

**STUDIES OF G-QUADRUPLEXES IN HUMAN CELLS: METHODS AND BIOLOGY**

by

YANG (SUNNY) YANG

B.ScH., Queen's University, 2013  
M.Sc., University of Toronto, 2015

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES  
(Pharmaceutical Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

December 2020

© Yang (Sunny) Yang, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Studies of G-quadruplexes in human cells: methods and biology

submitted by Yang (Sunny) Yang in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Pharmaceutical Sciences

**Examining Committee:**

Judy Wong, Associate Professor, Pharmaceutical Sciences, UBC

Supervisor

Peter Lansdorp, Professor, Medical Genetics, UBC

Supervisory Committee Member

Sean McKenna, Professor, Chemistry, University of Manitoba

External Examiner

LeAnn Howe, Professor, Biochemistry and Molecular Biology, UBC

University Examiner

Martin Hirst, Associate Professor, Microbiology and Immunology, UBC

University Examiner

**Additional Supervisory Committee Members:**

Brian Rodrigues, Professor, Pharmaceutical Sciences, UBC

Supervisory Committee Member

Corey Nislow, Professor, Pharmaceutical Sciences, UBC

Supervisory Committee Member

Adam Frankel, Associate Professor, Pharmaceutical Sciences, UBC

Supervisory Committee Member

## Abstract

While the predominant nucleic acid secondary structure is that of the canonical double-helix form, formations of other non-canonical structures can form under specific conditions, one of which is the highly stable four-stranded structure, called the G-quadruplex (G4). G4s are formed by Hoogsteen bonding of adjacent guanines into G-quartet units, which then  $\pi$ -stack with each other to form the overall columnar structure. Despite the observed high stability of G4s in *in vitro* settings, their existence in living cells and their biological roles remain unclear. My doctoral work's primary goal is to study the relevance of G4s in human cells by performing broad-scale investigations using G4-specific tools (antibodies and probes) coupled with cellular imaging and genomics techniques. The secondary goal is to provide new insights into the biology of G4 by using these newly developed methods. Through parallel optimizations of novel G4-detection methods, I discovered that specific combinations of G4-detection tools with corresponding methodologies are selectively suitable for studying G4-DNAs versus G4-RNAs. Coupled with high-throughput sequencing, genome- and transcriptome-wide profiling of G4s reveal that both G4-DNAs and G4-RNAs are widespread and transient in nature in G-rich regions. These techniques also offer a way to evaluate changes in the global G4 landscapes induced by the treatments of G4 ligands, chemical compounds designed to target and stabilize G4s. Despite the common belief that G4 formation solely results in negative consequences in cellular functions, recent evidence suggests that G4s can potentially have beneficial effects. Cancer cells that are telomerase-negative and utilize the alternative lengthening of telomere (ALT) mechanism for telomere maintenance harbour significantly longer telomeres, thus provide ideal models for studying telomeric G4s. I investigated G4s in ALT models to elucidate potential novel biological functions for G4s. Strikingly, results suggest that G4s may be adapted

by ALT+ cancers to initiate the ALT-specific telomere maintenance mechanism, supporting the view in which the existence of G4s can have both positive and negative consequences in context-specific manners. My doctoral work concludes that G4s are widespread and tightly regulated for optimal cellular functions in living human cells.

## **Lay Summary**

While the two-stranded double-helix is the predominant structure formed in nucleic acids (DNA and RNA), other structures can also be formed and play important biological roles. One such structure is called the G-quadruplex (G4s), a highly stable four-stranded structure. While there is strong evidence for the formation of G4s within test tubes, evidence for their formation in living cells remains limited. In my thesis, I aim to study the biological significance of G4s in human cells by first developing new tools and methods. I have successfully optimized a suite of methods for studying G4s on a large-scale using cellular imaging and genomics techniques. By combining with high throughput sequencing, widespread G4 formations have been observed, which provide evidence of G4 formations within living human cells. My thesis results contribute to establishing a framework for future investigations on the different biological roles of G4s.

## **Preface**

This thesis is composed of original published and unpublished work in which I am the primary contributor under the supervision of Dr. Judy Wong.

### **Chapter 3**

A version of this chapter has been published:

Yang, S. Y., Amor, S., Laguerre, A., Wong, J. M., & Monchaud, D. (2017). Real-time and quantitative fluorescent live-cell imaging with quadruplex-specific red-edge probe (G4-REP). *Biochimica Et Biophysica Acta (B.B.A.)-General Subjects*, 1861(5), 1312-1320. and reproduced by the permission of Elsevier.

I was performed all the confocal and live-cell imaging experiments and analyses (Fig. 3, 5, 6, cell imaging movies), as well as cytotoxicity studies (Supplementary Fig. 1). A.L and S.A. (ICMUB, CNRS, France) synthesized N-TASQ and performed multiphoton imaging (Fig. 2; Supplementary Fig. 2-6). D.M. contributed to the illustrations in Fig. 1. I co-wrote the manuscript with J.M.Y.W. and D.M.

### **Chapter 4**

A version of this chapter has been published:

Yang, S. Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A. G., Wong, J. M., & Monchaud, D. (2018). Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nature communications*, 9(1), 1-11. and reproduced under the Creative Commons BY License Attribution 4.0 International.

I designed the experiments with J.M.Y.W (UBC) and D.M. (ICMUB, CNRS, France). The in vitro experiments (Fig. 1; Supplementary Fig. 1-3, 11) were done by P.L. and D.M. (ICMUB, CNRS, France). S.C. and R.B. (CGFL, France) performed sequencing. I performed all cell-based investigations and sequencing analysis (Fig 2-5; Supplementary Fig. 4-10, Table 1-4). I also drafted the original manuscript with edits from G.R. (BC Cancer, Canada), J.M.Y.W. and D.M.

## **Chapter 5**

A version of this chapter is under submission:

Yang, S.Y., Chang, E.Y.C., Lim, J., Kwan, H., Stirling, P.C. and Wong, J.M.Y. Telomeric G-loops stimulate Alternative Lengthening-of-Telomeres. *In submission 2020-2021*.

I designed the experiments with J.M.Y.W. (UBC) and P.C.S. (BC Cancer, Canada). I optimized assays with help from H.K. and J.L. (UBC). PLA assays (Fig. 3-4) were performed by E.Y.C.C. and P.C.S. (BC Cancer, Canada). I performed all of the immunocytochemistry, chromatin-immunoprecipitation (ChIP), Re-ChIP, c-circle assay, overexpression and cytotoxicity experiments (Fig. 1-7; Supplementary Fig. 1-8). I also drafted the original manuscript with J.M.Y.W. with inputs from E.Y.C.C. and P.C.S.

## **Appendix 1**

A version of this chapter is published:

Amor, S., Yang, S. Y., Wong, J. M., & Monchaud, D. (2018). Cellular Detection of G-Quadruplexes by Optical Imaging Methods. *Current Protocols in Cell Biology*, 76(1), 4-33. and reproduced by the permission of Wiley.

I provided figures and wrote part of the protocol manuscript with S.A (ICMUB, CNRS, France), J.M.Y.W. (UBC) and D.M. (ICMUB, CNRS, France).

## **Appendix 2**

A version of this chapter is in progress.

I drafted the original manuscript with edits from J.M.Y.W (UBC).

## **Appendix 3**

This is an original and unpublished work.

I performed all the ChIP-seq, gene expression analyses and related experiments.

# Table of Contents

<b>Abstract</b> .....	<b>iii</b>
<b>Lay Summary</b> .....	<b>v</b>
<b>Preface</b> .....	<b>vi</b>
<b>Table of Contents</b> .....	<b>ix</b>
<b>List of Tables</b> .....	<b>xvii</b>
<b>List of Figures</b> .....	<b>xviii</b>
<b>List of Abbreviations</b> .....	<b>xx</b>
<b>Acknowledgements</b> .....	<b>xxv</b>
<b>Dedication</b> .....	<b>xxvi</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1    Introduction to Nucleic Acid Structures .....	1
1.1.1    Primary Structure .....	2
1.1.1.1    Nucleotide structure.....	2
1.1.2    Secondary Structure.....	3
1.1.2.1    Horizontal Base-Base Interaction: Hydrogen Bonding .....	3
1.1.2.2    Vertical Base-Base Interaction: Base Stacking, Hydrophobic Interactions and London Dispersion .....	4
1.1.3    Types of DNA and RNA Secondary/Tertiary Structures .....	6
1.1.3.1    Canonical Helical Structure – the Double Helix and its Polymorphs.....	6
1.1.3.2    Noncanonical Helical Structures – Triple helix, G-quadruplex and i-motif.....	7
1.1.3.3    Other Nucleic Acid Structures.....	9

1.1.3.4	Quaternary Structure and Beyond.....	10
1.2	Brief History of G4s and Key Discoveries .....	10
1.3	Structure of DNA and RNA G4s .....	12
1.3.1	1 <sup>st</sup> Dimension – Guanine base.....	13
1.3.2	2 <sup>nd</sup> Dimension – G-quartet .....	13
1.3.3	3 <sup>rd</sup> Dimension – G-quadruplex.....	17
1.3.3.1	Defining the Potential G4-forming Sequence.....	17
1.3.3.2	G-column .....	17
1.3.3.3	Loop.....	18
1.3.3.4	Metal Cations of the Central Channel.....	18
1.3.4	Structural diversity of G4-DNA.....	19
1.3.4.1	Canonical G4-DNA Structures.....	20
1.3.4.2	Noncanonical G4-DNA structures.....	21
1.3.5	Structural diversity of G4-RNA.....	22
1.3.5.1	Key differences between G4-DNA and G4-RNA.....	22
1.3.5.2	Noncanonical G4-RNA structures .....	23
1.3.6	G4-stabilizing Ligands (G4 ligands).....	23
1.3.6.1	G4 Ligand Design Considerations .....	24
1.3.6.2	Types of G4 ligands .....	25
1.3.6.3	Other G4-interacting molecules .....	29
1.4	Current strategies for studying G4s .....	30
1.4.1	<i>In vitro</i> structural validation methodologies .....	30
1.4.2	<i>In silico</i> bioinformatics methodologies.....	34

