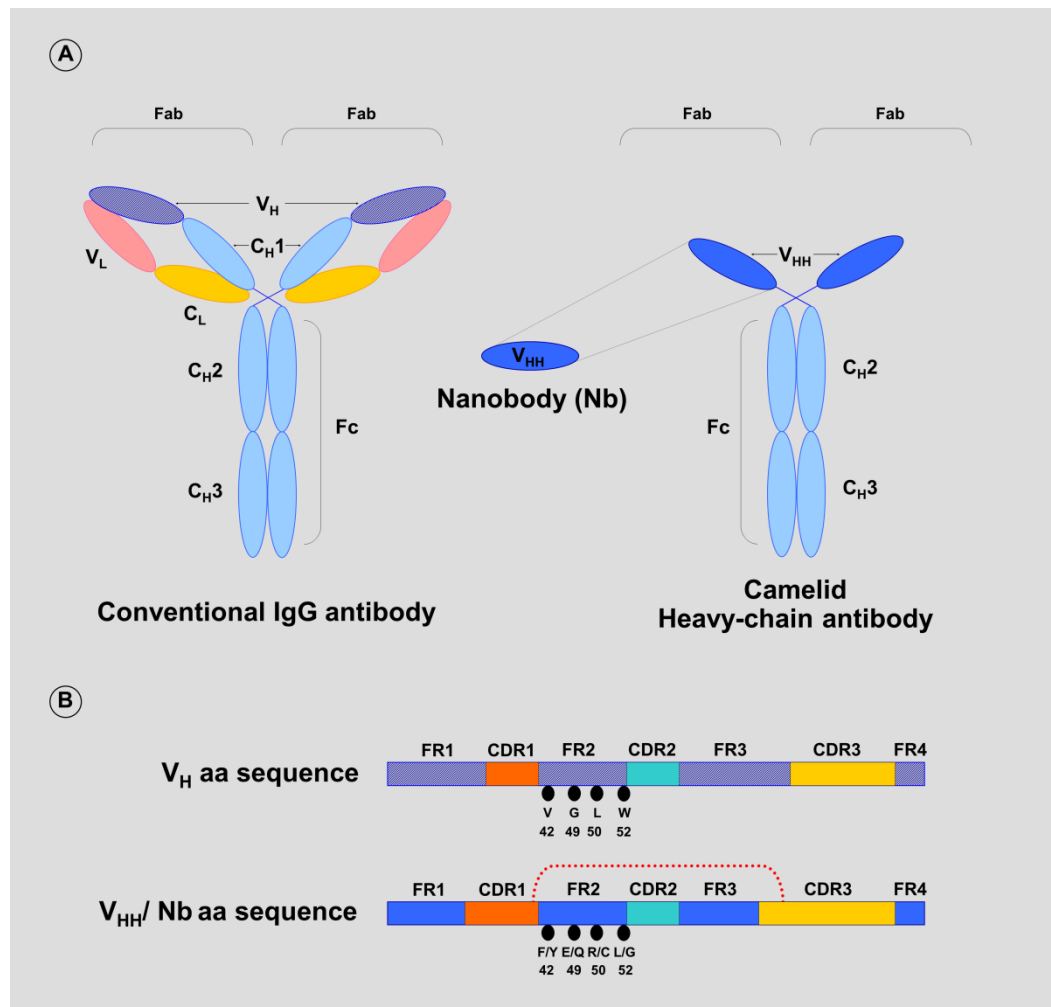
Figure 9. Schematic representation of nanobodies. (A) Composition of a classical antibody (left), a heavy-chain antibody present in camelids (right) and a single-domain antigen-binding entity derived from a heavy-chain antibody, the VHH or nanobody (middle). The domain structure within the heavy and light chains are shown. (B) Sequence organization of the VH and VHH with the four framework regions (FR) and the three complementary determining regions (CDRs). Within the dromedary VHH, the hallmark aminoacids in FR2 are given, and the inter-CDR disulfide bond is shown in orange. VL: variable domain of light-chain; CL: constant domain of light-chain; VH: variable domain of heavy-chain; CH: constant domain of heavy-chain; VHH: nanobody; Fab: fragment antigen biding; Fc: fragment crystallizable. Image obtained from Unciti-Broceta JD *et al.* Ther. Deliv. 2013, 4, 1321–1336.

OBJECTIVES

The general objective of this PhD project consists in studying the involvement of DNA G-quadruplexes in the progression of colorectal cancer, and searching for new therapeutic strategies based on ligands that target DNA G-quadruplexes. For that, the following specific objectives are proposed:

<u>OBJECTIVE 1:</u>

Study of DNA G-quadruplexes in the progression of colorectal cancer and analysis of their implication in tumor development and DNA damage.

<u>OBJECTIVE 2:</u>

Screening of chemical compounds in order to select ligands that target DNA G-quadruplexes and display antitumoral activity.

<u>OBJECTIVE 3:</u>

Rational production of nanobodies capable of specifically recognizing DNA G-quadruplexes harbored in oncogene promoters.

MATERIALS AND METHODS

1. Cell lines, patient samples and compounds

1.1. Cell lines

CRL1790, HCT116, HT29, SW480 and SW620 cell lines were purchased from American Type Culture Collection (ATCC). CRL1790 is a non-tumoral cell line. The remaining tumoral cell lines were genetically characterized respect to mutational status using the Cancer Cell Line Encyclopedia (CCLE) database.

1.2. Patient samples

Samples of patients with CRC were obtained from Hospital Universitario Virgen de las Nieves, Granada, Spain. Informed consent was obtained from all patients included in the study, which was approved by the local Ethical Committee of the University of Granada (Spain). Before treatment, tumor biopsies from 15 patients and histologically non-tumoral adjacent tissues from 7 patients were collected, and freshly frozen until RNA extraction. Both female (N = 10) and male (N = 12) donors were included, with a mean age of 59 years old (ranging from 51 to 66 years old). No statistically significant differences were found in terms of developmental stage, age or gender when comparing the two groups.

1.3. Naphthalene diimides (NDIs)

NDIs previously synthesized in a reported work (174, 175) were screened according to its antitumoral activity. Within the huge variety of NDI derivatives available, we selected carbohydrate-conjugated NDIs for our study in an attempt for a targeted antitumoral therapy according to previous results (174). Briefly, the carbohydrate-NDI conjugates were synthesized via the copper(I)-catalyzed azide alkyne cycloaddition (CuAAC) click reaction using NDI and sugar moieties as the azide and alkyne coupling partners respectively. Reverse phase preparative chromatography afforded the compounds in good yields (65-90%). All NDIs were dissolved in dimethyl sulfoxide (DMSO) (D8418, Sigma Aldrich), and stored at -20ºC.

1.4. Phenolic compounds

Five phenolic compounds available in a regular diet such as resveratrol, piceid, tyrosol, hydroxytyrosol, and gallic acid were screened according to its antitumoral activity. All of them were acquired from Sigma Aldrich (R5010 for resveratrol; 15721 for piceid; PHL80166 for tyrosol; H4291 for hydroxytyrosol; G7384 for gallic acid). Stock solutions were prepared in DMSO, and stored at -20ºC.

1.5. G4 ligands

Several synthetic compounds, including BMH21 (SLM1183, Sigma Aldrich), CX3543 (A12380, CliniSciences), CX5461 (HY-13323, MedChemExpress), and pyridostatin (CAY18013, Cayman Chemical) were used as established G4 ligands. Stock solutions were prepared in DMSO, and stored at -20ºC.

2. Cell culture experiments

2.1. Cell culture

Cells were cultured at 37ºC with 5% CO_2 atmosphere in medium supplemented with 10% fetal bovine serum (10270106, Gibco), 10 mg/mL penicillin (P0781, Sigma Aldrich), 10 mg/mL streptomycin (P0781, Sigma Aldrich), 100 mg/mL amphotericin (A2942, Sigma Aldrich), and 0.03% L-Glutamine (G8540, Sigma Aldrich), as recommended by the ATCC. For non-tumoral CRL1790, Minimum Essential Media (MEM) (M5650, Sigma Aldrich) was used, whilst Roswell Park Memorial Institute (RPMI) 1640 medium (L0501-500, Biowest) was used for the remaining cell lines. All procedures were carried out under aseptic conditions, and meeting biological safety requirements.

2.2. Cytotoxic assay

Cytotoxic activity was screened using Resazurin Fluorimetric Assay (R7017, Sigma Aldrich), according to the manufacturer's protocol. Cells seeded into 96-well plates ($8 \cdot 10^3$ cells/well) were incubated overnight for

attachment, and subsequently treated for 48 h with test compounds at decreasing concentrations via serial dilutions. A negative control with vehicle DMSO was included. Fluorescence was determined using Nanoquant Infinite M200 Pro multi-plate reader (Tecan). Half-maximal inhibitory concentration (IC_{50}) values were determined in triplicate by non-linear regression with Graphpad (Prism). Experiments were performed in biological triplicate.

2.3. Cell cycle analysis

Cell cycle analyses were carried out by flow cytometry with propidium iodide (PI) (P4864, Sigma Aldrich). Cells (10^6) were seeded into 10-cm culture dishes, incubated overnight for attachment, and three-times vigorously washed with phosphate-buffered saline (PBS) [137 mM sodium chloride (NaCl) (S9888, Sigma Aldrich), 2.7 mM potassium chloride (KCl) (P3911, Sigma Aldrich), 10 mM sodium phosphate (Na_3PO_4) (342483, Sigma Aldrich), 1.8 mM monobasic potassium phosphate (KH_2PO_4) (P0662, Sigma Aldrich)] prior to treatment. Cells were treated with test compounds at IC_{50} for 24 h or with the vehicle DMSO as control. In other experiments, to achieve cell cycle arrest, cells were treated with 2.5 mM thymidine (T1895, Sigma Aldrich) or serum-depleted medium for 24 h or 48 h. After treatment, cells were fixed with ice-cold 70% ethanol (ET00051000, Scharlab) on ice, and then stained with a freshly prepared solution of 0.04 mg/mL PI and 0.1 mg/mL ribonuclease A (19101, Qiagen) in PBS. Cytometry samples were incubated in darkness for 30 min at 37^oC. Cell cycle distribution was determined by an analytical DNA flow cytometer (FACSVerse, BD Biosciences) with instrument settings on low mode and FlowJo software.

2.4. Determination of intracellular localization of NDIs

Taking advantage of NDIs' strong emission in the red when excited with 594 nm-laser, the intracellular localization of uptaken NDIs was determined as previously described (174). Briefly, cells ($8 \cdot 10^4$ cells/well) were seeded on 13-mm circular coverslips and placed in 24-well plates. After overnight incubation for attachment, cells were treated with 5 µM of

testing compound for 1 h at 37ºC. Cells were then washed five times with PBS, and fixed with 4% (v/v) paraformaldehyde (PFA) (P6148, Sigma Aldrich) for 10 min at room temperature (RT). Finally, all coverslips were mounted onto slides (J2800AMNZ, Thermo Fisher Scientific) with Vectashield (H-1200, Vector) including 4',6-diamidino-2-phenylindole (DAPI) for nuclear counterstain. Images were acquired with a confocal microscope, and mean nuclear fluorescence intensity was measured using Fiji software (N > 50). To allow comparability, instrument settings were equally adjusted across the different cell lines.

2.5. NDIs uptake assay

Quantification of cellular uptake of NDIs in presence of GLUT1 transporter inhibitors such as BAY876 (SML1774, Sigma Aldrich) and WZB117 (400036, Sigma Aldrich) was achieved by fluorescence spectroscopy as previously described (174). In brief, cells ($2·10^5$ cells/well) were seeded in 12-well plates. After overnight incubation for attachment, cells were treated with 100 µM of a GLUT1 inhibitor for 1 h at 37ºC, and subsequently treated with 5 µM of testing compound for 2 h at 37ºC. Fluorescence intensity was detected with a Nanoquant Infinite M200 Pro multi-plate reader (Tecan) using 485 nm for excitation and 535 nm for emission; experiments were performed in triplicate. Fluorescence values were calculated as follows: % Fluorescence = A/B·100; where (A) corresponds to the fluorescence determined for each sample, and (B) is the fluorescence in the respective inhibitor-free control samples.

3. Protein techniques

3.1. BG4 expression

The scFv antibody BG4 specific for G4 structures was prepared using the expression vector pSANG10-3F-BG4 (plasmid no. 55756, Addgene) (Annex 1). This vector contains: (i) PelB sequence to direct the expressed BG4 to the bacterial periplasm, (ii) scFv sequence, (iii) 6xHis tag to allow purification, and (iv) FLAG tag serving as an epitope tag. Expression protocol was performed as previously described (25). Briefly,

BL21 (DE3) *Escherichia coli* (*E.coli*) were heat-shock transformed with pSANG10-3F-BG4 and grown in 2×YT medium [16 g/L tryptone (1616.00, Condalab), 10 g/L yeast extract (1702.00, Condalab), 5 g/L NaCl] supplemented with 50 µg/ml kanamycin (K1377, Sigma Aldrich) until OD600 = 0.6. Antibody expression was induced by addition of 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) (367-93-1, Sigma Aldrich) followed by incubation with shaking for 4 h at 37 °C. Then, cells were centrifuged and resuspended in TES buffer [50 mM Tris HCl at pH 8.0 (77-86-1, Scharlau), 20% (w/v) sucrose (S9378, Sigma Aldrich)] for 10 min on ice. The bacterial slurry was then two-fold diluted in water, and left on ice for 10 min before centrifugation at 16,000g for 30 min at 4 °C. Supernatant was collected and loaded onto HisTrap FF niquel spin columns (GE Healthcare) for immobilized metal affinity chromatography. After washing with 10 mM imidazole (I202, Sigma Aldrich) in PBS at pH 7.4, BG4 antibody was eluted in 250 mM imidazole in PBS at pH 7.4. Finally, BG4 was concentrated using an Spin 5K MWCO Concentrator (Agilent Technologies), and dialyzed against PBS buffer. BG4 aliquots were stored at -20 °C. Confirmation of BG4 expression and purification was achieved by western blot analysis with anti-histidine antibody (Annex 2).

3.2. Western blot

Protein extract from different experimental conditions was obtained using RIPA lysis buffer [50 mM Tris HCl, 150 mM NaCl, 1% (v/v) octylphenoxypolyethoxyethanol (NP40) (492016, Sigma Aldrich), 0.5% (w/v) sodium deoxycholate (D6750, Sigma Aldrich), 1 mM (ethylenedinitrilo)tetraacetic acid (EDTA) (E9884, Sigma Aldrich), 0.1% (w/v) sodium dodecyl sulfate (SDS) (436143, Sigma Aldrich) at pH 7.4]. Several protease inhibitors, including 10% (v/v) phenylmethylsulfonyl fluoride (PMSF) (P7626, Sigma Aldrich) and 1% (v/v) protease inhibitor cocktail (PIC) (P8340, Sigma Aldrich) were added. As a phosphatase inhibitor, 1% (v/v) sodium orthovanadate (S6508, Sigma Aldrich) was also included. Quantification of protein levels was achieved by Bradford method following manufacturer's protocol (500-0006, BioRad). Protein content was loaded on SDS-polyacrylamide gel (1610148, Biorad) at

required concentration for electrophoresis, and wet transferred to nitrocellulose membranes (66485, Pall corporation). Membranes were blocked with 5% semi-skimmed milk, and incubated overnight at 4ºC with primary antibodies. Then, membranes were three-times washed with 0.05% Tween 20 (P7949, Sigma Aldrich) in PBS (T-PBS) for 10 min and incubated with horseradish peroxidase (HRP)-labeled antibodies for 1 h at RT. After washing the membranes, luminol solution (1705060, Biorad) was added, and chemiluminiscence signals were measured using Image Quant LAS 4000 (GE Healthcare Life Sciences). Experiments were performed in biological triplicate. Protein levels were quantified by ImageJ. Antibodies together with respective dilutions used are listed in Annex 2.

3.3. Immunofluorescence assays

Cells were seeded on 13-mm circular coverslips, placed in 24-well plates, and incubated overnight for attachment. After exposure to different experimental conditions, fixation was performed with 4% (v/v) PFA for 10 min at RT, permeabilization with 0.1% (v/v) Triton X-100 (T8787, Sigma Aldrich) for 10 min at RT, and blocking with 10% bovine serum albumin (BSA) (A7906, Sigma Aldrich) containing 0.5% (v/v) Triton X-100 for 30 min at RT. Primary antibodies were incubated for 1 h at RT, and secondary antibodies for 30 min at 4 °C in darkness. Finally, all coverslips were mounted onto slides (J2800AMNZ, Thermo Fisher Scientific) with Vectashield (H-1200, Vector) including DAPI for nuclear counterstain. Following a different protocol, BG4 immunofluorescence was conducted as previously described (19). Images were acquired on a Confocal Zeiss LSM 710 inverted microscope with a 63x immersion objective. The images were captured from randomly selected fields of view. BG4 mean nuclear fluorescence intensity was quantified using Fiji software. Antibodies used are listed in Annex 2.

3.4. Autophagy and organelle assays

These experiments were performed at the Laboratory of Autophagy Signalling (Babraham Institute in Cambridge, United Kingdom) headed by Dr. Nicholas Ktistakis, thanks to a PhD international stay. Such stay was

achieved from 15/05/2021 to 15/07/21, and was funded by "Ayudas a la movilidad para estancias breves y traslados temporales para beneficiarios del programa de Formación del Profesorado Universitario (FPU)" from "Ministerio de Universidades, Gobierno de España".

The stable cell line HEK293-GFP-ATG13 and HEK293-GFP-LC3 were kindly provided by Dr. Nicholas Ktistakis to study autophagy dynamics. Several autophagy markers were analyzed by immunofluorescence assays upon treatment with test compound or vehicle for 24 h. Torkinib (PP242) (508770, Thermo Fisher Scientific), a mTOR inhibitor, was used as a positive control for induction of autophagy. Ivermectin (11412651, Thermo Fisher Scientific), an antiparasitic drug, was used as positive control for induction of selective autophagy. Staining was performed as described previously (176). In brief, cells seeded on glass coverslips in 12-well plates were fixed with 3.7% (v/v) formaldehyde (252549, Sigma Aldrich) for 10 min, and permeabilized with NETgel [150 mM NaCl, 5 mM EDTA, 50 mM Tris·HCl at pH 7.4, 0.05% (v/v) NP40, 0.25% (w/v) gelatin (G-6650, Sigma Aldrich), and 0.02% (w/v) sodium azide (S2002, Sigma Aldrich)] containing 0.25% NP40 for 10 min. Then, cells were stained for 30 min with primary antibodies, washed three times for 5 min with NETgel, stained for 30 min with secondary antibodies, and again washed three times for 5 min with NETgel. Finally, coverslips were mounted with Aqua Poly Mount mounting medium (18606, Polysciences). All steps were performed at RT. Images for quantification of puncta were captured with a a Zeiss Axio Imager D2 microscope equipped with a 63x immersion objective. Ten images were captured from randomly selected fields of view, and quantification of puncta was performed by ImageJ using "Cell counter" plugin. Antibodies used are listed in Annex 2.

Parental HEK293 cell line was kindly provided by Dr. Nicholas Ktistakis to analyze the status of different organelles after exposure to different experimental conditions. Immunofluorescence experiments were performed as described in the previous paragraph. Images were captured with a Zeiss Axio Imager D2 microscope equipped with a 63x immersion objective. The images were captured from randomly selected fields of

view. Fluorescence intensity was quantified using Fiji software. Antibodies used are listed in Annex 2.

4. DNA and RNA techniques

4.1. qRT-PCR

Total cellular RNA was isolated from cells using Trizol Reagent (15596, Invitrogen). Reverse transcription was conducted using RevertAid First Strand cDNA Synthesis Kit (K1622, Thermo Fisher Scientific) with random primers according to the manufacturer's protocol. Quantitative PCR (qPCR) was performed on 7900HT Fast Real-time PCR System (Applied Biosystems) with SYBR Green (4309155, Thermo Fisher Scientific), a final concentration of primers at 500 nM and using 1 µL of 100 ng/µL cDNA in 10 µL of reaction. Cycling conditions were 95 °C for 10 min, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C, and a final dissociation stage. Target mRNA levels were normalized to actin (ΔCt), and fold change was determined using the $2^{-\Delta\Delta CT}$ method. Experiments were conducted in biological triplicate. Primers used are listed in Annex 3.

4.2. Breaks labeling in situ and sequencing

Measuring the location and frequency of DNA double-strand breaks (DSBs) along the genome at single-nucleotide resolution was performed using breaks labeling in situ and sequencing (BLISS) protocol (177). The workflow is schematically described in Annex 4. In brief, CRL1790, SW480 and SW620 ($3 \cdot 10^5$) cells were attached onto 13-mm circular coverslips, placed in 24-well plates, and arrested at G0/G1 phase by serum deprivation for 48 h. Then, cells were fixed with 8% (v/v) metanol-free PFA (15710, Electron Microscopy Sciences) for 10 min at RT and lysed following the guidelines in the protocol (178). DSB ends were in situ blunted using Quick Blunting kit (E1201, New England Biolabs) for 1 h at RT, and tagged with double-strand DNA adapters purchased from Integrated DNA Technologies using T4 ligase (10799009001, Sigma Aldrich) overnight at RT. As described in literature (177), these adapters contain the T7 promoter sequence, the RA5 Illumina sequencing adapter,

a random stretch of 8-12 nucleotides that serves as a unique molecular identifier (UMI), and a sample barcode suitable for multiplexing. After removal of unligated adapters by successive washes with high-salt buffer (10 mM Tris HCl, 2 M NaCl, 2mM EDTA, 0.5% Triton X-100) for 1 h at 37 ⁰C, genomic DNA was extracted and sonicated to achieve a mean fragment size of 300-500 bp using Bioruptor Plus (Diagenode). Following sonication, the sequence immediately downstream to the tagged DSBs was transcribed via T7-mediated in vitro transcription with Megascript kit (AM1334, Thermo Fisher Scientific) overnight at 37 ⁰C. Resulting RNA was used for Illumina library preparation and sequencing. The RA3 Illumina sequencing adapter was ligated for 1 h at 28 ⁰C with T4 RNA ligase 2 truncated (M0242, New England Biolabs). Then, RNA was reverse transcribed with SuperScript III Reverse Transcriptase kit (18080044, Thermo Fisher Scientific) for 1 h at 50 ⁰C. Finally, the libraries were indexed and amplified using NEBNext High-Fidelity (M0541, New England Biolabs). All primers used are published in a previous work (178). Quality control of libraries was achieved using Bioanalyzer High Sensitivity DNA Assay (Agilent Technologies). Sequencing was conducted at Centre for Genomics and Oncological Research (Granada, Spain) using NextSeq High Output platform (Illumina), and the paired-end 75 bp format. A total of 20 million reads per sample were obtained.

Bioinformatic processing of sequencing data was achieved following the pipeline described before (177). The association between DSBs and G4s location was interrogated using G4s data (GSE63874) previously generated for the human genome (15). In this study, authors made separate bedgraph files available with the G4 density for each strand. We used the sum of the plus and minus strands in our analysis.

4.3. Sequencing of G4s

Chromatin isolated and purified for BG4 chromatin immunoprecipitation experiments (detailed in the next section) was subjected to PCR with specific primers to determine the sequence of G4s. Sequences of these primers are listed in Annex 3. The reactions were

performed in 1x PCR GC buffer, containing 0.5 µM of forward and reverse primers, 200 µM dNTPs, 3% DMSO, 1 U Phusion High-Fidelity polymerase (F530, Thermo Fisher Scientific) and 50 ng of DNA. PCR products were amplified in a Veriti Thermal Cycler (Applied Biosystems) with the following cycling conditions: 98 ºC for 30 s, followed by 30 cycles of 98 ºC for 10 s, 63 ºC for 30 s, and 72 ºC for 30 s, together with a final extension at 72 ºC for 10 min. The PCR products were purified with GenElute PCR Clean-up kit (NA1020, Sigma Aldrich), and sent to STAB Vida for Sanger sequencing. The mutational status of G4s was inferred by alignment of these sequences with the consensus one, using Multialin software (179).

4.4. BG4 chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) with BG4 antibody was performed as previously described (25) with some modifications. Briefly, crosslinking of $1 \cdot 10^7$ cells was performed via 1% (v/v) formaldehyde (F8775, Sigma Aldrich) treatment for 10 min at RT, and quenched with 0.12 M glycine (56-40-6, Scharlau) for 10 min at RT. Cells were harvested and nuclei were prepared by incubation with the lysis buffer [50 mM HEPES buffer at pH 7.5 (83264, Sigma Aldrich), 140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate] following the same protocol as in a previous publication (180). Next, the lysate was sonicated to obtain DNA fragments of 100–500 bp using Bioruptor Plus (Diagenode) by optimizing specific cycling conditions for each cell line. Chromatin was treated with 20 µg/mL RNase A (19101, Qiagen) to remove RNA G4s and DNA–RNA G4 hybrids, and blocked with 1% (w/v) BSA to reduce nonspecific interactions. Per ChIP reaction, ~500 ng of sonicated, cross-linked chromatin was incubated with 500 ng of recombinant BG4 for 1 h at 16 ºC with shaking at 1400 rpm. Per condition, two BG4-containing reactions were included, along with two BG4-free reactions in parallel as the Mock negative ChIP controls. Simultaneously, Pierce™ Protein A/G Magnetic Beads (88802, Thermo Fisher Scientific) were used to capture anti-FLAG antibody (F1804, Sigma Aldrich) following manufacturer's instructions. Then, these beads were added to each

sample and incubated for 1 h at 16 °C with shaking at 1400 rpm. Immobilized complexes were washed five times in high salt buffer to reduce nonspecific interactions at 4 °C. Elution of immunoprecipitated and chromatin crosslink reversal were performed by 0.25 mg/mL proteinase K (EO0491, Thermo Fisher Scientific) digestion for 1 h at 37 °C with shaking at 1400 rpm and 2 h at 65 °C with shaking at 1400 rpm. At this point, the two BG4-containing reactions were combined. In a separate manner, the two BG4-free reactions were combined (mock negative control). Finally, eluted DNA was purified using phenol:chloroform:isoamyl alcohol (25:24:1) protocol (P2069, Sigma Aldrich). Experiments were performed in biological triplicate.

The immunoprecipitated samples (with BG4 and without BG4 as mock) were subjected to G4 enrichment quantification via qPCR using SYBR Green (4309155, Thermo Fisher Scientific) with a 7900HT Fast Real-time PCR System (Applied Biosystems). Cycling conditions were 95 °C for 10 min followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C, and a final dissociation stage. Primer pairs that target *CMYC* and *KRAS* G4 regions were employed (Annex 3). Normalization of the data for each cell line and for each primer pair was performed by the fold enrichment method, where ChIP signals (with BG4) were divided by the no-antibody signals (without BG4), representing the ChIP signal as the fold increase in signal relative to the background signal. Experiments were conducted in triplicate.

4.5. POLR1A chromatin immunoprecipitation

ChIP with anti-POLR1A antibody was performed as previously described (181). Cells ($1 \cdot 10^7$) were crosslinked with 1% (v/v) formaldehyde (F8775, Sigma Aldrich) for 30 min at RT, and quenched with 0.125 M glycine (56-40-6, Scharlau) for 5 min at RT. Then, cells were lysed, chromatin was extracted as in a previous protocol (182), and sheared to 500-1,000 bp using Bioruptor Plus (Diagenode). In each ChIP reaction, 100 µg of sonicated DNA was immunoprecipitated with 5 µg of anti-POLR1A (sc-48385, Santa Cruz Biotechnology), and collected with

Pierce Protein A/G Magnetic Beads (88802, Thermo Fisher Scientific). DNA was finally purified using phenol:chloroform:isoamyl alcohol (25:24:1) protocol (P2069, Sigma Aldrich). Experiments were performed in biological triplicate.

The immunoprecipitated samples were analyzed by qPCR using SYBR Green (4309155, Thermo Fisher Scientific) with a 7900HT Fast Real-time PCR System (Applied Biosystems). Cycling conditions were 95 °C for 10 min followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C, and a final dissociation stage. Primers used for this study are listed in Annex 3. Normalization of the data was performed by the percent input method, where POLR1A ChIP signal was divided by the signalobtained from an input sample (representing the amount of chromatin used in the ChIP). Experiments were conducted in triplicate with 1% of starting chromatin as input.

5. G4 binding assays

5.1. G4s prefolding

G4-oligonucleotides listed in Annex 5 were purchased from Integrated DNA Technologies. All of them were dissolved in G4s buffer [10 mM potassium phosphate buffer at pH 7.0 [5.3 mM potassium phosphate dibasic (K_2HPO_4) (P9666, Sigma Aldrich) and 4.6 mM potassium phosphate monobasic (KH_2PO_4) (P9791, Sigma Aldrich)] containing 100 mM KCl]. G4-oligonucleotides were then heated at 95°C for 10 min, slowly cooled to RT, and stored at 4 °C.

5.2. Fluorescent intercalator displacement assay

We used TOPRO3 (T3605, Thermo Fisher Scientific) as a fluorescent intercalator for fluorescent intercalator displacement (FID) assays. In particular, 5 mM TOPRO3 was incubated with 10 mM pre-folded G4s (Annex 5) and exposed to the test compound in 96-well plates. TOPRO3 was excited at 642 nm and emission profile was monitored between 650–800 nm with Infinite M200 Plate Reader (Tecan). All assays

were conducted in triplicate. Fluorescence values were calculated as follows: % Fluorescence = A/B·100; where (A) is the fluorescence value in presence of the test compound and (B) corresponds to the fluorescence value in the test compound-free controls. The affinity was estimated by the DC_{50} value, which corresponds to the required concentration of the test compound to induce a 50% fluorescence decrease. DC_{50} values were determined by non-linear regression with Prism Graphpad using increasing concentrations of the test compound.

5.3. PCR-stop assay

The PCR stop assay was performed as previously described [183], with some modifications, and using a test oligonucleotide including the G4 sequence and a partially complementary oligonucleotide that hybridizes. Sequences of these oligonucleotides are listed in Annex 5. The reactions were performed in 1x PCR Combination buffer, containing 20 pmol of each pair of oligonucleotides, 0.2 mM dNTPs, 2.5 U Hot Start *Taq* polymerase (733-1331, VWR), and increasing amounts of the test compound. PCR products were amplified in a Veriti Thermal Cycler (Applied Biosystems) with the following cycling conditions: 95 ºC for 15 min, followed by 30 cycles of 95 ºC for 30 s, 58 ºC for 30 s, and 72 ºC for 30 s. Amplified products were resolved on 3% agarose gel in 1x TBE (100 mM Tris base, 100 mM boric acid (B6768, Sigma Aldrich), 2 mM EDTA), and stained with GelGreen (41005, Biotium). Gel Image was analyzed on ImageQuant LAS 4000. Three independent reactions were conducted per concentration.

5.4. Circular dichroism experiments

Circular dichroism (CD) experiments were performed at the "Departament of Physical Chemistry, University of Granada" (Granada, Spain). CD spectra were recorded at 25 °C on a JASCO 715 CD spectropolarimeter in G4s buffer conditions (10 mM potassium phosphate buffer containing 100 mM KCl at pH 7.0). The concentration of the prefolded G4 DNA was 10 µM, and the test compound was added at 100µM, and incubated overnight prior to register the new spectrum. The used wavelength range was 220-320 nm with 100 nm/min as scan speed.

The cuvette path length was 0.1 cm and three accumulation spectra were averaged for each measurement. The G4-oligonucleotides used in the current study are listed in Annex 5.

5.5. Ultraviolet-visible experiments

Ultraviolet-visible (UV−vis) experiments were performed at the "Departament of Physical Chemistry, University of Granada" (Granada, Spain). UV−vis absorption spectra were registered in a Varian Cary 50 UV-vis spectrophotometer at 25 °C. Concentration of the prefolded G4 DNA was 5 µM in G4s buffer (10 mM potassium phosphate buffer containing 100 mM KCl at pH 7.0). Once the DNA was placed in the cuvette, a concentrated solution of the test compound (1 mM) was routinely added, 1 µL each time, with a Hamilton syringe, and subsequently mixed with a pipette. After each addition of the test compound, a UV-vis spectrum was recorded. In total, 10 µL of the test compound solution were added, with a final ratio of 1:20 G4 DNA:compound. For the blank, the same experiment with the successive additions was repeated beginning just with buffer in the cuvette. Then, each titration spectrum was subtracted from its corresponding blank. The path length of the cuvette was 0.3 cm and the wavelength range used was 235-320 nm. Experiments were conducted in triplicate. The G4-oligonucleotides used in the current study are listed in Annex 5.

5.6. Nuclear magnetic resonance experiments

Nuclear magnetic resonance (NMR) experiments were performed at the "Institute of Physical Chemistry Rocasolano, CSIC" (Madrid, Spain) by Dr. Carlos González. G4-oligonucleotide listed in Annex 5 was resuspended in H_2O/D_2O 9:1 in 25 mM potassium phosphate at pH 7.0. NMR titrations were performed by adding increasing amounts of the test compound to the oligonucleotide solution at 100 µM. Different R ratios = [test compound] / [DNA] were considered (R= 0, 1, 2). NMR spectra were acquired in Bruker Avance spectrometer operating at 600 MHz, and processed with Topspin software. Water suppression was achieved by

including a WATERGATE module in the pulse sequence prior to acquisition.

5.7. In vitro transcription assays mediated by RNA polymerase I

In vitro transcription assays were performed at the "Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham" (Alabama, United States of America) by Professor David A. Schneider. A DNA template containing the *Saccharomyces cerevisiae* ribosomal DNA promoter (-250 to +56 with respect to the transcription start site) fused to a 700 base pair segment of the human ribosomal DNA starting at +3412 (containing the test G4 sequence) was synthesized (Genscript). This template was amplified by PCR and incorporated into a multi-round in vitro transcription assay for RNA Polymerase I. Transcription assays included all purified components (Pol I, core factor, Rrn3, and TBP) and were performed as previously described (184, 185). After preinitiation complexes were formed, the test compound (or vehicle) was added to final concentrations indicated. Transcription was initiated with substrate NTPs (200 µM ATP, GTP, CTP, 15 µM UTP, and 10 µCi α-^{32}P UTP) and reactions were halted by addition of 1 M ammonium acetate in 95% ethanol. RNA was precipitated, subjected to 5% denaturing polyacrylamide gel electrophoresis, and visualized by phosphorimaging. At least four independent reactions were conducted per concentration. The runoff RNA species were quantified using Quantity one software (BioRad).

5.8. Thioflavin T competition assay

Cells seeded into 13-mm circular coverslips and placed in 24-well plates (8·10^4 cells/well) were non-treated or treated with the test compound for 3 h. Then, cells were fixed for 10 min in cold methanol (ME03022500, Scharlau), rinsed twice with PBS, and incubated with 5 µM thioflavin T (ThT) (T3516, Sigma Aldrich) for 15 min at RT. Cellular nuclei were counterstained with PI and visualized by confocal microscopy as previously described (186). Images were acquired on a Confocal Zeiss LSM 710 inverted microscope with a 63x immersion objective. The images were captured from randomly selected fields of view.

6. In vivo experiments

6.1. Xenograft studies

NOD scid gamma (NSG) mice were purchased from CIBM-UGR and housed at the animal facility according to institutional guidelines (Approved Ethical Committee #152-CEEA-OH-2016). For xenograft generation, $1\cdot10^5$ SW480 cells in 100 µL PBS were subcutaneously injected in the flank of 8-weeks old female mice. Treatment started when tumors reached ~ 20 mm^3. Mice were randomly divided into two groups of seven mice each, and treated with either vehicle DMSO or 200 mg/kg of the test compound intraperitoneally every other day for 38 days. Animals were monitored every two days after cell injection until the end time point, when they were sacrificed, and tumors were dissected for further analysis. Specifically, tumor volumes were determined every two days using digital calipers according to the formula: In progress tumor volume = (π x length x width^2)/6 (187).

6.2. Immunohistochemical analysis of tumor sections

Fixation, paraffin-embedding, and sectioning of tumor samples were performed by the histopathology core service at the Centre for Genomics and Oncological Research (Granada, Spain). To evaluate the percentage of proliferating cells, tumor sections were immunostained with Ki67 and counterstained with hematoxylin and eosin at Atrys Health (Barcelona, Spain). The staining was visualized using NDP.view2 Viewing software (Hamamatsu) and Ki67 coverage was quantified on ten different images per tumor using Fiji software. To measure BG4 signal, tumor sections were dewaxed and rehydrated following standard methods. Epitope retrieval was performed at 100 °C for 20 min with citrate buffer at pH 6.0 (C9999, Sigma Aldrich) according to previous studies (82). After blocking, staining was achieved with BG4 antibody overnight at 4 ºC, following a 1 h incubation with anti-FLAG at RT, and a 30 min incubation with an anti-mouse antibody at RT in darkness. Slides were then counterstained for 5 min with DAPI to visualize the cell nuclei. Antifade Mowiol (81381, Sigma Aldrich) was used as mounting media. Images were acquired on a

Confocal Zeiss LSM 710 inverted microscope with a 63x immersion objective. BG4 mean nuclear fluorescence intensity was quantified using Fiji software (N > 2,000). The images were captured from randomly selected fields of view. Antibodies used are listed in Annex 2.

7. Generation and screening of nanobodies targeting G4s

These experiments were performed at the Laboratory of Biomedical Microbiology and Immunology (University of Veterinary Medicine and Pharmacy in Kosice, Slovakia) headed by Professor Mangesh Bhide, thanks to a PhD international stay. Such stay was achieved from 05/05/2019 to 28/06/19, and was funded by "Boehringer Travel Grants" from Boehringer Ingelheim Fonds.

7.1. In vitro immunization of llama lymphocytes

Llama (*Lama glama*) lymphocytes were in vitro immunized as previously described (188) with *CMYC* G4 oligonucleotide (189). The sequence of this oligonucleotide is listed in Annex 6. Heparinized blood was collected from a 4 years old healthy llama. Peripheral blood mononuclear cells (PBMCs) were immediately isolated by density centrifugation using Histopaque medium (10771, Sigma Aldrich) according to manufacturer's instructions. PBMCs in the buffy coat were washed with eRDF medium (RPMI:DMEM:F12 in the ratio of 2:1:1 as previously described) (190), and pelleted by centrifugation at 400g for 20 min at RT. Then, PBMCs were resuspended in 20 mM Leu-Leu methyl-ester hydrobromide (LLME) (L7393, Sigma Aldrich), and incubated for 20 min at RT to induce apoptosis of natural killer cells. PBMCs were harvested by centrifugation at 400g for 20 min, washed, and again resuspended in 1 mL of eRDF. Cell density was measured and adjusted to $1 \cdot 10^6$ cells/mL using eRDF. Cell suspensions (2 mL/well) were incubated overnight in 12-well plates at 37 °C with 1 ng/mL of interleukin-2 (IL-2) of llama (produced at laboratory of Dr. Mangesh Bhide and essential for B lymphocytes activation and differentiation), 1 ng/mL of interleukin-4 (IL-4) of llama (produced at laboratory of Dr. Mangesh Bhide and essential for B lymphocytes activation and differentiation), 0.25 µM Class A CpG

oligonucleotide ODN 2216 (tlrl-2216, InVivoGen) as a TLR9 agonist, and 20 µL/mL Mycokill (P11-016, PAA Laboratories). Antibody production was then induced by adding 5 µg of the prefolded *CMYC* G4 oligonucleotide as antigen. Cells were incubated with the antigen for 24 h at 37 ºC before adding 0.25 µM Class B CpG oligonucleotide ODN 2006 (tlrl-2006, InVivoGen), as a strong activator of B lymphocytes. The incubation was continued until 72 h and cell viability was checked every day under the microscope.