1.4.3	<i>In vitro</i> genomics methodologies .....	35
1.4.4	<i>In situ</i> cellular imaging methodologies .....	36
1.4.5	<i>In vivo</i> cellular imaging .....	37
1.4.6	<i>In vivo</i> genomics methodologies .....	38
1.5	Biological Relevance of G4 .....	40
1.5.1	Evolutionary Conservation of G4 motifs .....	40
1.5.2	G4-interacting Proteins .....	41
1.5.2.1	G4 Helicases .....	41
1.5.2.2	Other Non-helicase G4-interacting Proteins .....	45
1.5.3	Biological Functions of G4-DNAs .....	48
1.5.3.1	Telomere maintenance & integrity .....	48
1.5.3.2	Transcription .....	50
1.5.3.3	Genome instability .....	51
1.5.3.4	Epigenetic Control .....	51
1.5.4	Biological Functions of G4-RNAs .....	52
1.5.4.1	Translation .....	52
1.5.4.2	Other potential functions of G4-RNAs .....	53
1.5.5	G4 in human diseases & potential therapeutics .....	53
1.5.5.1	Cancer .....	54
1.5.5.2	Neurodegenerative diseases .....	54
1.5.5.3	G4 helicase dysfunction-related diseases .....	55
1.6	Rationale, Hypothesis & Aim .....	56
1.6.1	Research Questions of Interest .....	56

1.6.2	Hypothesis.....	59
1.6.3	Aims .....	60
<b>Chapter 2: Detailed Protocols: Materials and Methods.....</b>		<b>62</b>
2.1	Purification of the BG4 antibody.....	62
2.1.1	Reagents and Buffers .....	62
2.1.2	Procedures.....	63
2.2	Immunocytochemistry (ICC) Staining Standard Protocol.....	66
2.2.1	Reagents and Buffers .....	66
2.2.2	Standard Staining Procedure.....	67
2.2.3	G4-staining with BG4.....	68
2.2.4	G4-staining with N-TASQ.....	69
2.2.5	Imaging on Confocal Microscope.....	69
2.2.6	Automated foci quantification using CellProfiler .....	70
2.3	Real-time Live-cell imaging in Incucyte ZOOM.....	71
2.4	Chromatin-immunoprecipitation (ChIP) & Re-ChIP.....	72
2.4.1	Reagents and Buffers .....	72
2.4.2	Procedures.....	73
2.5	G4-RNA Precipitation (G4RP).....	77
2.6	C-circle Assay (CCA) with qPCR .....	77
2.6.1	Reagents and Buffers .....	77
2.6.2	Procedures.....	78
<b>Chapter 3: Real-time and quantitative fluorescent live-cell imaging methods</b>		
<b>(BBA-General Subjects, 2017) .....</b>		<b>81</b>

3.1	Summary .....	81
3.2	Introduction.....	82
3.3	Results.....	84
3.3.1	The red-edge effect: a mechanistic reminder.....	84
3.3.2	N-TASQ multimodality: multi-photon absorption and the red-edge effect.....	86
3.3.3	Disclosing the quadruplex entity of N-TASQ cellular targets: a quantitative analysis.....	90
3.3.4	N-TASQ, quadruplexes and the red-edge effect: what is next?.....	91
3.3.5	A spatiotemporal step beyond.....	93
3.4	Discussion .....	96
3.5	Methods.....	99
3.6	Supplementary Materials .....	101
<b>Chapter 4: Transcriptome-wide identification of transient RNA G-quadruplexes in human cells (<i>Nature Communications</i>, 2018).....</b>		<b>107</b>
4.1	Summary .....	107
4.2	Introduction.....	107
4.3	Results.....	110
4.3.1	BioTASQ selectively captures G4 targets <i>in vitro</i> .....	110
4.3.2	G4RP isolates G4-RNA targets from human cell extracts.....	113
4.3.3	G4RP-Seq identifies transcriptome-wide transient G4-RNAs .....	116
4.3.4	G4RP-seq identifies ligand-induced changes in G4-RNA profiles .....	120
4.4	Discussion .....	124
4.5	Methods.....	130

4.6	Supplementary Materials .....	134
<b>Chapter 5: Telomeric G-loops stimulate Alternative-lengthening-of-telomere cancers .....</b>		<b>156</b>
5.1	Summary .....	156
5.2	Introduction.....	156
5.3	Results.....	159
5.3.1	G4 and R-loop levels are elevated in ALT+ cells compared to TERT+ cells .....	159
5.3.2	Telomeric G4 and R-loop levels are correlated with telomere length in ALT+ cells and associated with DNA damage signals.....	161
5.3.3	Telomeric G4s and R-loops are spatially linked in ALT+ cells .....	164
5.3.4	Genetic and pharmacological manipulation reveals interdependency between G4s and R-loops.....	166
5.3.5	Telomeric G4 and R-loop cooperatively stimulate ALT activity. ....	169
5.3.6	Co-treatment with a G4 ligand and R-loop stabilizer leads to the synergism of cytotoxicity in ALT+ cells .....	170
5.4	Discussion .....	171
5.5	Methods.....	176
5.6	Supplementary Materials .....	183
<b>Chapter 6: Summary and Conclusions .....</b>		<b>190</b>
6.1	The optimization of G4 methods through comparison of different tools/methods.....	190
6.2	G-quadruplexes exist in nature and are widespread in the human genome and transcriptome.....	193
6.3	Limitations of tools and methods for G4 studies.....	195
6.4	The biological role of G4s – a matter of balance? .....	197

6.5	Relationship between G4s and other nucleic acid structures.....	200
6.6	Future considerations and applications.....	201
<b>Bibliography .....</b>		<b>204</b>
<b>Appendices.....</b>		<b>244</b>
Appendix A - Cellular detection of G-quadruplexes by optical imaging methods		
<i>(Current Protocols in Cell Biology, 2018)</i> .....		
		244
A.1	Summary .....	244
A.2	Introduction.....	245
A.3	Detection of Quadruplexes in cells: Confocal vs Two-photon Microscopes, Fixed vs Live-cell imaging .....	251
A.4	Detection of Quadruplexes <i>in cellulo</i> : Real-time Monitoring in Live Cells.....	263
A.5	Commentary .....	268
A.6	References.....	272
Appendix B - Transcriptomic Profiling of G-quadruplexes using G4RP: a detailed protocol and optimization guide .....		
		278
B.1	Summary .....	278
B.2	Introduction.....	278
B.3	Materials.....	285
B.4	Procedures.....	290
B.5	Troubleshooting .....	302
B.6	Anticipated Results .....	303
B.7	References.....	307

Appendix C - Genome-wide detection of G-quadruplexes and the effects of G4 ligands on transcription ..... 310

- C.1 Brief Introduction and Methods..... 310
- C.2 Preliminary ChIP-seq Results ..... 312
- C.3 ChIP-seq and RNA-seq intersection ..... 315
- C.4 Effects of G4 stabilization on transcriptome – a time course study ..... 317
- C.5 Future considerations for G4-ChIP..... 319
- C.6 References..... 319

## List of Tables

Table 1.1 Properties of Canonical G4 structures. ....	21
Table 1.2. CD Profile of different nucleic acid secondary structures. ....	32
Table 1.3. List of selected tools for studying G4s .....	38
Table 1.4. List of G4 helicases and their G4-related functions. ....	47
Table 1.5. List of non-helicase G4-interacting proteins .....	48
Table 1.6. List of G4 helicase-related human syndrome .....	56
Table 6.1 Optimization of G4 techniques through parallel comparisons of different combinations of tools with methods .....	192

## List of Figures

Figure 1.1. Base-base interactions. ....	6
Figure 1.2. Components of a G-quadruplex structure.....	13
Figure 1.3. Guanine and G-quartet Stacking modes and energetics. ....	16
Figure 1.4 Examples of noncanonical G4 structures. ....	22
Figure 3.1 Principles of the twice-as-smart quadruplex ligands.....	85
Figure 3.2 Multiphoton images with N-TASQ. ....	89
Figure 3.3 Quantitative analysis of the N-TASQ cellular targets. ....	91
Figure 3.4 Implementation of the red-edge effect for fluorescence cell imaging.....	92
Figure 3.5 Implementation of the red-edge effect for live cell imaging (part 1). ....	94
Figure 3.6 Implementation of the red-edge effect for live cell imaging (part 2). ....	97
Figure 4.1 Characterization of G4-specific affinity of BioTASQ.....	111
Figure 4.2 Isolation of G4 targets from human cell extracts using G4RP.....	115
Figure 4.3 Characterization of the baseline level of G4-RNA landscape using G4RP-seq.....	119
Figure 4.4 Characterization of the ligand-induced changes in the G4-RNA landscape. ....	123
Figure 4.5 Proposed model of G4 structural equilibrium of lncRNAs in the presence of G4-stabilizing and G4-destabilizing factors.....	126
Figure 5.1 ALT+ cells harbour higher level of telomeric G4s than TERT+ cells. ....	160
Figure 5.2 Levels of telomeric G4s and R-loops are higher in ALT+ cells and are broadly associated with telomere length and coupled with DNA damage signals. ....	163
Figure 5.3 TelG4 and TelR-loops are spatially linked.....	165
Figure 5.4 RNaseH-mediated reduction of R-loops decreases TelG4 levels.....	167

Figure 5.5 Small molecule-mediated inductions of TelG4s and TelR-loops show a reciprocal relationship..... 168

Figure 5.6 Induction of G4 and R-loops cooperatively stimulate ALT activity and leads to synergism of cytotoxicity..... 171

Figure 5.7 Proposed model of telomeric G-loop formation in ALT. .... 175

Figure 6.1 Balance of G4 formation (G4 homeostasis), depicted as a lever, for optimal cellular functions..... 200

## **List of Abbreviations**

ALT – Alternative-lengthening of telomeres

ATP – Adenosine triphosphate

ATRAX – Alpha-thalassemia/mental retardation, X-linked, Chromatin Remodeler

BCL2 – B-cell lymphoma 2

BIR – Break-induced replication

BITS – Break-induced telomere synthesis

BLM – Bloom syndrome RecQ like helicase

Bp – Base pair

BRCA – Breast Cancer, DNA Repair Associated

BRIP – BRCA1 interacting protein

BTR – BLM-TOP3A-RMI complex

CCA – C-circle assay

CD – Circular dichroism

ChIP – Chromatin-immunoprecipitation

ReChIP – sequential ChIP

CNBP – CCHC-type zinc finger nucleic acid binding protein

DAPI – 4',6-diamidino-2-phenylindole

DAXX – death domain associated protein

DDR – DNA damage response

DDX – DEAD-box helicase

DHX – DEAH-box helicase

DMS – Dimethyl sulfide

DNA – Deoxyribonucleic acid

DNMT – DNA methyltransferase

DNP - Deoxyribonucleoprotein

DSB – Double-strand break

eIF4A – Eukaryotic translation initiation factor 4A

ES – Enrichment score for BioTASQ-enrichment

ESI-MS – Electrospray ionization coupled with mass spectrometry

FANCM – Fanconi anemia group M protein

FGF2 – Basic fibroblast growth factor

FITC – Fluorescein isothiocyanate

FMRP - Fragile X mental retardation protein

FRET – Fluorescence resonance energy transfer

G4 – G-quadruplex

G4RP – G4-RNA-specific precipitation

GFP – Green fluorescent protein

GO – Gene ontology

HDR – Homology directed repair

hnRNP – Heterogeneous nuclear ribonucleoproteins

HPRT1 – Hypoxanthine phosphoribosyltransferase 1

HR – Homologous recombination

HRE – Hexanucleotide repeat expansion

ICC – Immunocytochemistry

IRES – Internal ribosome entry site

K<sup>+</sup> – Potassium

KIT – KIT proto-oncogene, receptor tyrosine kinase

KRAS – KIT proto-oncogene, GTPase

LD – Lethal dose

LED – Light-emitting diode

lncRNA – Long non-coding RNA

MALAT1 – Metastasis associated Lung adenocarcinoma transcript 1

MAZ – Myc-associated zinc finger protein

MeOH – Methanol

MiDAS – Mitotic DNA synthesis

MMP – Matrix metalloproteinase

MUS81 – Structure-specific endonuclease subunit

MYC – MYC proto-oncogene, BHLH transcription factor

Na<sup>+</sup> – Sodium

NAI – 2-methylnicotinic acid imidazolide

ncRNA – Non-coding RNA

NMR – Nuclear magnetic resonance

NPM1 – Nucleophosmin

NRAS – NRAS proto-oncogene, GTPase

Nt - Nucleotide

OB-fold – Oligonucleotide/oligosaccharide-binding

PARP – Poly (ADP-ribose) polymerase

pG4 – Potential G4-forming sequence

PIF1 – PIF1 ATP-dependent 5'-3' DNA Helicase

PLA – Proximity ligation assay

PML – promyelocytic leukaemia

POT1 – Protection of telomeres protein 1

qPCR – Quantitative polymerase chain reaction

REE – Red-edge effect

REP – Red-edge probe

RIP – RNA-immunoprecipitation

R-loop – DNA/RNA hybrid

RMI – RecQ-mediated genome instability protein

RMRP – RNA component of mitochondrial RNA processing endoribonuclease

RNA – Ribonucleic acids

RNP – Ribonucleoprotein

RPA – Replication protein A

RPL – Ribosomal protein

RPPH1 – Ribonuclease P RNA complement H1

RT – Reverse transcription

RTEL1 – Regulator of telomere elongation helicase 1

scFv – Single-chain variable fragment

SF - Superfamily

SP1 – Specificity protein 1

SRC – SRC proto-oncogene, non-receptor tyrosine kinase

SRSF – Serine and Arginine Rich Splicing Factor

TASQ – Template-assembled synthetic G-quartet

TDS – Thermal differential spectra

TelG4 – Telomeric G-quadruplex

TelR-loop – Telomeric R-loop

TERRA – Telomeric repeat-containing RNA

TERT+ – Telomerase-positive

TLS – Telomere length and silencing protein

TOS – Threshold overlap score

TP53 – Tumor protein p53

TPP1 – now known as ACD, Shelterin complex subunit and telomerase recruitment factor

TRF2/ TERF2 – Telomeric repeat binding factor 2

UV - Ultraviolet

VEGF – Vascular endothelial growth factor

WRN – Werner Syndrome RecQ like helicase

XIST – X inactive specific transcript

XPB/XPD – Xeroderma pigmentosum group B/D

XPF – Xeroderma pigmentosum group F

## Acknowledgements

This incredible PhD journey would not have been possible without ample support; thus, I have several individuals I would like to specifically thank.

First, I would like to thank my research supervisor (Dr. Judy Wong) for continually pushing me to sharpen my thinking and to produce high-quality work of strong scientific vigor. I would like to thank my supervisory committee (Drs. Brian Rodrigues, Corey Nislow, Peter Lansdorp, Adam Frankel, Paul O'Shea), whose expertise have been invaluable for guiding my doctoral work. I would also like to thank several research collaborators, including Drs. David Monchaud (Université de Bourgogne, France), Gordon Robertson (BCGSC) and Peter Stirling (BCCRC), for demonstrating the value of collaborative efforts in tackling tough scientific questions. I would like to express my gratitude to UBC and the Faculty of Pharmaceutical Sciences for providing the resources necessary for my graduate studies.

I have been very fortunate to have been surrounded by brilliant individuals who have supported me during my doctorate training. Thus, I would like to thank former (Jialin, Connor, Joanne, Alice, Diego) and current (Kat) Wong Lab members and volunteers (Ashley, Cecile) for their help during this period. I would like to thank my other officemates (Monica, Sophia) for their warm company. I am also grateful to several other graduate students for providing a friendly and supportive environment.

From high school to BScH to MSc to PhD, I have met so many wonderful individuals who have been instrumental in guiding me onto the path of scientific research. Thus, I am grateful to all my previous mentors. I would also like to thank all my friends who have supported me in both Vancouver and Toronto. Finally, I would like to thank my parents for supporting and putting up with me all these years.

***This work is dedicated to***  
*my family and friends for their endless support and encouragement.*

# Chapter 1: Introduction

## 1.1 Introduction to Nucleic Acid Structures

Nucleic acids are essential to all lifeforms and the most important biomolecules. They make up genetic materials, but also play roles as energy currency (adenosine triphosphate), hormonal messenger (cyclic AMP) and cofactors in enzymatic reactions (e.g. NAD<sup>+</sup>). Nucleic acid structures are divided into multiple levels, increasing order of magnitude: primary, secondary, tertiary, quaternary and beyond. Each level of organization can have an impact on the structural formation of the next level. Secondary and tertiary structures are crucial as they determine the 3-dimensional structure of the nucleic acid. The secondary structure of double helix (duplex) is the most predominant in nature, and thus considered the canonical structure. However, duplexes are not the only structures formed in nature, as scientists have also uncovered several other noncanonical structures. These noncanonical structures are typically much less thermodynamically stable than duplexes. One notable exception is that of the G-quadruplex (G4) as it is the only noncanonical structure with stability that can rival even that of the duplex under certain conditions (e.g. high concentration of cations, long stretches of guanines). Despite their stability and abundance, the biological roles of G4s remain mostly elusive. In my doctoral work, I aim at studying G4s to decipher their relevance in human biology using newly developed tools and methods. The study of G4s requires an understanding of the fundamental laws that govern all structures of nucleic acids. Thus, this section provides a brief overview of the fundamental concepts behind the physical-chemical interactions that make all these structures possible.

### 1.1.1 Primary Structure

The basic primary structure constitutes the linear sequence of nucleotides, basic building blocks of the polynucleotide (i.e. DNA and RNA). There are three components in a nucleotide: sugar (ribose or deoxyribose), heterocyclic bases and a phosphate group. In DNA, the nucleotide consists of deoxyribose sugar coupled with four different heterocyclic bases: adenine, guanine, cytosine, and thymine. RNA contains ribose sugar and uses the uracil base, instead of thymine. These bases are further classified under its ring systems where purines (adenine, guanine) have a two-ring system, and pyrimidines (cytosine, thymine) have a one-ring system (1).

#### 1.1.1.1 Nucleotide structure

Like any other molecule, the nucleotide's 3-dimensional structure is dependent on their bond lengths, bond angles, and rotations of groups of atoms around the bond. The partial double bond nature due to bases' properties as resonance structures drive them to adopt a planar rigid conformation. Another vital parameter to consider is the torsion angle, representing the rotational freedom, especially in the sugar-phosphate backbone. Due to steric hindrance, only a specific range of torsion angles are allowed in nucleotide, restricting particular conformations. Unlike the heterocyclic bases, the sugar is highly flexible and not planar; therefore, it can adopt two primary envelope conformation (also known as sugar puckering mode): C3'-endo (North puckering) and C2'-endo (South puckering) (1). C3'endo is more prevalent in A-form helix, and C2'-endo is more prevalent in B-form helix. As a result, the puckering mode contributes to preferences of DNA and RNA for specific secondary structures (i.e. B-form and A-form, respectively). The glycosyl bond between the base and sugar also provides another structural variation to consider with two main conformations: *anti* (bulky part of the base pointing away from the sugar) and *syn* (bulky part is closer). *Anti* conformation is generally favoured over that of *syn*, though there is an

exception for guanine, which can also shift to the *syn* conformation. Chemical modification of bases and sugar can also shift *syn/anti* equilibrium (1).

### **1.1.2 Secondary Structure**

The interactions between bases determine the secondary structure of nucleic acids. There are two kinds of interactions between bases: horizontal base-base interaction, which occurs in the plane of the bases due to the hydrogen bonding, and vertical base-base interaction (2), which is perpendicular to base planes due to the base stacking interaction, hydrophobic effects and London dispersion (Fig. 1.1).

#### **1.1.2.1 Horizontal Base-Base Interaction: Hydrogen Bonding**

A fundamental horizontal base-base interaction is base-pairing, where hydrogen bonding occurs between bases (2). Hydrogen bond formation is also cooperative in which charges of atoms involved become more polarized, leading to increased affinity for hydrogen bonding in surrounding atoms (3). The charge densities distribution dictates the capability of hydrogen bonding within bases. Hydrogen attached to amino groups in adenine, guanine, cytosine exhibits a weak acidity and thus act as hydrogen bond donors. "Amide-like" hydrogen atoms at uracil N<sub>3</sub> and guanine N<sub>1</sub> can also act as donors. On the other hand, hydrogen bond acceptors include N<sub>1</sub>, N<sub>3</sub> and N<sub>7</sub> in adenine, N<sub>3</sub>, N<sub>7</sub> and O<sub>6</sub> in guanine, O<sub>2</sub> and O<sub>4</sub> in uracil/thymine, and N<sub>3</sub> and O<sub>2</sub> in cytosine (1). The stability of specific base-base interactions is dependent on the number of hydrogen bonds and electronic complementarity; therefore, G-C base-pairs have the highest stability with three hydrogen bonds. Changes in pH can also cause protonation/deprotonation of bases, which can affect hydrogen bonding. Additionally, relocation of a proton leads a mixed population of species, tautomers, of which a shift in its equilibrium can impact hydrogen-bonding partners. Base-pairing displays two types of twofold symmetry: parallel or antiparallel.

In Watson-Crick base-pairing (often referred to as canonical base-pairing), which is the most predominant type of hydrogen bonding in nucleic acids in which adenine hydrogen-bonds with thymine/uracil and cytosine hydrogen-bonds with guanine, an antiparallel orientation is naturally adopted (with some exceptions as parallel orientation).