7.2. VHH amplification

Total RNA from immunized lymphocytes was isolated and treated with DNase I using RNeasy kit (74104, Qiagen) according to manufacturer's instructions. RNA was reverse transcribed to cDNA using RevertAid reverse transcriptase (EP0441, Thermo Fisher Scientific) following manufacturer's protocol. VHH repertoire was subjected to PCR with gene-specific primers that amplify VHH sequence and incorporate restriction sites of SfiI to allow the next cloning step. Amplification was performed in six independent reactions using High-fidelity Hot Start polymerase (PCR-205, Jena Bioscience) with the following cycling conditions: 95 ºC for 2 min, followed by 30 cycles of 95 ºC for 20 s, 56 ºC for 30 s, and 68 ºC for 1 min, and final extension at 68 ºC for 10 min. VHH amplicons from the six PCR reactions were purified by PCR Clean-up kit (740609.250, Macherey-Nagel) and mixed together. Primers used are listed in Annex 6.

7.3. Preparation of VHH library

Purified VHH amplicons were digested with restriction enzyme SfiI (R0123, New England Biolabs) for 1 h at 50 ºC. Digested DNA fragments were column purified by PCR Clean-up kit (740609.250, Macherey-Nagel) and ligated into SfiI digested phagemid pJB12 (kindly provided by Dr. Mangesh Bhide) using T4 DNA ligase (EL0011, Thermo Fisher Scientific). Such ligation leads to the expression of a VHH-gene III fusion applicable for phage display, due to read-through of the amber codons whenever expressed in strains with amber suppressor tRNA, like *E. coli XL1-Blue*.

After purification, ligation products were transformed into *E. coli XL1-Blue* (200236, New England Biolabs) by electroporation with Gene Pulser xCell™ (165-2660, Biorad) using a preset method for *E. coli* (voltage: 1.8 KV, capacitance: 25 μF, and resistance 200 Ω) in 1 mm cuvette. Transformants grew in LB agar [1% (w/v) tryptone (1616.00, Condalab), 0.5% (w/v) yeast extract (1702.00, Condalab), 1% (w/v) NaCl, 1.5% (w/v) agar (1800.00, Condalab)] plates containing 50 μg/mL tetracycline (T0150, Duchefa Biochemie BV) as selection antibiotic of *E. coli XL1-Blue*, and 50 μg/mL chloramphenicol (C0113, Duchefa Biochemie BV) as selection antibiotic of pJB12. Five randomly selected transformants were checked by PCR (colony PCR) using vector-specific primers (listed in Annex 6) to confirm the presence of cloned VHH sequence. Such PCR was performed with High-fidelity Hot Start polymerase (PCR-205, Jena Bioscience) and cycling parameters were: 95 ºC for 2 min, followed by 35 cycles of 95 ºC for 30 s, 65 ºC for 30 s, and 72 ºC for 1 min, and final extension at 72 ºC for 10 min. Finally, all colonies were scrapped and stored in LB medium [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl] with 50% glycerol at -80 °C for further experiments. To note, this suspension of *E. coli* is referred to as a "VHH-*E. coli* library", which contains a large repertoire of VHH variants cloned in pJB12 backbone into *E. coli XL1-Blue*.

7.4. Phage packaging

The VHH library was displayed on phage particles after helper phage infection of the transformed *E. coli XL1-Blue*. This method produces virus particles with the cloned VHHs at the viral surface due to their fusion with gene III (encoding pIII, a virion surface protein) of the phagemid pJB12 vector. For phage display, small aliquots of frozen VHH-*E. coli* library were inoculated in 2xYT medium [16 g/L tryptone (1616.00, Condalab), 10 g/L yeast extract (1702.00, Condalab), 5 g/L NaCl] supplemented with tetracycline 50 μg/mL (T0150, Duchefa Biochemie BV), chloramphenicol 50 μg/mL (C0113, Duchefa Biochemie BV), and 4% glucose (14431-43-7, Duchefa Biochemie BV) to obtain the initial OD_{600} ~ 0.1. The culture was incubated for 8 h with shaking at 200 rpm and at

37ºC until reaching the final OD_{600} = 0.5, which followed the addition of VCSM13 helper phages (200251, Agilent Technologies) to superinfect VHH-*E. coli* library with a ratio 1:5 (bacteria:phage). Next, the culture was incubated for 30 min at 37 °C (no shaking) and centrifuged at 3,500 g for 20 min to eliminate unbound phages in the supernatant. The resulting superinfected cell pellet was resuspended in 2xYT medium supplemented with 50 µg/mL tetracycline (T0150, Duchefa Biochemie BV), 50 µg/mL chloramphenicol (C0113, Duchefa Biochemie BV), and 50 µg/mL kanamycin (K0126, Duchefa Biochemie BV) as selection antibiotic of infected cells, and incubated overnight with shaking at 200 rpm and at 25 °C to allow phage escaping. Escaped phages were precipitated in 20% polyethylene glycol (P5413, Sigma Aldrich), 2.5 M NaCl, and resuspended in phage dilution buffer (10 mM Tris-HCl at pH 7.5, 20 mM NaCl, and 2 mM EDTA). To note, this suspension of VCSM13 phages is referred to as a "VHH-phage library", which contains a large repertoire of VHH variants cloned in pJB12 backbone into phages. The titration of phages was performed by a spectrophotometric method by measuring the absorbance at 269 and 320 nm (191). The number of phages per mL was calculated using the formula: $(A269-A320 * 6 \cdot 10^{16})$/(number of bases per virion).

7.5. Biopanning

The selection (biopanning) of phages expressing antigen specific VHH on pIII protein was performed with Pierce™ NeutrAvidin™ coated high capacity plates (15508, Thermo Fisher Scientific). First of all, VHH-phage library $(7 \cdot 10^{11}$ phages) were incubated with neutravidin coated wells for 1 h at RT and the supernatant was recovered. This step removes the phages that bind nonspecifically to neutravidin. Then, 5 µg of biotinylated *CMYC* G4 oligonucleotide (with the same sequence as that used for in vitro immunization and listed in Annex 6) was captured separately on new neutravidin coated wells following manufacturer's instructions. The VHH-phage library was then incubated with the G4 oligonucleotide immobilized on the neutravidin wells for 1 h at RT. After discarding the supernatant with non-specific phages, wells were vigorously washed 10 times as follows: nine washings with 0.1% T-PBS for 2 min at

RT, and one last washing with PBS for 2 min at RT. Finally, antigen specific phages were eluted by incubation with PBS containing 0.25 mg/mL trypsin at pH 7.4 (T1426-50, Sigma Aldrich) for 10 min at RT. Elution was achieved thanks to the presence of a "trypsin cleaving site" fused to VHH. Trypsin was immediately neutralized with 5% BSA (A7906, Sigma Aldrich). The number of eluted phage particles were calculated as described in subsection 7.4. These eluted phages were identified as phages from "1st round" of biopanning, and were amplified by superinfection of *E. coli XL-1 blue* and subsequent precipitation as described above. Then, the same protocol of biopanning was repeated to obtain eluted phages from "2nd round" of biopanning. In total, three rounds of biopanning were performed while repeating the same procedure in each round.

7.6. Quantitative evaluation of antigen specific phages

The antigen specific phages after each round of biopanning (as well as phages from "round zero" before biopanning) were evaluated by enzyme link immunosorbent assay (ELISA). Wells of Pierce™ NeutrAvidin™ coated high capacity plates (15508, Thermo Fisher Scientific) were coated with 5 µg of biotinylated G4 oligonucleotides or vehicle as blank for 90 min at RT. Sequence of these oligonucleotides are listed in Annex 6. Wells were then blocked with 1% BSA for 60 min at RT and subsequently washed three times with 0.1% T-PBS. Around 10^{10} phages were added and incubated for 60 min at 27 ºC with gentle agitation at 160 rpm. After three washings with 0.1% T-PBS, anti-phage (M13) antibody (1:1,000) (GE HealthCare, 27942001) was added and incubated for 60 min at 27 ºC with gentle agitation at 160 rpm. After three washings, Pierce™ Recombinant Protein A/G-HRP Conjugated (1:10,000) (Thermo Fisher Scientific, 32490) was then added and incubated for 60 min at 27 ºC with gentle agitation at 160 rpm. After three washings, ELISA development was carried out with 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) solution (1120452101, Sigma Aldrich) as substrate of HRP by incubation for 20 min in the dark. Finally, absorbance at 405 nm was determined using Nanoquant Infinite M200 Pro multi-plate reader

(Tecan). Experiments were performed in triplicate. Absorbance values were calculated as follows: Absorbance = A/B; where (A) corresponds to the absorbance determined in the presence of G4 oligonucleotides immobilized on neutravidin wells and (B) is the absorbance in the respective G4 oligonucleotide-free control wells. The threshold for considering antigen specificity was absorbance values higher than 2.5.

8. Use of bioinformatic software

In this section, only general bioinformatic software is included. Specific bioinformatic software used for the analysis of results obtained in BLISS is described in its respective section.

8.1. Cancer Cell Line Encyclopedia database

Cancer Cell Line Encyclopedia (192) (https://sites. broadinstitute. org/ccle/), that contains open access sequencing data for thousands of cancer cell line samples, was used to analyze the mutational status of several genes in tumoral cell lines.

8.2. Oncomine platform

Gene expression data from Oncomine platform (193) (https://www.oncomine.com/oncomine-platform-software) was subjected to different bioinformatic analysis. In particular, "TCGA Colorectal" dataset with non-tumoral (N = 22) and tumoral (N = 101) samples was used with the following filters: 1) "Cancer Type: Colorectal Cancer"; 2) "Gene: POLR1A/GLUT1/CMYC"; 3) "Data Type: mRNA"; 4) "Analysis Type: Cancer vs Normal Analysis", and 5) "Threshold Setting Condition (P<0.001, fold change >2, gene rank = top 10%)".

8.3. Primer-BLAST

Primer-BLAST (194) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to design PCR primers that are specific to intended targets. It is possible thanks to the employment of a global alignment algorithm to screen primers against user-selected database in order to

avoid primer pairs that can cause non-specific amplifications. The settings were conveniently modified for each individual search.

8.4. QGRS Mapper

QGRS Mapper program (195) (https://bioinformatics.ramapo.edu/ QGRS/index.php) was used to analyze the distribution of putative G4 forming sequences in given nucleotide sequences. For this search, the following settings were employed: "Maximum length = 30 bp", "Minimum G-group" = 3 bp", "Loop size = from 0 to 36 bp".

9. Statistical analysis

Statistical significance was assessed using Student's two-tailed t-test and two-way ANOVA with Graphpad (Prism). Both tests serve to compare the variance of two (t-test) or three or more (ANOVA) populations to determine if these populations are statistically different from each other. Independent samples were analyzed and normal distribution was assumed. Values represent mean ± standard deviation. For all tests, p-values below 0.05 were considered significant and expressed as follows: *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$.

RESULTS

<u>OBJECTIVE 1</u>: Study of DNA G4s in the progression of colorectal cancer and analysis of their implication in tumor development and DNA damage

Given the relationship between G4 structures and cancer disease, it is an important goal to unveil the role of G4s in colorectal carcinogenesis. In this section, we investigated the change in G4 levels in a cellular model of CRC progression, comparing not only between non-tumoral and tumoral stage, but also in metastasis. Moreover, we aimed to address the involvement of G4s in the abnormal regulation or function of CRC-relevant genes, as well as in genome instability.

Characterization of a cellular model of CRC progression and arrest at different cell cycle phases

To analyze G4s involvement in CRC carcinogenesis, we first established a cellular model to mimic CRC progression (Figure 10A). We used three different human cell lines: CRL1790 (colon epithelial cells) as non-tumoral stage, SW480 (Dukes' type B colorectal adenocarcinoma cells) representing the primary tumor stage, and SW620 (Dukes' type C colorectal adenocarcinoma derived from metastatic site cells) as the aggressive metastatic stage. To note, SW620 was established from a metastatic lymph node belonging to the same patient from whom the SW480 cell line was previously derived. According to the Cancer Cell Line Encyclopedia (CCLE) database, both tumoral cell lines were characterized by the typical pattern of genetic aberrations defining CRC (*APC*, *KRAS*, and *TP53* mutations) (Figure 10A).

Since G4 formation is modulated during cell cycle (19), we aimed to evaluate G4s abundance in cells arrested at G0/G1 and S phases. In this regard, we first performed cell cycle analysis after serum deprivation or 2.5 mM thymidine addition for 24 h and 48 h to induce cell cycle arrest. Serum starvation is widely used for synchronizing cells in G0/G1 phase due to serum contains growth factors indispensable for cyclins-mediated cell cycle transition (196). Instead, thymidine is a DNA synthesis inhibitor used

for S phase synchronization (196). Our results further validated such methodologies, showing an increase in cells arrested at G0/G1 after serum deprivation, whilst an increase in S phase-cells upon thymidine incubation, for all CRL1790, SW480, and SW620 cell lines (Figure 10B). Interestingly, tumoral cell lines were more sensitive to cell cycle arrest compared to non-tumoral CRL1790, whose cell population changes were less pronounced. In addition, cell cycle detention was progressive along time with a maximum at 48 h. For further experiments, CRL1790, SW480, and SW620 cells were arrested at G0/G1 or S phase upon serum starvation or 2.5 mM thymidine treatment respectively for 48 h.
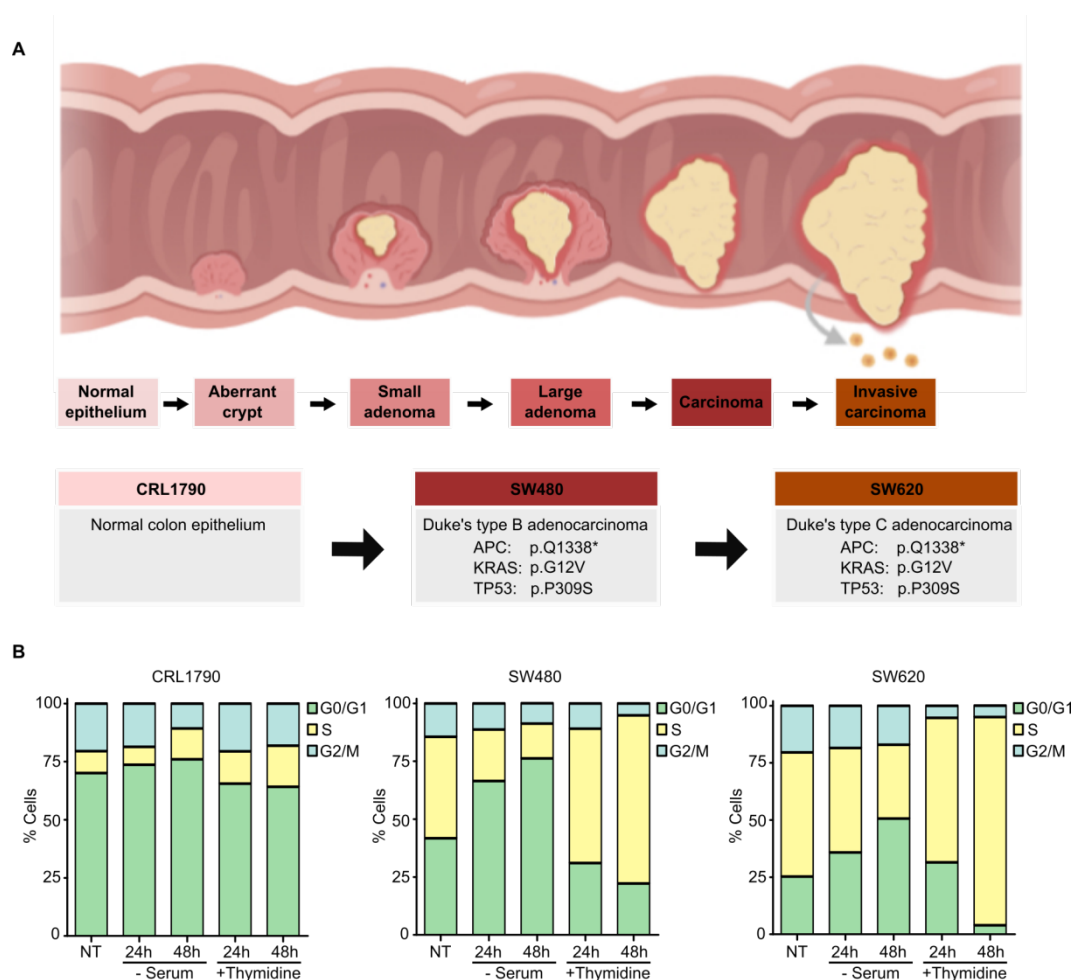
Figure 10. Characterization of a cellular model of CRC progression and arrest at different cell cycle phases. (A) CRC progression model. Cell lines used for this study, CRL1790, SW480, and SW620, mimic different stages in colorectal carcinogenesis and possess the typical pattern of genetic aberrations according to Cancer Cell Line Encyclopedia. (B) Stacked bar graph illustrating cell cycle distribution of CRL1790, SW480, and SW620 cells after propidium iodide flow cytometry analysis. The cells were

non-treated (NT) or incubated with serum-free medium or 2.5 mM thymidine for 24 h and 48 h to induce an arrest at G0/G1 or S phase respectively.

Expression and purification of BG4 antibody

In order to visualize DNA G4 structures in human cells we used the scFv antibody BG4 that had been previously reported (19). For that, BG4 antibody production was performed using the expression vector pSANG10-3F-BG4 as described before (25). Numerous colonies grew in kanamycin-containing plates, showing that BL21 (DE3) *E. coli* were successfully transformed with pSANG10-3F-BG4 (Figure 11A). No colonies appeared in the kanamycin plates containing non-transformed BL21 (DE3) *E. coli*, confirming that there was no contamination. Next, BG4 antibody (~33 kDa) was properly expressed and purified using immobilized metal affinity chromatography, as determined by western blot with anti-6x-histidine epitope tag antibody (Figure 11B). A total amount of 230 µg of BG4 was finally obtained. Small aliquots were stored at -20 ⁰C until use.
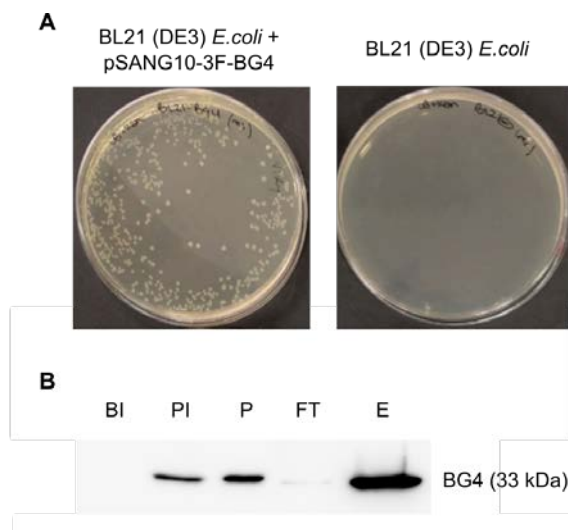
Figure 11. Expression and purification of BG4 antibody. (A) Colonies of BL21 (DE3) *E. coli* in kanamycin-containing plates after transformation with pSANG10-3F-BG4 or without transformation. (B) Western Blot analysis with anti-histidine to monitor expression and purification of BG4 antibody. Nomenclature of different samples is the following: BI refers to before induction of expression; PI to post induction of expression; P to periplasmic space prior purification; FT to discarded flow-through in chromatography; E to eluted BG4 after chromatography.

G4s abundance increases along CRC progression

We analyzed the levels of G4 structures in the cellular model of CRC progression both at G0/G1 and S phase. For this purpose, cells were incubated with serum-free medium or 2.5 mM thymidine for 48 h to induce the arrest at G0/G1 or S phase respectively, and subsequently subjected to BG4 immunofluorescence in presence of RNase (to avoid detection of RNA G4s). All cell lines showed punctated nuclear BG4 staining that indicates the presence of folded DNA G4s both at G0/G1 and S phases (Figure 12A). Nuclear BG4 signal was further quantified, revealing that G4s abundance significantly increased along CRC progression, regardless of the cell cycle phase (Figure 12B). In particular, CRL1790 non-tumoral epithelial cells showed the lowest G4 levels, which significantly increased in SW480 primary tumor cells, and were maximal in SW620 metastatic cells. In fact, G4s abundance was significantly higher in SW620 versus SW480, suggesting the implication of G4s in the metastatic process as well. As expected, BG4 staining at G0/G1 phase, when cellular processes are in a quiescent state, was lower than that at S phase, when DNA replication occurs. Therefore, G4s stabilization, which is maximal at S phase, increases along CRC progression and may be involved not only in carcinogenesis, but also in metastasis.
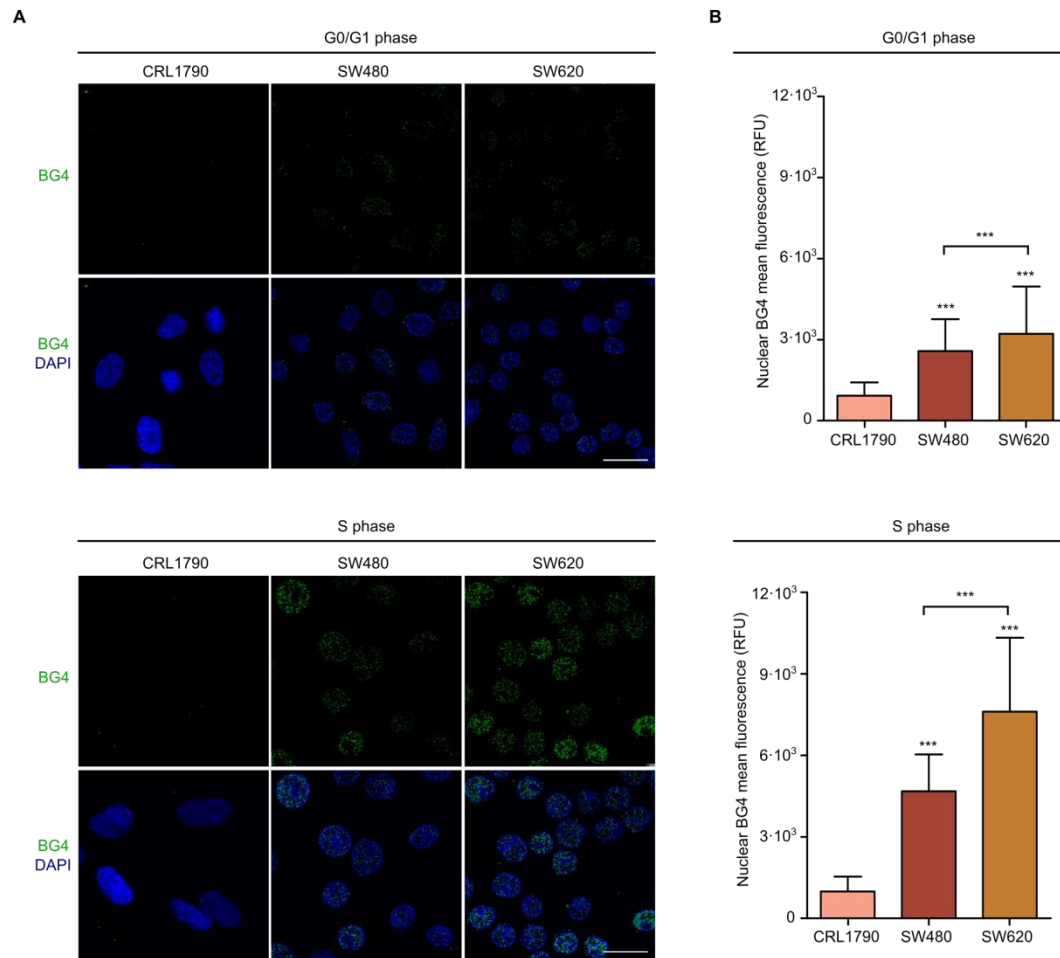
Figure 12. G4s abundance increases along CRC progression. (A) BG4 immunofluorescence images (green signal) of CRL1790, SW480, and SW620 cells arrested at G0/G1 or S phase. Nuclei are coloured in blue by counterstaining with DAPI. Scale bars, 20 μm. (B) Nuclear BG4 mean fluorescence level quantification from cells in (A) by Fiji analysis (N > 200). For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

G4 helicases are overexpressed along CRC progression

Within the cells, stable G4 structures can be enzymatically unfolded due to the action of some well-known DNA helicases which maintain genome stability during DNA replication and transcription (37). Based on the results obtained in the previous section, we hypothesized that the increase in genomic G4s along CRC progression could arise from the abnormal functioning of enzymes that process G4s. However, according to the Cancer Cell Line Encyclopedia database, no mutations were

previously reported on a panel of G4 helicases on SW480 or SW620 (data not shown). In addition, we measured the transcriptional levels by qRT-PCR of the panel of helicases involved in G4s-unfolding in CRL1790, SW480, and SW620 cells arrested at G0/G1 or S phase. Several G4 helicases, including *BLM*, *DDX5*, *DDX11*, *DHX9*, *DNA2*, *FANCJ*, *PIF1*, *TIMELESS*, and *WRN*, were significantly overexpressed in tumoral SW480 and SW620 cells compared to non-tumoral CRL1790 at G0/G1 phase (Figure 13A). The same occurred at S phase, when the differences in expression of helicases were even more marked (Figure 13B). The most remarkable changes were observed in *BLM*, *DDX11*, and *FANCJ* at both G0/G1 and S phases. Thereupon, helicases that resolve G4s are overexpressed along CRC progression, and thus, the increase of G4 structures in CRC is not explained by the downexpression of G4 helicases.
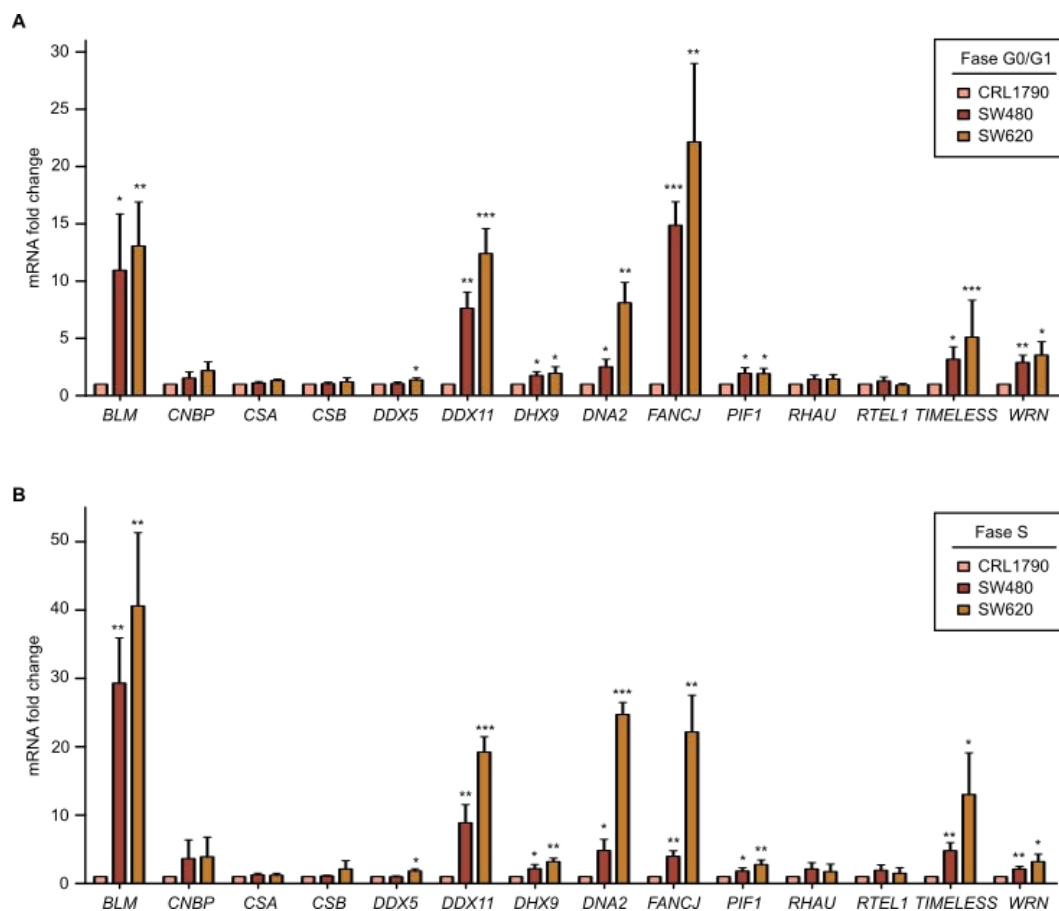
Figure 13. G4 helicases are overexpressed along CRC progression. (A) Transcriptional levels of a panel of G4 helicases obtained by qRT-PCR with CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. (B) The same experiment as in panel

(A) was performed with CRL1790, SW480, and SW620 cells arrested at S phase. Both experiments were conducted in biological triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

CMYC G4 unfolds along CRC progression

Several oncogenes that play key roles in CRC, including *CMYC* and *KRAS,* harbor G4s in their promoters or close to them, controlling their expression level (ref de mi review G4s). Based on this evidence, we investigated in more detail the status of G4s in *CMYC* and *KRAS* promoters to decipher their involvement in colorectal carcinogenesis. First, we analyzed whether these oncogenes were transcriptionally dysregulated measuring the transcriptional levels of *CMYC* and *KRAS* in the cellular model of CRC progression arrested at G0/G1. As expected, both tumoral SW480 and SW620 cells displayed a statistically significant and abnormal higher expression of *CMYC* and *KRAS* compared to non-tumoral CRL1790 (Figure 14A).

Then, we analyzed the mutational pattern of these G4s in the cellular model of CRC progression through DNA extraction, PCR amplification with specific primers and subsequent Sanger sequencing. The consensus sequence of G4s affecting *CMYC* and *KRAS* transcription was obtained from previous studies. In particular, one important G4 is located upstream of the P1 promoter of *CMYC* (197). In contrast, *KRAS* promoter contains three potential G4s that are referred to as "far", "mid", and "near", on the basis of their proximity to the transcription start site (198). No mutations that would disrupt the G4 structure were found in the G4s of *CMYC* and *KRAS* in any of the cell lines (Figure 14B).

In addition, we examined the folded/unfolded status of these G4s by BG4 chromatin immunoprecipitation and qPCR using primers for *CMYC* and *KRAS* G4s. For that, we arrested CRL1790, SW480, and SW620 at G0/G1 phase, and normalized the ChIP signals (with BG4 antibody) to mock control signals (without BG4 antibody) for each cell line to evaluate

the specific binding to target regions expressed as fold enrichment. As result, a high fold enrichment means that the G4 is specifically immunoprecipitated with BG4 antibody, and thus, the G4 is folded. Interestingly, ChIP followed by qPCR only showed a significant enrichment of *CMYC* G4 respect to the mock control in CRL1790, indicating that *CMYC* G4 was significantly folded in non-tumoral CRL1790 cells, but it was unfolded along CRC progression (Figure 14C). In contrast, *KRAS* G4 was not enriched in ChIP assays versus mock controls in any of the cell lines, showing that no significant differences were detected in the folding of *KRAS* G4. Altogether these results suggest that *CMYC* G4 is folded at the non-tumoral stage and it unfolds as CRC progresses, but such change is not driven by mutations in the G4 sequence that would disrupt the structure.
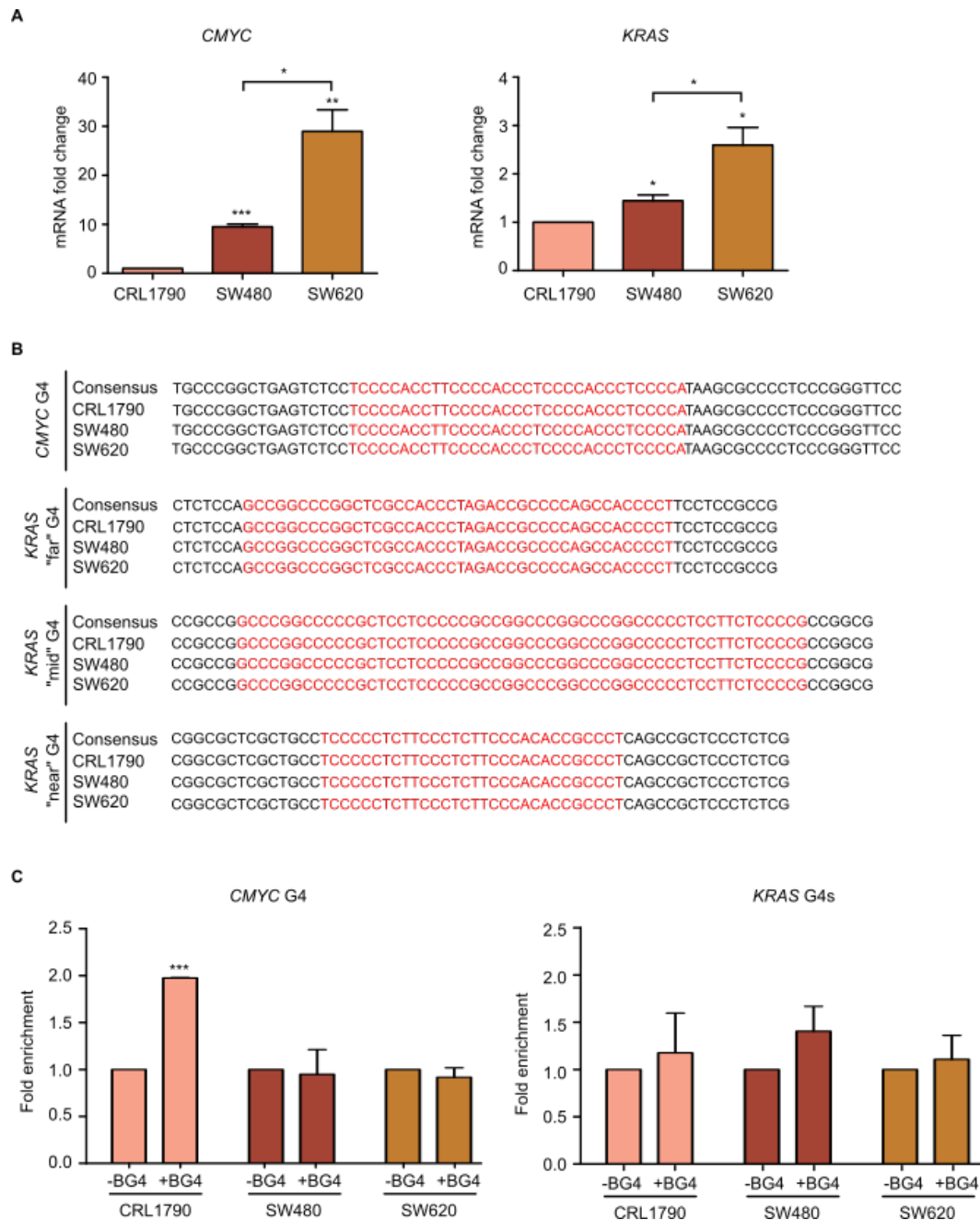
Figure 14. *CMYC* G4 unfolds along CRC progression. (A) *CMYC* and *KRAS* mRNA expression levels determined by qRT-PCR in CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. Experiments were conducted in biological triplicate. (B) Schematic representation of sequence alignment of G4s in *CMYC* and *KRAS* among the previously published consensus sequence and the respective sequences from the cellular model of CRC progression. G4s sequence is shown in red. (C) Fold enrichment corresponding to the BG4 ChIP-qPCR assay performed on CRL1790, SW480, and SW620 cells arrested at G0/G1 phase using primers that span the G4s of *CMYC* and *KRAS.* Normalization over a mock sample (without BG4) was performed for each cell line. Experiments were conducted in biological triplicate. For all tests, p-values below 0.05

were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

DNA damage increases along CRC progression

We anticipated that DNA damage accumulation would be higher along CRC progression, since genome instability is a hallmark of many cancers (83), and microsatellite instability testing is even used in CRC with diagnostic purposes (199). To investigate this hypothesis, we measured a marker of DNA damage response, such as phosphorylation of histone H2AX on Ser-139 (termed γH2AX), in the cellular model of CRC progression for both G0/G1 and S cell populations by western blot. In the same trend as G4s formation, γH2AX DNA damage response was remarkably induced throughout CRC progression at G0/G1 and S phase (Figure 15A). Consistent with these observations, γH2AX quantification yielded a statistically significant increase in SW480 and SW620 tumoral cells compared to non-tumoral CRL1790 arrested at G0/G1 or S phase (Figure 15B). In addition, we noticed that DNA damage response was higher in S phase, when DNA is being replicated. Within tumoral cells, DNA damage was significantly higher in metastatic SW620 cells versus SW480. Altogether, these results suggest that DNA damage is robustly induced during CRC malignant progression, rendering maximal levels at S phase.
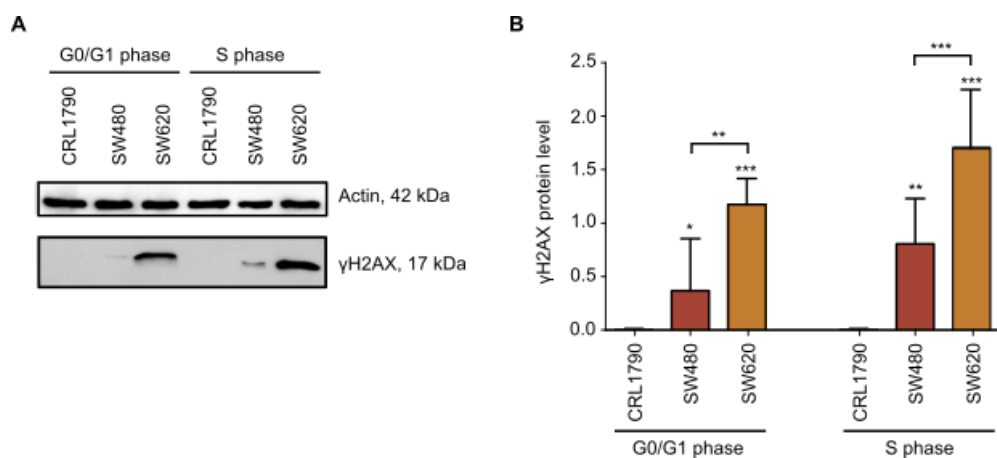
Figure 15 DNA damage increases along CRC progression (A) Western blot experiments in CRL1790, SW480, and SW620 cells arrested at G0/G1 or S phase to

determine protein levels of γH2AX as a marker of DNA damage and actin as housekeeping. Experiments were performed in biological triplicate and representative images are shown. (B) Quantification of γH2AX protein levels normalized to actin from cells in (A) by ImageJ. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

DNA damage is spread throughout the genome

To note, γH2AX does not label double-strand breaks (DSBs) directly, and γH2AX signal spreads tens of kilobases away from a single DSB as reported elsewhere (200). In an attempt to directly map DSBs at single-nucleotide resolution, we performed breaks labeling in situ and sequencing (BLISS) methodology (177) on CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. Consistent with γH2AX measurements in the previous section, estimation of total DSBs increased along CRC progression (Figure 16A). In BLISS methodology, multiple DSBs that map to the same location in different cells are labeled by distinct unique molecular identifiers (UMIs) (177). We further analyzed the BLISS results to test whether DSBs were repetitive at the same location. Most of DSBs were unique (only represented with 1 UMI) suggesting that DSBs did not accumulate at recurrent genomic locations in multiple cells (Figure 16B).