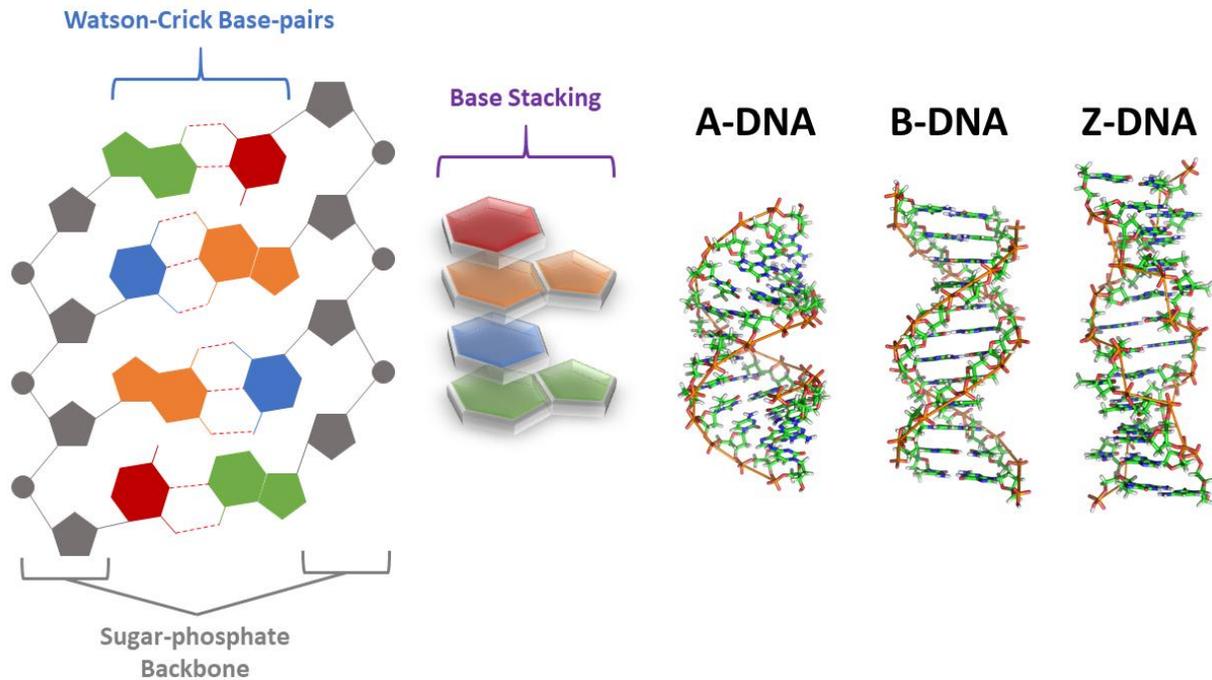
There are other noncanonical base-pairing modes. One such noncanonical base-pairing is the Hoogsteen base-pairing formed by Hoogsteen bonding, first discovered by Karst Hoogsteen, who found a disorder in the 1-methylthymine:9-methyladenine crystal structure in which 10-13% of 1-methylthymine were rotated  $180^\circ$  about its  $C_6-N_3$  axis (4). This unique bond formation allows the  $N_7$  of a purine (rotated  $180^\circ$ ) to act as a hydrogen bond acceptor (1). In general, for both Watson-Crick and Hoogsteen base-pairs, at least two cyclic hydrogen bonds are required for stability. Hoogsteen bonding greatly expands the number of possible structures by allowing for a more twisted non-linear arrangement. Thus, noncanonical nucleic acid structures other than the Watson-Crick duplex, including triplexes, G-quadruplexes and i-motifs, can form naturally (5). Another type of non-Watson-Crick base pairing of importance is the wobble base-pair, which allows base-pairing of G-U, I-U, I-A and I-C, which are essential for RNA structures, especially that of tRNAs (6).

#### **1.1.2.2 Vertical Base-Base Interaction: Base Stacking, Hydrophobic Interactions and London Dispersion**

While horizontal base-base interactions are interactions in the plane with the bases, vertical base-base interactions are analogous to a roll of coins stacked up. Base stacking (also known as  $\pi$ -stacking) is a non-covalent interaction between aromatic rings (1,2,7). Due to the partial double-bond properties (alternating single and double bonds) of the bases' aromatic rings, the  $\pi$  orbital overlaps forming  $\pi$  conjugation, which is very stable due to the greater freedom of movement of

electrons (8). This system creates  $\pi$ -electron clouds on either side of the aromatic ring plane. The  $\pi$  orbitals of adjacent bases can further overlap, giving greater freedom for electrons; thus, more base stacking would provide higher stability. A twist due to the slight displacement or tilt between the bases can maximize overlap and overcome steric hindrance (9). This twist's unevenness gives rise to major and minor grooves, alternating pockets in the sugar-phosphate backbone. These twists are not only present in Watson-crick duplex but also other noncanonical structures as well. Other than the  $\pi$ -electron systems, hydrophobic interactions and London dispersion are key contributors to the vertical base-base interactions. Since water is a significant component of life, hydrophobic interactions are essential to the proper formation of nearly all biological structures, including proteins and nucleic acids. Hydrophobic interaction promotes interactions between nonpolar groups, which contributes to the stability of secondary and tertiary structures. Water molecules remake new hydrogen bonds creating an ice-like cage or clathrate cage around the hydrophobic molecule (10). However, based on the law of thermodynamics, this formation is unfavourable due to the loss of entropy. If the hydrophobic molecules aggregate, this releases the ordered water molecules from clathrate cages, increasing overall entropy. While hydrophobic interactions significantly contribute to base stacking, they do not explain base preferences for stacking, such as seen for purines over pyrimidines. Thus, another force to consider is the London Dispersion force (also known as Van der Waals forces or dipole-induced dipole moments), the weakest intermolecular force but nonetheless a contributor to vertical base-base interactions (1). Asymmetric electronic charge distribution within atomic groups from electron fluctuation creates temporary dipoles that polarize neighbouring molecules, producing parallel dipoles and attracting each other. Overall, vertical base-base interactions result from a

combination of contributors including  $\pi$ -electron system, hydrophobic interactions and London dispersion forces.



**Figure 1.1. Base-base interactions.**

(Left) Watson-Crick base-pairing horizontal base-base interaction (Center) Base-stacking vertical base-base interaction (Right) Double-helix polymorphs: A-DNA, B-DNA, and Z-DNA.

### 1.1.3 Types of DNA and RNA Secondary/Tertiary Structures

The horizontal and vertical base-base interactions allow nucleic acids to form a wide range of secondary/tertiary structures. While the most dominant form of secondary structure is the double helix (duplex), other noncanonical structures are also essential and can play critical biological roles.

#### 1.1.3.1 Canonical Helical Structure – the Double Helix and its Polymorphs

One of the most significant discoveries regarding nucleic acid structures was the DNA double-helix model by the works of James Watson and Francis Crick (11). Multiple observations made the model possible, including Chargaff's base-pairing data, Astbury's X-ray photographs,

Gulland's electro-titration studies, and Franklin/Wilkin's X-ray diffraction data (12). We now know that this model represents only one type of double-helical structure, the B-form DNA. DNA can take a wide range of double-helical structures from A-, B-, C-, D-, E-, P-, S- and Z-forms, whereas RNA is restricted to form A-type (12-13). A-, B, and Z-forms are the only ones found currently in nature, while the other types have only been observed in a synthetic setting. The major types, A-forms and B-forms, are both right-handed helices but different in sugar pucker mode (C3'-endo in A-type and C2'-endo in B-type) and base stacking overlaps (12). In general, B-types are more structurally flexible. The twist of B-type DNA also forms the minor (narrower) and major grooves (wider and displays more interaction potential). On the other hand, A-types are more compressed with deeper/narrower major groove and wide/shallow minor groove. DNA structures are highly polymorphic and can shift between A-type and B-type depending on the hydration level of which dehydration conditions favour A-type. Z-type DNA is very peculiar as it still has Watson-Crick base-pairing but has a left-handed helix, instead of the more common right-handed helix (12). Overall, the type of double-helical structure depends on DNA or RNA, the hydration condition, and other environmental factors (i.e. different levels of salt and other compounds).

#### **1.1.3.2 Noncanonical Helical Structures – Triple helix, G-quadruplex and i-motif**

Homopolymer nucleotides are polynucleotides with consecutive runs of the same base. While usually less thermodynamically stable, bases can associate with bases of the same type, which allows the formation of noncanonical helical structures (14). For example, poly-A tails of mRNA and poly-C RNAs can adopt single-stranded helices.

Triple-helix DNA (or triplex DNA or H-DNA) can be formed from polypurines or polypyrimidines when a third strand associates with the B-type duplex with mirror repeat

symmetry via Hoogsteen bonding (15). Two strands associate by Watson-Crick base-pairing, whereas the other associate with one of the two strands via Hoogsteen bonding. This interaction forms two types of orientation: antiparallel motif formed by triplets G:G-C (guanine associating with G-C duplex) and A:A-T, and parallel motif formed by triplets C+:G-C (requiring protonated cytosine to associate with the G-C duplex) and T:A-T. The third strand can either originate from the same strand (intramolecular) or different strand (intermolecular). Triplex DNA has been associated with gene expression regulation and genome instability (15). Furthermore, triplexes can form unique biological structures, known as displacement loop (D-loop), found at the telomere and locations of DNA repair. Additionally, triplex formed in long non-coding RNA (lncRNAs) may increase RNA stability (15). Other than triple-stranded structures, nucleic acids can also take on four-stranded structures.

The most well-studied noncanonical nucleic acid structure is the four-stranded structures, G-Quadruplex (G4), which has gained considerable attention in recent years. The next section of the chapter discusses G4s in more detail (Section 1.2). In short, G4s rely on the ability of guanines to form noncanonical Hoogsteen base-pairing with themselves. Another four-stranded structure of recent interest is the i-motif, which is similar to G4 but is instead formed by cytosines. A caveat of i-motif is that it requires hemi-protonated cytosines ( $C^+$ ) to form Hoogsteen bonds between C- $C^+$  and thus requires a lower pH (~5-6) for stability (although formation is still possible at near-neutral pH) (16). Generally, at least five consecutive cytosines are required to form a stable i-motif. While DNA and RNA can both form stable triplexes and G4s, i-motif RNAs are much less stable than its DNA counterpart.

DNAs and RNAs can cooperate to form a duplex, triplex, and quadruplex of DNA/RNA hybrid type. DNA/RNA hybrid, R-loop, is formed by DNA-RNA hybridization on one strand

displaces the opposing DNA, forming a displaced loop region (17). DNA/RNA hybrid triplexes may play a role in lncRNA functions by allowing proper association with genomic DNA targets (18). Likewise, DNA/RNA hybrid G4s, while less characterized, have also been observed.

### **1.1.3.3 Other Nucleic Acid Structures**

The diversity of nucleic acid structures extends beyond the ones mentioned above. Complex branched structures such as the three-way junctions (3WJ), and four-way junctions (4WJ or cruciform) can form in sequences containing inverted repeats (19). 3WJ may have functional roles in replication (i.e. replication fork) in DNA and splicing and translating RNA. On the other hand, a well-known DNA 4WJ is the Holliday junction, an essential intermediate in homologous recombination.

RNAs can take a wide range of other structures, which can be essential for proper function. The most common one is the stem-loop, formed when the RNA strand folds back onto itself and forms a duplex segment or "stem" with the unpaired segment forming the "loop." Co-axial stacking, where two RNA duplexes interact to form a contiguous helix, allows more complex tertiary RNA structures to form, including the two most common motifs: the pseudoknot and kissing stem-loop (20). Pseudoknots form between two stem-loops in which half of one stem intercalates between two halves of another stem, forming a knot-like shape. Kissing stem-loop forms when the unpaired regions of two stem-loops base pair with each other. Proper formation of these RNA structures is necessary for forming higher-level order structures through their interaction with other nucleic acids or proteins, which can then mediate certain biological functions.

#### 1.1.3.4 Quaternary Structure and Beyond

Quaternary structures and higher-ordered structures result from nucleic acid-protein interactions to form nucleoproteins with essential biological functions (1). The most well-characterized example of a deoxyribonucleoprotein (DNP; a complex of DNA and protein) is the nucleosome, a subunit of the chromatin, in which the nuclear DNA associate with the octameric histone proteins. On the other hand, an example of ribonucleoprotein (RNP; a complex of RNA and protein) is the ribosome, a complex structure composed of ribosomal RNA (rRNA) and ribosomal proteins, forming the core component of the translational apparatus. When studying nucleic acid structures, it is also important to consider their interactions with other molecules and biomolecules, particularly those of proteins and the environment (Eg. salt condition, hydration conditions, pH level).

### 1.2 Brief History of G4s and Key Discoveries

While G-quadruplex (G4) is the most well-characterized non-duplex secondary/tertiary nucleic acid structure, much of our current knowledge has been obtained only in the last two decades. Evidence of G4s predates the discovery of the Watson Crick duplex, and the first report on G4 came from the observation made by a Norwegian chemist Ivar Bang in 1910, where he found that guanylic acid forms gels at high concentrations (21). At that time, many biochemists/chemists considered this phenomenon merely an *in vitro* artifact due to the artificial conditions in test tubes. It was not until 1962, when Gellert's team worked with fibre X-ray diffraction data on the guanylic acid, that the gel-forming property of guanylic acid was explained (22). Shortly after, both DNA and RNA were confirmed to be able to adopt this G-quartet structure. However, due to the attention drawn by Watson and Crick's duplex discovery,

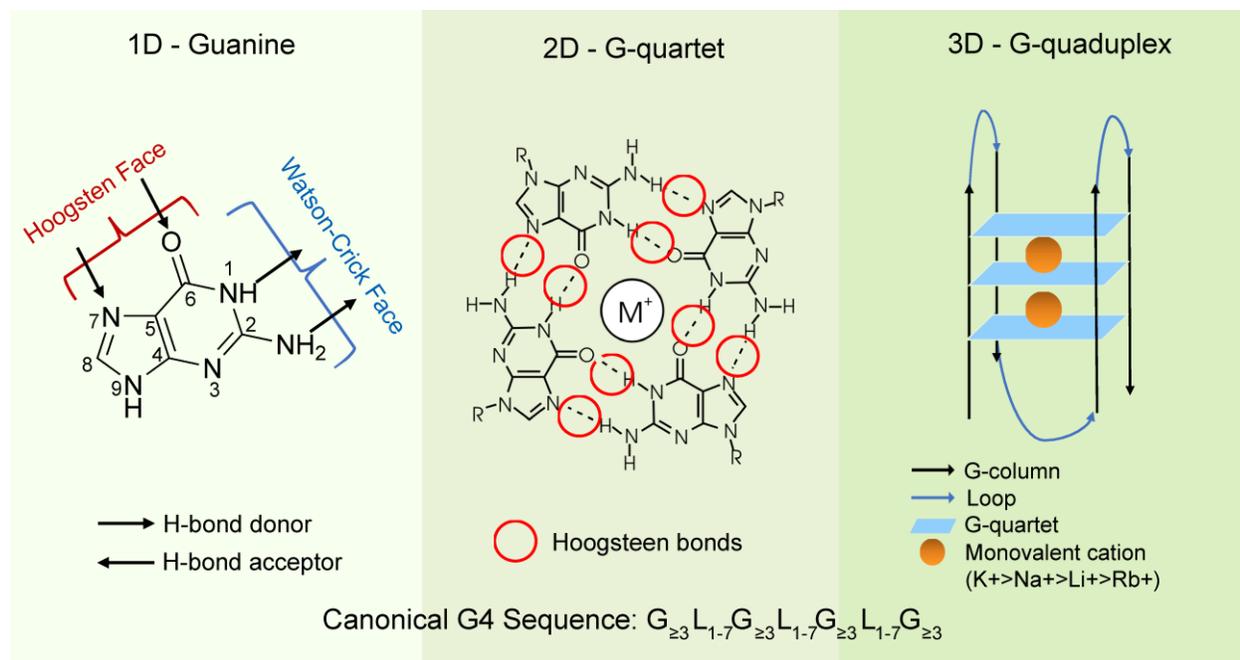
research in G4s had been relatively slow during that period. Interest in G4 was rekindled in the late 1980s and early 1990s when G4s were found in DNA sequences of biological importance. Multiple groups discovered that G-rich telomeric sequences could form a four-stranded structure known as a G-quadruplex (23-25).

Around the same time, the rapidly expanding telomere and telomerase research also contributed to an early interest in G4s. In the early 1990s, chemical compounds that can stabilize G4s, referred to as G4 ligands, were shown to inhibit telomerase activity *in vitro* (26). The relevance of telomerase in cancer led to a heightened interest in the chemistry field to find more specific and potent G4-stabilizing ligands with anti-cancer potential. To understand ligand binding, the parallel investigation into the different types of G4 structures uncovered a wide range of possible conformations and structural polymorphism in the 1990s (27-28). G4 structural diversity continued to be a popular topic due to discoveries of several noncanonical G4 structures, greatly expanding the repertoire of possible G4 conformations. In the early 2000s, folded G4 structures were then found to inhibit transcription of an oncogene, *MYC*, which spurred interest in a G4-mediated anti-cancer mechanism other than telomerase inhibition (29). Helicases, such as BLM (30) and WRN (31), were soon discovered to have the ability to unfold G4s, showing that there are biological mechanisms for downregulating these structures. The successful sequencing of the human genome in 2001 allowed G4 studies on a genomic scale (32). While the telomere contains a significant portion of G4-rich repeats, other genome regions also contain potential G4-forming sequences. Screening of different sequences for the stable formation of G4s in Neidle's group allowed them to define the general canonical G4 sequence (33). Using this G4 sequence and bioinformatics analyses on the human genome, Balasubramanian's group found more than 375000 potential G4-forming (pG4) sequences (34-

35). Importantly, they found that many of these pG4 sequences were enriched in oncogene promoters, further highlighting the therapeutic potential of G4 ligands against cancer. While enthusiasm for research in G4 continued to grow, the lack of tools for G4s made it challenging to study these structures in a biologically meaningful manner. Furthermore, there was still much skepticism as to whether G4s exist in cells or whether they are merely *in vitro* artifacts. Thus, the next phase of research, initiated in the 2010s, focused on finding tools, such as probes and antibodies, for detecting G4s in a cellular context. Two candidates, IH6 and BG4, soon emerged. IH6 is a murine monoclonal antibody produced via hybridoma technology (36), while BG4 is a single-chain fragment antibody isolated from the phage display library through selection from a panel of folded G4 oligonucleotides (37). Several chemical probes were also developed for detecting G4s. While these tools and methodologies are useful, they all have certain limitations. Therefore, there is still an ongoing effort to detect G4s in an unbiased or at least less-biased manner.

### **1.3 Structure of DNA and RNA G4s**

The combined observation from the X-ray diffraction data and gel mobility shift assays first confirmed that G4s are four-stranded structures (22, 23). The geometry of the G4 structure can be visualized by dividing into three parts in terms of dimensions: 1) first dimension being a guanine base; 2) the second dimension being a G-quartet unit; 3) the third dimension being the overall G4 four-stranded columnar structure (Fig. 1.2).



**Figure 1.2. Components of a G-quadruplex structure.**

(Left) 1-dimension: guanine base (Middle) 2-dimension: G-quartet unit (Right) 3-dimension: overall G-quadruplex structure

### 1.3.1 1<sup>st</sup> Dimension – Guanine base

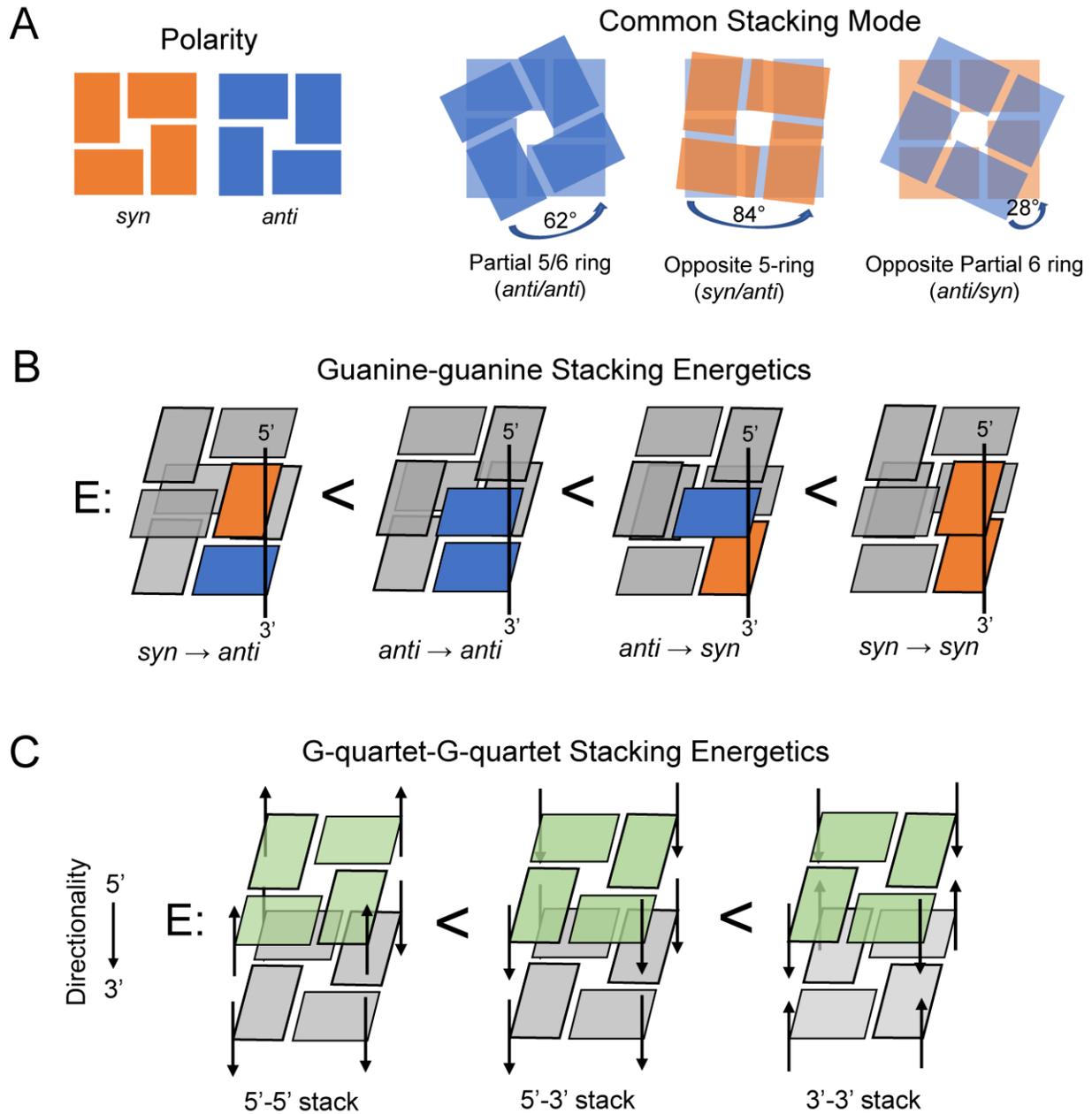
The absolute prerequisite for the formation of G4 is the presence of guanine. Guanine, unlike other bases, has a unique property that allows hydrogen bonding on two adjacent surfaces, the Watson-Crick and the Hoogsteen face. Guanine contains two hydrogen bond acceptors (at the N7 and O6 positions) and two hydrogen bond donors (at the N1 and N2 positions), which can participate in Hoogsteen bonding (1). In the 3D structure of G-quadruplex, the guanine base can represent the zero-dimensional object (0D) as a point of interest. The interaction between a pair of guanines through two Hoogsteen bonding creates the 1D structure.