We continued the analysis focused on DSBs that reoccur at the same genomic location with ≥4 UMIs. Interestingly, the estimation of recurrent DSBs at the same genomic location minimally overlapped among different cell lines (Figure 16C). Thus, we hypothesized that recurrent DSBs that map at different locations across different CRC stages could be affecting key genes in cancer. To evaluate this hypothesis, we performed an individual analysis of recurrent DSBs. We identified that centromeres were the most susceptible regions for DNA damage accumulation, and this susceptibility remarkably increased along CRC progression, being maximal in metastatic SW620 cells (Figure 16C). In contrast, DSBs mapping at gene loci decreased along CRC progression (Figure 16C). Based on literature, we identified recurrent DSBs in some cancer-related genes for both tumoral SW480, and SW620 cell lines

(Figure 16C). In addition, the number of DSBs across the whole gene body of these cancer-related genes increased along CRC progression (Figure 16D). Altogether these results suggest that DSBs are widely spread throughout the genome, and recurrent DSBs seem to accumulate in cancer-related genes in tumoral SW480 and SW620 rather than in non-tumoral CRL1790 cells.
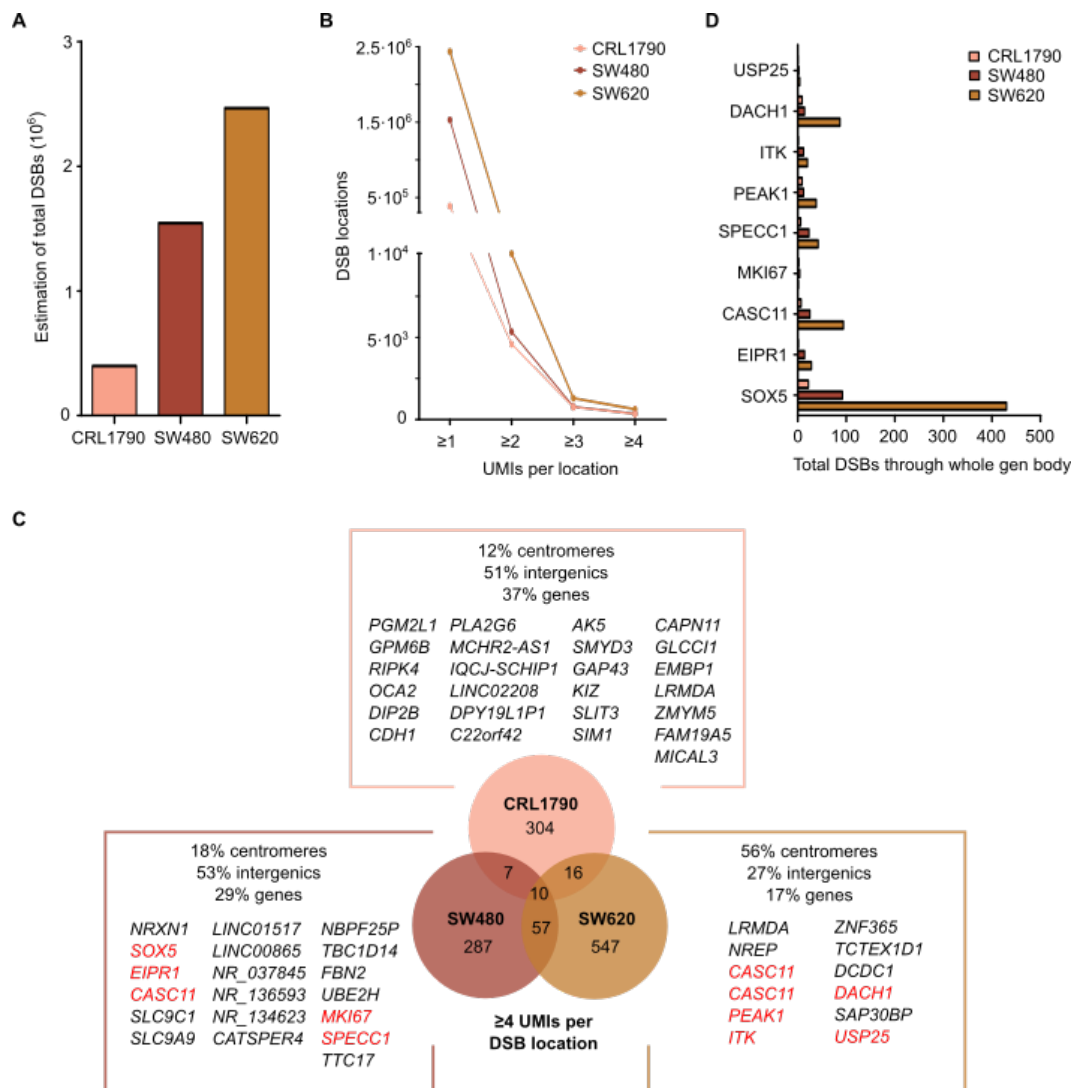
Figure 16. DNA damage is spread throughout the genome. (A) Total estimation of double-strand breaks (DSBs) by BLISS methodology in CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. (B) Number of DSB locations by filtering on the minimum number of UMIs per DSB in CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. (C) Venn diagram representing the overlap of recurrent DSBs with ≥4 UMIs in the cellular model of CRC progression arrested at G0/G1 phase. Percentage of DSBs with ≥4 UMIs that accumulate in different genomic regions (including centromeres, genes and intergenic loci) are shown. In particular, genes enriched in recurrent DSBs with ≥4 UMIs

are listed, and cancer-related genes are in red. (D) Estimation of total DSBs through the whole gene body in cancer-related genes identified in (C) for CRL1790, SW480, and SW620 cells arrested at G0/G1 phase.

DNA damage is associated with G4s presence

Accumulation of DNA damage is a common consequence of G4s stabilization because G4s can interfere with the progression of DNA replication forks (55). In this context, we investigated the relationship between G4s and DNA damage throughout CRC progression. To assess, we overlaid G4s data previously used and accepted as reference (15) with DSBs obtained by BLISS method on CRL1790, SW480, and SW620 cells arrested at G0/G1 phase. In particular, we focused on DSBs with ≥2 UMIs (reoccurring more than twice). Interestingly, DSBs were strongly enriched in the neighborhood of the G4s in all cell lines (Figure 17A). We repeated the analysis using 25 bp or 250 bp windows tiling the genome. With these windows, DSBs that overlapped with at least one other ±25 bp or ±250 bp were merged. As result, we further confirmed that DSBs preferentially occurred in the proximity of G4 sites independently of the 25 bp (Figure 17B) or 250 bp (Figure 17C) window considered. When the window included more distance, the estimation of total DSBs with ≥2 UMIs increased as expected, and the differences in total DSBs with ≥2 UMIs among cell lines were more pronounced. Thereupon, DSBs are accumulated in the vicinity of G4 sequences, suggesting the association between genome instability and G4s presence.
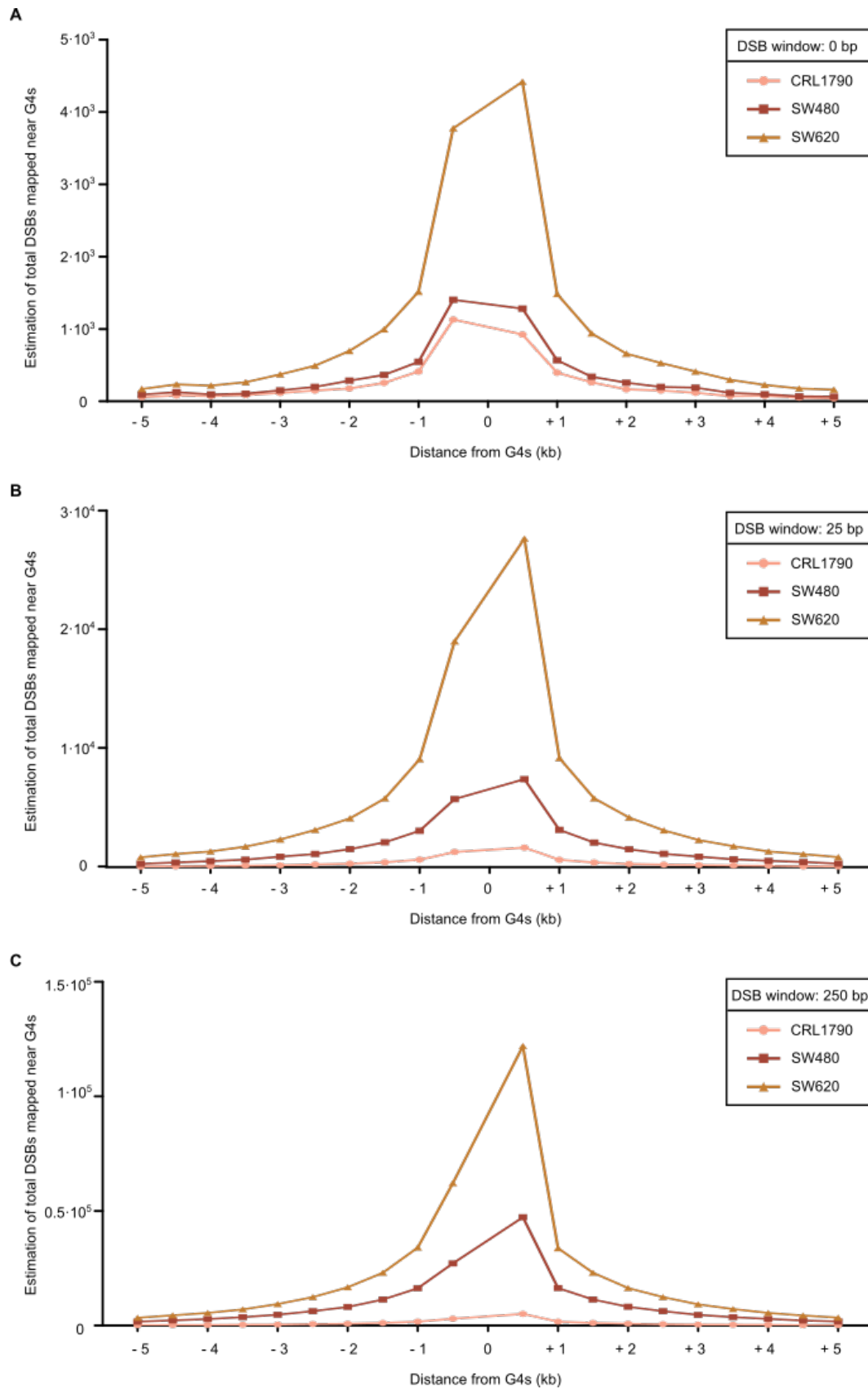
Figure 17. DNA damage is associated with G4s presence. (A) Total estimation of double-strand breaks (DSBs) with ≥2 UMIs in CRL1790, SW480, and SW620 cells arrested at G0/G1 phase obtained by BLISS that map around G4 sites. (B) The same

analysis as in (A), but using DSB windows of 25 bp to merge overlapping DSBs in this distance. (C) The same analysis as in (A), but using DSB windows of 250 bp to merge overlapping DSBs in this distance.

G4 ligands stabilize G4s and induce DNA damage and cell death along CRC progression

Cell cultures with increased G4 levels are supposed to be more sensitive to small molecules targeting G4s according to a previous work (65). In order to evaluate their therapeutic potential in CRC, we analyzed the cytotoxic potency of several established, yet structurally distinct, G4 ligands including BMH21 (201), CX3543 (104), CX5461 (202), and pyridostatin (19). For that, we determined the half-maximal inhibitory concentration (IC_{50}) values in the cellular model of CRC progression upon treatment for 48 h with increasing concentrations from $1 \cdot 10^{-5}$ μM to 50 μM. All G4 ligands induced cell cytotoxicity, suggesting that G4s could act as therapeutic targets in CRC (Table 1). However, G4 ligands lacked selectivity, and tumoral cells were slightly more resistant to treatment with G4 ligands than non-tumoral cells.

G4 ligand	Cell line	IC_{50} (μM)
BMH21	CRL1790	22.9 ± 0.3
	SW480	> 50
	SW620	49.9 ± 1.2
CX3543	CRL1790	37.0 ± 0.8
	SW480	> 50
	SW620	38.3 ± 1.1
CX5461	CRL1790	30.2 ± 0.7
	SW480	> 50
	SW620	43.8 ± 1.4
Pyridostatin	CRL1790	32.6 ± 0.3
	SW480	> 50
	SW620	42.0 ± 0.8

Table 1. IC_{50} values for established G4 ligands in the cellular model of CRC progression. IC_{50} values represent G4 ligand concentration inhibiting cell growth by 50% and are expressed as mean ± standard deviation. Experiments were performed in biological triplicate.

Among the tested G4 ligands, we selected BMH21, displaying the lowest IC_{50} value, for further studies. We determined G4 levels by BG4 immunofluorescence and DNA damage response by γH2AX western blot after treatment with 1 µM BMH21 for 3 h. We observed a robust increase in BG4 signal indistinctly in all cell lines upon BMH21 treatment, which indicates that BMH21 stabilized G4s in a non-selective manner (Figure 18A). In agreement, BMH21 induced the DNA damage response in all cell lines without selectivity, as measured by γH2AX levels (Figure 18B). Therefore, treatment with G4 ligands leads to an increase in G4s, DNA damage and cell death in the cellular model of CRC progression, but these effects are non-selective for tumoral cells.
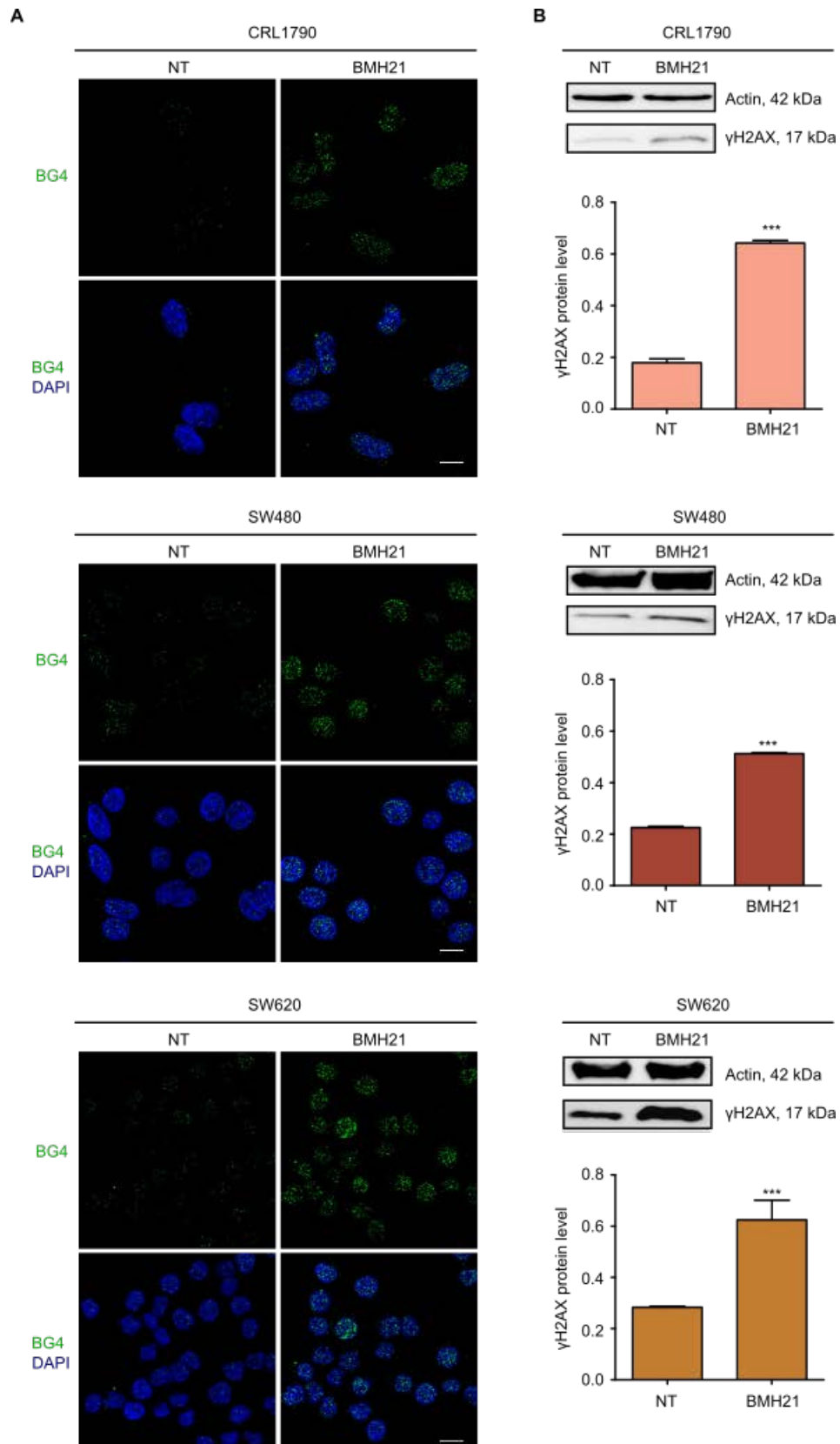
Figure 18. G4 ligands stabilize G4s and induce DNA damage and cell death along CRC progression. (A) BG4 immunofluorescence images of CRL1790, SW480, and SW620 cells non-treated (NT) or treated with BMH21 1 µM for 3 h. Nuclei are coloured in

blue by counterstaining with DAPI. Representative images are shown. Scale bars, 10 µm. (B) Western blot experiments in CRL1790, SW480, and SW620 cells non-treated (NT) or treated with BMH21 1 µM for 3 h to determine protein levels of γH2AX as a marker of DNA damage response and actin as housekeeping gene. Experiments were performed in biological triplicate and representative images are shown. Protein levels of γH2AX protein levels were quantified and normalized to actin. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

OBJECTIVE 2: Screening of chemical compounds to select ligands that target DNA G4s and display antitumoral activity

G4 structures are promising targets for design of drugs in cancer therapy. Nevertheless, when we analyzed the targeting and cytotoxicity of four typical G4 ligands for CRC in the previous section, we detected that their selectivity towards tumoral cells is limited. In this context, we aimed to explore the therapeutic potential of other structurally distinct compounds that interact with G4s in CRC, focusing on selective G4 ligands to avoid undesirable side effects. To that end, we investigated both synthetic and natural compounds.

OBJECTIVE 2.A: Screening of synthetic compounds

Identification of two naphthalene diimides (T1 and T5) with selective antitumor activity in the cellular model of CRC

Within the plethora of existing G4 ligands, we focused our attention on naphthalene diimides (NDIs) given their potential for chemical variability. In particular, we screened several carbohydrate-conjugated NDIs, which were previously synthesized in a reported work, in an attempt for a targeted antitumoral therapy (174, 175). For that, we used the cellular model with three different human cell lines including CRL1790, SW480, and SW620 to mimic colorectal carcinogenesis (Figure 10A). We determined the IC_{50} values upon treatment for 48 h with increasing concentrations from $2 \cdot 10^{-5}$ µM to 100 µM. Selection criteria for anticancer agents included cytotoxic potency and differential activity against tumoral versus normal cell lines (203). Among seven NDI derivatives tested, only two met the selection criteria (Table 2). Aglycone-NDI (T1) and β-Lact-C-di-NDI (T5) inhibited SW480 and SW620 cell growth with half-maximal inhibitory concentration (IC_{50}) values lower than 30 µM, and were much less cytotoxic to CRL1790, showing selectivity indexes greater than 2.0 (both compounds were over twice more cytotoxic to the tumor cells as compared with the normal cells). Interestingly, T5 displayed a higher selectivity for tumoral cells.

Treatment	Cell line	IC$_{50}$ (µM)	Selectivity Index
(T1) Aglycone-NDI 25	CRL1790	22.38 ± 2.64	-
	SW480	6.81 ± 0.64	3.29
	SW620	5.36 ± 0.61	4.18
(T2) α-Man-C2-di-NDI	CRL1790	> 100	-
	SW480	> 100	1
	SW620	37.15 ± 5.45	2.69
(T3) α-Man-C-di-NDI	CRL1790	23.93 ± 3.49	-
	SW480	20.18 ± 0.74	1.19
	SW620	11.64 ± 1.25	2.06
(T4) β-Lact-C2-tri-NDI	CRL1790	> 100	-
	SW480	> 100	-
	SW620	> 100	-
(T5) β-Lact-C-di-NDI	CRL1790	50.46 ± 2.94	-
	SW480	5.62 ± 0.57	8.98
	SW620	7.94 ± 0.72	6.36
(T6) β-Malt-C-tri-NDI	CRL1790	> 100	-
	SW480	> 100	-
	SW620	> 100	-
(T7) β-Man-TEG-di-NDI	CRL1790	> 100	-
	SW480	70.79 ± 3.25	1.41
	SW620	53.33 ± 4.81	1.87

Table 2. IC$_{50}$ values for established G4 ligands in the cellular model of CRC progression. IC$_{50}$ values represent NDI concentration inhibiting cell growth by 50% and are expressed as mean ± standard deviation. Selectivity index is the ratio of IC$_{50}$ values in non-tumoral and cancer cell lines. Experiments were performed in biological triplicate. The nomenclature given for each NDI derivative corresponds to that used in a previously reported study (175).

Cytotoxic activity of both T1 and T5, regardless of *TP53* mutational status, was confirmed in other CRC cell lines, such as HCT116 (*TP53* wild

type) and HT29 (*TP53* pP309S; C -> T) (Figure 19A). To note, T1 is an aglycone-NDI, while T5 is an NDI conjugated with a carbohydrate, lactose in particular (Figure 19B). Next, we extended the findings by performing cell-cycle assays. Both NDIs triggered an S phase arrest and a G2/M increase (Figure 19C). Propidium iodide cell-cycle analysis (Figure 19D) indicated that these effects occurred to a lower extent with T5. Based on these results, both T1 and T5 NDIs affect cell cycle and were considered leading antitumor candidates for subsequent analysis.
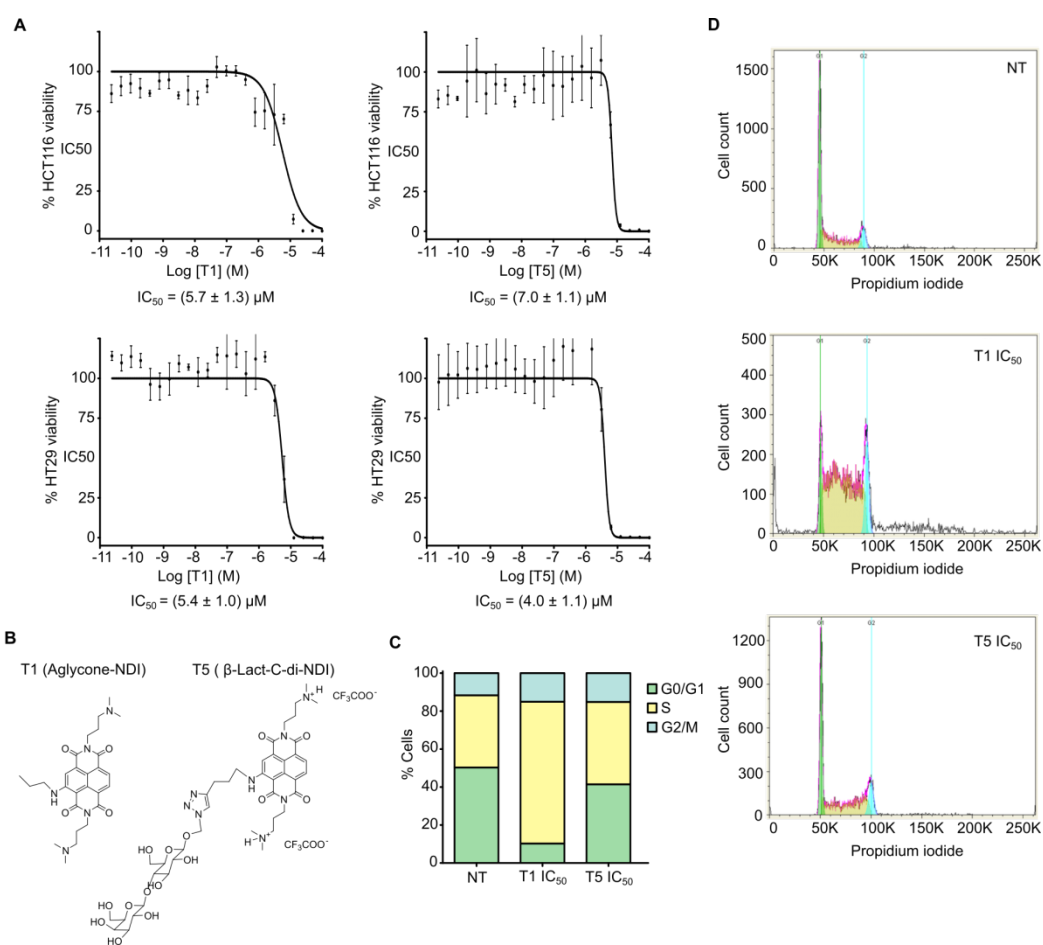
Figure 19. Identification of two naphthalene diimides (T1 and T5) with selective antitumor activity. (A) IC$_{50}$ determination of T1 and T5 in HCT116 and HT29 cell lines. Values are expressed as mean ± standard deviation. Experiments were conducted in biological triplicate. (B). Chemical structure of T1 (aglycone-NDI) and T5 (β-Lact-C-di-NDI) selected as antitumoral candidates. (C and D) Propidium iodide flow cytometry analysis of SW480 cells treated with DMSO (non-treated, NT) or treated with T1 IC$_{50}$ or T5 IC$_{50}$ for 24 h. (C) Stacked bar graph illustrating cell-cycle distribution. (D) Representative histograms are shown.

T1 and T5 cause nucleolus disintegration, loss of Pol I catalytic subunit A, and autophagy

T1 and T5 derivatives belong to the NDI family, known to be G4 ligands. However, NDI binding to G4s in ribosomal DNA (rDNA) has not been previously studied. To explore this hypothesis, we performed an immunofluorescence of SW480 cells treated with T1/T5 IC_{50} for 24 h to determine the intracellular localization of nucleolar proteins indicative of nucleolus status (204). Markedly, both derivatives caused segregation of nucleolar structures, including translocation of the granular component protein nucleolin (NCL) to the nucleoplasm, segregation of fibrillarin (FBL) to the nucleolar periphery caps, and loss of Pol I catalytic subunit A (POLR1A) (Figure 20A). Therefore, T1 and T5 cause a remarkable nucleolar stress.

To confirm these alterations in POLR1A protein levels, we performed a western blot of SW480 cell lysates treated with T1/T5 IC_{50} for 3, 6, and 24 h. T1 and T5 markedly induced a POLR1A clearance (Figure 20B). Quantitatively, POLR1A was downregulated around 3, 10, and 40 times in cells treated for 3, 6, and 24 h respectively (Figure 20C). We then checked whether such POLR1A degradation at 3 h was impaired after treatment with the proteasome inhibitor MG132 (10 µM, 6 h). We proved that inhibition of ubiquitination by MG132 rescued POLR1A degradation caused by T1 and T5 (Figure 20D). However, POLR1A clearance observed at the protein level was not associated with a decrease in POLR1A mRNA levels after treatment with T1/T5 IC_{50} for 3 h (Figure 20E). Next, to explore the relationship between nucleolar stress and autophagy (205), we treated SW480 cells with T1/T5 IC_{50} for 24 h, and assessed LC3 protein levels as an indicator of autophagy. Although T1 exerted a more potent effect, both NDIs led to the conversion of LC3-I to LC3-II, thus inducing autophagy (Figure 20F). We quantitatively confirmed that LC3-II production upon T1/T5 treatment for 24 h was statistically significant (Figure 20G). Altogether, these results show that T1 and T5 induce proteasome-dependent POLR1A degradation, leading to cell death by autophagy.
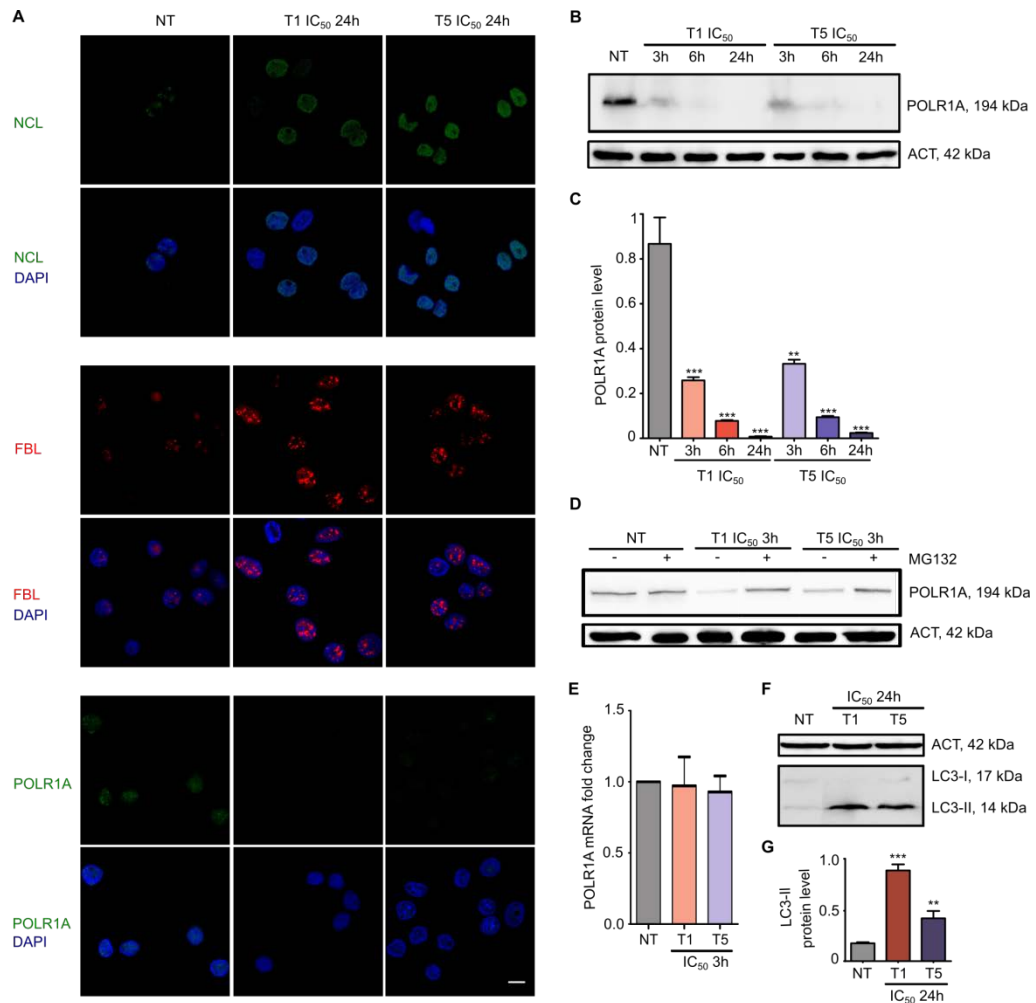
Figure 20. T1 and T5 cause nucleolus disintegration, loss of Pol I catalytic subunit A, and autophagy. (A) SW480 cells were treated with DMSO (NT) or treated with T1 or T5 IC_{50} for 24 h and stained for nucleolin (NCL), fibrillarin (FBL), and POLR1A by immunofluorescence. Merged image with DAPI for DNA staining is shown below. Representative images are shown. Scale bar, 10 µm. (B) Western blot analysis performed in SW480 cells after treatment with T1 or T5 IC_{50} for 3, 6, or 24 h, or treated with DMSO (NT) to determine protein levels of POLR1A and actin as housekeeping gene. Experiments were conducted in biological triplicate. (C) Quantification of POLR1A protein levels normalized to actin data in (B). (D) POLR1A and actin immunoblotting analysis of SW480 cells preincubated with the proteasome inhibitor MG132 (10 µM) for 6 h and subsequently treated with DMSO (NT) or treated with T1 or T5 IC_{50} for 3 h. Experiments were conducted in biological triplicate. (E) qRT-PCR analysis of SW480 cells non-treated (NT) or treated with T1 or T5 IC_{50} for 3 h to determine POLR1A mRNA levels. Experiments were performed in biological triplicate. (F) Western blot analysis of SW480 cells treated with T1 or T5 IC_{50} for 24 h or DMSO (NT) to determine actin and LC3 (both LC3-I and LC3-II) protein levels as an autophagy marker. Experiments were conducted in biological triplicate. (G) Quantification of LC3-II protein levels normalized to actin of data

in (F). For all tests, p-values below 0.05 were considered significant and expressed as follows: *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$.

T1 and T5 inhibit rRNA synthesis and cause POLR1A disassembly from rDNA prior to POLR1A degradation

Since disruption of the nucleolar structure is a cellular hallmark of rRNA transcription impairment (204), we aimed to determine whether T1 and T5 affected cellular transcription by Pol I using qRT-PCR and considering the short-lived 5' external transcribed spacer (*5'ETS*) of the pre-RNA (Figure 21A), whose abundance is generally reflective of the rRNA synthesis rate (206). We observed a drastic and significant decrease of *5'ETS* transcripts (Figure 21B) in SW480 cells treated with T1/T5 IC$_{50}$ for 3 h. To further assess the extent to which rRNA synthesis was inhibited by T1 and T5, we analyzed Pol I transcription in SW480 cells using both compounds at doses below their IC$_{50}$ (0.5, 1, and 2.5 μM) for 3 h. Concentrations of 0.5 μM for T1 and 1 μM for T5 showed a 10- and 5-fold inhibition of rRNA synthesis, respectively (Figure 21C). In addition, we excluded that neither T1 0.5 μM nor T5 1 μM for 3 h had inhibitory effect on Pol II-driven transcription of G4-enriched genes through quantification of *BCL2*, *CMYB*, *CMYC*, and *KRAS* expression levels by qRT-PCR (Figure 21D). Therefore, a prominent inhibition of Pol I transcription is observed after NDI treatment for 3 h with an inhibitory concentration of 0.5 μM for T1 and 1 μM for T5.

Next, to analyze the kinetics of the cellular response to T1 and T5, we treated SW480 cells with T1/T5 at inhibitory concentrations for 0.5, 1 and 3 h, and measured rRNA synthesis. A significant decrease in *5'ETS* transcript was observed after treatment, with a minimum inhibition at 30 min and a maximum inhibition at 3h (Figure 21E). As rDNA transcription occurs via modulation of the assembly efficiency of Pol I transcription complex subunits on rDNA (207), we tested whether T1 0.5 μM or T5 1 μM for 3 h could be altering POLR1A engagement with rDNA by POLR1A-chromatin immunoprecipitation and qPCR using primers for different

regions of the rDNA gene body. As a consequence of T1 and T5 treatment, POLR1A association with rDNA was significantly affected throughout the rDNA gene (Figure 21F). Altogether, these results are consistent with a rapid kinetics regarding the inhibition of rRNA synthesis and a POLR1A disengagement from rDNA, posing the ribosomal locus as the main target of T1 and T5.

Our next aim was to assess the induction of POLR1A degradation at the inhibitory concentration (lower than IC$_{50}$). For this purpose, protein extracts of SW480 cells treated with T1 0.5 µM and T5 1 µM for 3 h were subjected to western blot analysis. Surprisingly, POLR1A protein was not degraded at these doses (Figure 21G). Thereupon, within 3 h after T1 0.5 µM and T5 1 µM treatment, the inhibition kinetics of rRNA synthesis and POLR1A disengagement from rDNA is faster than POLR1A clearance, which could be considered a downstream effect.
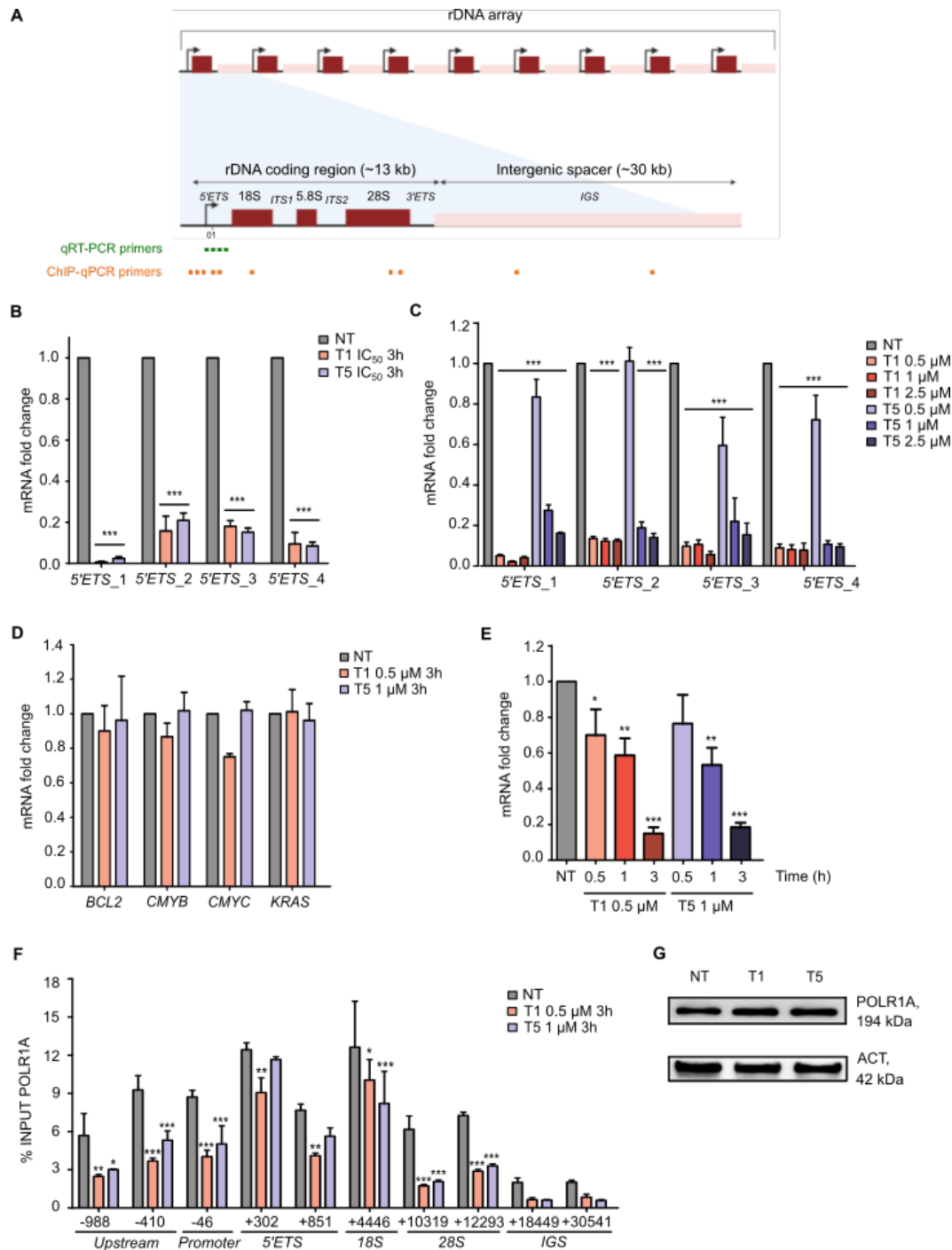
Figure 21. T1 and T5 inhibit rRNA synthesis and cause POLR1A disassembly from rDNA prior to POLR1A degradation. (A) Diagram of human tandem arrangement of rDNA repeated units, each containing an rRNA coding region (red) and an intergenic spacer (pink). Location of qRT-PCR and chromatin immunoprecipitation (ChIP)-qPCR primers used for this study is marked in green and orange, respectively. (B) SW480 cells were treated with T1 or T5 IC$_{50}$ for 3 h and rDNA transcription was analyzed by qRT-PCR using four primer sets for short-lived *5'ETS* rRNA. (C) rDNA transcription analysis (as in B) in SW480 cells treated with T1 or T5 at lower doses than IC$_{50}$ (0.5, 1, and 2.5 μM) for 3 h. (D) qRT-PCR analysis of SW480 cells non-treated (NT) or treated with T1 0.5 μM or T5

1 µM for 3 h to determine mRNA levels of several Pol II-transcribed genes. (E) Kinetic analysis of rDNA transcription by qRT-PCR using *5'ETS*_1 primers in SW480 cells treated with T1 0.5 µM or T5 1 µM for 0.5, 1, and 3 h. (F) ChIP-qPCR analyses of POLR1A binding to rDNA in SW480 cells treated with DMSO (NT) or treated with T1 0.5 µM or T5 1 µM for 3 h. Primer locations and associated rDNA regions are shown on the x axis. (G) POLR1A and actin (housekeeping) protein abundance determined by western blot in SW480 cells treated either with T1 0.5 µM or T5 1 µM for 3 h. All experiments were performed in biological triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$.