### 1.3.2 2<sup>nd</sup> Dimension – G-quartet

To understand the 2<sup>nd</sup> dimension (2D) of a G4, envision a square composed of four 1D objects (i.e. guanine-guanine interactions). Four guanines align in a square conformation in which N1

and N2 of one guanine form hydrogen bonds with O6 and N7 of another guanine, respectively, at an average distance of 2.84 Å for the N1-O6 bond and 2.87 Å for the N2-N7 bond (38) (Fig. 1.2). These pairings create 8 Hoogsteen hydrogen bonds between the four guanines to form the G-quartet, the basic planar unit of a G4 structure. Like other nucleic acid bases, guanines can adopt two different conformations based on the orientation of the base with the sugar: *syn* (same side) or *anti* (opposite side) (1). However, unlike other bases which greatly favours the *anti*-conformation, guanine can stably adopt *syn*-conformation as well. Computational modelling and prediction allow calculations for energetically favourable structures defining preference for particular conformations. The four guanines in the G-quartet can theoretically take on any of the six combinations of *anti*- and *syn*- (4 *anti*-, 4 *syn*-, 3 *anti*-/1 *syn*-, 1 *anti*-/3 *syn*- and 2 *anti*-/2 *syn*- (2 types: opposing or adjacent side)) which provide conformational flexibility and polymorphism (39). The presence of the *anti*- or *syn*- conformation is an essential consideration due to their influence on the sugar-phosphate backbone strand directionality. The center of the G-quartet is highly electronegative due to the four O6 carbonyl oxygen atoms. Thus, for a stable structure, a metal ion, usually a monovalent cation (i.e. K<sup>+</sup>, Na<sup>+</sup>), situates in the center. G-quartet units can stack on each other for further stability due to the vertical-vertical base interaction via base-base  $\pi$ -stacking (38), with more stacks creating a more stable structure by promoting more overlap for the  $\pi$ -electron system. Although often depicted as straight stacking in simplified diagrams, rather than stacking with exact alignment, each G-quartet is slightly shifted or staggered to relieve steric hindrance (40). Analogous to the head and tail sides of a coin, G-quartets can also have different polarities based on the direction of the Hoogsteen bond donor-to-acceptor pattern. Two G-quartets can stack in one of two ways: same-polarity (e.g. *anti/anti*) and opposite-polarity (e.g. *syn/anti* and *anti/syn*) (Fig. 1.3A). There are three common stacking modes for G4s in terms of

the overlap geometry: Partial 5/6 ring (*anti/anti*, 64° rotation), 5-ring (*syn/anti* 84° rotation) and partial 6-ring (*anti/syn*, 28° rotation) with stacking distance ranging from 3-3.6 Å. In terms of the stability, the internal energies for two guanines stacking in the 5'-3' direction from lowest to highest (or can also be interpreted as more readily formed to least readily formed) is as followed:  $\text{syn} \rightarrow \text{anti} < \text{anti} \rightarrow \text{anti} < \text{anti} \rightarrow \text{syn} < \text{syn} \rightarrow \text{syn}$  (39) (Fig. 1.3B). Thus, for most cases, *syn*-guanine (or *anti*-guanine) is often situated on the 5'-side with *anti*-guanine on the 3'-side. In terms of the entire G-quartet, for the energetics between two G-quartet stacks (or dimerization between two parallel orientations), the internal energies from lowest to highest are as followed:  $5'-5' < 5'-3' < 3'-3'$  (Fig. 1.3C). There is a higher preference for the 5' faces to stack over the 3' faces in the event of dimerization (41).



**Figure 1.3. Guanine and G-quartet Stacking modes and energetics.**

A) Polarity of the G-quartet (*syn* or *anti*) and the three stacking modes. B) Guanine stacking energetics C) Bimolecular G-quartet stacking energetics. Ranked from left to right is internal energies lowest to higher, which corresponds with highest to lowest for stability.

### 1.3.3 3<sup>rd</sup> Dimension – G-quadruplex

The three-dimensional structure of a G4 composes of three parts: G-quartet (2D planar) units, G-columns, and the loops (Fig. 1.2). The central channel is occupied by metal ions to coordinate the electronegativity of the central channel.

#### 1.3.3.1 Defining the Potential G4-forming Sequence

Specific sequences can fold more readily into G4s than others. Early studies by screening a range of oligonucleotides with different alterations in sequences revealed certain rules for forming stable G4s (38). The general G4 sequence to form a canonical G4 structure is defined as  $G_{\geq 3}L_{1-7}G_{\geq 3}L_{1-7}G_{\geq 3}L_{1-7}G_{\geq 3}$ , comprised of four G-tracts (consecutive guanines) with intervening loop regions of variable length and sequence identity. G-tracts usually require 3 or more consecutive guanines for stability, leading to stacks of 3 or more G-quartets. Shorter loop regions generally lead to a more stable structure and thus are defined as shorter than or equal to 7nt. In the last decade, noncanonical G4s have been discovered to form, which do not abide by the previously defined general G4 consensus sequence, and which have greatly expanded the number of possible potential G4-forming sequences and structural diversity. Some relatively common noncanonical G4 structures include the two-G-quartet, bulge, and long loops, all of which are usually of lower stability (42,43) (described in Section 1.3.4.2).

#### 1.3.3.2 G-column

The G-columns form the four pillar-like structures consisting of the sugar-phosphate backbone of the four strands of nucleic acids (39). In canonical G4 structures, the G-column is continuous (attached to every guanine participating in the G-quartet stacking). The strand orientation of the G-column can be the same, which would result in a parallel conformation. The G-columns can also be mixed and arranged as a (3+1) Hybrid, *aabb*-antiparallel or *abab*-antiparallel

conformation. Additionally, the G-columns can be discontinuous in certain noncanonical structures, such as G4s, with the bulge structure (43). In general, G4s formed by discontinuous G-columns are much less stable than the canonical G4s.

### **1.3.3.3 Loop**

The loop regions of the G4s can act as the connection of any two G-columns. The existence of the loop is not mandatory, and the number of loops depends on whether the G4 is intramolecular (i.e. unimolecular) or intermolecular (i.e. bimolecular, tetramolecular). Tetramolecular G4s would not require loops since each strand can supply the G-columns without any connections between them, whereas bimolecular G4 typically would have two loops (or three in special cases), each connecting two G-columns. Intramolecular requires three loops. There are three types of loops seen in canonical G4 structures, including the propeller (also known as strand-reversal, which links the adjacent parallel G-columns), edgewise (which links adjacent antiparallel G-columns) and diagonal (also known as lateral, which links non-adjacent G-columns) (44). Noncanonical structures adopt V-shaped loops and even peripheral secondary structures such as triads and duplexes (39).

### **1.3.3.4 Metal Cations of the Central Channel**

The G4 structure contains a central channel of high electronegativity, and thus cations are required for charge screening and interactions with other components of the structures (e.g. loops and grooves). Monovalent and divalent cations can fulfil this role including  $\text{Li}^+$ ,  $\text{Rb}^+$ ,  $\text{Cs}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{NH}_4^+$ ,  $\text{Sr}^{2+}$ ,  $\text{Pb}^{2+}$ , and  $\text{Ba}^{2+}$  (45). The cations' stabilizing ability on G4s from the highest to lowest stability is as followed:  $\text{K}^+ > \text{Ca}^{2+} > \text{Na}^+ > \text{Mg}^{2+} > \text{Li}^+$ . On the other hand, some divalent cations could destabilize G4s under certain conditions, including  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Co}^{2+}$ , and  $\text{Zn}^{2+}$ . Of the large list of cations,  $\text{Na}^+$  and  $\text{K}^+$  are considered the most important cations of interest due to their

high concentrations in biologically relevant environments. The metal ion plays a central role in the G4 structure's stability, which may be even greater than hydrogen bonding and  $\pi$ -stacking interactions. The key consideration for a good G4-stabilizing monovalent cation is the ionic radius. It cannot be too large, which would destabilize the G4, or too small, which is insufficient to coordinate with negative charges of the guanine O6 atoms.  $\text{Na}^+$  has a relatively smaller ionic radius and thus can situate within the plane of the G-quartet.  $\text{K}^+$  has a larger ionic radius and thus situates between the planes of two G-quartet. Despite a larger ionic radius,  $\text{K}^+$  has an advantage over  $\text{Na}^+$  by coordinating with loops and water molecules and requiring less dehydration energy cost (46).  $\text{Li}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$  are poorer G4-stabilizing monovalent cations, mainly due to suboptimal ionic radius (either too small or too large). While not as prevalent in biological conditions, divalent cations are potentially good stabilizers of G4s due to their increased charge density. However, one caveat of a divalent cation is a strong electrostatic repulsion between the divalent cations. The monovalent cation placement in the central channel of a G4 is dynamic. Despite being responsible for charge coordination and thus G4 stability, these cations can be exchanged with the bulk solution at a relatively high rate without necessarily requiring the unfolding of the G4s (46).