T5 possesses a tumor-selective effect depending partly on GLUT1 overexpression

We sought to identify if the inhibition of rRNA synthesis was maintained across the cellular model of CRC. CRL1790, SW480, and SW620 cells treated with T1 0.5 µM, T5 1 µM, or CX5461 1 µM (as a positive control for Pol I inhibition) for 3 h were subjected to Pol I transcription analysis (Figure 22A). As a result, T1 significantly inhibited rRNA synthesis in all cell lines to the same extent. Conversely, T5 showed a significant and selective effect for tumoral cell lines SW480 and SW620, while rRNA levels remained unchanged in CRL1790 non-tumoral cells. CX5461 was confirmed to act as a non-selective RNA Pol I inhibitor in all cell lines. Next, we investigated stabilization properties of G4s for T1 and T5 in a cellular environment. It was achieved by immunofluorescence with the G4-selective antibody BG4 in CRL1790, SW480, and SW620 cells after incubation with T1 0.5 µM or T5 1 µM for 3 h. T1 induced a notorious increase of nuclear BG4 signal in all cell lines, while this effect was only observed for tumoral SW480 and SW620 cells upon T5 treatment (Figure 22B). In accordance, BG4 mean fluorescence quantification (N > 100) rendered a significant increase after T1 and T5 treatment, indicating that these compounds strongly stabilized G4 structures, except for T5 in CRL1790 cells (where significant differences were not observed) (Figure 22C). Therefore, T5 offers the best therapeutic window with a predominant effect on tumoral cells.

NDIs are fluorescent molecules whose excitation and emission maxima take place at 595 and 661 nm, respectively (174). To further investigate the differences in tumoral selectivity between both compounds, we analyzed T1 and T5 uptake in the cellular model of CRC by fluorescence confocal microscopy, and quantified the uptake profiles by measuring the nuclear fluorescence. No differences were found in cellular uptake between non-tumoral and tumoral cells for T1, while T5 showed a 4-fold greater entrance into tumoral cells (SW480) in relation to non-tumoral ones (Figure 22D). As mentioned before, T1 is an aglycone-NDI while T5 is a lactose-conjugated NDI, and such chemical differences could affect their uptake rates. Then, we hypothesized that T5 could be preferably translocated into cancer cells through glucose transporters (GLUTs). Since GLUT1 is frequently upregulated during oncogenesis (208), we quantified the entrance of both NDI derivatives in SW480 cells in the presence of two different GLUT1 inhibitors: BAY876 (209) and WZB117 (210). T1 uptake was not affected after exposure to GLUT1 inhibitors. Notwithstanding, GLUT1 inhibition caused a significant decrease in T5 cellular uptake (Figure 22E). Therefore, T5 is preferably taken up by tumoral cells, at least in part, through GLUT1. However, additional uptake mechanisms associated with cancer cells must be involved. Due to its higher tumor selectivity, we further continued our study with T5.
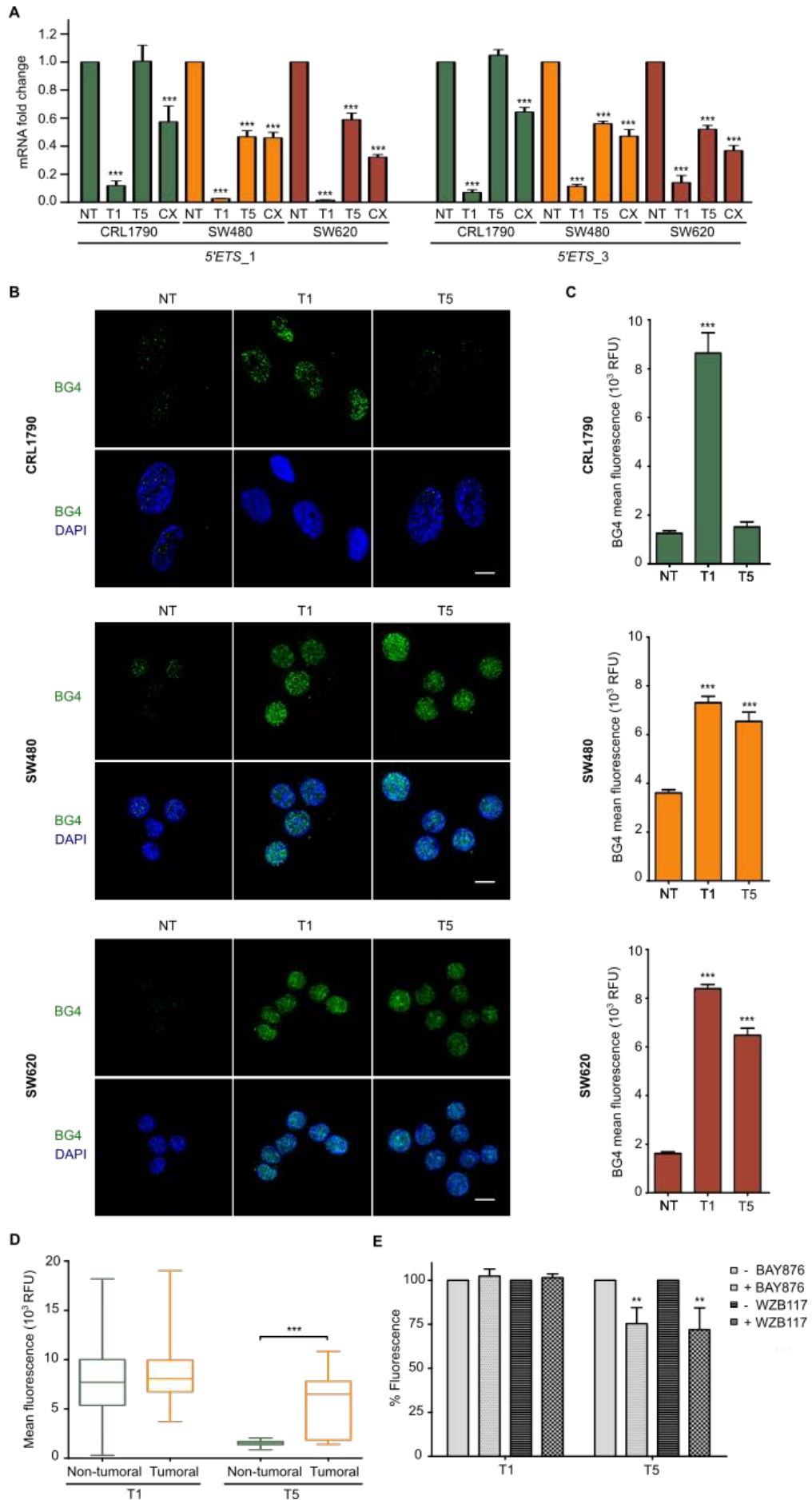
Figure 22. T5 possesses a tumor-selective effect depending partly on GLUT1 overexpression. (A) *5'ETS* rRNA fold change across different cell lines in the cellular model of CRC determined in triplicate by qRT-PCR using *5'ETS*_1 and *5'ETS*_3 primers after T1 0.5 µM, T5 1 µM, or CX5461 10 µM treatment for 3 h or non-treated (NT). Experiments were conducted in biological triplicate. (B) BG4 immunofluorescence images of CRL1790, SW480, and SW620 cells non-treated (NT) or treated with T1 0.5 µM or T5 1 µM for 3 h. Nuclei are colored blue by counterstaining with DAPI. Representative images are shown. Scale bars, 10 µm. (C) Nuclear BG4 fluorescence level quantification from cells in (B) by Fiji analysis (N > 100). (D) NDI nuclear uptake quantification from cells treated with T1 or T5 5 µM by fluorescence confocal microscopy and subsequent Fiji analysis (N > 50). (E) T1 and T5 uptake quantification in SW480 cells in the absence or presence of BAY876 and WZB117 as GLUT1 inhibitors by fluorescence spectroscopy. Experiments were conducted in biological triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

T5 exerts its effect through binding to specific G4s in rDNA

Using the software QGRS mapper, we found 12 putative G4-forming sequences in the human ribosomal *5'ETS* DNA region. Inhibition of the rDNA transcription by T5 prompted us to examine whether T5 interacted with these G4-forming sequences by TOPRO3 fluorescent intercalator displacement (FID) assay. Interestingly, T5 5 µM significantly exhibited a preferential binding to G4_F5 and G4_R4, decreasing the fluorescence percentage down to 34.5% and 70.6%, respectively (Figure 23A). Moreover, we performed FID titration assays using T5 from 0.02 to 100 µM with the selected G4_F5 and G4_R4 to quantify the concentration required to displace TOPRO3 from the DNA matrix by 50% (DC_{50} values) (211). To note, DC_{50} values were at micromolar range and T5 DC_{50} value for G4_F5 was significantly lower than for G4_R4 (Figure 23B). Therefore, we continued our study with G4_F5. Guanine imino signals between 10.5 and 12 ppm in the NMR spectrum clearly confirmed the formation of a G4 structure in G4_F5 (Figure 23C). Upon T5 addition, NMR signals changed due to interaction between T5 and the G4_F5. The general signal broadening suggested the formation of higher-order structures resulting from ligand-induced association of G4s. This behavior is common in many

G4 ligands. To further assess the mode of action, we evaluated the ability of T5-stabilized G4_F5 sequence to stall a DNA polymerase by an in vitro DNA polymerase extension assay. As negative control, we used a mutant sequence incapable of G4 formation (MUT_F5). T5 demonstrated a selective and dose-dependent inhibition of PCR amplification for the wild type G4_F5 (Figure 23D). In contrast, T5 had no effect on the DNA polymerase stalling in the mutant sequence. Overall, these results confirm that T5 exerts the inhibition of rRNA synthesis by a high-affinity binding to specific G4s in rDNA.

To test effects of T5 on Pol I transcription in vitro, a DNA template that fused the yeast rDNA promoter to the human rDNA sequence carrying the G4_F5 sequence or a mutated G4_F5 sequence were synthesized. These templates were included in fully reconstituted in vitro transcription assays using purified yeast components in multiple rounds (184, 185). Promoter-dependent transcription of the linear templates yielded 756 nt runoff RNA products. RNA accumulation was significantly inhibited by addition of T5 in reactions, including the G4_F5 template (Figures 23E and 23F). Reactions using the mutant form of the G4 sequence were less affected by T5. Therefore, T5 directly inhibits transcription by RNA Pol I in vitro.

Next, we investigated in a cellular environment the stabilization properties of T5 in rDNA-associated G4s by a competition experiment in the presence of thioflavin T (ThT), a fluorescent light-up probe of ribosomal G4s in the nucleolus (186). SW480 cells non-treated or treated with T5 1 µM for 3 h were subjected to ThT fluorescence confocal microscopy. In non-treated cells, ThT foci mainly accumulated in the nucleoli. In contrast, the ThT-stained foci sharply decreased upon treatment with T5, indicating that T5 displaced ThT from the nucleolar G4s (Figure 23G). These results further sustain that T5 exhibits a G4-binding pattern inside cells, and mainly targets G4s in the nucleolar rDNA at physiological conditions.
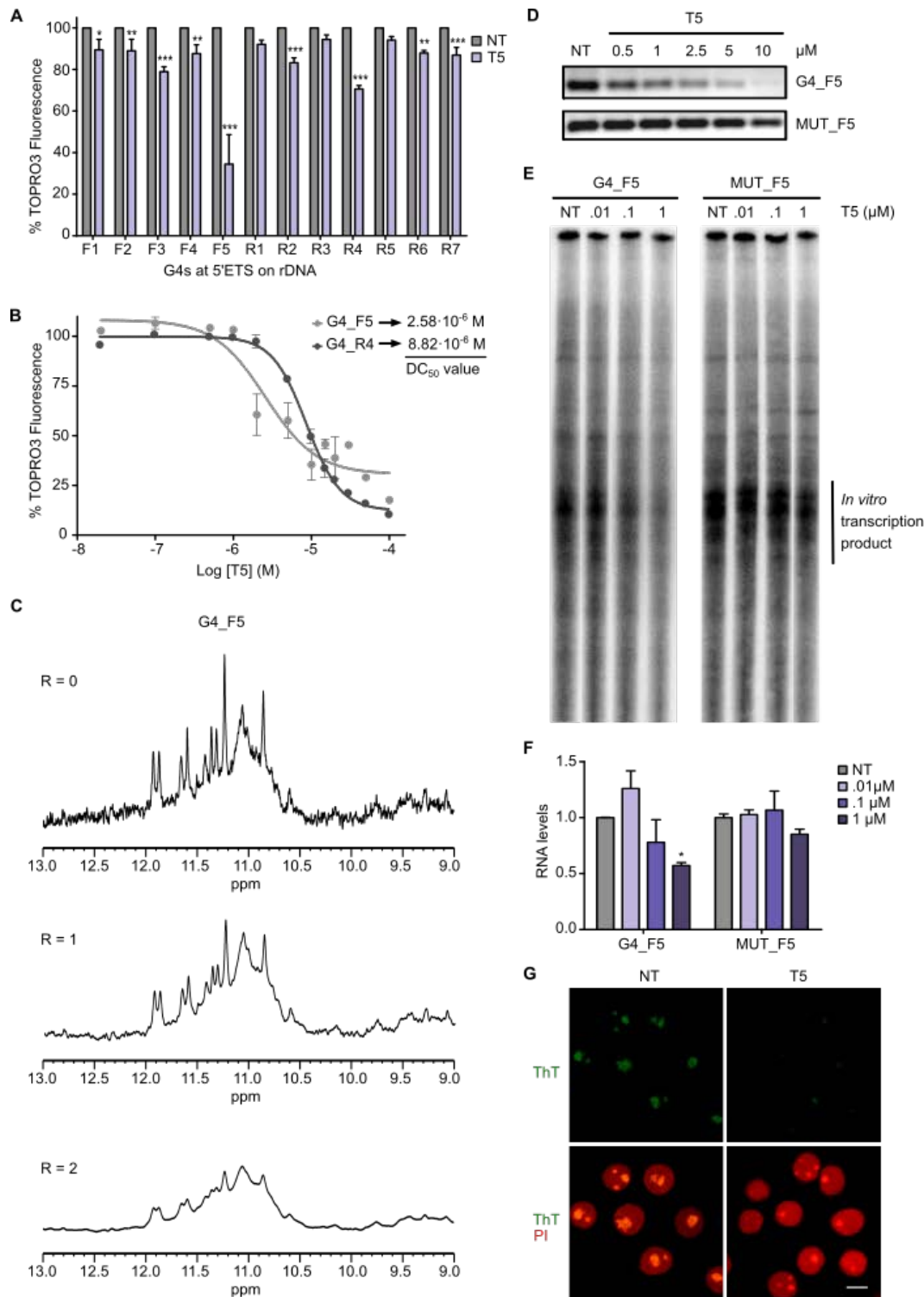
Figure 23. T5 exerts its effect through binding to specific G4s in rDNA. (A) TOPRO3 FID assay using oligonucleotides from putative G4 sequences found in *5'ETS* of rDNA to determine the fluorescence percentage in the absence (NT) or presence of T5 5 µM. Experiments were performed in triplicate. (B) FID titration assay with increasing concentrations of T5 to determine DC_{50} values in the selected putative G4s. Experiments were conducted in triplicate. (C) Exchangeable proton region of the NMR spectra of G4_F5 with T5 at different R = [T5]/[DNA] ratios. (D) PCR stop assay to determine the

effect of T5 on the stabilization of the G4-forming candidate G4_F5 and mutant MUT_F5 with increasing T5 concentrations. Three independent reactions were conducted per concentration and representative lanes are displayed. (E) Pol I in vitro transcription elongation assay of the yeast/human rDNA fusion template carrying the G4_F5 sequence and a mutated MUT_F5 sequence. Each reaction was performed at least four independent times and representative lanes are shown for display purposes. (F) Quantification of RNA levels from in vitro transcription experiments in (E). (G) Fluorescence confocal imaging of SW480 cells pre-stained with thioflavin T non-treated (NT) or treated with T5 1 µM for 3 h. Nuclei are colored in red by counterstaining with propidium iodide (PI). Representative images are shown. Scale bar, 10 µm. For all tests, p-values below 0.05 were considered significant and expressed as follows: *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$.

T5 could be explored as a therapeutic agent for patients with CRC

Since POLR1A activity rate is proportional to cell proliferation (212), we postulated that *POLR1A* expression was increased along colorectal carcinogenesis to meet the increasing demands for protein synthesis, and thus, it could be considered an attainable antitumor target in CRC. In particular, we determined *POLR1A* expression level in the cellular model of CRC by qRT-PCR. The Ct is the threshold cycle of detection and ΔCt values show relative gene expression using actin as housekeeping gene. As expected, ΔCt *POLR1A* decreased in CRC meaning that *POLR1A* expression was higher in both tumoral cells, SW480 and SW620, versus CRL1790 normal cells (Figure 24A). We also appreciated that T1 and T5 IC_{50} values were correlated to *POLR1A* expression with a Pearson R value of 0.87 (p = 0.0254). Overall, these results confirm that specific POLR1A inhibition by T1 and T5 represents a remarkable therapeutic opportunity in CRC. To further understand the selective mode of action of T5 in CRC, we aimed to analyze *GLUT1* expression levels in the cellular model of CRC. Both SW480 and SW620 cancer cells showed a significantly higher *GLUT1* expression (lower ΔCt *GLUT1*) in comparison with CRL1790 normal cells (Figure 24A). These results suggest that T5 is a more promising leading compound since its cellular uptake and inhibition

of rRNA synthesis is partially restricted to tumoral cells owing to *GLUT1* overexpression.

Finally, we aimed to assess the translational potential of T5 in a patient cohort with CRC. To explore POLR1A inhibitors as a therapeutic strategy in CRC, *POLR1A* qRT-PCR analysis was performed in 15 CRC tumoral samples and 7 colorectal biopsies derived from non-tumoral adjacent tissue in patients with CRC. *POLR1A* mRNA expression was significantly upregulated (lower ΔCt *POLR1A*) in CRC compared with non-tumoral tissues (Figure 24B). Furthermore, since *GLUT1* overexpression would be exploitable for a targeted T5 treatment, we also analyzed *GLUT1* expression in patients with CRC by qRT-PCR. *GLUT1* expression level was significantly increased (lower ΔCt *GLUT1*) in tumoral compared with non-tumoral tissues (Figure 24B). *POLR1A* and *GLUT1* expression patterns observed in the patient cohort with CRC were further validated by a bioinformatic analysis from the Oncomine database (Figure 24C). These findings reveal that POLR1A constitutes a clinically attainable target in CRC, while *GLUT1* overexpression could be beneficial for a translational and targeted T5 treatment, evading undesirable side effects.
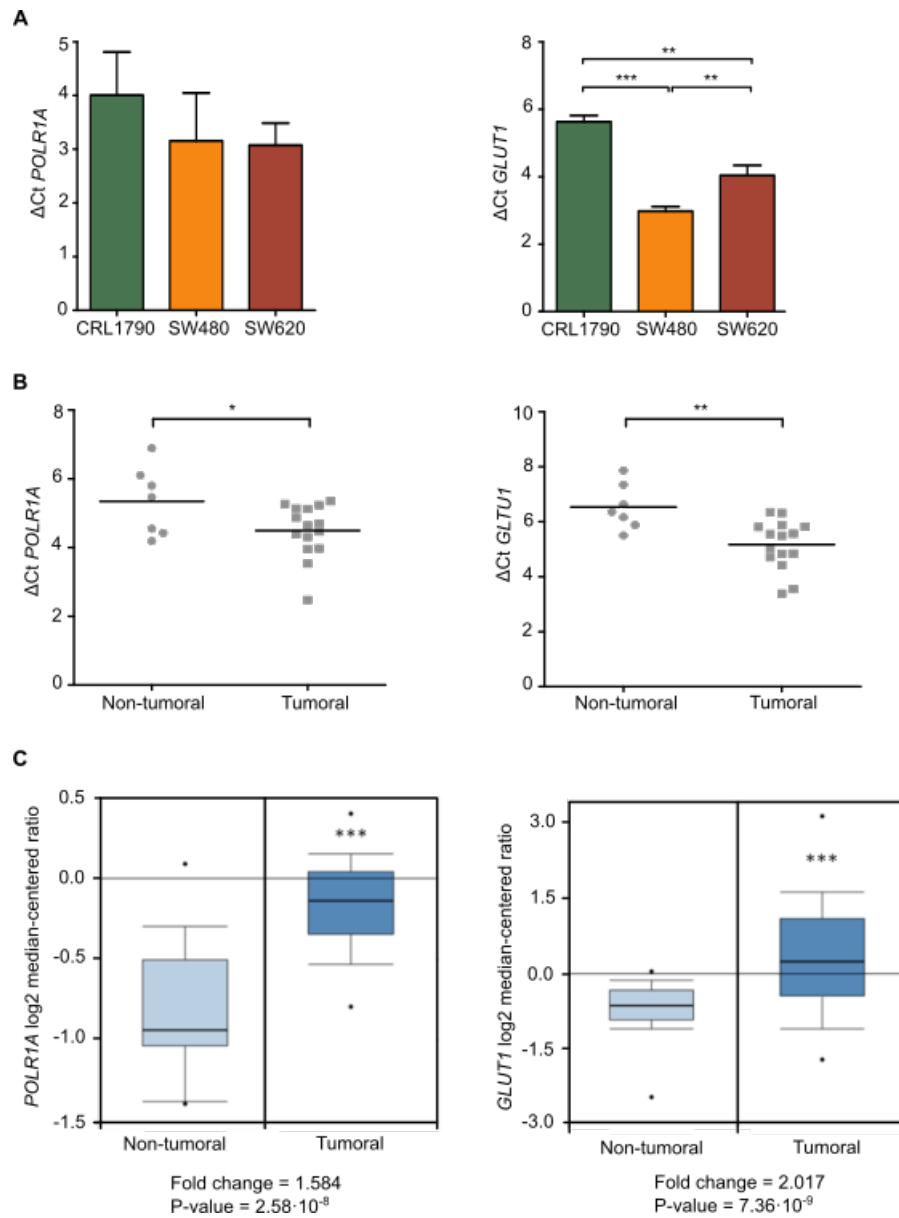
Figure 24. T5 could be explored as a therapeutic agent for patients with CRC. (A) ΔCt results for *POLR1A* and *GLUT1* expression in the cellular model of CRC obtained by qRT-PCR. (B) ΔCt results for *POLR1A* and *GLUT1* expression in the cohort of patients with CRC, obtained by qRT-PCR. (C) Analysis of *POLR1A* and *GLUT1* expression levels in Oncomine database with "TCGA Colorectal" dataset including non-tumoral (N = 22) and tumoral (N = 101) samples. All qRT-PCR experiments were conducted in triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

OBJECTIVE 2.B: **Screening of natural compounds**

Gallic acid shows anticancer activity in vitro

Natural products, such as phenolic compounds, have attracted attention for their anticancer properties. In order to identify potential drugs for CRC, we determined cytotoxic activity of five natural phenols including resveratrol (RSV), piceid (PIC), tyrosol (TYR), hydroxytyrosol (HTYR), and gallic acid (GA). We used three human cell lines to mimic CRC progression in the cellular model (Figure 10A). Cytotoxic potency was analyzed by determination of IC_{50} values upon 48 h treatment with phenolic compounds at increasing concentrations from $2 \cdot 10^{-5}$ µM to 100 µM (Table 3). Only GA inhibited cell growth in SW480 and SW620 at clinically-relevant concentrations (IC_{50} values lower than 30 µM) (203). Furthermore, GA exhibited a higher selectivity for cancer cells with minimal affectation of non-tumoral CRL1790 cells (IC_{50} values higher than 100 µM). Based on these results, GA was selected for further studies.

Treatment	Cell line	IC_{50} (µM)	Selectivity Index
(RSV) Resveratrol	CRL1790	> 100	-
	SW480	> 100	-
	SW620	> 100	-
(PIC) Piceid	CRL1790	> 100	-
	SW480	> 100	-
	SW620	> 100	-
(TYR) Tyrosol	CRL1790	> 100	-
	SW480	> 100	-
	SW620	> 100	-
(HTYR) Hydroxytyrosol	CRL1790	> 100	-
	SW480	> 100	-
	SW620	71.94 ± 3.52	> 1.39
(GA) Gallic acid	CRL1790	> 100	-
	SW480	22.39 ± 2.12	> 4.47
	SW620	11.83 ± 1.54	> 8.45

Table 3. IC$_{50}$ values for natural phenolic compounds in the cellular model of CRC progression. IC$_{50}$ values represent phenol concentration inhibiting cell growth by 50% and are expressed as mean ± standard deviation. Selectivity index is the ratio of IC$_{50}$ values in non-tumoral and cancer cell lines. Experiments were performed in biological triplicate.

Gallic acid induces cell cycle arrest and nucleolus disintegration

GA (3,4,5-trihydroxybenzoic acid) is a naturally occurring triphenolic compound with low molecular weight (Figure 25A). GA is widely present in the plant kingdom and largely found in different food sources (213). We started the study of the antitumoral effect of GA by analyzing its effect on the cell cycle in SW480 cells by flow cytometry with propidium iodide staining (Figure 25B). After GA IC$_{50}$ treatment during 24 h, frequency of cells at different stages of cell cycle changed respect to non-treated cells (45.34 % of cells at G1, 34.16 % at S, 7.65 % at G2/M). GA induced an increase at S and G2/M phases (28.30 % of cells at G1, 43.82 % at S, 14.46 % at G2/M). These results suggested that GA might alter DNA replication triggering S and G2/M phase cell cycle arrest.

In the previous section, we had described that NDIs are DNA-binding compounds that affect the cell cycle in a similar way to GA, increasing the S and G2/M phases. However, GA binding to DNA had not been previously reported. To explore this hypothesis, we determined the intracellular localization of nucleolar proteins by immunofluorescence of SW480 cells treated with GA IC$_{50}$ for 6 h, as we had observed in the previous section that NDIs affect nucleolar organization. We analyzed nucleolin, fibrillarin, and Pol I catalytic subunit A, all of them indicative of nucleolus status (Figure 25C). GA altered localization of nucleolar structures, including translocation of nucleolin to nucleoplasm and segregation of fibrillarin to nucleolar periphery caps. However, significant changes in POLR1A were not observed. Altogether these results imply that GA causes a remarkable cell cycle arrest and nucleolar stress.
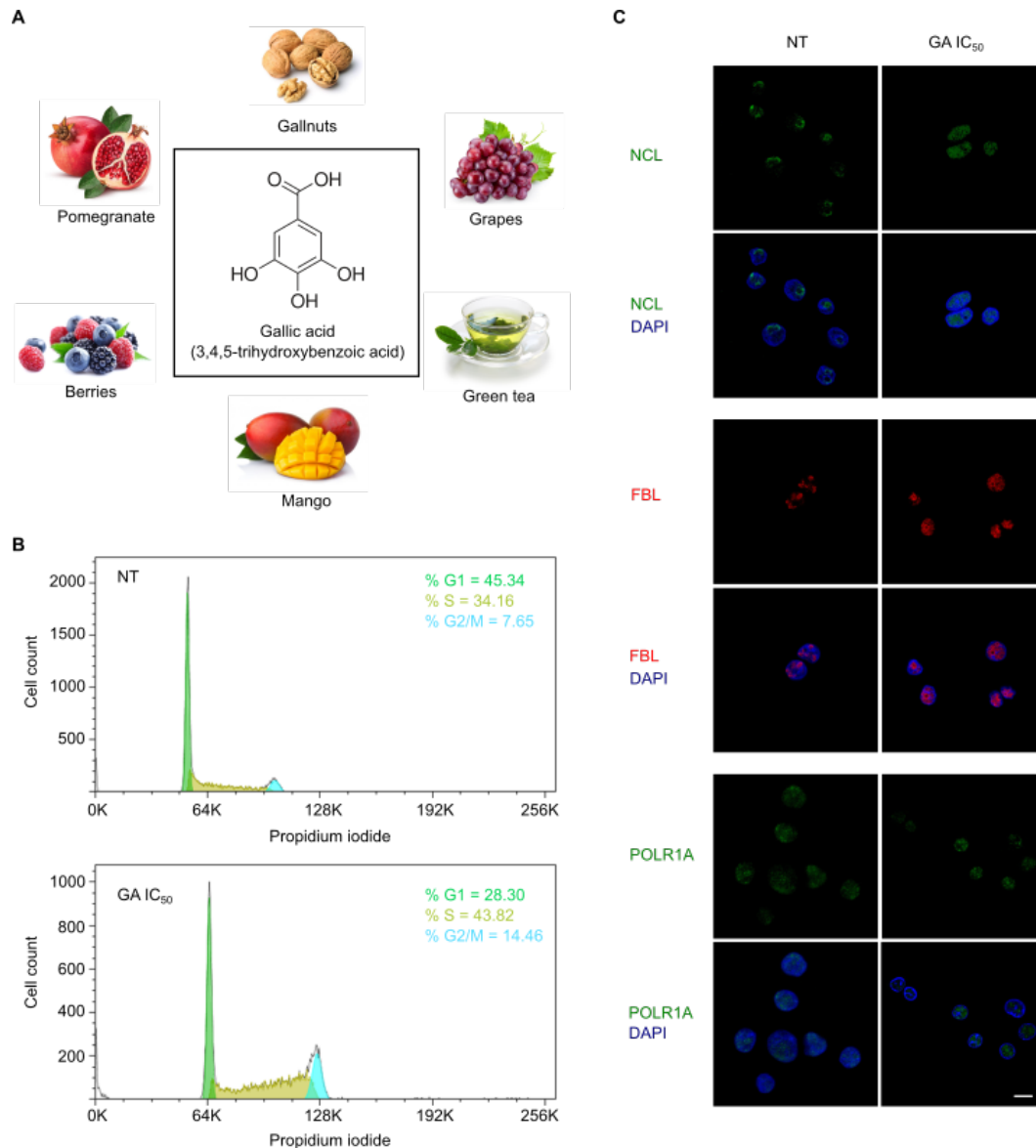
Figure 25. Gallic acid induces cell cycle arrest and nucleolus disintegration. (A) Chemical structure of GA selected as antitumoral candidate in this study and natural sources where it can be found. (B) Histograms from propidium iodide flow cytometry analysis of SW480 cells treated with DMSO (non-treated, NT) or treated with GA IC$_{50}$ for 24 h. (C) Immunofluorescence images of SW480 cells treated with vehicle DMSO (NT) or treated with GA IC$_{50}$ for 6 h, and stained for nucleolin (NCL), fibrillarin (FBL) or POLR1A. Merged images with DAPI for DNA counterstaining are also displayed. Representative images are shown. Scale bar, 10 μm.

Gallic acid stabilizes G4s, inhibiting the transcription of several oncogenes, and induces DNA damage

As mentioned in the previous section, nucleolus disintegration is a hallmark of ribosomal RNA transcription blockage by some compounds targeting DNA. Specifically, these compounds bind to G4s and are shown to modulate transcription (89). In particular, transcription of several oncogenes (including *BCL2*, *CMYB*, *CMYC*, *KRAS*, and *VEGFA*) is thought to be controlled by stabilization of G4s, and ribosomal DNA gene also harbors G4 sequences which impair ribosomal RNA synthesis (84). In this context, we analyzed whether GA affects the transcription of G4-enriched genes by qRT-PCR. In the case of ribosomal DNA, we measured the short-lived 5' external transcribed spacer (*5'ETS*) of the pre-RNA, whose abundance reflects ribosomal RNA synthesis rate (206). Treatment of SW480 cells with GA IC_{50} for 6 h resulted in a significant downregulation of several genes which contain G4s (Figure 26A). The well-known G4 ligand, CX5461, also influenced the expression levels of some of these genes.

Next, we investigated G4-stabilization properties of GA in a cellular environment by immunofluorescence with the G4 selective antibody BG4 (19) in SW480 cells. GA IC_{50} induced a notorious and significant increase of nuclear BG4 signal after treatment for 6 h, suggesting that GA strongly trapped G4 structures (Figure 26B and 26C). A similar effect was observed using CX5461 as a positive control. Then, we investigated DNA damage response because the induction of double-strand breaks is a well-known effect associated with G4s stabilization (55). To this end, we measured the phosphorylation of histone H2AX on Ser-139 (γH2AX), as DNA damage marker, by western blot. In the same way that occurred upon exposure to CX5461, GA IC_{50} incubation for 6 h significantly induced DNA damage (Figure 26D and 26E). Therefore, GA acts as a G4 ligand inducing downregulation of several G4-enriched oncogenes and DNA damage.
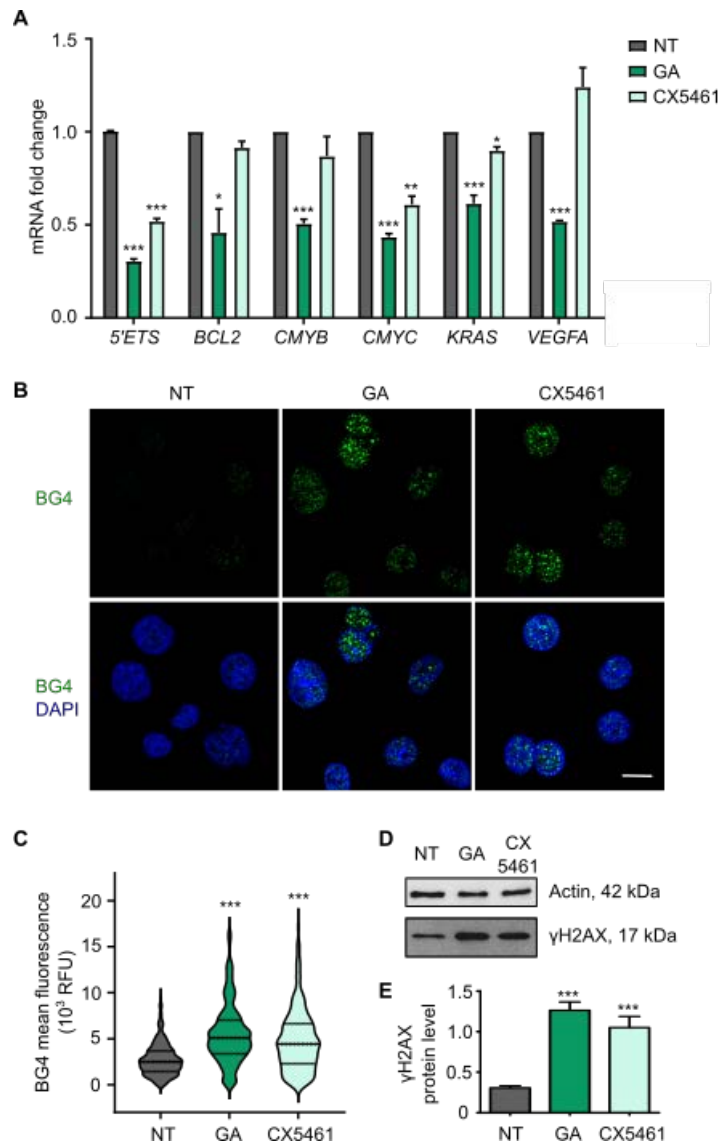
Figure 26. Gallic acid stabilizes G4s, inhibiting the transcription of several oncogenes, and induces DNA damage. (A) SW480 cells were treated with vehicle DMSO (NT), GA IC_{50} or CX5461 10 µM for 6 h and transcription of different G4-enriched genes was analyzed by qRT-PCR. Experiments were performed in biological triplicate. (B) Immunofluorescence images of SW480 cells treated with vehicle (NT), GA IC_{50} or CX5461 10 µM for 6 h and stained with the G4-selective antibody, BG4. Merged images with DAPI for DNA counterstaining are also displayed. Scale bar, 10 µm. (C) Quantification of nuclear BG4 mean fluorescence intensity from cells in (B) by Fiji analysis (N > 250). (D) Western blot experiments in SW480 cells upon treatment with vehicle (NT), GA IC_{50} or CX5461 10 µM for 6 h to determine protein levels of γH2AX as a marker of DNA damage and actin as housekeeping gene. Experiments were performed in biological triplicate and representative images are shown (E) Quantification of γH2AX protein levels normalized to actin of data in (D) by ImageJ. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

Gallic acid interacts with G4s in *5'ETS* and *CMYC*

The transcriptional inhibition of genes that harbor G4s in their promoters prompted us to examine whether GA interacted with some of these G4s using TOPRO3 fluorescent intercalator displacement (FID) assay. FID experiments are based on the displacement of a DNA light-up probe (TOPRO3) from G4 DNA upon competition with G4-binding ligands (183). For that, we used G4 sequences which were fully characterized in previous studies such as *5'ETS* (214), *BCL2* (215), *CMYB*, *CMYC*, and *KRAS* (197), *VEGFA* (216), and telomeric sequences (217). GA 10 µM exhibited a significant binding to G4s found in *5'ETS* and in the promoter of *CMYC*, decreasing the fluorescence percentage down to 84.2% and 84.3% respectively (Figure 27A). The stabilization of these G4 structures by GA was further investigated by a PCR-stop assay using a test oligonucleotide, which includes the target G4 sequence, and a complementary oligonucleotide that partially hybridizes to the test oligonucleotide. The experiment was planned because the specific binding of ligands with intramolecular G4 structures blocks the action of the DNA polymerase, and the final double-stranded DNA PCR product is not detected. In this regard, GA inhibited the accumulation of amplified products when added to PCR reactions including G4 sequences for *5'ETS* and *CMYC* at 100 µM and 50 µM respectively (Figure 27B). In contrast, after DMSO (vehicle) treatment at the same dilution as GA, no inhibition was observed even at the highest concentration.

In order to understand the effect of GA on the *5'ETS* and *CMYC* G4 conformation, circular dichroism (CD) studies were performed. Both G4s showed a positive band around 260 nm and a negative band at 240 nm, indicating the existence of a parallel G4 conformation. Upon addition of 100 µM GA, the intensity of the positive CD band of both *5'ETS* and *CMYC* G4s decreased (Figure 27C), in a similar way as is described elsewhere (218). Further binding studies were carried out using UV−vis spectroscopy. When GA was added to the *5'ETS* G4, the complex peaked around 254 nm and displayed hypochromicity (Figure 27D). In addition, when GA was added to *CMYC* G4, the UV absorption spectra exhibited

both hypochromicity in the peak at 254 nm together with an isosbestic point at 297 nm, indicating the existence of two different species in equilibrium with each other. Altogether, these results confirm that GA interacts with *5'ETS* and *CMYC* G4s.
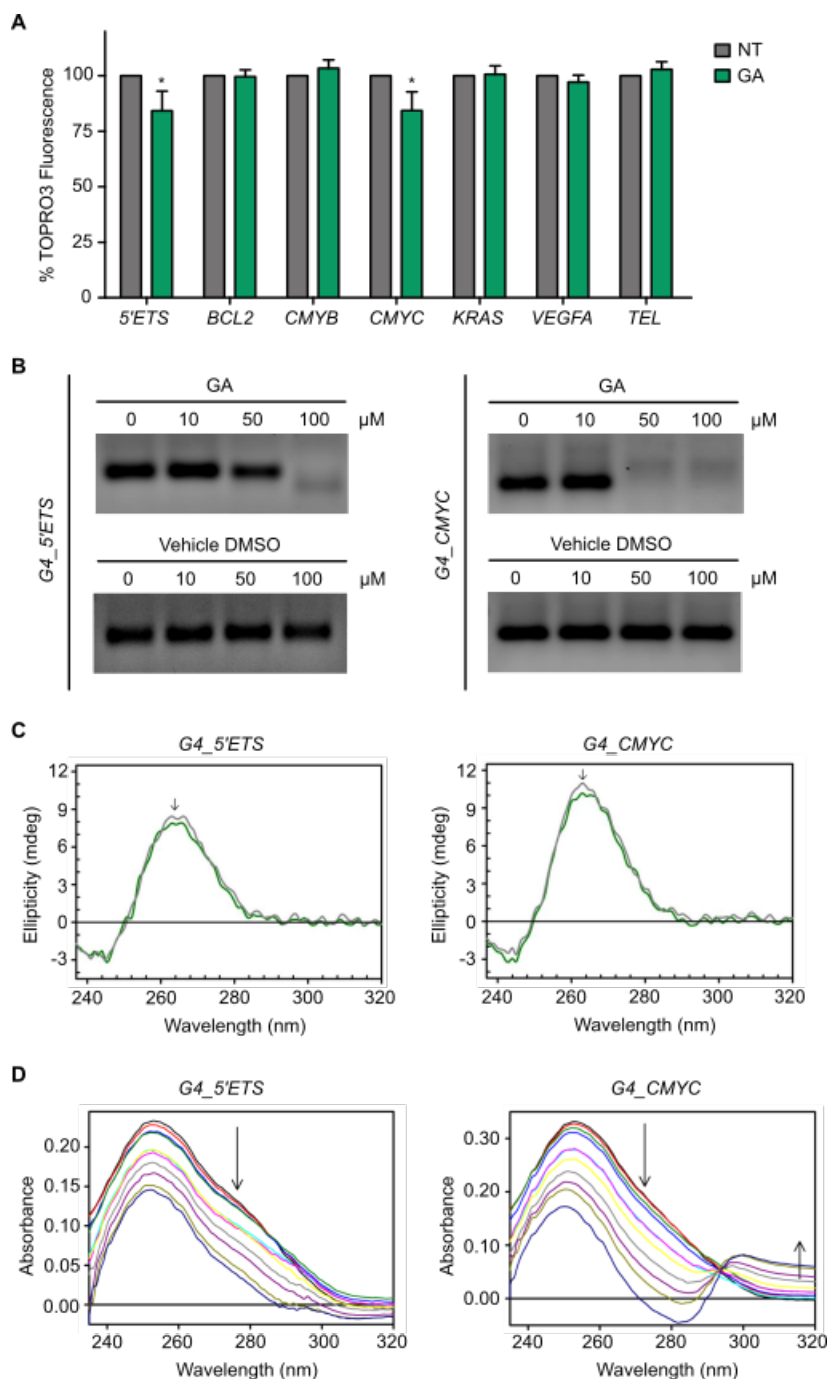
Figure 27. Gallic acid interacts with G4s in *5'ETS* and *CMYC*. (A) FID assay using different G4-containing oligonucleotides to determine the TOPRO3 fluorescence percentage in the absence (NT) or presence of GA 10 µM. Experiments were performed in triplicate. (B) Effect of increasing concentrations of GA or the corresponding vehicle

DMSO on a PCR-stop assay including the G4-containing oligonucleotide of *5'ETS* and *CMYC*. Three independent reactions were conducted per concentration and representative lanes are displayed. (C) CD spectra obtained with the G4s formed by *5'ETS* and *CMYC* in the absence or presence of GA 100 µM. The arrows indicate the direction of movement of CD peaks upon addition of GA. (D) UV−vis spectra of the G4s formed by *5'ETS* and *CMYC* obtained upon the addition of increments of GA 10 µM up to 100 µM as final concentration. The arrows indicate the direction in which the absorption peak moves after interaction of GA with G4. Experiments were performed in triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

Gallic acid induces autophagy and affects the endocytic pathway

Natural phenolic compounds are extensively associated with autophagy (219). In fact, phenols exhibit their antitumoral effects through upregulating autophagy, the non-apoptotic mode of cell death (220). In addition, several G4 ligands are reported to induce autophagy (221). In this sense, we further explored the impact of GA on autophagy pathway. For that, we performed an immunofluorescence of stable HEK293-GFP-ATG13 and HEK293-GFP-LC3 cells treated with GA IC$_{50}$ for 24 h to determine the protein level of several autophagy markers including ATG13 and WIPI2 as early autophagy makers, LC3 as a late autophagy maker, as well as PTBK1 to measure selective autophagy. In these experiments, we used PP242 (an mTOR inhibitor) at 1 µM for 1 h as a positive control for induction of autophagy, and 20 µM ivermectin (an antiparasitic drug) for 1 h as inductor of selective autophagy. We quantified the number of puncta per cell (N > 50) for each autophagy marker, and GA significantly increased the level of both early (ATG13 and WIPI2) and late (LC3) autophagy markers in a similar way than PP242 (Figure 28A). However, no significant changes were detected in PTBK1 levels upon GA treatment, in contrast to what occurred after treatment with ivermectin as positive control for selective autophagy. These results indicate that GA induces autophagy without affecting selective autophagy cascade.

Autophagy plays an essential quality control function in the cell by promoting turnover of long-lived organelles (222). In this sense, we investigated the status of organelles in HEK293 cells treated with GA IC$_{50}$ for 24 h by immunofluorescence to detect calnexin (in endoplasmic reticulum), EEA1 (in endosomes), giantin (in Golgi apparatus), LAMPI (in lysosomes), SERCA2 (in endoplasmic reticulum), and TOM20 (in mitochondria). No remarkable differences were observed in organelles after treatment with GA, except in the case of EEA1, whose fluorescent signal increased (Figure 28B). Next, we further analyzed the endocytic pathway upon treatment with GA IC$_{50}$ for a short time (6 h) by immunofluorescence and using additional markers that included EEA1, RAB5, and SNX1 for early endosomes, along with RAB11 for recycling endosomes. Changes in fluorescent signal were notoriously observed only for EEA1 marker (Figure 28C). In fact, EEA1 mean fluorescence quantification of (N > 50) revealed a significant increase after GA treatment (Figure 28D). Thereupon, GA enhances EEA1 recruitment onto membranes of newly formed endocytic vesicles, thus affecting early endosome population. Such affection may be interconnected with autophagy induction.
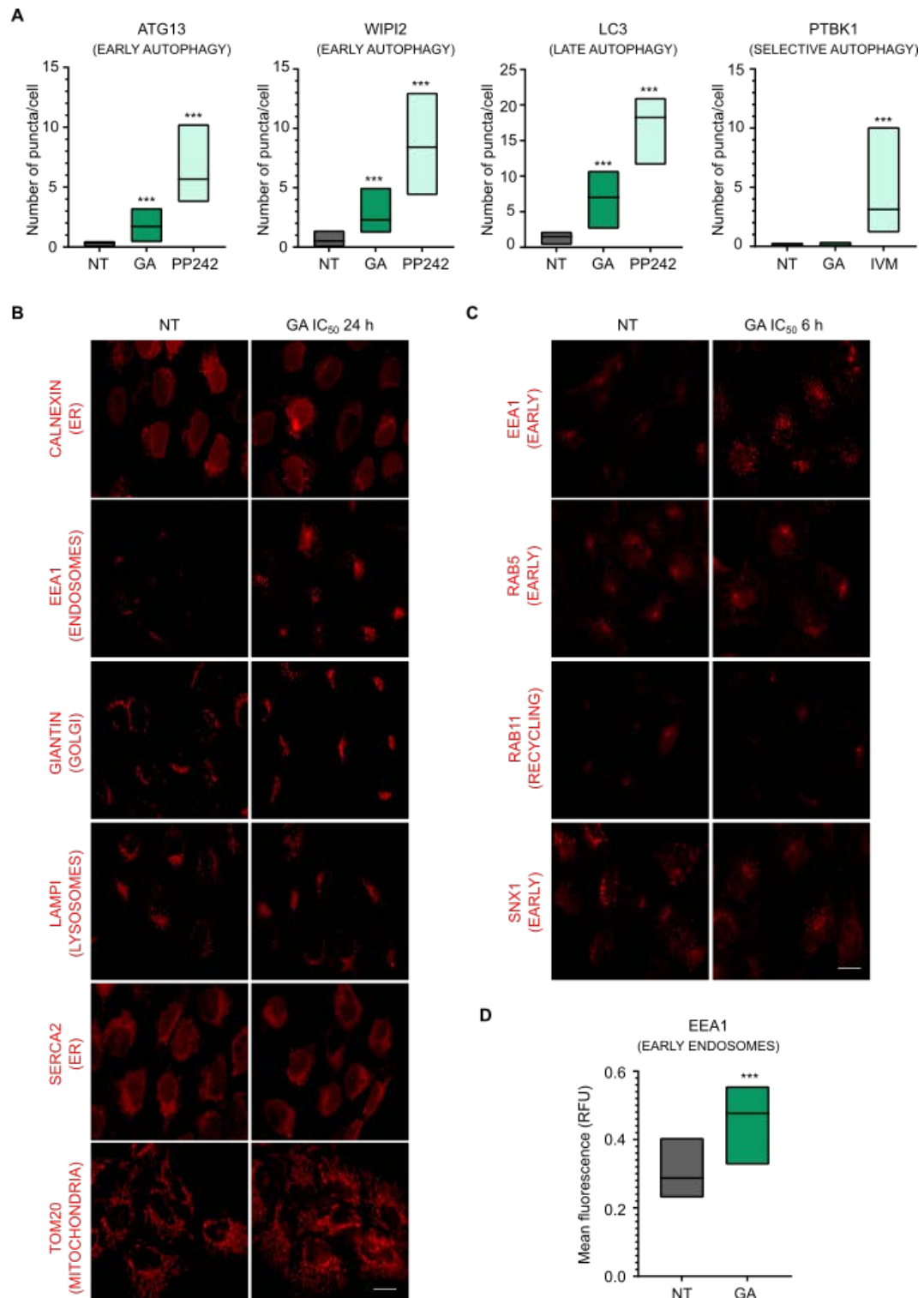
Figure 28. Gallic acid induces autophagy and affects the endocytic pathway. (A) Quantification by Fiji analysis (N > 50) of the number of puncta per cell of ATG13, WIPI2, LC3, and PTBK1 from GFP-ATG13 and HEK293-GFP-LC3 cells non-treated (NT) or treated with GA IC_{50} for 24 h or PP242 1 μM for 1 h or Ivermectin (IVM) 20 μM for 1 h. (B) Immunofluorescence images of HEK293 cells non-treated (NT) or treated with GA IC_{50} for 24 h using different antibodies to detect calnexin, EEA1, giantin, LAMPI, SERCA2, and TOM20. Representative images are shown. Scale bars, 20 μm (ER: endoplasmic

reticulum). (C) Immunofluorescence images of HEK293 cells non-treated (NT) or treated with GA IC$_{50}$ for 6 h to detect different markers of endocytosis including EEA1, RAB5, RAB11, and SNX1. Scale bars, 20 μm. (D) EEA1 mean fluorescence quantification from cells in (C) by Fiji (N > 50). For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

Gallic acid blocks tumor progression and stabilizes G4 structures in vivo

We then investigated the activity of GA in a mouse xenograft model of CRC to determine if the in vitro activity of GA was paralleled in vivo (Figure 29A). According to previous reports with SW480 cells (223), we executed xenograft studies by subcutaneous injection in NOD scid gamma (NSG) mice. Intraperitoneal treatment started when tumors reached ~ 20 mm^3 (35 days post-injection), and a therapeutic schedule with either vehicle or 200 mg/kg of GA every other day for 38 days was explored based on previous studies (224). Initially, each group included seven animals but two mice from the control group were excluded because the tumors developed lately. Therefore, five mice were considered in the control group and seven in the GA-treated one. No body weight reduction or adverse effects such as tumor ulceration were observed at any time during the study (data not shown). Over the course of the experiment, GA caused a robust blockade on the progression of tumor xenografts in treated mices compared to the vehicle control group (Figure 29B). In particular, significant differences on tumor volume were observed since the 23rd day after initiation of the treatment.

Histopathological analyses of tumors from animals sacrificed at the ending point were also conducted. Immunofluorescence analysis with BG4 showed a significant increase in nuclear BG4 signal in tumors from GA-treated animals (Figure 29C and 29D), which confirmed that GA also had a direct action at G4 sites in vivo. Furthermore, immunohistochemical assessment of the antigen Ki67 was used to estimate cell proliferation. The results demonstrated that the coverage of Ki67 was significantly decreased in tumors originated in GA-treated mice, indicating that tumors

were less proliferative after treatment with GA (Figure 29E and 29F). Finally, gene expression data obtained by qRT-PCR from excised tumors showed that, upon GA treatment, *5'ETS* and *CMYC* were significantly downregulated in responder mice (Figure 29G). Altogether our data strongly suggest that GA stabilizes G4 structures in vivo leading to an inhibition of tumor growth in CRC xenografts.
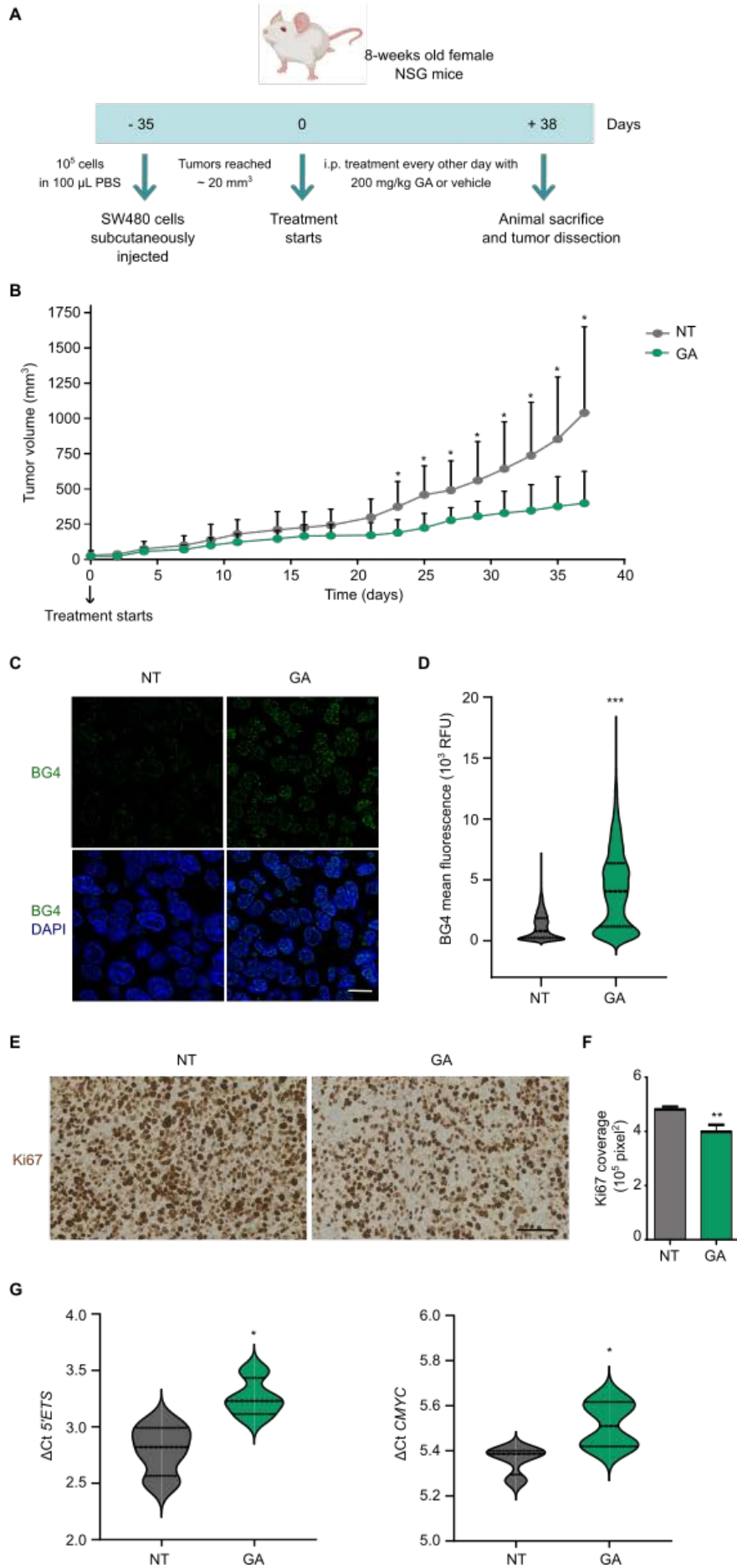
Figure 29. Gallic acid blocks tumor progression and stabilizes G4s in vivo. (A) Timeline of xenograft experiments. SW480 cells were subcutaneously injected in NOD scid gamma (NSG) mice. When tumors reached ~ 20 mm^3, mice were intraperitoneally (i.p.) treated with DMSO vehicle control (N=7) or 200 mg/kg GA (N=7) every other day for 38 days. (B) Tumor volume of SW480 xenograft mice treated with DMSO vehicle control (N=5) or 200 mg/kg GA (N=7) every other day for 38 days. Each point represents the mean tumor volume of the group ± standard deviation (only the positive bars are shown). (C) Representative images of BG4 immunofluorescence in tumor sections from control and GA-treated xenografts. Merged images with DAPI for DNA counterstaining are also shown. Scale bar, 10 μm. (D) Quantification of nuclear BG4 mean fluorescence intensity from tumor sections in (C) by Fiji analysis (N > 2,000). (E) Representative images of Ki67 staining in tumor sections from control and GA-treated mice. Hematoxylin and eosin were used as counterstaining. Scale bar, 100 μm. (F) Quantification of Ki67 coverage from tumor sections in (E) by Fiji analysis (ten different images per tumor). (G) ΔCt results for *5'ETS* and *CMYC* expression obtained by qRT-PCR from control and responder mice. Experiments were performed in triplicate. For all tests, p-values below 0.05 were considered significant and expressed as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

Gallic acid could be explored as a therapeutic agent for patients with CRC

Cancer cells overexpress ribosomal machinery (225) and *CMYC* (226) to meet their requirements for limitless proliferation. Therefore, downregulation of *5'ETS* and *CMYC* by GA could be a feasible strategy for CRC treatment. In order to assess the translational potential of GA for patients with CRC, we determined *POLR1A* (that transcribes ribosomal gene) and *CMYC* expression levels in a patient cohort with CRC by qRT-PCR. In particular, 15 CRC tumoral samples and 7 colorectal biopsies derived from non-tumoral adjacent tissue were included in this study. The analysis was performed by ΔCt method, considering that Ct is the threshold cycle of detection, and ΔCt values show relative gene expression using actin as housekeeping gene. As expected, ΔCt *CMYC* decreased in CRC meaning that *CMYC* was overexpressed in colorectal tumors compared with non-tumoral tissues (Figure 30A). *POLR1A* overexpression in CRC had already been detected in the previous section (Figure 24B). *POLR1A* and *CMYC* expression patterns observed in these

samples were further validated with a larger cohort of patients with CRC by bioinformatic analyses from the Oncomine database. The expression levels of *CMYC,* measured by log2 median-centered ratios, was significantly higher in colon adenocarcinoma than that in the non-tumoral tissues (Figure 30B). In the same cohort, *POLR1A* overexpression in CRC had already been identified in the previous section (Figure 24C). These findings reveal that *POLR1A* and *CMYC* constitute clinically attainable targets in CRC, positioning GA as a candidate for CRC treatment that requires translational exploration in the future.
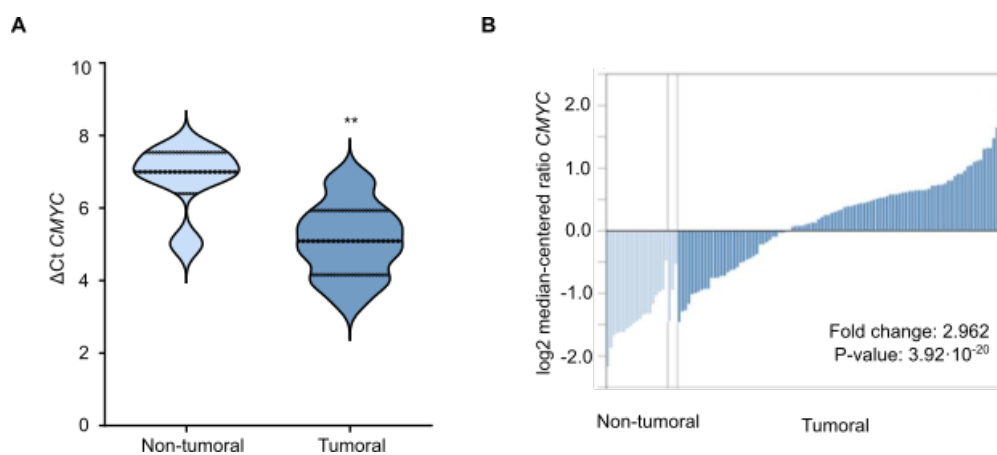
Figure 30. Gallic acid could be explored as a therapeutic agent for patients with CRC. (A) ΔCt results for *CMYC* expression in the cohort of patients with CRC, obtained by qRT-PCR. Experiments were conducted in biological triplicate. (B) Analysis of *CMYC* expression levels in Oncomine database with "TCGA Colorectal" dataset including non-tumoral (N = 22) and tumoral (N = 101) samples. For all tests, p-values below 0.05 were considered significant and expressed as follows: *$p < 0.05$; **$p < 0.01$ and ***$p < 0.001$.

OBJECTIVE 3: **Rational production of nanobodies capable of specifically recognizing DNA G4s harbored in oncogene promoters**

The major limitation for the clinical application of G4 ligands is directly related to selectivity because most compounds indiscriminately bind to different G4s (89). Within this framework, we aimed to obtain a selective ligand that interacts with a specific G4 in particular. Based on the advantages of nanobodies, we planned to produce nanobodies that specifically target G4s with a therapeutic purpose. In particular, we were interested in the generation of nanobodies against the G4 harbored in *CMYC* gene promoter intending to control the transcriptional levels of such oncogene. As we observed in Figure 14C, *CMYC* G4 plays a key role in the dysregulation of *CMYC* expression in colorectal carcinogenesis.

Generation of nanobodies targeting G4s

In an attempt to overcome limitations in active immunization of camelid animals, we performed the in vitro stimulation of naïve llama B-lymphocytes with G4s in the presence of recombinant llama interleukins according to a methodology previously described (188) (Figure 31A). Peripheral blood mononuclear cells (PBMCs) from llama were stimulated with 5 µg of *CMYC* G4 (189). Throughout the in vitro immunization protocol, cell viability was evaluated everyday microscopically, wherein all cells were viable. Similarly, there was no reduction in the cell number during and at the end of in vitro immunization. Total RNA was isolated and reverse transcribed, and VHH fragments were amplified using 30 cycles of PCR to retain the maximum possible diversity of the VHH gene (Figure 31B). VHH fragments were ligated into the phagemid pJB12, and a library with the VHH repertoire fused to the pIII gene was generated in *E. coli XL1-Blue* (VHH-*E. coli* library). Several clones were picked randomly and checked by PCR with vector specific primers, confirming that the repertoire of VHH fragments was correctly inserted (~500 bp) (Figure 31C). Part of the VHH-*E. coli* library was then amplified and infected with VCSM13 helper phages to render a VHH-phage library composed of phage particles

that possess the cloned VHHs fused to pIII proten at the viral surface. The titration of this VHH-phage library was $2.4 \cdot 10^{14}$.
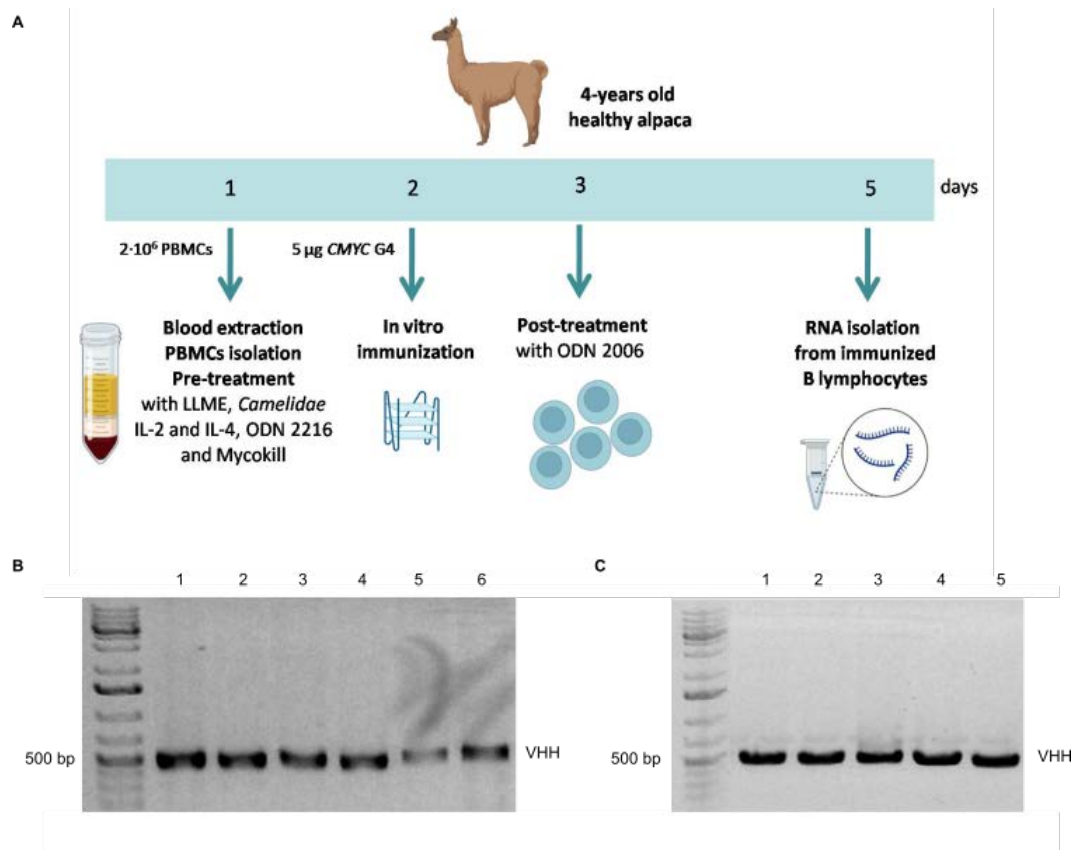
Figure 31. Generation of nanobodies targeting G4s. (A) Timeline of in vitro immunization of llama lymphocytes with *CMYC* G4 (PBMCs: peripheral blood mononuclear cell; LLME: Leu-Leu methyl-ester hydrobromide; IL-2: interleukin-2; IL-4: interleukin-4; ODN 2216: Class A CpG oligonucleotide ODN 2216; ODN 2006: Class B CpG oligonucleotide ODN 2006). (B) VHH fragments amplified from in vitro immunized lymphocytes in six independent PCR reactions. (C) VHH amplicons obtained from five different *E. coli XL1-Blue* clones electropored with VHH library. PCR amplification was performed to confirm the correct insertion of VHH in the pJB12 vector.

Screening of nanobodies targeting G4s

The presence of antigen specific phages were evaluated by enzyme link immunosorbent assay (ELISA) using neutravidin-wells coated with biotinylated G4 oligonucleotides or vehicle. First of all, the VHH–phage library was screened against different G4s by quantitative ELISA in order to determine the selectivity of VHHs or cross-reaction with different

G4 structures. For that, we captured biotinylated G4 oligonucleotides, including telomeric, *BCL2*, *VEGFA*, *KRAS*, *CMYB*, and *CMYC* (197), in neutravidin plates. The threshold for considering antigen specificity was absorbance values higher than 2.5. Apart from *CMYC* G4, VHHs also recognized additional parallel G4s like those in *VEGFA* and *KRAS*, together with mixed parallel/antiparallel G4s of *BCL2* and telomeres (Figure 32A). However, no cross-reaction was detected with the G4 harbored in *CMYB,* which adopts a tetrad:heptad structure. As expected, no signal was noticed in the control without G4. Therefore, the initial VHH-phage library seems to contain VHHs that target both parallel and antiparallel G4 structures.

To specifically select the VHHs with affinity to the *CMYC* G4, we performed three successive rounds of biopanning. For this end, escaped phages were incubated with *CMYC* G4, the same antigen used for in vitro immunization, to capture the antigen binding phages. Quantitative ELISA confirmed the binding ability of VHHs expressed on phages to *CMYC* G4, and such specific VHHs progressively enriched after each round of biopanning (Figure 32B). These results indicate that, upon three rounds of biopanning, VHHs targeting *CMYC* G4 are obtained. However, individual selection of these VHHs and further selectivity studies are required in the future.
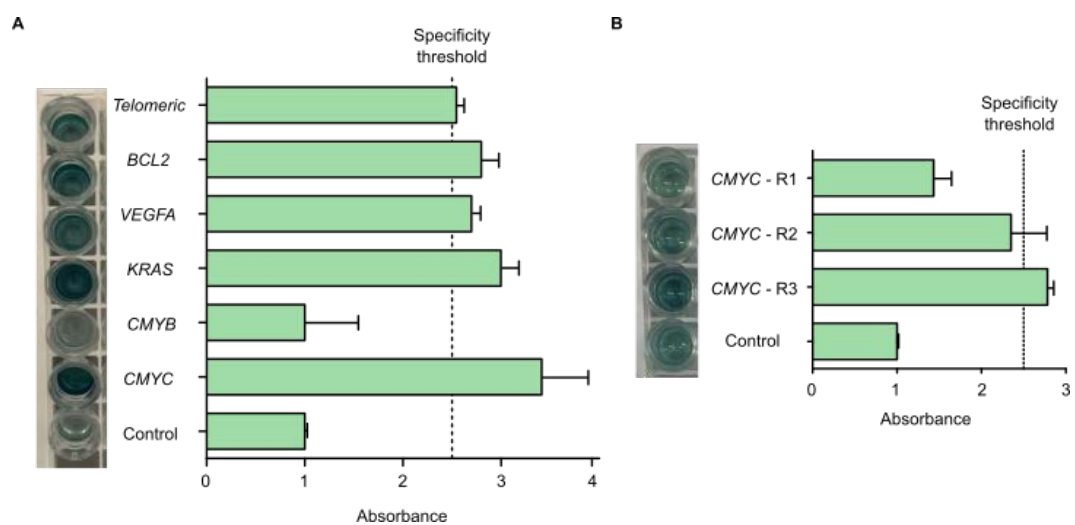
Figure 32. Screening of nanobodies targeting G4s. (A) Screening of VHH–phage library (before biopanning) against different immobilized G4s by quantitative ELISA. The

specificity threshold was set at 2.5. Experiments were performed in triplicate and representative images are shown. (B) Evaluation of specific phages to *CMYC* G4 after each round of biopanning by quantitative ELISA with immobilized *CMYC* G4. The specificity threshold was set at 2.5. Experiments were performed in triplicate and representative images are shown (R1: phages eluted from first round of biopanning; R2: phages eluted from second round of biopanning; R3: phages eluted from third round of biopanning).

DISCUSSION

The study of G4s originated from curiosity-driven structural investigations. For many decades, formation of G4s was merely considered as an in vitro phenomenon with limited impact on biology, but it has progressed to the point that G4s are now considered fundamental features of the genome and epigenome, as well as essential modulators of transcriptome (227). However, the number of known DNA G4 structures is still very limited when compared to the large number of potential G4-forming sequences predicted by bioinformatic analyses (36). This observation suggests that some G4 topologies are still unexplored, waiting to be discovered. Furthermore, this PhD thesis is exclusively focused on DNA G4s. The existence of additional non-canonical structures in nucleic acids (i.e., RNA G4s, DNA i-Motifs, DNA triplexes, etc.), and the orchestration of all these elements would provide mechanistically relevant information in the future.

DNA G4s act as multi-faceted regulatory elements in a complex cellular environment. Thanks to recent development of highly sensitive genome-, transcriptome-, and proteome-wide approaches, we are currently witnessing an explosion of studies that shed light on the specific mechanisms underlying their biological roles (46). However, whilst mechanistically informative, the main limitation of most studies is their reliance on synthetic G4 oligonucleotides, that may not realistically represent the G4 status in a cellular and chromatin context. Moreover, multiple works use ligands that artificially stabilize G4 structures and may displace natural G4-binding partners.

One of the current challenges in the G4 field is to elucidate the molecular details of how G4 formation is regulated. G4 dynamics seems to be mediated by association of G4s with proteins within live cells (228). Nevertheless, the many G4-protein interactions that have been revealed need to be further characterized in order to demonstrate where and when G4s naturally exist in the all complexity of a cellular, and ultimately, in an in vivo environment. For example, G4 presence may be tightly controlled at particular cell states. In fact, imbalance in G4 dynamics contribute to cancer development (82).

OBJECTIVE 1: **Study of DNA G4s in the progression of colorectal cancer and analysis of their implication in tumor development and DNA damage**

Cancer is a major disease that poses a serious threat to human life and health. G4s are extensively associated with cancer, playing an important role in telomere maintenance and control of gene expression of several oncogenes and tumor suppressors (84). Elevated G4 formation is a characteristic of human liver and stomach cancer because a significantly higher number of G4-positive nuclei has been identified in these tumors compared to background non-neoplastic tissues (82). The landscape of G4 structures in humans has been investigated in a variety of different cell lines, not only in non-tumoral cells such as B lymphocytes (15), but also in breast cancer (65), osteosarcoma, and leukemia (25). However, the involvement of G4s in CRC has not been studied before. In addition, the landscape of G4s in metastasis has not been deciphered yet.

In this context, we measured G4 levels along CRC progression using a cellular model to mimic the three key stages in colorectal carcinogenesis: non-tumoral, primary tumor, and metastatic. Interestingly, G4 levels increased as CRC progressed, being maximal in metastasis. We anticipated that G4 formation could be more likely to occur during DNA replication, since the associated mechanisms necessitate that duplex strands become separated at replication forks, where single-stranded DNA may fold more easily into secondary structures (19). In fact, comparing the number of G4 structures at different phases of the cell cycle, we demonstrated that G4s were higher accumulated at S phase than G0/G1 phase in all cell lines. Therefore, this maximum signal at S phase was consistent with the replication-dependent formation of G4 structures.

In addition, we investigated the reason explaining such increase in G4 levels in cancer. We hypothesized that an increase in genomic G4s in CRC may arise from mutations and/or a lower expression in enzymes that process G4s. In contrast, after analyzing a panel of G4 helicases, no mutations were detected, and these enzymes were overexpressed along

CRC progression. In fact, G4-resolving helicases are upregulated in various cancers, suggesting their requirement for rapidly proliferating tumoral cells to cope with abundant replicative demands (229). Therefore, helicase deficiency could not explain the increase in G4 levels. Searching for additional reasons, we discarded a higher replication rate in tumoral cells because we had performed the experiments with all cell lines arrested at the same phase of the cell cycle. Furthermore, differences in methylation pattern among the cell lines could explain the observed changes in G4 levels. Genomes in non-tumoral tissue are generally hypermethylated and stable, whilst stochastic hypomethylation occurs during tumorigenesis, offering a favorable environment in which G4s can fold (53). But it is also reported that G4s mold the DNA methylome by inhibition of DNA methyltransferase (73). In this sense, both processes seem to be concomitant and we would not be able to prove which come first. Nevertheless, other causes could aid to explain the increase in G4 levels. The transcriptional addiction in cancer (230) could influence G4 levels because a higher transcription rate implies the transient separation of double-stranded DNA, thus favoring G4 formation. Moreover, G4-binding proteins naturally present within cells may promote guanine-rich strands to fold into stable G4s. Interestingly, nucleolin, one of the G4-binders, is overexpressed in tumor cells (231). Altogether, these actions may take part in an interconnected regulatory loop and require further investigation in the future.

Despite the potential association or causative link between G4s and cancer development, the underlying mechanisms remain poorly understood. To establish how G4 DNA structures are related to CRC biology, we investigated G4 status in relevant genes for CRC etiopathogenesis, such as *CMYC* and *KRAS*. In the same way as previously described (226), we confirmed that *CMYC* was overexpressed in CRC with the cellular model. *KRAS* oncogene harbors gain-of-function mutations widely described in literature (138), and we observed that *KRAS* was also overexpressed in CRC with the cellular model. In this scenario, we analyzed whether G4s could be affecting *CMYC* and *KRAS* gene

expression levels. It has been reported that approximately 30% of tumors contain specific G-to-A mutations, not present in the surrounding normal tissue, that destabilize the parallel G4 in *CMYC* (232). However, we did not detect G4-disturbing mutations for *CMYC* or *KRAS* in any of the cell lines. We then applied BG4 ChIP-qPCR methodology to map G4 formation in *CMYC* and *KRAS* in the cellular model. Interestingly, we demonstrated that the G4 involved in *CMYC* silencing was folded in non-tumoral cells but unfolded in CRC, which could give rise to the increase in basal transcriptional activity of *CMYC* observed in the CRC cell lines. In this sense, we pose *CMYC* G4 as responsible for transcriptional repression of *CMYC* at the non-tumoral stage, as widely described (57). Future research may reveal how *CMYC* G4 is unfolded during malignant transition. In contrast, no changes were appreciated in G4s folding of the *KRAS* promoter, suggesting that mechanisms of control different from G4s may be involved.

G4 structures, if not resolved during DNA replication, represent fragile sites that promote genome instability, which is a well-known hallmark of cancer (233). We confirmed this observation on the CRC cellular model. DNA damage increased along CRC progression and DNA damage levels were higher in S phase compared to those in G0/G1 phase. We explain that because when DNA is being replicated at S phase, natural errors may occur. Alternatively, such increase in DNA damage could be due to the concomitant maximal abundance of G4s structures during S phase. To decipher this hypothetical association between G4s and DNA damage, we mapped DSBs at nucleotide resolution by BLISS methodology, and we overlaid the resulting DSBs with G4s data previously used as reference. Interestingly, we found that DSBs were specifically enriched at G4 sequences and neighboring regions. In fact, some researchers have used the genome-wide distribution of DNA damage sites to identify clusters of sequences with a propensity for G4 formation (55). To go further, DNA damage imposed by G4s presence may be associated with the acquisition of cancer-relevant copy number alterations and mutations over time (53).