#### **1.3.4 Structural diversity of G4-DNA**

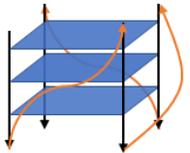
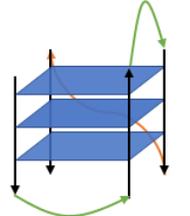
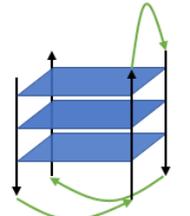
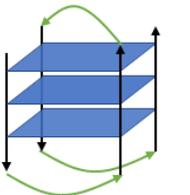
DNA G-quadruplexes (G4-DNAs) can take on a wide range of conformations. As briefly mentioned earlier, G4s can form canonical and noncanonical structures depending on whether the G-columns (or G-tracts for the sequence) are continuous or discontinuous. Canonical G4 structures include the parallel, (3+1) hybrid, *aabb*-antiparallel or *abab*-antiparallel conformations (Table 1-1). On the other hand, noncanonical G4 structures include 2-G-quartet, discontinuous G-columns (bulge and v-shape loop), long loops and peripheral structures (Fig. 1.4).

Additionally, G4-DNA can be formed as an intramolecular or intermolecular (bimolecular or tetramolecular) structure, further expanding the structural diversity of G4-DNA.

#### 1.3.4.1 Canonical G4-DNA Structures

Parallel G4 is the most stable and common G4 structure formed with 4-*anti* conformation with the same partial 5/6 ring stacking mode (39). Parallel G4, formed as an intramolecular structure, has three loops, all of which are propeller-type. (3+1) Hybrid conformation forms in two ways as either a 3-*syn*/1-*anti* or 3-*anti*/1-*syn* orientation. Thus, (3+1) Hybrid displays a mix of *same partial 5/6 ring* for same-polarity stacking and *opposite 5-ring* for opposite-polarity stacking. Loop patterns of intramolecular (3+1) Hybrid conformation are either edgewise-edgewise-propeller or propeller-edgewise-edgewise (39). Anti-parallel can also be formed in two different ways depending on whether the orientation of strands matches adjacently or diagonally: *abab*-antiparallel (also known as "chair") or *aabb*-antiparallel (also known as "basket"). The *abab*-antiparallel conformation has three loops of edgewise type and a *syn-anti-syn-anti* orientation. On the other hand, the *aabb*-antiparallel has a *syn-syn-anti-anti* orientation with three loops: two loops of edgewise and one loop running diagonally across the end G-quartet. Both types have either the opposite 5-ring or opposite 6-ring stacking type (39).

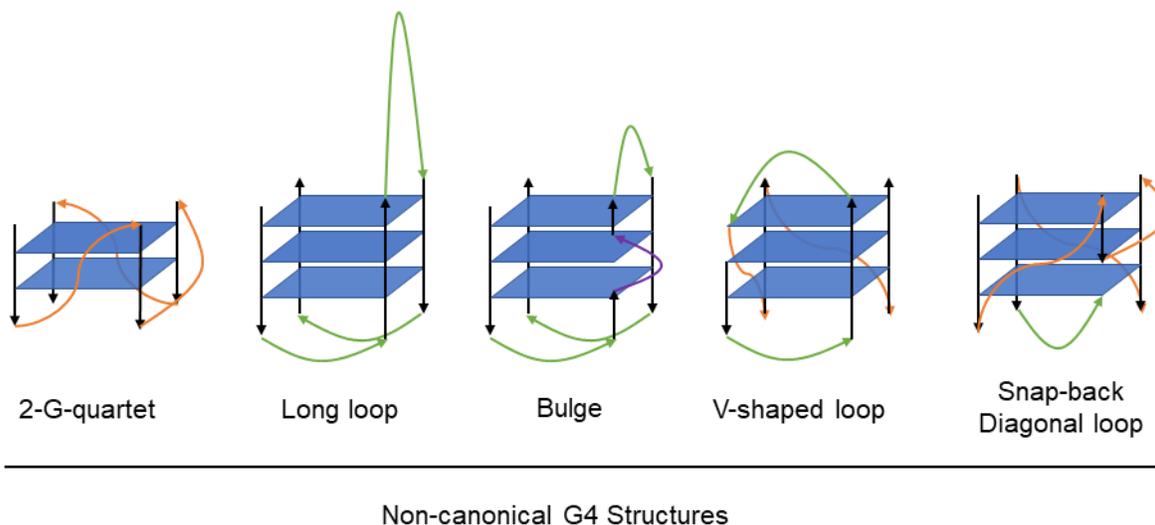
**Table 1.1 Properties of Canonical G4 structures.**

<b>G4 Structural Conformations</b>  	<b>Parallel</b>  	<b>(3+1) Hybrid</b>  	<b>Abab Antiparallel ("Chair")</b>  	<b>Aabb Antiparallel ("Basket")</b>  
<b>G-quartet Orientation(s)</b>	<ul style="list-style-type: none"> <li>• <i>anti-anti-anti-anti</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>anti-anti-anti-syn</i></li> <li>• <i>syn-syn-syn-anti</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>syn-anti-syn-anti</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>syn-syn-anti-anti</i></li> </ul>
<b>Stacking mode</b>	<ul style="list-style-type: none"> <li>• <i>-same partial 5/6 ring</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>-same partial 5/6 ring</i></li> <li>• <i>-opposite 5-ring</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>-opposite 5-ring</i></li> <li>• <i>-opposite 6-ring</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>-opposite 5-ring</i></li> <li>• <i>-opposite 6-ring</i></li> </ul>
<b>Loop pattern(s)</b>	<ul style="list-style-type: none"> <li>• Propeller only</li> </ul>	<ul style="list-style-type: none"> <li>• Edgewise-propeller-propeller</li> <li>• Propeller-propeller-edgewise</li> </ul>	<ul style="list-style-type: none"> <li>• Edgewise-Edgewise-Edgewise</li> </ul>	<ul style="list-style-type: none"> <li>• Edgewise-Edgewise-Edgewise</li> </ul>

### 1.3.4.2 Noncanonical G4-DNA structures

Noncanonical G4 structures have discontinuous G4-columns and/or short G-tracts and/or long loops, making them generally less stable while giving rise to a plethora of different possible conformations. G-tracts of only 2 guanines form the 2-G-quartet structure. While less stable, these 2-G-quartet motifs are much more common and stable under the optimized conditions (e.g. shorter loops, high cation concentration). Long loop G4 structures have loop regions that exceed the 7nt limit set for the canonical sequence. The bulge structure forms when there is one or more interrupted G-tract (forming discontinuous G-column) within the G4 sequence. For example, rather than a canonical G4 sequence GGGnnnGGGnnnGGGnnnGGG, a bulge structure can form from GGGnnnG**n**GGnnnGGGnnnGGG. A bulge loop can be displaced to allow the guanines to complete the formation of the G-quartets. Other types of structure with discontinuous G4-columns include the V-shape loop, in which the 5' end of the strand starts from the guanine of the middle G-quartet, and the snap-back diagonal loop, in which a 3' end guanine diagonally

loop to complete the end G-quartet formation. Peripheral structures can also help stabilize G4s through interactions within the loops. Non-G-quartet planar structures, such as triads, formed by the loop region, can help contribute to stacking interactions. The loops may also form stable duplex structures, which helps stabilize the linked G4 structure.



**Figure 1.4** Examples of noncanonical G4 structures.

### 1.3.5 Structural diversity of G4-RNA

Characterization of the G4-RNA structure has unveiled some surprising results. Comparisons with RNA counterparts of well-established G4-DNA sequences can be performed using thermal-melting and CD spectroscopy. Compared to G4-DNAs, G4-RNAs have a much lower structural diversity and relatively higher stability. G4-RNAs are nearly entirely restricted to parallel conformations preferring shorter loops. Some less stable G4-DNA sequences have significantly higher stability when transcribed into their G4-RNA forms (47).

#### 1.3.5.1 Key differences between G4-DNA and G4-RNA

Despite shared characteristics between G4-DNAs and G4-RNAs, the subtle differences between DNA and RNA significantly contribute to the difference seen between their conformational

flexibility and stability. The ribose 2'-hydroxyl groups in RNA leads to steric hindrance that restricts the guanines to *anti*-conformations (i.e. parallel conformation). Furthermore, this extra hydroxyl group in RNA also provides an additional hydrogen bond, which increases overall G4-RNA stability (48). The lack of the methyl group of uracil in RNA compared to thymine in DNA allows the folded G4s to be less hydrated and stable through higher hydrophobic interactions (49). Importantly, this suggests that G4-RNAs are less dependent on environmental factors than G4-DNA and have structures that may be easier to predict and target due to the parallel conformation restriction.

### 1.3.5.2 Noncanonical G4-RNA structures

While canonical G4-RNA structures are essentially all-parallel forms, there are a few rare noncanonical structures under special conditions. Of note, the two-G-quartet G4-RNA structure can be very stable but is much more dependent on  $K^+$  concentration than G4-RNAs of three G-quartet stacks or more (50). Recently, antiparallel G4-RNAs have been observed, which require the presence of 8-bromoguanosine derivatives in the sequence; however, the biological relevance is uncertain due to the specific conditional requirements (51). Also, synthetic RNA aptamers can form noncanonical G4-RNAs, including the FMRP aptamer, Spinach, Mango and related derivatives. FMRP-binding RNA aptamer showed several discontinuous G-columns and bulges, whereas RNA-aptamer fluorescent trackers (e.g. Spinach and Mango) also display interesting G4 conformations with several bulges and *syn*-guanine orientations (52).

### 1.3.6 G4-stabilizing Ligands (G4 ligands)

After discovering that sequences of biological importance such as the telomere and oncogene promoters can fold into G4s, the development of G4-interacting chemical compounds became a popular research direction for many chemists worldwide. Since nearly all G4-interacting

chemical compounds are G4 stabilizers, and only a few display G4-destabilizing effects, G4 ligands usually refer to chemical compounds of G4-stabilizing effect. Nearly 1000 different chemical compounds to date have been studied for the ability to stabilize G4s (53). Despite many known G4 ligands, only a handful of them is commonly used in biological research, and even fewer have reached clinical trials. There are still ongoing issues with these ligands related to their lack of specificity and poorly optimized pharmacokinetic profiles (likely due to high lipophilicity). Thus, designing such molecules requires careful consideration.