Our work is pioneer in demonstrating the association between G4s and DNA damage using BLISS methodology to map DSBs. Previous studies have obtained similar findings but using indirect measurements of DNA damage, such as ChIP-seq or immunofluorescence with antibodies against γH2AX or RAD51 (202). Nevertheless, the main limitation of our study resides in the use of a G4 map as reference, which derives from a distinct cell line with different experimental conditions (15). Since a substantial number of G4 sites are cell-type dependent and hence are dependent on the cellular state (25), our study could be improved by overlapping DSBs with BG4 ChIP peaks obtained with the same cells at identical conditions.

Based on BLISS results, most of DSBs were unique (only represented with 1 UMI), suggesting that DSBs are widely spread through the genome and do not accumulate at recurrent genomic locations in multiple cells. This phenomenon could be explained due to the dynamic nature of G4 structures in absence of particular G4-stabilizing ligands. From our point of view, the characteristic dynamism of G4s would have repercussions on the dynamism of DSBs. Moreover, we would have to consider that part of these detected DSBs arise as a result of normal cellular processes where the genome is continuously and randomly damaged. In particular, we found that DSBs mapped both at intergenic regions and genes. Interestingly, some of these genes have been extensively related to cancer in the literature. For instance, *CASC11,* which activates the WNT/β-catenin pathway to promote growth and metastasis in CRC (234), accumulated DNA damage in both tumoral SW480 and SW620 cells. *CASC11* is located in chromosome 8, just upstream of the *CMYC* gene. Their proximity raises curiosity for future studies focused on the effects imposed by the well-characterized G4 harbored in *CMYC* promoter on the proximal regions. Moreover, a variable percentage of DSBs mapped at centromeric regions, confirming that centromeres are repetitive fragile sites for genome instability (235). However, G4s have not been described in centromeres so far, and thus, DSBs can not be considered as consequence of G4 formation in these

genomic loci for the moment. We observed an increase in centromeric DSBs along CRC progression. Taking into account that centromeres are required for faithful genome inheritance at each round of cell division, the accumulation of DSBs in centromeres may be associated with aberrant rearrangements that frequently exist in cancer.

Unrepaired DSBs are cytotoxic genetic lesions that lead to lethality (236). Therefore, overstabilization of G4s with G4-binding ligands results in DNA damage induction at first, and cell death at the end. In fact, this mechanism of genome instability and cell killing by G4 ligands has been particularly effective in tumors deficient for DNA damage repair (202). In this way, G4 structures are considered attractive molecular targets for cancer therapeutics. In order to potentially identify treatment strategies in CRC, we explored the cytotoxic potency of several established, yet structurally distinct, small molecules targeting G4s such as BMH21, CX3543, CX5461, and pyridostatin. These compounds target multiple G4 structures and are considered as G4 ligands of broad spectrum (55). According to previous studies, we expected that tumoral cells, with higher G4 levels, would respond better to G4 ligand treatment (65). In turn, upon treatment with these G4 ligands in the cellular model of CRC, cancer cells were not more sensitive to the treatment. Therefore, G4s level is not sufficient to predict G4 sensitivity and could not be used as a predictive biomarker for G4 ligand therapy. Furthermore, after BMH21 treatment, G4 stabilization and induction of DNA damage was confirmed in all cell lines without selectivity. In fact, non-tumoral cells were slightly more sensitive to G4 ligands. As discussed above, G4 helicases maintain genome stability during DNA replication and transcription (45). In this sense, we propose that lower expression levels of G4 helicases in non-tumoral cells may be associated with higher accumulation of DNA damage and higher induction of cell death after treatment with G4 ligands. Regardless, more selective G4 ligands are required to be exploitable for CRC treatment.

OBJECTIVE 2: Screening of chemical compounds to select ligands that target DNA G4s and display antitumoral activity

OBJECTIVE 2.A: Screening of synthetic compounds

We assessed the potential therapeutic role of NDI G4 ligands in CRC, where only compounds interacting with the *KRAS* promoter have been previously evaluated (237). Using the cellular model to simulate colorectal carcinogenesis, we identified the G4 ligands T1 and T5, which exhibited a potent and selective effect inhibiting cell growth in vitro. T1 and T5 belong to the NDI family, a class of small molecules capable of stabilizing G4s with high affinity (98). They are composed of a single naphthalene core that can be substituted with alternative side chains, from alkyls to aminoacids (99), and carbohydrates (174), resulting in an extensive list of NDI derivatives. Interplay of different binding modes of NDIs to G4s with different topologies has been confirmed, with end-stacking always operative as the predominant binding event (238). Traditionally, NDIs have been considered to bind primarily to telomeric G4s and inhibit telomerase activity (98). More recent studies have designated additional targets for G4 binding, including the promoters of *BCL2* and *CMYC* oncogenes (239). However, our work provides evidence of NDI derivatives targeting G4s in ribosomal DNA and impacting on Pol I transcription. In fact, to date most research has focused on ex vivo G4-binding assays with large amounts of NDIs. Very little is known about the primary role and mechanism of NDIs on tumoral cells at smaller concentration ranges. We demonstrated that T1 and T5 NDI derivatives inhibited rRNA synthesis in a cellular context at concentrations even lower than IC_{50}. This gap between rRNA synthesis inhibition and its effect on cell proliferation may be due to the long half-life of human ribosomes (240).

In an attempt to study the precise manner in which T5 inhibits rDNA transcription, we identified specific rDNA sequences capable of forming stable G4s structures and mainly targeted by T5 in physiological conditions. We propose a mechanism whereby binding of T5 to rDNA G4s

causes the disengagement of POLR1A subunit and the consequent inhibition of transcription by Pol I. These rapid kinetic effects occur before the decrease in the abundance of POLR1A. Next, in response to the blockade of POLR1A near G4 sites, POLR1A is marked for proteasome-mediated clearance as a downstream effect. Finally, T5 causes cell death by autophagy.

The above-mentioned activity is strikingly similar to other structurally distinct compounds described as Pol I inhibitors, such as BMH21, CX5461, and CX3543. To note, redundancy in therapeutic strategies targeting Pol I is not only useful, but essential for clinical development. Notwithstanding, the proposed mechanism for T5 is a distinctive feature among the previously described Pol I inhibitors. Within them, BMH21 has been suggested to intercalate within GC-rich regions in rDNA and promote degradation of POLR1A subunit (181). Alternatively, previous studies have also shown that CX5461 exerts its effect through targeting the SL1 transcription factor, indispensable to recruit the Pol I preinitiation complex (105). Conversely, CX3543 has been considered a nucleolus targeting agent that disrupts nucleolin/rDNA G4 complexes, thereby inhibiting Pol I transcription (104). In fact, a recent study has proposed a role for both CX compounds as general G4 binders and has remarked that both are mechanistically different from BMH21, which revealed no detectable G4 binding (202). More recently, BMH21 has been reported to bind to G4s in the *CMYC* promoter (201). In this context, there is no strong evidence connecting Pol I inhibitors with a direct stabilization of rDNA G4s. At this point, it is worth noting the fact that T5 specifically targeted G4s in rDNA backbone at the micromolar range, inducing a rapid inhibition of Pol I-mediated transcription as a consequence. In accordance with its preferential binding to rDNA G4s, T5 displaced ThT from nucleolar G4s in a cellular environment.

Selectivity is a major concern in cancer management to decrease adverse effects frequently attributed to classical cytotoxic drugs. Generally, two key processes contribute to a targeted therapy in cancer:

exploiting unique features for tumor cells and enhancing drug uptake by malignant cells.

Firstly, Pol I is a highly active enzyme responsible for synthesizing rRNAs, the most abundant RNA species in the cell, constituting the rate-limiting step in ribosome biogenesis (212). Since cancer cells possess an increased ribosynthetic activity to meet their demands for increased protein synthesis, it is easily concluded that Pol I constitutes an emerging target in cancer therapeutics. Moreover, we show in this study that *POLR1A* was overexpressed both in tumoral and metastatic cell lines, and in CRC tumor biopsies, compared with the normal cell line and non-tumoral colon tissues, confirming that *POLR1A* expression rate is proportional to cell growth. These results are consistent with data showing deregulation of Pol I transcription in carcinogenesis (225). In fact, *POLR1A* has been previously identified among other candidates as a biomarker in CRC (241). Despite evidence, the clinical significance of Pol I has not been exploited before in CRC. On the other hand, we determined that *POLR1A* expression levels were inversely correlated with IC_{50} values in the cellular model of CRC progression. Searching for personalized medicine, these results could open up avenues to perform *POLR1A* expression analysis before treatment to potentially select the best responders to POLR1A inhibitors, such as T5.

Secondly, there is no denying that carbohydrate-NDI conjugation in T5 contributed to promoting its selective entry into tumoral cells. It is widely established that tumor cells prioritize glucose uptake via increasing the number of GLUT receptors, and coordinate an enhanced glucose entry with increased glycolysis (Warburg effect) to sustain a demanding and uncontrolled proliferation (242). Our study reveals that T5 was selectively taken up by cancer cells where GLUT1 seemed to play a major role, although the influence of different GLUT isoforms still needs further investigations. Additional mechanisms could be also involved, including asialoglycoprotein receptors, which are overexpressed in tumoral tissues, and specifically recognize galactose residues harbored in the lactose residue of T5 (243). The analysis of samples collected from patients with

CRC and the cellular model demonstrated that *GLUT1* was overexpressed in cancer compared with non-tumoral specimens. In fact, *GLUT1* is frequently upregulated during oncogenesis in many different types of tissues (208), and *GLUT1* overexpression has been previously considered as an unfavorable prognostic biomarker in patients with CRC (244). Altogether, these data highly reinforce that T5 takes advantage of the unique traits of cancer cells, and elicits a potent and selective antitumor activity, while having minimal effect on non-malignant cells. Therefore, T5 would result in less genotoxic side effects than conventional therapy. In the future, we would aim to continue our study by exploring the anticancer potential of T5 in vivo.

OBJECTIVE 2.B: Screening of natural compounds

A huge volume of literature data suggests that a diet rich in fruits and vegetables could reduce the incidence of CRC. This effect has been mostly attributed to phenolic compounds (245). Apart from chemoprevention, the development of anticancer therapies involving natural phenols has undergone exponential growth in recent years (246). However, underlying mechanisms of phenolic compounds are not fully understood. Here, we studied the antitumor potential in CRC of five different phenolic compounds. Among them, we selected GA as a promising candidate and provide detailed evidence of its mechanism of action via binding to G4s.

GA is a phenolic molecule widely present in varied food sources, with a mean content of 1.75 mg/100 g and 8.25 mg/100 g in different fruits and vegetables respectively (213). Among other phenols that we tested, we selected GA due to its potent and selective antitumoral effect in CRC. Such difference in efficacy seems to be due to the variations in their structures as well as their molecular targets. In fact, compounds with a greater number of hydroxylic groups exhibit better anticancer activity compared to those without hydroxyl groups or compounds with methoxy (-OCH_3) moieties. In this regard, GA, which possesses three hydroxyl groups attached to 3, 4, and 5 positions of a benzoic acid core, is reported

to be more effective versus other phenols (166). Moreover, it has been shown that GA suppresses cell growth not only in CRC (247), but also in other types of cancer (248).

In agreement with previous studies demonstrating that phenolic compounds are associated with the dysregulation of the cell cycle (249), our results indicated that GA induced an arrest at S and G2/M phases. However, our work includes relevant findings in several ways. Firstly, we identified that GA caused nucleolar stress, and secondly, that GA induced down-regulation of G4-containing genes. These effects prompted us to further examine the role of GA as a G4-ligand. Thirdly, we confirmed that GA stabilized G4s in a cellular environment both in vitro and in xenograft sections by immunofluorescence with BG4 antibody. Therefore, to the best of our knowledge, our work is the first demonstration that a natural phenol binds to G4s in human cancer cell lines, paving the way for future studies.

In addition, we demonstrated GA binding to G4s present in *5'ETS* (the same G4 targeted by NDIs) and in the promoter of *CMYC* through biophysical studies. Based on the most simplistic model, G4s are considered as repressors of transcription by preventing polymerase processivity (250). Hence, GA, once inside the cell, would bind to the G4 found in *5'ETS* and *CMYC*, which could explain the downregulation of *5'ETS* and *CMYC* upon GA treatment that we observed both in vitro and in vivo. Strikingly, G4s harbored in *5'ETS* and *CMYC* have in common that both adopt a parallel structure, and both are biologically relevant substrates of nucleolin, the most abundant nucleolar phosphoprotein (197). However, although GA also inhibited the transcription of other G4-enriched oncogenes such as *BCL2*, *CMYB*, *KRAS*, and *VEGFA*, we were not able to identify what G4s were involved in these regulation loops and further investigation is required. Compelling research has suggested that G4s may not only be involved in proximal transcriptional control, but also part of long-distance epigenetic mechanisms (46). Therefore, we must consider the G4 not as an isolated entity within a specific genomic location. Instead, G4s are part of an interconnected network of interactions with other biomolecules in living cells (46).

In accordance with its G4-stabilizing properties, GA shared some effects with other well-established G4 ligands. Consistent with a previous study (251), our results indicated that GA induced DNA damage. In fact, many G4-stabilizing ligands produce DNA damage in the vicinity of G4-forming sequences (55). Furthermore, GA induced bona fide autophagy, a similar effect than that exerted by the G4 ligand CX5461 (252). Consistent with these results, additional G4 ligands are reported to tip the balance between apoptosis and autophagy, favoring the last one (221). However, the molecular mechanism explaining how autophagy is regulated by G4s requires further investigation. Recently, it has been demonstrated that expression of *ATG7*, a gene that is critical for autophagy, is controlled by DNA G4s (253). In the case of GA, it inhibited transcription mediated by Pol I, which finally would lead to the arrest of biosynthetic processes. We speculate that such stress prompted by GA may be sufficient to induce autophagy. Moreover, accumulating evidence shows that autophagy and endocytosis are interdependent, since both pathways intersect at different stages during vesicle formation, fusion and trafficking, and share parts of the molecular machinery (254). We also observed that GA impacted on the endocytic pathway. In the literature, there is no evidence for modulation of endocytosis with phenolic compounds so far. Nevertheless, the trafficking events within the endosomal network are controlled by a subset of kinases like p38 mitogen-activated protein kinase (MAPK) (255). Considering that GA is described to modulate p38 MAPK signalling (256), we hypothesize that GA may affect endosomal populations through this mechanism. Regardless, the association between G4s, endocytic pathway, and autophagy remains waiting to be discovered.

Interestingly, the anticancer activities of GA have been widely disclosed in the literature before. In CRC, most studies have mainly attributed its anticancer effects to the generation of reactive oxygen species and induction of apoptosis (247, 257). In addition, GA is reported to inhibit angiogenesis through suppression of VEGF secretion in ovarian cancer (258). Moreover, GA is shown to impose anti-inflammatory effects in prostate cancer through inhibition of the expression of many cytokines

(259). However, altogether these experiments have been carried out after exposure to GA for a long time (24, 48 or even 72 h). Based on our observations, we suggest that part of the previously reported effects may be considered as downstream and indirect processes that derive from the role of GA as a G4-stabilizing ligand. Furthermore, epigenome-modifying abilities of GA have been observed in tobacco-associated cancers where GA reduces DNA methyltransferases activity within one week (260). It is possible that such effects would be explained since GA increases the percentage of stabilized G4s, and these structures themselves mold the DNA methylome by sequestering methyltransferase DNMT1 (73). Still, how these processes are so carefully orchestrated within the cells through G4 targeting with GA requires further investigation.

From a translational point of view, we demonstrated that patients with CRC overexpressed *POLR1A* and *CMYC*, and thus, we propose that G4-mediated downregulation of ribosomal and *CMYC* genes exerted by GA would constitute an attainable approach for CRC treatment. On behalf of the in vivo experiments, treatment with GA successfully reduced tumor growth in CRC xenografts at dosage levels that did not cause observable damage to major organs. Thereupon, simultaneous targeting of multiple fundamental CRC pathways by G4 stabilization constitutes an advantageous approach, although variability of potency and selectivity among different G4s and pathways remains unclear. Notwithstanding their promising role for cancer treatment, polyphenols often have a poor bioavailability when administered as pure active principles, representing an important limit to their use (246). Their bio-transformation, at colon level, by the heterogeneity of human gut microbiota, also leaves open enormous spaces for further research (261). In particular, two bacterial strains have specifically been identified as producers of GA in humans (262). Bioavailability, and thus efficacy, of these compounds has been improved by their administration in combination with other phytochemicals or even in nanotechnology-based formulations (263). Moreover, the possibility of combining conventional chemotherapeutic drugs with polyphenols has offered valuable advantages, such as the building of

more efficient anticancer therapies with less side effects on the health of patients (245).

Undoubtedly, our work stands out the implication of nutrigenomics in cancer treatment. Nutrigenomics refers to the use of biochemistry, physiology, nutrition, genomics, proteomics, metabolomics, transcriptomics, and epigenomics to seek and explain the existing reciprocal interactions between genes and nutrients at molecular level (264). We revealed here how a natural bioactive compound that we consume in our regular diet, GA, was able to affect gene expression by interaction with G4s. The stabilization properties of GA were inferred from experiments with cells in vitro, and most importantly, with animals, being possibly extrapolated to humans. Through this underlying mechanism, GA is directly involved in nutrigenomics, which ultimately govern human health and disease.

OBJECTIVE 3: Rational production of nanobodies capable of specifically recognizing DNA G4s harbored in oncogene promoters

Camelid nanobodies constitute a therapeutic approach to treat a wide range of diseases ranging from immune, bone, blood and neurological disorders; infectious diseases, and cancer (171). Nanobodies are able to recognize their antigen with high affinity and specificity. Potential targets include cell surface proteins, cytokines, other secreted proteins, and even intracellular proteins (265). However, DNA-targeting nanobodies have not been developed yet. The advent of nanobodies technology, together with the relevant functions displayed by G4s in living cells, increased our interest in the development of nanobodies capable of specifically target G4 structures. In particular, we aimed to generate nanobodies against the G4 that controls *CMYC* expression in CRC. This challenging approach constitutes the first G4-targeting in a specific way with nanobodies all over the world.

To date, a very limited number of antibodies targeting G4s have been generated. The first antibody was Sty49, a single-chain variable fragment antibody that revealed G4 formation at telomeres of ciliates (18). One decade later, the single-chain variable fragment antibody BG4 emerged to reveal G4s formation in human cells (19). In the following years, alternative G4-specific antibodies like IgG 1H6 (20), and single-chain variable fragment antibody D1 (21) appeared. Notwithstanding, these antibodies show substantial binding to several types of G4 structures, and also display detectable cross-reactivity to other secondary structures in DNA (266). In this context, we aimed to improve the specificity using nanobodies.

We faced several methodological difficulties in the production of nanobodies against DNA G4 structures. In a conventional manner, nanobodies production requires active immunization of *Camelidae* animals, which is laborious, time consuming, and costly. Another downside of active immunization is that it becomes extremely complicated when antigens are of high pathogenicity, toxicity or represents non-

immunogenic molecules. Despite of existing natural antibodies against double-stranded DNA (i.e., in lupus disease), DNA immunogenicity is low, and several approaches, such as hapten-conjugation, can be applied to improve it (267). However, we decided to immunize with the G4 alone to avoid the possible production of nanobodies against the hapten. In an attempt to overcome limitations of active immunization, we performed the in vitro stimulation of naïve llama B-lymphocytes with *CMYC* G4. Such pipeline reduces the cost substantially required for maintenance of camelid herd for active immunization, and can be achieved within a week (188). Two crucial components are required for successful in vitro immunization, the Leu-Leu methyl ester hydrobromide (LLME) and interleukins (ILs). Peripheral blood mononuclear cells (PBMCs) contain lysosome-rich cytolytic T cell subpopulation, which is removed from culture before in vitro immunization by treatment with LLME. Species-specificity of ILs was solved by the production of IL-2 and IL-4 specific for the given camelid species at the laboratory of Dr. Mangesh Bhide.

After in vitro immunization, we fortunately generated a library of nanobodies against *CMYC* G4, although some nanobodies showed cross-reactivity with other types of G4 structures. The ideal is that G4 ligands would display selectivity among different G4 topologies. In the future, structure-activity relationship studies could significantly improve the physicochemical properties of ligands, and enable the optimum trade-off between affinity and selectivity. In this regard, it has been reported that mitigating the affinity of the binding core for G4s results in an increased selectivity and sensitivity (268). Furthermore, it has been described the ligand design and development to acquire specificity and selectivity without compromising affinity (88). However, general rules to get a highly selective G4 ligand are not deciphered yet.

Thanks to biopanning rounds, we enriched the repertoire of nanobodies targeting *CMYC* G4. Henceforward, good binders would have to be individually selected and fully characterized in antigen-binding and epitope recognition. The selected nanobodies would have to be cloned in a suitable expression vector, and then, transformed into *E.coli WK6* to

obtain an optimal expression and production of soluble nanobodies. Finally, candidate nanobodies would be purified by immobilized metal affinity chromatography because of their fusion with a histidine-tag.

In the best case, these nanobodies could be used for CRC therapy. For that, nanobodies would have to be delivered intracellularly. Several strategies have been explored for their intracellular transportation (269). For instance, lentiviral vectors can be engineered to encode nanobodies, thus producing intracellular nanobodies that could associate with intracellular antigens. The stability of nanobodies is perfectly suited for the production of intrabodies that require their expression in the reducing environment of the cytoplasm. In fact, Bax-specific intrabodies have been developed to prevent the oxidative stress-induced apoptosis (270). Another prominent strategy for the intracellular delivery of nanobodies consists of decorating them with cell-penetrating peptides (CPPs), which are composed of short amino acid sequences rich in arginine, and mediate the direct cross of the cell plasma membrane (271). The targeting of intracellular targets may also proceed by taking advantage of bacterial type III secretion system (T3S). Gram-negative bacteria use a specialized secretion apparatus known as the T3S system to inject proteins directly into the eukaryotic cells. It has been demonstrated that *E. coli* bacteria carrying a T3S can successfully translocate nanobodies into mammalian cells (272). All these strategies would be further explored to achieve the antitumoral use of nanobodies targeting *CMYC* G4 within a cellular context.

Overall, in this PhD thesis, several G4 ligands were investigated as therapeutic approaches for CRC. Interestingly, the obtained results may be extrapolated to other types of cancer. Although further experiments are required, the present PhD thesis opens avenues for application of these promising G4 ligands not only for CRC, but also for other types of tumors.

Despite the existence of thousands of small molecules targeting G4s (89), their clinical application has thus far been limited due to several reasons. Firstly, the variety of binding sites for these ligands and the differences in their effects on the G4 structures make it difficult to unravel how G4s influence biological function, i.e., whether the stabilization or destabilization of G4s promotes or inhibits gene expression. Secondly, the correlation between stabilization in vitro and cell activity is not straightforward. In particular, a G4 target characterized in vitro may not be the sole G4 targeted in cells. Furthermore, there is also inherent cell variability, which has an impact on the relationship between in vitro and in vivo results. A further point to be addressed for the majority of ligands described thus far is that they are generally characterized by high-molecular weights and protonated side chains, which may affect their cellular uptake. However, the major limitation for the clinical application of G4 ligands remains to be directly related to selectivity. In fact, the selectivity pattern of G4 ligands seems to be dose-dependent. Although global or multiple G4 targeting approaches may be effective, targets need to be clearly defined in advance. Other conceivable obstacles are the potential side effects of the ligands on normal tissues, that highlight the importance of selecting G4 ligands according to their selectivity indexes. Moreover, the predictive response biomarkers need to be identified if a personalized anticancer management is to be achieved.

More efforts should be devoted to improve selectivity and reduce side effects derived from the treatment with G4 ligands. Given the rapid development of G4 ligands, we are confident that the limitations could be overcome in the forthcoming years. In this sense, the G4-mediated antitumoral effects reported herein may pave the way for cutting-edge therapeutic approaches in the future treatment of human cancer.

CONCLUSIONS/ CONCLUSIONES

1. The cellular model of CRC progression that we have established is proven to be an useful model to analyze the implication of G4s in colorectal carcinogenesis, their association with DNA damage, and the therapeutic potential and selectivity of G4 ligands.

2. Stabilization of G4s and DNA damage increase along CRC progression. The same occurs with the expression level of helicases involved in G4s unwinding, which implies that there is no direct correlation between the levels of these helicases and the abundance of G4 structures. DNA damage accumulates at the vicinity of G4s, establishing a connection between both processes.

3. G4s dynamics may affect the expression level of key genes in CRC. The G4 at the promoter region of *CMYC* oncogene is folded at the non-tumoral stage but unfolded at CRC, which confirms its repressor role in the expression of *CMYC* in non-tumoral cells.

4. NDIs possess antitumoral activity by a mode of action that involves ribosomal G4s targeting and impairment of transcription mediated by RNA polymerase I. Carbohydrate-conjugation in NDI derivative T5 improves its selectivity for tumoral cells due to its preferential uptake through overexpressed glucose transporter 1. This strategy could be exploited for targeted therapy in cancer.

5. GA, a natural phenolic compound present in our diet, exerts antitumoral activity by interaction with G4s both in vitro and in vivo. In particular, GA binds to G4s in ribosomal and *CMYC* genes, inhibiting their transcription. These results demonstrate the importance of diet in the prevention and treatment of CRC.

6. In patient cohorts with CRC, *CMYC*, *GLUT1*, and *POLR1A* are transcriptionally overexpressed in tumoral samples versus non-tumoral ones. Therefore, T5 and GA could be explored as therapeutic agents for patients with CRC.

7. A library of nanobodies that recognize *CMYC* G4 is produced, which constitutes the first G4-targeting strategy with nanobodies worldwide. Thanks to the elevated specificity of nanobodies, we expect that our approach could be applicable to specifically inhibit *CMYC* expression in CRC.

1. El modelo celular de progresión de CRC que hemos establecido demuestra ser un modelo útil para analizar la implicación de los G4s en la carcinogénesis colorrectal, su asociación con el daño en el DNA, además del potencial terapéutico y la selectividad de ligandos de G4s.

2. La estabilización de los G4s y el daño en el DNA aumenta a lo largo de la progresión de CRC. Lo mismo ocurre con el nivel de expresión de helicasas involucradas en desenrollar G4s, lo que implica que no hay una correlación directa entre estas helicasas y la abundancia de estructuras de G4s. El daño en el DNA se acumula en la vecindad de los G4s, estableciéndose una conexión entre ambos procesos.

3. La dinámica de los G4s puede afectar el nivel de expresión de genes clave en el CRC. El G4 en la región promotora del oncogen *CMYC* se encuentra plegado en el estadío no tumoral pero está desplegado en CRC, lo que confirma su papel represor en la expresión de *CMYC* en células no tumorales.

4. Los NDIs poseen actividad antitumoral mediante un mecanismo de acción que implica su interacción con G4s del ADN ribosómico y la disfunción en la transcripción mediada por la RNA polimerasa I. La conjugación con carbohidrato en el derivado de NDI T5 mejora su selectividad para células tumorales debido a su captación preferente a través de transportadores de glucosa 1 sobreexpresados. Esta estrategia podría explotarse para la terapia dirigida en cáncer.

5. El GA, un compuesto fenólico natural presente en nuestra dieta, ejerce actividad antitumoral mediante interacción con G4s tanto in vitro como in vivo. En concreto, el GA se une a G4s en genes ribosómicos y en *CMYC*, inhibiendo su transcripción. Estos resultados demuestran la importancia de la dieta en la prevención y el tratamiento del CRC.

6. En cohortes de pacientes con CRC, *CMYC*, *GLUT1* y *POLR1A* están transcripcionalmente sobreexpresados en muestras tumorales frente a no tumorales. Por lo tanto, T5 y GA podrían explorarse como agentes terapéuticos para pacientes con CRC.

7. Se produce una batería de nanoanticuerpos que reconocen al G4 de *CMYC*, lo que constituye la primera estrategia dirigida a G4s con nanoanticuerpos en el mundo. Gracias a la elevada especificidad de los nanoanticuerpos, esperamos que nuestra aproximación pueda aplicarse para inhibir específicamente la expresión de *CMYC* en CRC.

REFERENCES

1. Watson,J.D. and Crick,F.H.C. (1953) Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nat. 1953 1714361*, **171**, 964–967.

2. Bang, I. (1910) Untersuchungen über die Guanylsäure. *Biochem.,* **26**, 293–311.

3. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.*, **48**, 2013–2018.

4. Neidle,S. and Balasubramanian,S. (2006) Quadruplex Nucleic Acids. *RSC Biomol. Sci.*, **10**, 1039-1045.

5. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.

6. Singh,A. and Kukreti,S. (2018) A triple stranded G-quadruplex formation in the promoter region of human myosin β(Myh7) gene. *J. Biomol. Struct. Dyn.*, **36**, 2773–2786.

7. Li,X.M., Zheng,K.W., Zhang,J.Y., Liu,H.H., He,Y. De, Yuan,B.F., Hao,Y.H. and Tan,Z. (2015) Guanine-vacancy-bearing G-quadruplexes responsive to guanine derivatives. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 14581–14586.

8. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadraplexes. *Biochemistry*, **47**, 689–697.

9. Lombardi,E.P. and No-Vallejo,A.L. (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.*, **48**, 1603.

10. Varizhuk,A., Ischenko,D., Tsvetkov,V., Novikov,R., Kulemin,N., Kaluzhny,D., Vlasenok,M., Naumov,V., Smirnov,I. and Pozmogova,G. (2017) The expanding repertoire of G4 DNA structures. *Biochimie*, **135**, 54–62.

11. Mathad,R.I. and Yang,D. (2011) G-quadruplex structures and G-quadruplex-interactive compounds. *Methods Mol. Biol.*, **735**, 77–96.

12. del Villar-Guerra,R., Trent,J.O. and Chaires,J.B. (2018) G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **57**, 7171–7175.

13. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.

14. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.

15. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.

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

17. Dabrowiak,J.C., Goodisman,J. and Ward,B. (1997) Quantitative DNA footprinting. *Methods Mol. Biol.*, **90**, 23–42.

18. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci.*, **98**, 8572–8577.

19. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

20. Henderson,A., Wu,Y., Huang,Y.C., Chavez,E.A., Platt,J.,

Johnson,F.B., Brosh,R.M., Sen,D., Lansdorp,P.M. and Lansdorp,P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–9.

21. Liu,H.Y., Zhao,Q., Zhang,T.P., Wu,Y., Xiong,Y.X., Wang,S.K., Ge,Y.L., He,J.H., Lv,P., Ou,T.M., *et al.* (2016) Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.*, **23**, 1261–1270.

22. Di Antonio,M., Ponjavic,A., Radzevičius,A., Ranasinghe,R.T., Catalano,M., Zhang,X., Shen,J., Needham,L.M., Lee,S.F., Klenerman,D., *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.*, **12**, 832–837.

23. Lefebvre,J., Guetta,C., Poyer,F., Mahuteau-Betzer,F. and Teulade-Fichou,M.-P. (2017) Copper-Alkyne Complexation Responsible for the Nucleolar Localization of Quadruplex Nucleic Acid Drugs Labeled by Click Reactions. *Angew. Chemie*, **129**, 11523–11527.

24. Summers,P.A., Lewis,B.W., Gonzalez-Garcia,J., Porreca,R.M., Lim,A.H.M., Cadinu,P., Martin-Pintado,N., Mann,D.J., Edel,J.B., Vannier,J.B., *et al.* (2021) Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.*, **12**, 1–11.

25. Hänsel-Hertsch,R., Spiegel,J., Marsico,G., Tannahill,D. and Balasubramanian,S. (2018) Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.*, **13**, 551–564.

26. Law,M.J., Lower,K.M., Voon,H.P.J., Hughes,J.R., Garrick,D., Viprakasit,V., Mitson,M., De Gobbi,M., Marra,M., Morris,A., *et al.* (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*, **143**, 367–378.

27. Gray,L.T., Vallur,A.C., Eddy,J. and Maizels,N. (2014) G quadruplexes

are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol. 2014 104*, **10**, 313–318.

28. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nat. 1988 3346180*, **334**, 364–366.

29. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.

30. Shen,W., Gorelick,R.J. and Bambara,R.A. (2011) HIV-1 nucleocapsid protein increases strand transfer recombination by promoting dimeric G-quartet formation. *J. Biol. Chem.*, **286**, 29838–29847.

31. Cahoon,L.A. and Seifert,H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science*, **325**, 764–767.

32. Smargiasso,N., Gabelica,V., Damblon,C., Rosu,F., De Pauw,E., Teulade-Fichou,M.P., Rowe,J.A. and Claessens,A. (2009) Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes. *BMC Genomics*, **10**, 1–12.

33. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., Antonio,M. Di and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.

34. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.

35. Hänsel-Hertsch,R., Beraldi,D., Lensing,S. V, Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M., *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.

36. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

37. Mendoza,O., Bourdoncle,A., Boulé,J.B., Brosh,R.M. and Mergny,J.L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.

38. Zoroddu,M.A., Aaseth,J., Crisponi,G., Medici,S., Peana,M. and Nurchi,V.M. (2019) The essential metals for humans: a brief overview. *J. Inorg. Biochem.*, **195**, 120–129.

39. Selvam,S., Koirala,D., Yu,Z. and Mao,H. (2014) Quantification of topological coupling between DNA superhelicity and G-quadruplex formation. *J. Am. Chem. Soc.*, **136**, 13967–13970.

40. Shrestha,P., Jonchhe,S., Emura,T., Hidaka,K., Endo,M., Sugiyama,H. and Mao,H. (2017) Confined space facilitates G-quadruplex formation. *Nat. Nanotechnol. 2017 126*, **12**, 582–588.

41. Kouzine,F., Liu,J., Sanford,S., Chung,H.J. and Levens,D. (2004) The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol. 2004 1111*, **11**, 1092–1100.

42. Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618.

43. Cui,Y., Kong,D., Ghimire,C., Xu,C. and Mao,H. (2016) Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry*, **55**, 2291–2299.

44. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.

45. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and

Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol. 2020 218*, **21**, 459–474.

46. Robinson,J., Raguseo,F., Nuccio,S.P., Liano,D. and Di Antonio,M. (2021) DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.*, **49**, 8419–8431.

47. Webb,C.J., Wu,Y. and Zakian,V.A. (2013) DNA Repair at Telomeres: Keeping the Ends Intact. *Cold Spring Harb. Perspect. Biol.*, **5**.

48. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. *Cell*, **51**, 899–908.

49. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DMA structures. *Nature*, **350**, 718–720.

50. Moye,A.L., Porter,K.C., Cohen,S.B., Phan,T., Zyner,K.G., Sasaki,N., Lovrecz,G.O., Beck,J.L. and Bryan,T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**.

51. Monsen,R.C., DeLeeuw,L., Dean,W.L., Gray,R.D., Sabo,T.M., Chakravarthy,S., Chaires,J.B. and Trent,J.O. (2020) The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res.*, 10.1093/nar/gkaa107.

52. Puget,N., Miller,K.M. and Legube,G. (2019) Non-canonical DNA/RNA structures during Transcription-Coupled Double-Strand Break Repair: Roadblocks or Bona fide repair intermediates? *DNA Repair (Amst).*, **81**.

53. De,S. and Michor,F. (2011) DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.*, **18**, 950–955.

54. Georgakopoulos-Soares,I., Morganella,S., Jain,N., Hemberg,M. and Nik-Zainal,S. (2018) Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.*, **28**, 1264–1271.

55. Rodriguez,R., Miller,K.M., Forment,J. V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.

56. Wang,Y., Yang,J., Wild,A.T., Wu,W.H., Shah,R., Danussi,C., Riggins,G.J., Kannan,K., Sulman,E.P., Chan,T.A., *et al.* (2019) G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat. Commun. 2019 101*, **10**, 1–14.

57. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) 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.*, **99**, 11593–11598.

58. Gromak,N., West,S. and Proudfoot,N.J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol. Cell. Biol.*, **26**, 3986–3996.

59. Aguilera,A. and García-Muse,T. (2012) R Loops: From Transcription Byproducts to Threats to Genome Stability. *Mol. Cell*, **46**, 115–124.

60. Zheng,K.W., Xiao,S., Liu,J.Q., Zhang,J.Y., Hao,Y.H. and Tan,Z. (2013) Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.,* **41**, 5533–5541.

61. Skourti-Stathaki,K., Proudfoot,N.J. and Gromak,N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell*, **42**, 794–805.

62. Agarwal,T., Roy,S., Kumar,S., Chakraborty,T.K. and Maiti,S. (2014) In the sense of transcription regulation by G-quadruplexes: Asymmetric effects in sense and antisense strands. *Biochemistry*, **53**, 3711–3718.

63. Lee,C.Y., McNerney,C., Ma,K., Zhao,W., Wang,A. and Myong,S. (2020) R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat. Commun. 2020 111*, **11**, 1–15.

64. Du,Z., Zhao,Y. and Li,N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **18**, 233.

65. Hänsel-Hertsch,R., Simeone,A., Shea,A., Hui,W.W.I., Zyner,K.G., Marsico,G., Rueda,O.M., Bruna,A., Martin,A., Zhang,X., *et al.* (2020) Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat. Genet. 2020 529*, **52**, 878–883.

66. Raiber,E.A., Kranaster,R., Lam,E., Nikan,M. and Balasubramanian,S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.*, **40**, 1499–1508.

67. Cogoi,S., Paramasivam,M., Membrino,A., Yokoyama,K.K. and Xodo,L.E. (2010) The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.*, **285**, 22003–22016.

68. Li,L., Williams,P., Ren,W., Wang,M.Y., Gao,Z., Miao,W., Huang,M., Song,J. and Wang,Y. (2020) YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol. 2020 172*, **17**, 161–168.

69. Hou,Y., Li,F., Zhang,R., Li,S., Liu,H., Qin,Z.S. and Sun,X. (2019) Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics*, **14**, 894–911.

70. Williams,J.D., Houserova,D., Johnson,B.R., Dyniewski,B., Berroyer,A.,

French,H., Barchie,A.A., Bilbrey,D.D., Demeis,J.D., Ghee,K.R., *et al.* (2020) Characterization of long G4-rich enhancer-associated genomic regions engaging in a novel loop:loop 'G4 Kissing' interaction. *Nucleic Acids Res.*, **48**, 5907–5925.

71. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.

72. Cree,S.L., Fredericks,R., Miller,A., Pearce,F.G., Filichev,V., Fee,C. and Kennedy,M.A. (2016) DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett.*, **590**, 2870–2883.

73. Mao,S.Q., Ghanbarian,A.T., Spiegel,J., Martínez Cuesta,S., Beraldi,D., Di Antonio,M., Marsico,G., Hänsel-Hertsch,R., Tannahill,D. and Balasubramanian,S. (2018) DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol. 2018 2510*, **25**, 951–957.

74. Saha,D., Singh,A., Hussain,T., Srivastava,V., Sengupta,S., Kar,A., Dhapola,P., Dhople,V., Ummanni,R. and Chowdhury,S. (2017) Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J. Biol. Chem.*, **292**, 15205–15215.

75. Pavlova,I.I., Tsvetkov,V.B., Isaakova,E.A., Severov,V. V., Khomyakova,E.A., Lacis,I.A., Lazarev,V.N., Lagarkova,M.A., Pozmogova,G.E. and Varizhuk,A.M. (2020) Transcription-facilitating histone chaperons interact with genomic and synthetic G4 structures. *Int. J. Biol. Macromol.*, **160**, 1144–1157.

76. Steenken,S. and Jovanovic,S. V. (1997) How Easily Oxidizable Is DNA? One-Electron Reduction Potentials of Adenosine and Guanosine Radicals in Aqueous Solution. *J. Am. Chem. Soc.*, **119**, 617–618.

77. Fleming,A.M., Ding,Y. and Burrows,C.J. (2017) Oxidative DNA

damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 2604–2609.

78. Fleming,A.M., Zhu,J., Ding,Y. and Burrows,C.J. (2019) Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. *Nucleic Acids Res.*, **47**, 5049–5060.

79. Fleming,A.M., Zhu,J., Ding,Y. and Burrows,C.J. (2017) 8-Oxo-7,8-dihydroguanine in the Context of a Gene Promoter G-Quadruplex Is an On-Off Switch for Transcription. *ACS Chem. Biol.*, **12**, 2417–2426.

80. Edwards,A.D., Marecki,J.C., Byrd,A.K., Gao,J. and Raney,K.D. (2021) G-Quadruplex loops regulate PARP-1 enzymatic activation. *Nucleic Acids Res.*, **49**, 416–431.

81. Maizels,N. (2015) G4-associated human diseases. *EMBO Rep.*, **16**, 910–922.

82. Biffi,G., Tannahill,D., Miller,J., Howat,W.J. and Balasubramanian,S. (2014) Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLoS One*, **9**, e102711.

83. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: The next generation. *Cell*, **144**, 646–674.

84. Sanchez-Martin,V., Lopez-Pujante,C., Soriano-Rodriguez,M. and Garcia-Salcedo,J.A. (2020) An updated focus on quadruplex structures as potential therapeutic targets in cancer. *Int. J. Mol. Sci.*, **21**, 1–24.

85. Maciejowski,J. and De Lange,T. (2017) Telomeres in cancer: Tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.*, **18**, 175–186.

86. Matsui,A., Ihara,T., Suda,H., Mikami,H. and Semba,K. (2013) Gene amplification: Mechanisms and involvement in cancer. *Biomol. Concepts*, **4**, 567–582.

87. Neidle,S. (2017) Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.*, **1**, 1–10.

88. Asamitsu,S., Bando,T. and Sugiyama,H. (2019) Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chem. - A Eur. J.*, **25**, 417–430.

89. Sanchez-Martin,V., Soriano,M. and Garcia-Salcedo,J.A. (2021) Quadruplex ligands in cancer therapy. *Cancers (Basel).*, **13**.

90. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–75.

91. Sun,D., Thompson,B., Cathers,B.E., Salazar,M., Kerwin,S.M., Trent,J.O., Jenkins,T.C., Neidle,S. and Hurley,L.H. (1997) Inhibition of human telomerase by a G-Quadruplex-Interactive compound. *J. Med. Chem.*, **40**, 2113–2116.

92. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.

93. Marchetti,C., Zyner,K.G., Ohnmacht,S.A., Robson,M., Haider,S.M., Morton,J.P., Marsico,G., Vo,T., Laughlin-Toth,S., Ahmed,A.A., *et al.* (2018) Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. *J. Med. Chem.*, **61**, 2500–2517.

94. Huang,A., Garraway,L.A., Ashworth,A. and Weber,B. (2019) Synthetic lethality as an engine for cancer drug target discovery. *Nat. Rev. Drug Discov. 2019 191*, **19**, 23–38.

95. Xu,H., Di Antonio,M., McKinney,S., Mathew,V., Ho,B., O'Neil,N.J., Santos,N. Dos, Silvester,J., Wei,V., Garcia,J., *et al.* (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.*, **8**, 14432.

96. McLuckie,K.I.E., Di Antonio,M., Zecchini,H., Xian,J., Caldas,C., Krippendorff,B.F., Tannahill,D., Lowe,C. and Balasubramanian,S. (2013) G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.*, **135**, 9640–9643.

97. Marchetti,C., Minarini,A., Tumiatti,V., Moraca,F., Parrotta,L., Alcaro,S., Rigo,R., Sissi,C., Gunaratnam,M., Ohnmacht,S.A., *et al.* (2015) Macrocyclic naphthalene diimides as G-quadruplex binders. *Bioorganic Med. Chem.*, **23**, 3819–3830.

98. Cuenca,F., Greciano,O., Gunaratnam,M., Haider,S., Munnur,D., Nanjunda,R., Wilson,W.D. and Neidle,S. (2008) Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorganic Med. Chem. Lett.*, **18**, 1668–1673.

99. Răsădean,D.M., Sheng,B., Dash,J. and Pantoş,G.D. (2017) Amino-Acid-Derived Naphthalenediimides as Versatile G-Quadruplex Binders. *Chem. - A Eur. J.*, **23**, 8491–8499.

100. Gunaratnam,M., Collie,G.W., Reszka,A.P., Todd,A.K., Parkinson,G.N. and Neidle,S. (2018) A naphthalene diimide G-quadruplex ligand inhibits cell growth and down-regulates BCL-2 expression in an imatinib-resistant gastrointestinal cancer cell line. *Bioorganic Med. Chem.*, **26**, 2958–2964.

101. Lago,S., Nadai,M., Ruggiero,E., Tassinari,M., Marušič,M., Tosoni,B., Frasson,I., Cernilogar,F.M., Pirota,V., Doria,F., *et al.* (2021) The MDM2 inducible promoter folds into four-tetrad antiparallel G-quadruplexes targetable to fight malignant liposarcoma . *Nucleic Acids Res.*, **49**, 847–863.

102. Ohnmacht,S.A., Marchetti,C., Gunaratnam,M., Besser,R.J., Haider,S.M., Di Vita,G., Lowe,H.L., Mellinas-Gomez,M., Diocou,S., Robson,M., *et al.* (2015) A G-quadruplex-binding compound showing anti-tumour activity in an in vivo model for pancreatic cancer. *Sci. Rep.*, **5**, 1–11.

103. Kim,M.Y., Duan,W., Gleason-Guzman,M. and Hurley,L.H. (2003) Design, synthesis, and biological evaluation of a series of fluoroquinoanthroxazines with contrasting dual mechanisms of action against topoisomerase II and G-quadruplexes. *J. Med. Chem.*, **46**, 571–583.

104. Drygin,D., Siddiqui-Jain,A., O'Brien,S., Schwaebe,M., Lin,A., Bliesath,J., Ho,C.B., Proffitt,C., Trent,K., Whitten,J.P., *et al.* (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res.*, **69**, 7653–7661.

105. Drygin,D., Lin,A., Bliesath,J., Ho,C.B., O'Brien,S.E., Proffitt,C., Omori,M., Haddach,M., Schwaebe,M.K., Siddiqui-Jain,A., *et al.* (2011) Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res.*, **71**, 1418–1430.

106. Jackobel,A.J., Zeberl,B.J., Glover,D.M., Fakhouri,A.M. and Knutson,B.A. (2019) DNA binding preferences of S. cerevisiae RNA polymerase I Core Factor reveal a preference for the GC-minor groove and a conserved binding mechanism. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1862**, 194408.

107. Sung,H., Ferlay,J., Siegel,R.L., Laversanne,M., Soerjomataram,I., Jemal,A. and Bray,F. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.*, **71**, 209–249.

108. Rawla,P., Sunkara,T. and Barsouk,A. (2019) Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Prz. Gastroenterol.*, **14**, 89–103.

109. Schoen,R.E., Razzak,A., Yu,K.J., Berndt,S.I., Firl,K., Riley,T.L. and Pinsky,P.F. (2015) Incidence and Mortality of Colorectal Cancer in Individuals with a Family History of Colorectal Cancer. *Gastroenterology*, **149**, 1438-1445.e1.

110. Czene,K., Lichtenstein,P. and Hemminki,K. (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int. J. Cancer*, **99**, 260–266.

111. Jiao,S., Peters,U., Berndt,S., Brenner,H., Butterbach,K., Caan,B.J., Carlson,C.S., Chan,A.T., Chang-Claude,J., Chanock,S., *et al.* (2014) Estimating the heritability of colorectal cancer. *Hum. Mol. Genet.*, **23**, 3898–3905.

112. Syngal,S., Brand,R.E., Church,J.M., Giardiello,F.M., Hampel,H.L. and Burt,R.W. (2015) ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am. J. Gastroenterol.*, **110**, 223–262.

113. Vasen,H.F.A., Tomlinson,I. and Castells,A. (2015) Clinical management of hereditary colorectal cancer syndromes. *Nat. Rev. Gastroenterol. Hepatol.*, **12**, 88–97.

114. Dekker,E., Tanis,P.J., Vleugels,J.L.A., Kasi,P.M. and Wallace,M.B. (2019) Colorectal cancer. *Lancet*, **394**, 1467–1480.

115. Jess,T., Rungoe,C. and Peyrin-Biroulet,L. (2012) Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clin. Gastroenterol. Hepatol.*, **10**, 639–645.

116. Cottet,V., Jooste,V., Fournel,I., Bouvier,A.M., Faivre,J. and Bonithon-Kopp,C. (2012) Long-term risk of colorectal cancer after adenoma removal: a population-based cohort study. *Gut*, **61**, 1180–1186.

117. Oruc,Z. and Kaplan,M.A. (2019) Effect of exercise on colorectal cancer prevention and treatment. *World J. Gastrointest. Oncol.*, **11**, 348–366.

118. Botteri,E., Iodice,S., Bagnardi,V., Raimondi,S., Lowenfels,A.B. and Maisonneuve,P. (2008) Smoking and Colorectal Cancer. *Jama*, **300**, 2765.

119. Cai,S., Li,Y., Ding,Y., Chen,K. and Jin,M. (2014) Alcohol drinking and the risk of colorectal cancer death: A meta-analysis. *Eur. J. Cancer Prev.*, **23**, 532–539.

120. Chan,D.S.M., Lau,R., Aune,D., Vieira,R., Greenwood,D.C., Kampman,E. and Norat,T. (2011) Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. *PLoS One*, **6**.

121. Song Mingyang, Garrett Wendy S,C.A.T. (2015) Nutrients, foods, and colorectal cancer prevention. *Gastroenterology*, **148**, 1244–60.

122. Botteri,E., Støer,N.C., Sakshaug,S., Graff-Iversen,S., Vangen,S., Hofvind,S., De Lange,T., Bagnardi,V., Ursin,G. and Weiderpass,E. (2017) Menopausal hormone therapy and colorectal cancer: a linkage between nationwide registries in Norway. *BMJ Open*, **7**.

123. Dobrzycka,M., Spychalski,P., Łachiński,A.J., Kobiela,P., Jędrusik,P. and Kobiela,J. (2020) Statins and Colorectal Cancer - A Systematic Review. *Exp. Clin. Endocrinol. Diabetes*, **128**, 255–262.

124. Bibbins-Domingo,K., Grossman,D.C., Curry,S.J., Davidson,K.W., Epling,J.W., García,F.A.R., Gillman,M., Harper,D.M., Kemper,A.R., Krist,A.H., *et al.* (2016) Aspirin Use for the Primary Prevention of Cardiovascular Disease and Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann. Intern. Med.*, **164**, 836–845.

125. Aleksandrova,K., Pischon,T., Jenab,M., Bueno-de-Mesquita,H.B., Fedirko,V., Norat,T., Romaguera,D., Knüppel,S., Boutron-Ruault,M.C., Dossus,L., *et al.* (2014) Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. *BMC Med.*, **12**.

126. Cardoso,R., Guo,F., Heisser,T., Hackl,M., Ihle,P., De Schutter,H., Van Damme,N., Valerianova,Z., Atanasov,T., Májek,O., *et al.* (2021) Colorectal cancer incidence, mortality, and stage distribution in

European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol.*, **22**, 1002–1013.

127. Hewitson,P., Glasziou,P., Watson,E., Towler,B. and Irwig,L. (2008) Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *Am. J. Gastroenterol.*, **103**, 1541–1549.

128. Miller,E.A., Pinsky,P.F., Schoen,R.E., Prorok,P.C. and Church,T.R. (2019) Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: long-term follow-up of the randomised US PLCO cancer screening trial. *lancet. Gastroenterol. Hepatol.*, **4**, 101–110.

129. Brenner,H., Stock,C. and Hoffmeister,M. (2014) Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ*, **348**.

130. Zeki,S.S., Graham,T.A. and Wright,N.A. (2011) Stem cells and their implications for colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.*, **8**, 90–100.

131. Kuipers,E.J., Grady,W.M., Lieberman,D., Seufferlein,T., Sung,J.J., Boelens,P.G., Van De Velde,C.J.H. and Watanabe,T. (2015) Colorectal cancer. *Nat. Rev. Dis. Prim.*, **1**, 1–25.

132. Muzny,D.M., Bainbridge,M.N., Chang,K., Dinh,H.H., Drummond,J.A., Fowler,G., Kovar,C.L., Lewis,L.R., Morgan,M.B., Newsham,I.F., *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330–337.

133. Huang,D., Sun,W., Zhou,Y., Li,P., Chen,F., Chen,H., Xia,D., Xu,E., Lai,M., Wu,Y., *et al.* (2018) Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer Metastasis Rev.*, **37**, 173–187.

134. Erisman,M.D., Rothberg,P.G., Diehl,R.E., Morse,C.C.,

Spandorfer,J.M. and Astrin,S.M. (1985) Deregulation of c-myc gene expression in human colon carcinoma is not accompanied by amplification or rearrangement of the gene. *Mol. Cell. Biol.*, **5**, 1969–1976.

135. Worthley,D.L., Whitehall,V.L.J., Buttenshaw,R.L., Irahara,N., Greco,S.A., Ramsnes,I., Mallitt,K.A., Le Leu,R.K., Winter,J., Hu,Y., *et al.* (2010) DNA methylation within the normal colorectal mucosa is associated with pathway-specific predisposition to cancer. *Oncogene*, **29**, 1653–1662.

136. Yamauchi,M., Morikawa,T., Kuchiba,A., Imamura,Y., Qian,Z.R., Nishihara,R., Liao,X., Waldron,L., Hoshida,Y., Huttenhower,C., *et al.* (2012) Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut*, **61**, 847–854.

137. Kostic,A.D., Gevers,D., Pedamallu,C.S., Michaud,M., Duke,F., Earl,A.M., Ojesina,A.I., Jung,J., Bass,A.J., Tabernero,J., *et al.* (2012) Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.*, **22**, 292–298.

138. Hong,S.N. (2018) Genetic and epigenetic alterations of colorectal cancer. *Intest. Res.*, **16**, 327–337.

139. Fijten,G.H., Starmans,R., Muris,J.W., Schouten,H.J., Blijham,G.H. and Knottnerus,J.A. (1995) Predictive value of signs and symptoms for colorectal cancer in patients with rectal bleeding in general practice. *Fam. Pract.*, **12**, 279–286.

140. Kaminski,M.F., Regula,J., Kraszewska,E., Polkowski,M., Wojciechowska,U., Didkowska,J., Zwierko,M., Rupinski,M., Nowacki,M.P. and Butruk,E. (2010) Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *N. Engl. J. Med.*, **362**, 1795–1803.

141. Beets-Tan,R.G.H., Lambregts,D.M.J., Maas,M., Bipat,S., Barbaro,B.,

Curvo-Semedo,L., Fenlon,H.M., Gollub,M.J., Gourtsoyianni,S., Halligan,S., *et al.* (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur. Radiol.*, **28**, 1465–1475.

142. Church,T.R., Wandell,M., Lofton-Day,C., Mongin,S.J., Burger,M., Payne,S.R., Castaños-Vélez,E., Blumenstein,B.A., Rösch,T., Osborn,N., *et al.* (2014) Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*, **63**, 317–325.

143. Ahlquist,D.A., Sargent,D.J., Loprinzi,C.L., Levin,T.R., Rex,D.K., Ahnen,D.J., Knigge,K., Lance,M.P., Burgart,L.J., Hamilton,S.R., *et al.* (2008) Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Ann. Intern. Med.*, **149**, 441–450.

144. Imperiale,T.F., Ransohoff,D.F., Itzkowitz,S.H., Levin,T.R., Lavin,P., Lidgard,G.P., Ahlquist,D.A. and Berger,B.M. (2014) Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *N. Engl. J. Med.*, **370**, 1287–1297.

145. Jayanna,M., Burgess,N.G., Singh,R., Hourigan,L.F., Brown,G.J., Zanati,S.A., Moss,A., Lim,J., Sonson,R., Williams,S.J., *et al.* (2016) Cost Analysis of Endoscopic Mucosal Resection vs Surgery for Large Laterally Spreading Colorectal Lesions. *Clin. Gastroenterol. Hepatol.*, **14**, 271-278.e2.

146. Abdel-Rahman,O. and Cheung,W.Y. (2018) Integrating Systemic Therapies into the Multimodality Treatment of Resectable Colorectal Liver Metastases. *Gastroenterol. Res. Pract.*, **2018**.

147. Glynne-Jones,R., Wyrwicz,L., Tiret,E., Brown,G., Rödel,C., Cervantes,A. and Arnold,D. (2017) Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.*, **28**, iv22–iv40.

148. Rullier,E., Rouanet,P., Tuech,J.J., Valverde,A., Lelong,B., Rivoire,M., Faucheron,J.L., Jafari,M., Portier,G., Meunier,B., *et al.* (2017) Organ preservation for rectal cancer (GRECCAR 2): a prospective, randomised, open-label, multicentre, phase 3 trial. *Lancet (London, England)*, **390**, 469–479.

149. Vodenkova,S., Buchler,T., Cervena,K., Veskrnova,V., Vodicka,P. and Vymetalkova,V. (2020) 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacol. Ther.*, **206**.

150. Haller,D.G., Tabernero,J., Maroun,J., De Braud,F., Price,T., Van Cutsem,E., Hill,M., Gilberg,F., Rittweger,K. and Schmoll,H.J. (2011) Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer. *J. Clin. Oncol.*, **29**, 1465–1471.

151. Arango,D., Wilson,A.J., Shi,Q., Corner,G.A., Arañes,M.J., Nicholas,C., Lesser,M., Mariadason,J.M. and Augenlicht,L.H. (2004) Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *Br. J. Cancer*, **91**, 1931–1946.

152. Douillard,J.Y., Cunningham,D., Roth,A.D., Navarro,M., James,R.D., Karasek,P., Jandik,P., Iveson,T., Carmichael,J., Alakl,M., *et al.* (2000) Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet*, **355**, 1041–1047.

153. Fuchs,C., Mitchell,E.P. and Hoff,P.M. (2006) Irinotecan in the treatment of colorectal cancer. *Cancer Treat. Rev.*, **32**, 491–503.

154. Piawah,S. and Venook,A.P. (2019) Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer*, **125**, 4139–4147.

155. Li,Q.H., Wang,Y.Z., Tu,J., Liu,C.W., Yuan,Y.J., Lin,R., He,W.L., Cai,S.R., He,Y.L. and Ye,J.N. (2020) Anti-EGFR therapy in metastatic

colorectal cancer: mechanisms and potential regimens of drug resistance. *Gastroenterol. Rep.*, **8**, 179–191.

156. Seeber,A., Gunsilius,E., Gastl,G. and Pircher,A. (2018) Anti-Angiogenics: Their Value in Colorectal Cancer Therapy. *Oncol. Res. Treat.*, **41**, 188–193.

157. Adenis,A., de la Fouchardiere,C., Paule,B., Burtin,P., Tougeron,D., Wallet,J., Dourthe,L.M., Etienne,P.L., Mineur,L., Clisant,S., *et al.* (2016) Survival, safety, and prognostic factors for outcome with Regorafenib in patients with metastatic colorectal cancer refractory to standard therapies: results from a multicenter study (REBECCA) nested within a compassionate use program. *BMC Cancer*, **16**.

158. Mayer,R.J., Van Cutsem,E., Falcone,A., Yoshino,T., Garcia-Carbonero,R., Mizunuma,N., Yamazaki,K., Shimada,Y., Tabernero,J., Komatsu,Y., *et al.* (2015) Randomized Trial of TAS-102 for Refractory Metastatic Colorectal Cancer. *N. Engl. J. Med.*, **372**, 1909–1919.

159. Roth,A.D., Tejpar,S., Delorenzi,M., Yan,P., Fiocca,R., Klingbiel,D., Dietrich,D., Biesmans,B., Bodoky,G., Barone,C., *et al.* (2010) Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J. Clin. Oncol.*, **28**, 466–474.

160. Kopetz,S., Grothey,A., Yaeger,R., Van Cutsem,E., Desai,J., Yoshino,T., Wasan,H., Ciardiello,F., Loupakis,F., Hong,Y.S., *et al.* (2019) Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer . *N. Engl. J. Med.*, **381**, 1632–1643.

161. Le,D.T., Uram,J.N., Wang,H., Bartlett,B.R., Kemberling,H., Eyring,A.D., Skora,A.D., Luber,B.S., Azad,N.S., Laheru,D., *et al.* (2015) PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N. Engl. J. Med.*, **372**, 2509–2520.

162. Wahle,K.W.J., Brown,I., Rotondo,D. and Heys,S.D. (2010) Plant phenolics in the prevention and treatment of cancer. *Adv. Exp. Med.*

Biol., **698**, 36–51.

163. Rice-Evans,C.A., Miller,N.J. and Paganga,G. (1997) Antioxidant properties of phenolic compounds. *Trends Plant Sci.*, **2**, 152–159.

164. Tsao,R. (2010) Chemistry and biochemistry of dietary polyphenols. *Nutrients*, **2**, 1231–1246.

165. Ls,R., Nja,S., Ncp,S., Mc,M. and Aj,T. (2016) Anticancer Properties of Phenolic Acids in Colon Cancer – A Review. *J. Nutr. Food Sci.*, **6**.

166. Anantharaju,P.G., Gowda,P.C., Vimalambike,M.G. and Madhunapantula,S. V. (2016) An overview on the role of dietary phenolics for the treatment of cancers. *Nutr. J.*, **15**, 1–16.

167. Mueller-Harvey,I., Feucht,W., Polster,J., Trnková,L., Burgos,P., Parker,A.W. and Botchway,S.W. (2012) Two-photon excitation with pico-second fluorescence lifetime imaging to detect nuclear association of flavanols. *Anal. Chim. Acta*, **719**, 68–75.

168. S,M., F,B., C,B. and M,C. (1993) Conventional antibodies: requirements and methods for their optimization. *Int. J. Biol. Markers*, **8**, 143–146.

169. Muyldermans,S., Baral,T.N., Retamozzo,V.C., De Baetselier,P., De Genst,E., Kinne,J., Leonhardt,H., Magez,S., Nguyen,V.K., Revets,H., *et al.* (2009) Camelid immunoglobulins and nanobody technology. *Vet. Immunol. Immunopathol.*, **128**, 178–183.

170. Hamers-Casterman,C., Atarhouch,T., Muyldermans,S., Robinson,G., Hammers,C., Songa,E.B., Bendahman,N. and Hammers,R. (1993) Naturally occurring antibodies devoid of light chains. *Nature*, **363**, 446–448.

171. Unciti-Broceta,J.D., Del Castillo,T., Soriano,M., Magez,S. and Garcia-Salcedo,J.A. (2013) Novel therapy based on camelid nanobodies. *Ther. Deliv.*, **4**, 1321–1336.

172. Harmsen,M.M., Van Solt,C.B., Van Zijderveld-Van Bemmel,A.M.,

Niewold,T.A. and Van Zijderveld,F.G. (2006) Selection and optimization of proteolytically stable llama single-domain antibody fragments for oral immunotherapy. *Appl. Microbiol. Biotechnol.*, **72**, 544–551.

173. Vincke,C., Loris,R., Saerens,D., Martinez-Rodriguez,S., Muyldermans,S. and Conrath,K. (2009) General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J. Biol. Chem.*, **284**, 3273–3284.

174. Arévalo-Ruiz,M., Doria,F., Belmonte-Reche,E., De Rache,A., Campos-Salinas,J., Lucas,R., Falomir,E., Carda,M., Pérez-Victoria,J.M., Mergny,J.L., *et al.* (2017) Synthesis, Binding Properties, and Differences in Cell Uptake of G-Quadruplex Ligands Based on Carbohydrate Naphthalene Diimide Conjugates. *Chem. - A Eur. J.*, **23**, 2157–2164.

175. Zuffo,M., Stucchi,A., Campos-Salinas,J., Cabello-Donayre,M., Martínez-García,M., Belmonte-Reche,E., Pérez-Victoria,J.M., Mergny,J.L., Freccero,M., Morales,J.C., *et al.* (2019) Carbohydrate-naphthalene diimide conjugates as potential antiparasitic drugs: Synthesis, evaluation and structure-activity studies. *Eur. J. Med. Chem.*, **163**, 54–66.

176. Karanasios,E., Stapleton,E., Manifava,M. and Ktistakis,N.T. (2014) Imaging autophagy. *Curr. Protoc. Cytom.*, 10.1002/0471142956.cy1234s69.

177. Yan,W.X., Mirzazadeh,R., Garnerone,S., Scott,D., Schneider,M.W., Kallas,T., Custodio,J., Wernersson,E., Li,Y., Gao,L., *et al.* (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.*, **8**, 1–9.

178. Winston X. Yan, Reza Mirzazadeh, Silvano Garnerone, David Scott, Martin W. Schneider, Tomasz Kallas, Joaquin Custodio, Erik Wernersson, Yinqing Li, Linyi Gao, Yana Federova, Bernd Zetsche, Feng Zhang,M.B.& N.C. (2017) Breaks labeling in situ and

sequencing: Protocol exchange. *Protoc. Exch.*, 10.1038/protex.2017.018.

179. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, **16**, 10881–10890.

180. Schmidt,D., Wilson,M.D., Spyrou,C., Brown,G.D., Hadfield,J. and Odom,D.T. (2009) ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods*, **48**, 240–248.

181. Wei,T., Najmi,S.M., Liu,H., Peltonen,K., Kucerova,A., Schneider,D.A. and Laiho,M. (2018) Small-Molecule Targeting of RNA Polymerase I Activates a Conserved Transcription Elongation Checkpoint. *Cell Rep.*, **23**, 404–414.

182. Denissov,S., Lessard,F., Mayer,C., Stefanovsky,V., Van Driel,M., Grummt,I., Moss,T. and Stunnenberg,H.G. (2011) A model for the topology of active ribosomal RNA genes. *EMBO Rep.*, **12**, 231–237.

183. Tawani,A., Mishra,S.K. and Kumar,A. (2017) Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. *Sci. Rep.*, **7**, 1–13.

184. Keener,J., Josaitis,C.A., Dodd,J.A. and Nomura,M. (1998) Reconstitution of yeast RNA polymerase I transcription in vitro from purified components: TATA-binding protein is not required for basal transcription. *J. Biol. Chem.*, **273**, 33795–33802.

185. Bedwell,G.J., Appling,F.D., Anderson,S.J. and Schneider,D.A. (2012) Efficient transcription by RNA polymerase I using recombinant core factor. *Gene*, **492**, 94–99.

186. Zhang,S., Sun,H., Chen,H., Li,Q., Guan,A., Wang,L., Shi,Y., Xu,S., Liu,M. and Tang,Y. (2018) Direct visualization of nucleolar G-quadruplexes in live cells by using a fluorescent light-up probe. *Biochim. Biophys. Acta - Gen. Subj.*, **1862**, 1101–1106.

187. Peris-Torres,C., Plaza-Calonge,M.D.C., López-Domínguez,R.,

Domínguez-García,S., Barrientos-Durán,A., Carmona-Sáez,P. and Rodríguez-Manzaneque,J.C. (2020) Extracellular protease adamts1 is required at early stages of human uveal melanoma development by inducing stemness and endothelial-like features on tumor cells. *Cancers (Basel).*, **12**, 1–20.

188. Comor,L., Dolinska,S., Bhide,K., Pulzova,L., Jiménez-Munguía,I., Bencurova,E., Flachbartova,Z., Potocnakova,L., Kanova,E. and Bhide,M. (2017) Joining the in vitro immunization of alpaca lymphocytes and phage display: rapid and cost effective pipeline for sdAb synthesis. *Microb. Cell Fact.*, **16**, 13.

189. Calabrese,D.R., Chen,X., Leon,E.C., Gaikwad,S.M., Phyo,Z., Hewitt,W.M., Alden,S., Hilimire,T.A., He,F., Michalowski,A.M., *et al.* (2018) Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.*, **9**, 1–15.

190. Chua,F., Oh,S.K.W., Yap,M. and Teo,W.K. (1994) Enhanced IgG production in eRDF media with and without serum. A comparative study. *J. Immunol. Methods*, **167**, 109–119.

191. L. A. Day and R. Luke Wiseman (1978) A Comparison of DNA Packaging in the Virions of fd, Xf, and Pf1. *Cold Spring Harb. Monogr. Arch.*, **08**, 605–625.

192. Ghandi,M., Huang,F.W., Jané-Valbuena,J., Kryukov,G. V., Lo,C.C., McDonald,E.R., Barretina,J., Gelfand,E.T., Bielski,C.M., Li,H., *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, **569**, 503–508.

193. Rhodes,D.R., Yu,J., Shanker,K., Deshpande,N., Varambally,R., Ghosh,D., Barrette,T., Pandey,A. and Chinnaiyan,A.M. (2004) ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia*, **6**, 1–6.

194. Ye,J., Coulouris,G., Zaretskaya,I., Cutcutache,I., Rozen,S. and Madden,T.L. (2012) Primer-BLAST: A tool to design target-specific

primers for polymerase chain reaction. *BMC Bioinformatics*, **13**.

195. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, 676–682.

196. Davis,P.K., Ho,A. and Dowdy,S.F. (2001) Cell-Cycle Synchronization of Mammalian Cells. *Biol. Methods*, **30**, 1322–1330.

197. González,V., Guo,K., Hurley,L. and Sun,D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.

198. Kaiser,C.E., Van Ert,N.A., Agrawal,P., Chawla,R., Yang,D. and Hurley,L.H. (2017) Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.*, **139**, 8522–8536.

199. Chang,L., Chang,M., Chang,H.M. and Chang,F. (2017) Expending Role of Microsatellite Instability in Diagnosis and Treatment of Colorectal Cancers. *J. Gastrointest. Cancer*, **48**, 305–313.

200. Iacovoni,J.S., Caron,P., Lassadi,I., Nicolas,E., Massip,L., Trouche,D. and Legube,G. (2010) High-resolution profiling of γh2AX around DNA double strand breaks in the mammalian genome. *EMBO J.*, **29**, 1446–1457.

201. Musso,L., Mazzini,S., Rossini,A., Castagnoli,L., Scaglioni,L., Artali,R., Di Nicola,M., Zunino,F. and Dallavalle,S. (2018) c-MYC G-quadruplex binding by the RNA polymerase I inhibitor BMH-21 and analogues revealed by a combined NMR and biochemical Approach. *Biochim. Biophys. Acta - Gen. Subj.*, **1862**, 615–629.

202. Xu,H., Di Antonio,M., McKinney,S., Mathew,V., Ho,B., O'Neil,N.J., Santos,N. Dos, Silvester,J., Wei,V., Garcia,J., *et al.* (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.*, **8**.

203. Hostettmann,K. (1991) Assays for Bioactivity. Method in Plant Biochemistry. *New Phytol.*, **6**, 474.

204. Hernandez-Verdun,D. (2006) Nucleolus: From structure to dynamics. *Histochem. Cell Biol.*, **125**, 127–137.

205. Katagiri,N., Kuroda,T., Kishimoto,H., Hayashi,Y., Kumazawa,T. and Kimura,K. (2015) The nucleolar protein nucleophosmin is essential for autophagy induced by inhibiting Pol i transcription. *Sci. Rep.*, **5**, 1–9.

206. Popov,A., Smirnov,E., Kováčik,L., Raška,O., Hagen,G., Stixová,L. and Raška,I. (2013) Duration of the first steps of the human rRNA processing. *Nucl. (United States)*, **4**, 134–141.

207. Gorski,S.A., Snyder,S.K., John,S., Grummt,I. and Misteli,T. (2008) Modulation of RNA Polymerase Assembly Dynamics in Transcriptional Regulation. *Mol. Cell*, **30**, 486–497.

208. Ganapathy,V., Thangaraju,M. and Prasad,P.D. (2009) Nutrient transporters in cancer: Relevance to Warburg hypothesis and beyond. *Pharmacol. Ther.*, **121**, 29–40.

209. Siebeneicher,H., Cleve,A., Rehwinkel,H., Neuhaus,R., Heisler,I., Müller,T., Bauser,M. and Buchmann,B. (2016) Identification and Optimization of the First Highly Selective GLUT1 Inhibitor BAY-876. *ChemMedChem*, **11**, 2261–2271.

210. Liu,Y., Cao,Y., Zhang,W., Bergmeier,S., Qian,Y., Akbar,H., Colvin,R., Ding,J., Tong,L., Wu,S., *et al.* (2012) A small-molecule inhibitor of glucose transporter 1 downregulates glycolysis, induces cell-cycle arrest, and inhibits cancer cell growth in vitro and in vivo. *Mol. Cancer Ther.*, **11**, 1672–1682.

211. Largy,E., Hamon,F. and Teulade-Fichou,M.P. (2011) Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.*, **400**, 3419–3427.

212. Pelletier,J., Thomas,G. and Volarevic,S. (2017) Ribosome biogenesis in cancer: players and therapeutic avenues. *Nat. Publ. Gr.*, **18**, 51–63.

213. Neveu,V., Perez-Jiménez,J., Vos,F., Crespy,V., du Chaffaut,L., Mennen,L., Knox,C., Eisner,R., Cruz,J., Wishart,D., *et al.* (2010) Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database (Oxford).*, **2010**, 1–9.

214. Sanchez-Martin,V., Schneider,D.A., Ortiz-Gonzalez,M., Soriano-Lerma,A., Linde-Rodriguez,A., Perez-Carrasco,V., Gutierrez-Fernandez,J., Cuadros,M., González,C., Soriano,M., *et al.* (2021) Targeting ribosomal G-quadruplexes with naphthalene-diimides as RNA polymerase I inhibitors for colorectal cancer treatment. *Cell Chem. Biol.*, 10.1016/j.chembiol.2021.05.021.

215. Agrawal,P., Lin,C., Mathad,R.I., Carver,M. and Yang,D. (2014) The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in k+ solution. *J. Am. Chem. Soc.*, **136**, 1750–1753.

216. Sun,D., Liu,W.J., Guo,K., Rusche,J.J., Ebbinghaus,S., Gokhale,V. and Hurley,L.H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol. Cancer Ther.*, **7**, 880–889.

217. Phan,A.T. (2010) Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.

218. Yadav,K., Meka,P.N.R., Sadhu,S., Guggilapu,S.D., Kovvuri,J., Kamal,A., Srinivas,R., Devayani,P., Babu,B.N. and Nagesh,N. (2017) Telomerase Inhibition and Human Telomeric G-Quadruplex DNA Stabilization by a β-Carboline-Benzimidazole Derivative at Low Concentrations. *Biochemistry*, **56**, 4392–4404.

219. Hasima,N. and Ozpolat,B. (2014) Regulation of autophagy by polyphenolic compounds as a potential therapeutic strategy for

cancer. *Cell Death Dis.*, **5**, 1–13.

220. Kiruthiga,C., Devi,K.P., Nabavi,S.M. and Bishayee,A. (2020) Autophagy: A potential therapeutic target of polyphenols in hepatocellular carcinoma. *Cancers (Basel).*, **12**, 1–31.

221. Zhou,W.J., Deng,R., Zhang,X.Y., Feng,G.K., Gu,L.Q. and Zhu,X.F. (2009) G-quadruplex ligand SYUIQ-5 induces autophagy by telomere damage and TRF2 delocalization in cancer cells. *Mol. Cancer Ther.*, **8**, 3203–3213.

222. Lyndsay Murrow and Jayanta Debnath (2013) Autophagy As A Stress Response And Quality Control Mechanism—Implications for Cell Injury and Human Disease. *Annu. Rev. Pathol.*, **8**, 105–137.

223. Shionoya,Y., Kanaseki,T., Miyamoto,S., Tokita,S., Hongo,A., Kikuchi,Y., Kochin,V., Watanabe,K., Horibe,R., Saijo,H., *et al.* (2017) Loss of tapasin in human lung and colon cancer cells and escape from tumor-associated antigen-specific CTL recognition. *Oncoimmunology*, **6**, 1–11.

224. Phan,A.N.H., Hua,T.N.M., Kim,M. kyu, Vo,V.T.A., Choi,J.W., Kim,H.W., Rho,J.K., Kim,K.W. and Jeong,Y. (2016) Gallic acid inhibition of Src-Stat3 signaling overcomes acquired resistance to EGF receptor tyrosine kinase inhibitors in advanced non-small cell lung cancer. *Oncotarget*, **7**, 54702–54713.

225. Drygin,D., Rice,W.G. and Grummt,I. (2010) The RNA Polymerase I Transcription Machinery: An Emerging Target for the Treatment of Cancer. *Annu. Rev. Pharmacol. Toxicol.*, **50**, 131–156.

226. Rochlitz,C.F., Herrmann,R. and De Kant,E. (1996) Overexpression and amplification of c-myc during progression of human colorectal cancer. *Oncology*, **53**, 448–454.

227. Spiegel,J., Adhikari,S. and Balasubramanian,S. (2020) The Structure and Function of DNA G-Quadruplexes. *Trends Chem.*, **2**, 123–136.

228. Lago,S., Tosoni,E., Nadai,M., Palumbo,M. and Richter,S.N. (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.*, **1861**, 1371–1381.

229. Brosh,R.M. (2013) DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer*, **13**, 542–558.

230. Bradner,J.E., Hnisz,D. and Young,R.A. (2017) Transcriptional Addiction in Cancer. *Cell*, **168**, 629–643.

231. Brignole,C., Bensa,V., Fonseca,N.A., Del Zotto,G., Bruno,S., Cruz,A.F., Malaguti,F., Carlini,B., Morandi,F., Calarco,E., *et al.* (2021) Cell surface Nucleolin represents a novel cellular target for neuroblastoma therapy. *J. Exp. Clin. Cancer Res.*, **40**, 1–13.

232. Bulgakov,O. V, Wen,X., Woodruff,M.L., Yang,J., Fain,G.L. and Sandberg,M.A. (2004) Mutations in the G-quadruplex silencer element and their relationship to c-MYC overexpression, NM23 repression, and therapeutic rescue. *Proc. Natl. Acad. Sci.*, **101**, 6140–5.

233. Negrini,S., Gorgoulis,V.G. and Halazonetis,T.D. (2010) Genomic instability an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.*, **11**, 220–228.

234. Zhang,Z., Zhou,C., Chang,Y., Zhang,Z., Hu,Y., Zhang,F., Lu,Y., Zheng,L., Zhang,W., Li,X., *et al.* (2016) Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/β-catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett.*, **376**, 62–73.

235. Black,E.M. and Giunta,S. (2018) Repetitive fragile sites: Centromere satellite DNA as a source of genome instability in human diseases. *Genes (Basel).*, **9**.

236. Roos,W.P. and Kaina,B. (2006) DNA damage-induced cell death by apoptosis. *Trends Mol. Med.*, **12**, 440–450.

237. Lavrado,J., Brito,H., Borralho,P.M., Ohnmacht,S.A., Kim,N.S., Leitão,C., Pisco,S., Gunaratnam,M., Rodrigues,C.M.P., Moreira,R., *et al.* (2015) KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. *Sci. Rep.*, **5**, 1–10.

238. Platella,C., Trajkovski,M., Doria,F., Freccero,M., Montesarchio,D. and Plavec,J. (2020) On the interaction of an anticancer trisubstituted naphthalene diimide with G-quadruplexes of different topologies: a structural insight. *Nucleic Acids Res.*, 10.1093/nar/gkaa1001.

239. Pirota,V., Nadai,M., Doria,F. and Richter,S.N. (2019) Naphthalene diimides as multimodal G-quadruplex-selective ligands. *Molecules*, **24**.

240. Boisvert,F.M., Ahmad,Y., Gierliński,M., Charrière,F., Lamont,D., Scott,M., Barton,G. and Lamond,A.I. (2012) A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell. Proteomics*, **11**, 1–15.

241. Palma,P., Cano,C., Conde-Muiño,R., Comino,A., Bueno,P., Ferrón,J.A. and Cuadros,M. (2014) Expression Profiling of Rectal Tumors Defines Response to Neoadjuvant Treatment Related Genes. *PLoS One*, **9**, e112189.

242. Thorens,B. and Mueckler,M. (2010) Glucose transporters in the 21st Century. *Am. J. Physiol. - Endocrinol. Metab.*, **298**, 141–145.

243. Ma,Y., Chen,H., Su,S., Wang,T., Zhang,C., Fida,G., Cui,S., Zhao,J. and Gu,Y. (2015) Galactose as broad ligand for multiple tumor imaging and therapy. *J. Cancer*, **6**, 658–670.

244. Yang,J., Wen,J., Tian,T., Lu,Z., Wang,Y., Wang,Z., Wang,X. and Yang,Y. (2017) GLUT-1 overexpression as an unfavorable prognostic biomarker in patients with colorectal cancer. *Oncotarget*, **8**, 11788–11796.

245. Briguglio,G., Costa,C., Pollicino,M., Giambò,F., Catania,S. and Fenga,C. (2020) Polyphenols in cancer prevention: New insights. *Int.*

J. Funct. Nutr., **1**, 1–11.

246. Russo,G.L., Tedesco,I., Spagnuolo,C. and Russo,M. (2017) Antioxidant polyphenols in cancer treatment: Friend, foe or foil? *Semin. Cancer Biol.*, **46**, 1–13.

247. Lin,X., Wang,G., Liu,P., Han,L., Wang,T., Chen,K. and Gao,Y. (2021) Gallic acid suppresses colon cancer proliferation by inhibiting SRC and EGFR phosphorylation. *Exp. Ther. Med.*, **21**, 1–11.

248. Wang,K., Zhu,X., Zhang,K., Zhu,L. and Zhou,F. (2014) Investigation of Gallic acid induced anticancer effect in human breast carcinoma MCF-7 cells. *J. Biochem. Mol. Toxicol.*, **28**, 387–393.

249. Jafari,S., Saeidnia,S. and Abdollahi,M. (2014) Role of Natural Phenolic Compounds in Cancer Chemoprevention via Regulation of the Cell Cycle. *Curr. Pharm. Biotechnol.*, **15**, 409–421.

250. Kim,N. (2017) The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.*, **26**, 2898–2917.

251. Weng,S.W., Hsu,S.C., Liu,H.C., Ji,B.C., Lien,J.C., Yu,F.U.S., Liu,K.C., Lai,K.C., Lin,J.P. and Chung,J.G. (2015) Gallic acid induces DNA damage and inhibits DNA repair-associated protein expression in human oral cancer SCC-4 Cells. *Anticancer Res.*, **35**, 2077–2084.

252. Li,L., Li,Y., Zhao,J., Fan,S., Wang,L. and Li,X. (2016) CX-5461 induces autophagy and inhibits tumor growth via mammalian target of rapamycin-related signaling pathways in osteosarcoma. *Onco. Targets. Ther.*, **9**, 5985–5997.

253. Lejault,P., Moruno-Manchon,J.F., Vemu,S.M., Honarpisheh,P., Zhu,L., Kim,N., Urayama,A., Monchaud,D., McCullough,L.D. and Tsvetkov,A.S. (2020) Regulation of autophagy by DNA G-quadruplexes. *Autophagy*, **16**, 2252–2259.

254. Birgisdottir,Å.B. and Johansen,T. (2020) Autophagy and endocytosis – interconnections and interdependencies. *J. Cell Sci.*, **133**.

255. Fratti,R.A., Chua,J. and Deretic,V. (2003) Induction of p38 Mitogen-activated Protein Kinase Reduces Early Endosome Autoantigen 1 (EEA1) Recruitment to Phagosomal Membranes. *J. Biol. Chem.*, **278**, 46961–46967.

256. Ahad,A., Ahsan,H., Mujeeb,M. and Siddiqui,W.A. (2015) Gallic acid ameliorates renal functions by inhibiting the activation of p38 MAPK in experimentally induced type 2 diabetic rats and cultured rat proximal tubular epithelial cells. *Chem. Biol. Interact.*, **240**, 292–303.

257. Subramanian,A.P., Jaganathan,S.K., Mandal,M., Supriyanto,E. and Muhamad,I.I. (2016) Gallic acid induced apoptotic events in HCT-15 colon cancer cells. *World J. Gastroenterol.*, **22**, 3952–3961.

258. He,Z., Chen,A.Y., Rojanasakul,Y., Rankin,G.O. and Chen,Y.C. (2016) Gallic acid, a phenolic compound, exerts anti-angiogenic effects via the PTEN/AKT/HIF-1α/VEGF signaling pathway in ovarian cancer cells. *Oncol. Rep.*, **35**, 291–297.

259. Heidarian,E., Keloushadi,M., Ghatreh-Samani,K. and Valipour,P. (2016) The reduction of IL-6 gene expression, pAKT, pERK1/2, pSTAT3 signaling pathways and invasion activity by gallic acid in prostate cancer PC3 cells. *Biomed. Pharmacother.*, **84**, 264–269.

260. Weng,Y.P., Hung,P.F., Ku,W.Y., Chang,C.Y., Wu,B.H., Wu,M.H., Yao,J.Y., Yang,J.R. and Lee,C.H. (2018) The inhibitory activity of gallic acid against DNA methylation: application of gallic acid on epigenetic therapy of human cancers. *Oncotarget*, **9**, 361.

261. Cueva,C., Silva,M., Pinillos,I., Bartolomé,B. and Moreno-Arribas,M.V. (2020) Interplay between dietary polyphenols and oral and gut microbiota in the development of colorectal cancer. *Nutrients*, **12**.

262. Chen,H., Hayek,S., Rivera Guzman,J., Gillitt,N.D., Ibrahim,S.A., Jobin,C. and Sang,S. (2012) The microbiota is essential for the generation of black tea theaflavins-derived metabolites. *PLoS One*, **7**.

263. Silva,C., Correia-Branco,A., Andrade,N., Ferreira,A.C., Soares,M.L.,

Sonveaux,P., Stephenne,J. and Martel,F. (2019) Selective pro-apoptotic and antimigratory effects of polyphenol complex catechin:lysine 1:2 in breast, pancreatic and colorectal cancer cell lines. *Eur. J. Pharmacol.*, **859**, 172533.

264. Sales,N.M.R., Pelegrini,P.B. and Goersch,M.C. (2014) Nutrigenomics: Definitions and advances of this new science. *J. Nutr. Metab.*, **2014**.

265. Siontorou,C.G. (2013) Nanobodies as novel agents for disease diagnosis and therapy. *Int. J. Nanomedicine*, **8**, 4215.

266. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.

267. Eusébio,D., Neves,A.R., Costa,D., Biswas,S., Alves,G., Cui,Z. and Sousa,Â. (2021) Methods to improve the immunogenicity of plasmid DNA vaccines. *Drug Discov. Today*, **26**, 2575–2592.

268. Zuffo,M., Guédin,A., Leriche,E.D., Doria,F., Pirota,V., Gabelica,V., Mergny,J.L. and Freccero,M. (2018) More is not always better: Finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.*, **46**.

269. Mackenzie,C.R., Burrone,O.R., Muyldermans,S., Hu,Y. and Liu,C. (2017) Nanobody-Based Delivery Systems for Diagnosis and Targeted Tumor Therapy. *Front. Immunol.*, **8**, 1442.

270. Gueorguieva,D., Li,S., Walsh,N., Mukerji,A., Tanha,J., Pandey,S., Gueorguieva,D., Li,S., Walsh,N., Mukerji,A., *et al.* (2006) Identification of single-domain, Bax-specific intrabodies that confer resistance to mammalian cells against oxidative-stress-induced apoptosis. *FASEB J.*, **20**, 2636–2638.

271. Herce,H.D., Schumacher,D., Schneider,A.F.L., Ludwig,A.K., Mann,F.A., Fillies,M., Kasper,M.A., Reinke,S., Krause,E.,

Leonhardt,H., *et al.* (2017) Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nat. Chem.*, **9**, 762–771.

272. Blanco-Toribio,A., Muyldermans,S., Frankel,G. and Fernández,L.Á. (2010) Direct injection of functional single-domain antibodies from E. coli into human cells. *PLoS One*, **5**.

ANNEXES

Annex 1. Map of plasmid vectors used in this study.
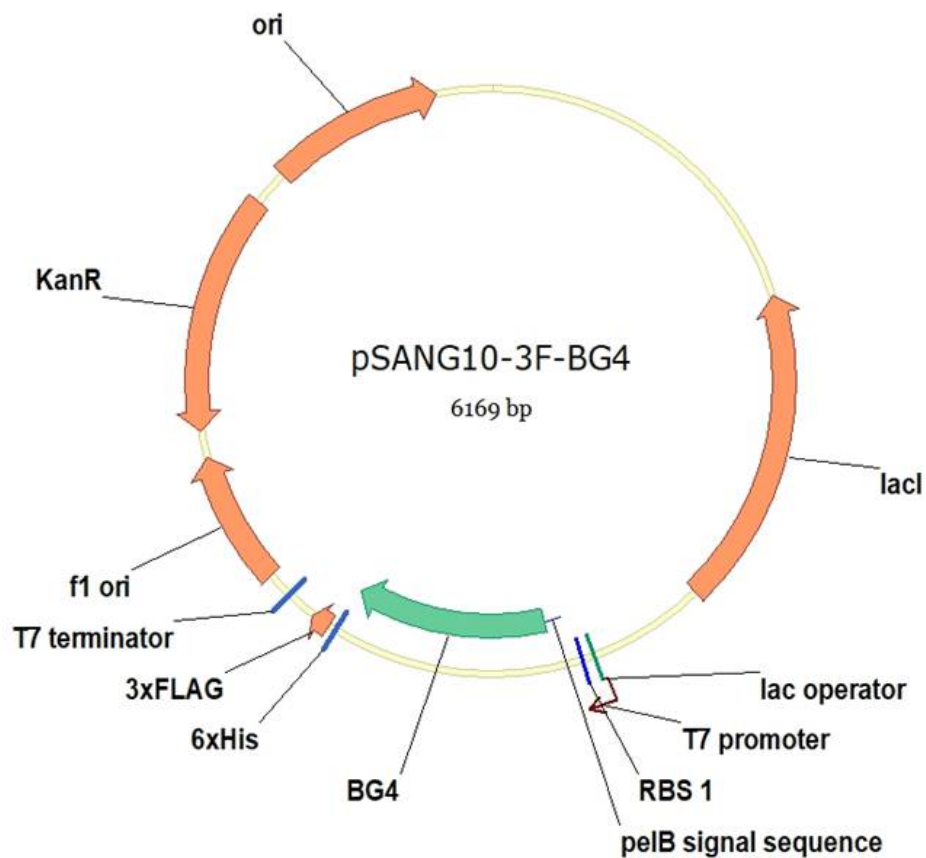
Map of pSANG10-3F-BG4 vector used for BG4 expression as described in the text. The vector contains a kanamycin resistance cassette (KanR), the origins for phage replication (f1 ori) and plasmid replication (ori). In addition, the plasmid contains lacI repressor gene, T7 promoter that allows for expression, a ribosomal binding site (RBS), and T7 terminator. BG4 sequence is 5' fused to pelB (pectate lyase gene of Erwinia carotovora) leader sequence to direct the protein to the bacterial periplasm. In addition, BG4 is fused in frame to a histidine-tag for purification purposes and to FLAG-tag for antibody recognition. Image made with Vector NTI software using the sequence in Addgene (Plasmid #55756).
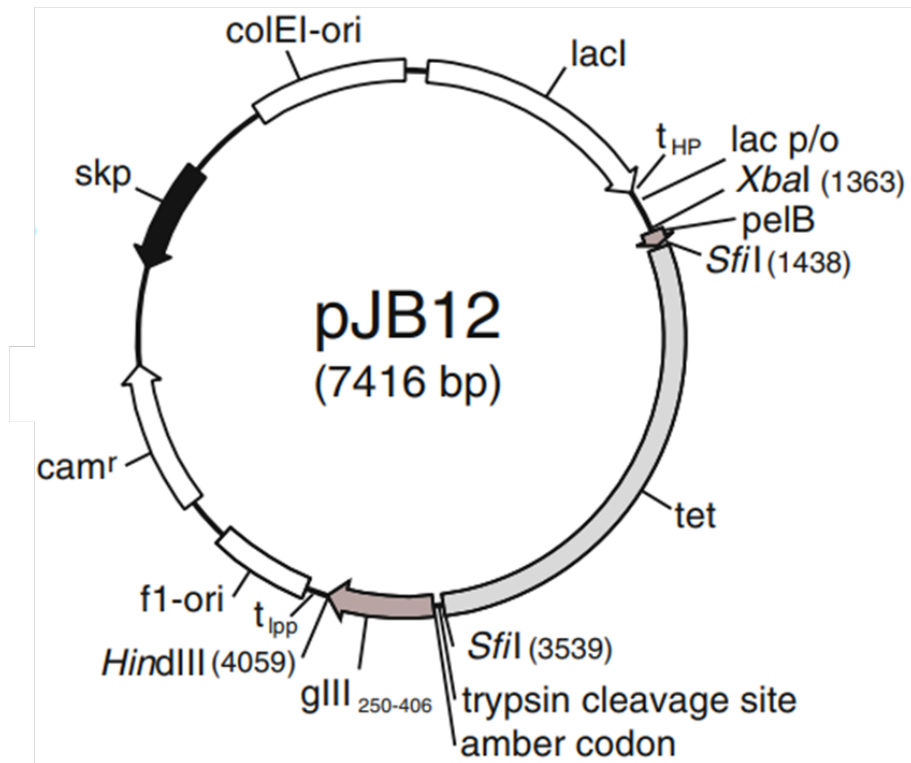
Map of pJB12 vector used for nanobodies production as described in the text. The pJB12 vector contains a chloramphenicol resistance cassette (cam^r) and additionally a tetracycline resistance cassette which will be replaced by the VHH gene (the tet cassette allows the monitoring of complete SfiI digested vector by plating of transformed cells on tetracycline plates). This vector contains the origins for phage replication (f1-ori) and plasmid replication (colEI-ori). In addition, pJB12 contains the lacI repressor gene, a strong upstream terminator (t$_{HP}$) to avoid read-through and premature expression, the lac promoter/operator and the pelB (pectate lyase gene of *Erwinia carotovora*) leader sequence (modified to contain a SfiI site), as well as a downstream terminator (t$_{lpp}$). The antibody gene is fused in frame to gIII$_{250-406}$ for phage display purposes, through an amber codon. Furthermore, it contains a trypsin cleavage site that can be conveniently used for selection of high-affinity binders. The co-expressed periplasmic protein Skp, encoded on this vector, increases the functional yield of antibody fragments expressed in the periplasm without the need of cotransformation with another plasmid coding for further chaperones. Image obtained from Schaefer JV *et al*. Antibody Engineering 2010.

Annex 2. Antibodies used in this study. List of all antibodies and respective dilutions used for immunofluorescence (IF) and western blot (WB) experiments. The antibodies are alphabetically sorted by name.

Name	Source	Commercial reference	Study	Dilution
α-Actin	Mouse	Sigma Aldrich (A5441)	IF	-
			WB	1:5000
α-BG4	ScFv	Non-commercial	IF	20 nM
			WB	-
α-Calnexin	Rabbit	GeneTex (109669)	IF	1:150
			WB	-
α-EEA1	Mouse	BD Transduction Laboratories (610457)	IF	1:70
			WB	-
α-Fibrillarin	Rabbit	Abcam (ab5821)	IF	1:100
			WB	-
α-FLAG	Mouse	Sigma Aldrich (F1804)	IF	1:1000
			WB	-
α-Giantin	Mouse	Abcam (ab-37266)	IF	1:200
			WB	-
α-γH2AX	Rabbit	Cell Signaling Technology (9718)	IF	-
			WB	1:800

			IF	-
α-Histidine	Mouse	Santa Cruz Biotechnology (sc-8036)	WB	1:1000
α-LAMPI	Mouse	Pharmingen (34201A)	IF	1:100
			WB	-
α-LC3	Rabbit	MBL (PD014)	IF	-
			WB	1:1000
α-Mouse Alexa Fluor 488	Goat	Invitrogen (A-11001)	IF	1:500
			WB	-
α-Mouse HRP conjugated	Goat	Promega (W4021)	IF	-
			WB	1:2500
α-Nuclelolin	Mouse	Invitrogen (39-6400)	IF	1:100
			WB	-
α-POLR1A	Mouse	Santa Cruz Biotechnology (sc-48385)	IF	1:100
			WB	1:100
α-pTBK1	Rabbit	Cell Signalling (5483)	IF	1:50
			WB	-
α-Rabbit Alexa Fluor 555	Donkey	Invitrogen (A-31572)	IF	1:1000
			WB	-
α-Rabbit	Donkey	Invitrogen (SA1-200)	IF	-

|				
HRP conjugated			WB	1:2500
α-RAB5	Rabbit	Cell Signalling (3547)	IF	1:80
			WB	-
α-RAB11	Rabbit	Cell Signalling (55892)	IF	1:80
			WB	-
α-SERCA2	Mouse	Novus Biological (NB300-581)	IF	1:150
			WB	-
α-SNX1	Mouse	BD Transduction Laboratories (611482)	IF	1:80
			WB	-
α-TOM20	Rabbit	Santa Cruz Biotechnology (sc-11415)	IF	1:200
			WB	-
α-WIPI2	Mouse	BioRad (MCA5780GA)	IF	1:200
			WB	-

Annex 3. PCR primers used in this study. List of all forward (FW) and reverse (RV) primers and their respective sequences used for PCR. The primers are alphabetically sorted by name.

Name		Gene	Sequence (5'->3')
5'ETS_1 (qRT-PCR)	FW	*5'ETS rDNA: +302+548*	GTGCGTGTCAGGCGTTCT
	RV		GGGAGAGGAGCAGACGAG
5'ETS_2 (qRT-PCR)	FW	*5'ETS rDNA: +851+961*	GAACGGTGGTGTGTCGTT
	RV		GCGTCTCGTCTCGTCTCACT
5'ETS_3 (qRT-PCR)	FW	*5'ETS rDNA: +1297+1483*	CAGGTGTTTCCTCGTACCG
	RV		GCTACCATAACGGAGGCAGA
5'ETS_4 (qRT-PCR)	FW	*5'ETS rDNA: +1317+1497*	CCTTCCCCAGGCGTCCCTCG
	RV		GGCAGCGCTACCATAACGGA
-988 (ChIP-qPCR)	FW	*Upstream rDNA:-988-798*	GCTTCTCGACTCACGGTTTC
	RV		GGAGCTCTGCCTAGCTCACA
-410 (ChIP-qPCR)	FW	*Upstream rDNA:-410-272*	GATCCTTTCTGGCGAGTCC
	RV		GGAGCCGGAAGCATTTTC
-46 (ChIP-qPCR)	FW	*Promoter rDNA: -46+13*	GGTATATCTTTCGCTCCGAG
	RV		AGCGACAGGTCGCCAGAGGA
+4446 (ChIP-qPCR)	FW	*18S rDNA: +4446+4612*	CCCGAAGCGTTTACTTTGAA
	RV		CGGTCCAAGAATTTCACCTA

+10319 (ChIP-qPCR)	FW	*28S rDNA: +10319+10450*	GAACTTTGAAGGCCGAAGTG
	RV		ATCTGAACCCGACTCCCTTT
+12293 (ChIP-qPCR)	FW	*28S rDNA: +12293+12472*	TGGGTTTTAAGCAGGAGGTG
	RV		AACCTGTCTCACGACGGTCT
+18449 (ChIP-qPCR)	FW	*IGS rDNA: +18449+18572*	TGGTGGGATTGGTCTCTCTC
	RV		CAGCCTGCGTACTGTGAAAA
+30541 (ChIP-qPCR)	FW	*IGS rDNA: +30541+30621*	ACTGGCGAGTTGATTTCTGG
	RV		CGAGACAGTCGAGGGAGAAG
ACTIN (qRT-PCR)	FW	*ACTIN*	TGCGTCTGGACCTGGCTGGC
	RV		GCCTCAGGGCAGCGGAACCG
BCL2 (qRT-PCR)	FW	*BCL2*	CTGCACCTGACGCCCTTCACC
	RV		CACATGACCCCACCGAACTCAAAGA
CMYB (qRT-PCR)	FW	*CMYB*	ACCATGACTATGATGGGCTGC
	RV		TCCCCAAGTGACGCTTTCC
CMYC (qRT-PCR)	FW	*CMYC*	CGTCCTCGGATTCTCTGCTC
	RV		GCCTGCCTCTTTTCCACAGA
CMYC_G4 (sequencing)	FW	*CMYC G4*	TAGGCGCGCGTAGTTAATTC
	RV		CGGAGATTAGCGAGAGAGGA

CMYC_G4 (ChIP-qPCR)	FW	*CMYC G4*	CTCTCTCGCTAATCTCCGCC
	RV		CGTCCAGACCCTCGCATTAT
GLUT1 (qRT-PCR)	FW	*GLUT1*	CTGCTCATCAACCGCAAC
	RV		CTTCTTCTCCCGCATCATCT
KRAS (qRT-PCR)	FW	*KRAS*	GACTGAATATAAACTTGTGGTAGTTGGA
	RV		CATATTCGTCCACAAAATGATTCTG
KRAS_G4 (sequencing)	FW	*KRAS G4*	TAGGCGCGCGTAGTTAATTC
	RV		TTCCGCGCTCGATTCTTCTT
KRAS_G4 (ChIP-qPCR)	FW	*KRAS G4*	AAGAAGAATCGAGCGCGGAA
	RV		AAATCGAGCTCCGAGCACAC
POLR1A (qRT-PCR)	FW	*POLR1A*	GGTGAAGCCAAAGGCAGATGT
	RV		TATGATGCGGCTTCTGGCAGGT
VEGFA (qRT-PCR)	FW	*VEGFA*	CTACCTCCACCATGCCAAGT
	RV		GCAGTAGCTGCGCTGATAGA

Annex 4. Schematic representation of breaks labeling in situ and sequencing methodology. The workflow starts by fixing cells onto coverslips in 24-well plates. Double-strand break (DSB) ends are then in situ blunted and tagged with double-strand DNA adapters containing components described in the boxed legend and in the text (UMI refers to unique molecular identifier). Tagged DSB ends are linearly amplified using in vitro transcription and the resulting RNA is used for Illumina library preparation and sequencing. Image made with Biorender (https://biorender.com) and adapted from Yan W *et al*. Nat. Commun. 2017, 8, 15058.
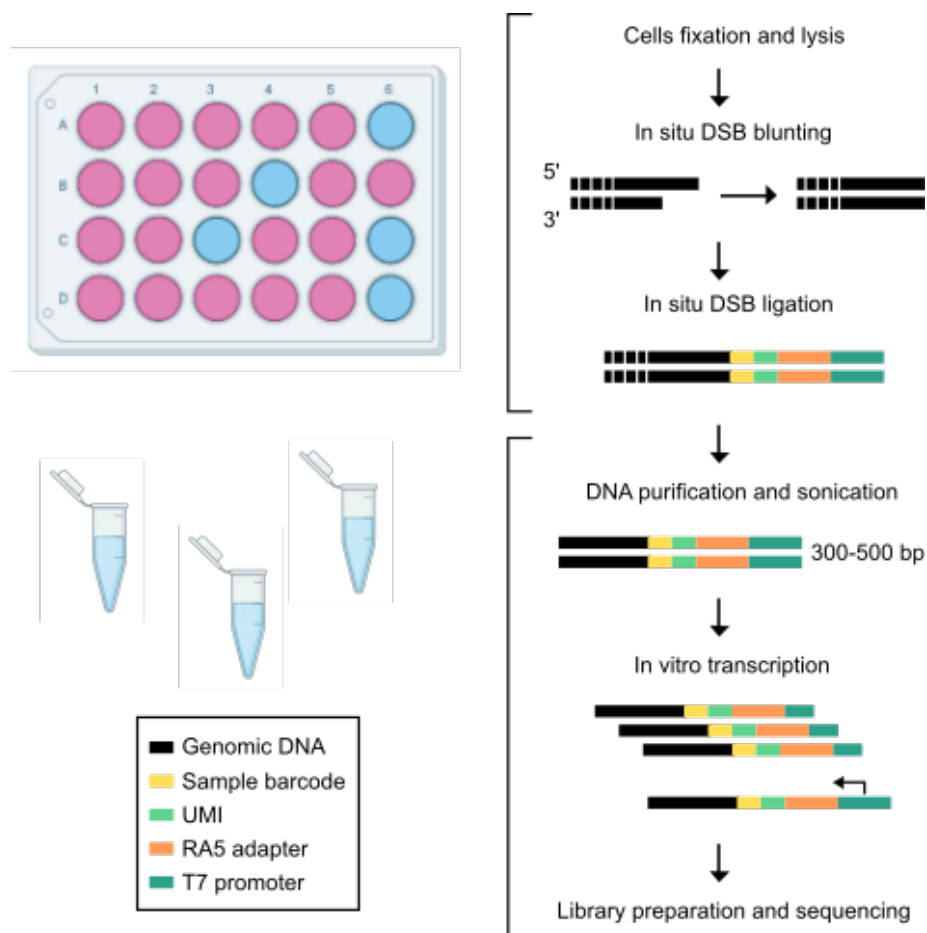
Annex 5. G4-oligonucleotides used in this study. List of all G4-oligonucleotides and their respective sequences used for fluorescent intercalator displacement (FID), PCR-stop, nuclear magnetic resonance (NMR), circular dichroism (CD) and ultraviolet-visible (UV-vis) experiments. The oligonucleotides are alphabetically sorted by name.

G4-oligonucleotides for FID assay	
Name	Sequence (5'->3')
5'ETS_F1	GGGTGGACGGGGGGGCCTGGTGGG
5'ETS_F2	GGGTTGGGGGGGAGAAGCGAGGG
5'ETS_F3	GGGGGGAGCCGCGGGGATCGCCGAGGG
5'ETS_F4	GGGGTGGGGCCCGGGCCGGGG
5'ETS_F5	GGGGGGCGGGTGGTTGGG
5'ETS_R1	GGGAGGGAGCGAGCGGGCGCGGG
5'ETS_R2	GGGACCGGTGGGGCCGGGGCGGGG
5'ETS_R3	GGGAGGGACCACCGGGCCGCGCTCGGG
5'ETS_R4	GGGCGGCGGGCGGGGAAGAGGG
5'ETS_R5	GGGCGAGGGCCGGGGACCGCGAGGG
5'ETS_R6	GGGTGGGAGCGCCGGGCCCGGCCCGGCGGG
5'ETS_R7	GGGACGCCTGGGGAAGGGAGGGGG
BCL2	TAGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG

CMYB	TAGGAGGAGGAGGTCACGGAGGAGGAGGAGAAGG AGGAGGAGGAAA
CMYC	GGGGCGCTTATGGGGAGGGTGGGGAGGGTGGGGA AGGTGGGGAGGAG
KRAS	TAGGGCGGTGTGGGAAGAGGGAAGAGGGGGGAGGC AG
VEGFA	GGGGCGGGCCGGGGGCGGGG
TEL	TAGGGTTAGGGTTAGGGTTAGGGT

G4-oligonucleotides for PCR-stop assay	
Name	Sequence (5'->3')
5'ETS_FW	TCGCGTGGGGGGCGGGTGGTTGGG
5'ETSMUT_FW	TCGCGTCCACTTCAAGTGGTTGGG
5'ETS_RV	TTCTCGTCCCAACCAC
CMYC_FW	GGGGCGCTTATGGGGAGGGTGGGGAGGGTGGGGA AGGTGGGGAGGAG
CMYC_RV	TTCTCGTCTCCTCCCC

G4-oligonucleotides for NMR experiments	
Name	Sequence (5'->3')
5'ETS	TGGGGGGCGGGTGGTTGGGT

G4-oligonucleotides for CD experiments	
Name	Sequence (5'->3')

5'ETS	GGGGGGCGGGTGGTTGGG
CMYC	TGGGGAGGGTGGGGAGGGTGGGGAAGG
G4-oligonucleotides for UV-vis experiments	
Name	Sequence (5'->3')
5'ETS	GGGGGGCGGGTGGTTGGG
CMYC	TGGGGAGGGTGGGGAGGGTGGGGAAGG

Annex 6. Oligonucleotides used for production and selection of nanobodies in this study. List of all oligonucleotides and their respective sequences. The oligonucleotides are sorted according to their sequential use. TEG refers to the 15 carbon arm spacer employed for biotinylation to avoid hindrance issues.

Name	Sequence (5'->3')
G4_CMYC for in vitro immunization	TTGGGGAGGGTGGGGAGGGTGGGGAAGG
Primer for RT-PCR	CCAGCGGCCGCTSWGGAGACR
Forward primer for PCR of VHH region	GCGGCCCAGCCGGCCGCCSAGGTGSAGSTSSWGSMGTC
Reverse primer for PCR of VHH region	AAAGGCCCCAGAGGCCGATSWGGAGACRGTGACCWGGGTCC
Forward primer for checking the ligation of VHH into pJB12	ATGAAATACCTATTGCCTACGGCAG
Reverse primer for checking the ligation of VHH into pJB12	CATAATCAAAATCACCGGAACCAGAG
Biotinylated G4_CMYC for biopanning and ELISA	BIOTIN-TEG-TTGGGGAGGGTGGGGAGGGTGGGGAAGG
Biotinylated G4_BCL2 for ELISA	BIOTIN-TEG-TAGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG
Biotinylated G4_CMYB for ELISA	BIOTIN-TEG-TAGGAGGAGGAGGTCACGGAGGAGGAGGAGAAGGAGGAGGAGGAAA

Biotinylated G4_KRAS for ELISA	BIOTIN-TEG- TAGGGCGGTGTGGGAAGAGGGAAGAG GGGGAGGCAG
Biotinylated G4_TEL for ELISA	BIOTIN-TEG- TAGGGTTAGGGTTAGGGTTAGGGT
Biotinylated G4_VEGFA for ELISA	BIOTIN-TEG- TCGGGGCGGGCCGGGGGCGGGGTCC CGGCGGGGCGGAGC

PUBLICATIONS

Sanchez-Martin V, Schneider DA, Ortiz-Gonzalez M, Soriano-Lerma A, Linde-Rodriguez A, Perez-Carrasco V, Gutierrez-Fernandez J, Cuadros M, Morales JC, González C, Soriano M, Garcia-Salcedo JA. Targeting ribosomal G-quadruplexes with naphthalene-diimides as RNA polymerase I inhibitors for colorectal cancer treatment. Cell Chem Biol. 2021 Dec 16;28(12):1807. doi: 10.1016/j.chembiol.2021.12.007. Erratum for: Cell Chem Biol. 2021 Nov 18;28(11):1590-1601.e4. PMID: 34919843.

CellPress **Cell Chemical Biology**

Article

Targeting ribosomal G-quadruplexes with naphthalene-diimides as RNA polymerase I inhibitors for colorectal cancer treatment

Victoria Sanchez-Martin,[1,2,3] David A. Schneider,[4] Matilde Ortiz-Gonzalez,[1,5] Ana Soriano-Lerma,[1,6] Angel Linde-Rodriguez,[1,2] Virginia Perez-Carrasco,[1,2] Jose Gutierrez-Fernandez,[2,7] Marta Cuadros,[1,3] Juan C. Morales,[8] Carlos González,[9] Miguel Soriano[1,5,10] and Jose A. Garcia-Salcedo[1,2,11,*]

[1]GENYO. Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, Granada 18016, Spain
[2]Microbiology Unit, Biosanitary Research Institute IBS.Granada, University Hospital Virgen de las Nieves, Granada 18014, Spain
[3]Department of Biochemistry, Molecular Biology III and Immunology, University of Granada, Granada 18016, Spain
[4]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[5]Centre for Intensive Mediterranean Agrosystems and Agri-food Biotechnology (CIAIMBITAL), University of Almeria, Almeria 04001, Spain
[6]Department of Physiology, University of Granada, Granada 18011, Spain
[7]Department of Microbiology, University of Granada, Granada 18011, Spain
[8]Department of Biochemistry and Molecular Pharmacology, Instituto de Parasitología y Biomedicina López Neyra, CSIC, PTS Granada, Avda. del Conocimiento, 17, 18016 Armilla, Granada, Spain
[9]Instituto de Química Física "Rocasolano", CSIC, Madrid 28006, Spain
[10]These authors contributed equally
[11]Lead contact
*Correspondence: jags@genyo.es
https://doi.org/10.1016/j.chembiol.2021.05.021

Sanchez-Martin V, Soriano M, Garcia-Salcedo JA. Quadruplex Ligands in Cancer Therapy. Cancers (Basel). 2021 Jun 24;13(13):3156. doi: 10.3390/cancers13133156. PMID: 34202648; PMCID: PMC8267697.

cancers MDPI

Review

Quadruplex Ligands in Cancer Therapy

Victoria Sanchez-Martin [1,2,3], Miguel Soriano [1,4,*] and Jose Antonio Garcia-Salcedo [1,2,*]

1   Centre for Genomics and Oncological Research, Pfizer-University of Granada-Andalusian Regional Government, PTS Granada, 18016 Granada, Spain; victoria.sanchez@genyo.es
2   Microbiology Unit, Biosanitary Research Institute IBS, University Hospital Virgen de las Nieves, 18014 Granada, Spain
3   Department of Biochemistry, Molecular Biology III and Immunology, University of Granada, 18016 Granada, Spain
4   Centre for Intensive Mediterranean Agrosystems and Agri-Food Biotechnology (CIAMBITAL), University of Almeria, 04001 Almeria, Spain
*   Correspondence: msoriano@ual.es (M.S.); jags@genyo.es (J.A.G.-S.); Tel.: +34-958715500 (M.S.); +34-958715500 (J.A.G.-S.)

Sanchez-Martin V, Lopez-Pujante C, Soriano-Rodriguez M, Garcia-Salcedo JA. An Updated Focus on Quadruplex Structures as Potential Therapeutic Targets in Cancer. Int J Mol Sci. 2020 Nov 24;21(23):8900. doi: 10.3390/ijms21238900. PMID: 33255335; PMCID: PMC7734589.

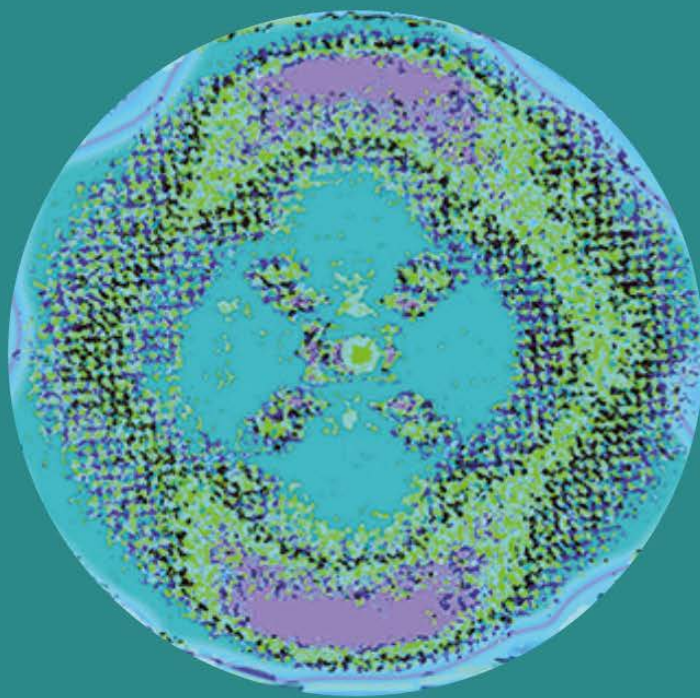International Journal of *Molecular Sciences* MDPI
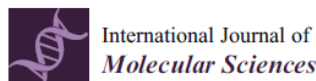
Review

An Updated Focus on Quadruplex Structures as Potential Therapeutic Targets in Cancer

Victoria Sanchez-Martin [1,2,3], Carmen Lopez-Pujante [1], Miguel Soriano-Rodriguez [1,4,*] and Jose A. Garcia-Salcedo [1,2,*]

1   GENYO, Centre for Genomics and Oncological Research, Pfizer/University of Granada/Andalusian Regional Government, PTS Granada, 18016 Granada, Spain; victoria.sanchez@genyo.es (V.S.-M.); carmenlopu@correo.ugr.es (C.L.-P.)
2   Microbiology Unit, University Hospital Virgen de las Nieves, Biosanitary Research Institute IBS, Granada, 18014 Granada, Spain
3   Department of Biochemistry, Molecular Biology III and Immunology, University of Granada, 18016 Granada, Spain
4   Centre for Intensive Mediterranean Agrosystems and Agri-food Biotechnology (CIAMBITAL), University of Almeria, 04001 Almeria, Spain
*   Correspondence: msoriano@ual.es (M.S.-R.); jags@genyo.es (J.A.G.-S.); Tel.: +34-958715500 (M.S.-R.); +34-958715500 (J.A.G.-S.)

Other publications to which I have contributed during my PhD studies:

- Soriano-Lerma A, García-Burgos M, Alférez MJM, Pérez-Carrasco V, **Sanchez-Martin V**, Linde-Rodríguez Á, Ortiz-González M, Soriano M, García-Salcedo JA, López-Aliaga I. Gut microbiome-short-chain fatty acids interplay in the context of iron deficiency anaemia. Eur J Nutr. 2022 Feb;61(1):399-412. doi: 10.1007/ s00394-021-02645-6. Epub 2021 Aug 12. PMID: 34383140.

- Cobo F, Pérez-Carrasco V, **Sanchez-Martin V**, García-Salcedo JA, Martín EB, Navarro-Marí JM. A rare cause of bacteremia due to Porphyromonas asaccharolytica in a patient with necrotizing fasciitis. Anaerobe. 2021 Oct;71:102442. doi: 10.1016/ j.anaerobe.2021.102442. Epub 2021 Sep 2. PMID: 34481990.

- Cobo F, Pérez-Carrasco V, González A, **Sanchez-Martin V**, García-Salcedo JA, Navarro-Marí JM. Bacteremia caused by Anaerococcus SPP: Is this an underdiagnosed infection? Anaerobe. 2021 Aug;70:102405. doi: 10.1016/j.anaerobe.2021. 102405. Epub 2021 Jun 18. PMID: 34153469.

- Soriano-Lerma A, Pérez-Carrasco V, Sánchez-Marañón M, Ortiz-González M, **Sanchez-Martin V**, Gijón J, Navarro-Mari JM, García-Salcedo JA, Soriano M. Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. Sci Rep. 2020 Aug 12;10(1):13637. doi: 10.1038/s41598-020-70141-8. PMID: 32788589; PMCID: PMC7423937.

Granada, 2022