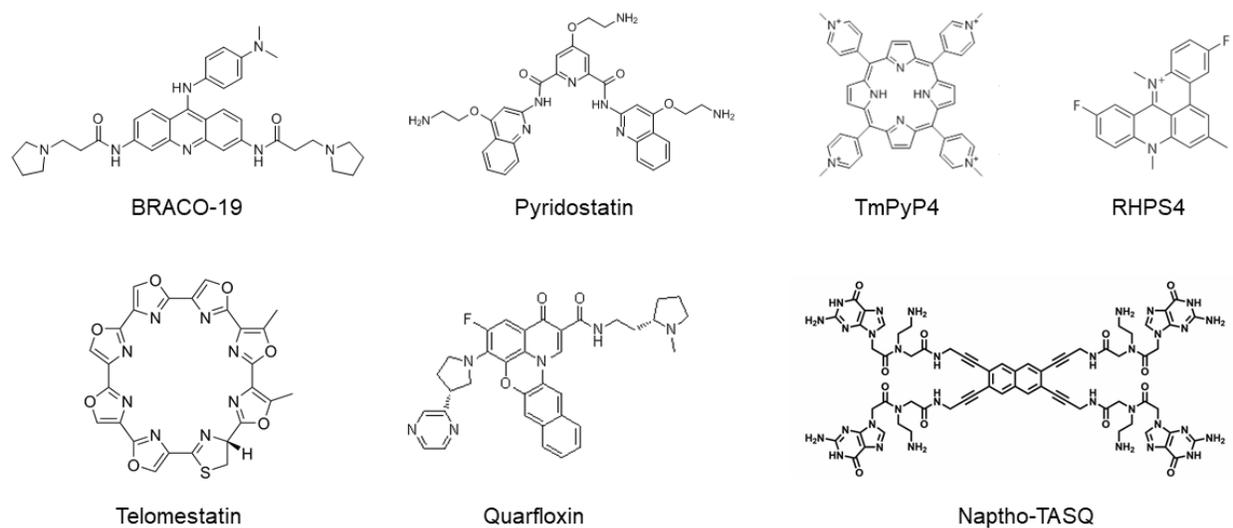
#### **1.3.6.1 G4 Ligand Design Considerations**

The high specificity of G4 ligands is essential to avoid off-target effects (especially duplex), which can cause unintended toxicities and side effects. G-quadruplexes have multiple potential interaction sites, including the end G-quartet, the middle G-quartet, the grooves, G-columns (phosphate backbone), loops and the central channel (54). The most common type of ligand-target interaction for G4 ligands is using planar aromatic rings to interact with G4s primarily through  $\pi$ -stacking with the end G-quartet. These planar G4 ligands may also, albeit to a lesser degree, intercalate between G-quartets. Positively charged amino groups can improve binding by interacting with negatively charged regions such as the groove and phosphate backbone.

Attached side chains can interact with grooves, loops and the negatively charged backbone of the G-columns. Since several of these compounds are very lipophilic, some added side chains can enhance water solubility and improve the ligand's pharmacokinetic profiles. Additionally, positively charged side chains or metal cations may also interact with the highly electronegative central channels of G4s (54).

### 1.3.6.2 Types of G4 ligands

There are no standard classifications for G4 ligands. Generally, G4 ligands have a central core of aromatic nature (or scaffold) with modifications to increase water solubility. Broadly, these chemical compounds can be classified by their chemical design: aromatic core with protonable side chain, N-methylated aromatic core, metallo-organic, neutral macrocyclic, fluoroquinolone derivatives (CX series), other "borrowed" planar compounds (which are initially developed for a different therapeutic target but later discovered to have a secondary effect of stabilizing G4s) and finally the synthetic G-quartet biomimetic G4 ligand (55-56).



**Figure 1.5.** Examples of common G4 ligands.

#### Aromatic core with protonable side chain

Aromatic core with a protonable side chain allows the molecule to be sufficiently water-soluble. A classic example of this type of G4 ligand is BRACO-19 (Fig. 1.5). BRACO-19 contains three protonable side chains that can also interact with the G4 target grooves and increase water solubility (57). Another compound, pyridostatin (PDS) (Fig. 1.5), was developed using similar strategies by having an aromatic core with amino group side chains attached to improve

electrostatic interactions (58). PDS is more selective for G4-DNA over G4-RNA, whereas a carboxylated derivative, carboxypyridostatin (c-PDS), is more selective for G4-RNA (59).

#### N-methylated aromatic core

N-methylated ligands have enhanced water solubility and electrostatic interactions without the need for additional side chains. The classic example of this type of G4 ligand is TmPyP4 (Fig. 1.5), a compound with a porphyrin core as the aromatic scaffold with four N-methylated pyridyl groups (60). TmPyP4 can interact with a G4 target through stacking with the end G-quartet or by intercalation. Unfortunately, TmPyP4 has low selectivity for G4s over duplexes, limiting its application as a G4 ligand. Several other porphyrin-based G4 ligands (TMPyPz, Se2SAP) have also been studied with improved selectivity but are difficult to synthesize (55). RHPS4 (Fig. 1.5), a pentacyclic acridinium which is also N-methylated, exhibits strong G4-stabilizing ability by stacking to the end G-quartets of the G4 target (61). Likewise, Phen-DC3, a phenanthroline analogue with a relatively simpler synthesis process, shows exceptional G4 selectivity (55).

#### Metallo-organic

To build upon the success of TmPyP4 as a strong G4 stabilizer while improving its poor G4 selectivity, metal ions were inserted into the TmPyP4 central cavity. The metal center is especially useful as it could be positioned over the cation channel to improve stacking interactions. One such compound is the Mn-TmPyP4, which shows a much better G4 selectivity over duplex DNA. Other compounds with similar setup include Ni(II)-salphen and Mn(III)-porphyrin, which also show excellent G4 selectivity (55).

### Neutral Macrocyclic

Not all G4 ligands are necessarily synthetic. Some neutral ligands have been discovered from natural sources such as telomestatin, isolated from the bacterium, *Streptomyces annulatus*.

Telomestatin is a relatively large molecule comprised of a large polyheteroaromatic ring. While it displays one of the highest G4 selectivity as neutral molecules have negligible interactions with duplex DNA, the complexity of this compound makes large-scale synthesis or harvest impractical (55).

### Quinolone

Quinolones, a common class of antibiotics, may also have G4-stabilizing effects. Quinolones contain a bicyclic core, 4-quinolone, which can act as a scaffold for G4 binding. Some of the quinolone derivatives, especially the CX series, include G4 ligands that have reached the furthest in clinical development. Quarfloxin (CX-3543) is the first G4 ligand to reach Phase II clinical trial as an anti-cancer treatment (62). It is also an RNA polymerase I inhibitor by disruption of the G4-nucleolin complex, resulting in ribosome biogenesis inhibition. Despite good toxicity profiles, quarfloxin did not proceed past Phase II due to bioavailability issues (63). CX-5461 is another quinolone with RNA Pol I inhibiting and G4-stabilizing effects. Furthermore, CX-5461 also showed selective toxicity against BRCA-deficient cancer cells, further expanding its clinical relevance (64). It is currently in Phase I/II clinical trials with favourable results and optimistic potential of reaching further into the clinical development pipeline.

### Other planar compounds with secondary G4-stabilizing effects

While several G4 ligands are novel synthetic compounds, some older drugs may have similar aromatic structures that could act as G4-binding scaffolds. The classical compound crystal violet,

used in Gram staining and as a topical antiseptic in the clinic, has been shown to have G4-binding capabilities and the potential as a G4-reporter, due to its enhanced fluorescence upon binding anti-parallel G4s (65). Well-characterized anthracyclines used in chemotherapy treatment, such as daunorubicin, doxorubicin and nemorubicin, have also been found to exhibit G4-stabilizing abilities (66). Natural alkaloids used in herbal supplements, such as berberine, have displayed G4-stabilizing capabilities (67). Intriguingly, the increasing number of these naturally-occurring G4 ligands discovered suggests that they are more ubiquitous in the environment than previously thought.

### Synthetic G-quartet

Another strategy is to use biomimetic interaction to achieve higher specificity, which led to the conception of a new class of novel G4 ligand called the Template-Assembled Synthetic G-quartets (TASQs) (68). The basic design of TASQ molecules includes a central core with four arms, each holding guanine. Upon association with a G4-target, the individual guanines self-assemble into a synthetic G-quartet, which then  $\pi$ -stack with the end G-quartet. This strategy offers high specificity due to the use of natural quartet-quartet interaction. TASQ is very flexible as the central core is changeable to accommodate different applications. Multiple TASQ generations have been developed, including DOTASQ (1<sup>st</sup> gen) (69), PorphyrSQ (2<sup>nd</sup> gen) (70), PyroTASQ (3<sup>rd</sup> gen) (71) and Naphtho-TASQ (N-TASQ; 4<sup>th</sup> gen) (72). PyroTASQ and N-TASQ can act as G4 probes due to a fluorescent moiety's attachment to the central core. In my thesis, I describe the development of novel methodologies using the two newest generations (4<sup>th</sup> and 5<sup>th</sup>) of TASQs, Naphtho-TASQ (N-TASQ) (Chapter 3) and BioTASQ (Chapter 4).

### 1.3.6.3 Other G4-interacting molecules

While not strictly defined as G4 ligands (chemical compounds that stabilizes G4), some molecules can also mimic or destabilize G4s.

#### G4 aptamers

G4 aptamers are G-rich DNA oligonucleotide, of which several members exhibit some incredible therapeutic potential. Of particular interest is the G4 aptamer AS1411, a synthetic 26-base DNA oligonucleotide with the sequence, GGTGGTGGTGGTTGTGGTGGTGGTGG, which can itself fold into a G4 and dimerize with itself by further G-quartet stacking (73). AS1411 may target cancer cells specifically by interacting with cancer-associated surface nucleolin. AS1411 went on to further testing in Phase I and II clinical trials for treating advanced cancers. While it had exceptional safety profiles, its pharmacological efficacy profiles are suboptimal with highly variable patient responses. There is still ongoing research in optimizing this aptamer for future clinical use.

#### G4-destabilizers

A chemical compound that destabilizes G4s is very rare as the specific targeting of G4s itself likely promotes stabilization. Nevertheless, a few special cases of G4 destabilization have been documented. TmPyP4, a classical G4 ligand, has been found to destabilize G4-RNAs in *MT3-MMP* mRNA (74) and *C9orf72* transcript (75). Recently, a stiff-stilbene G4 ligand showed stabilization in potassium buffer but unexpectedly showed destabilization when in sodium buffer (76). These observations emphasize that while G4 destabilization is rare, it is not impossible and can occur under specialized conditions (i.e. specific sequences and the presence of favourable environmental factors).

## Heme: the endogenous G4-interacting molecule

Fe(III) heme can bind either G4-DNA or G4-RNA to form an active G4-hemin

DNAzyme/RNAzyme that can catalyze oxygen transfer reactions (77). These complexes could be valuable biocatalysts for the development of biosensors and biomaterials. However, the biological relevance of G4-hemin complexes remains to be elucidated, although their formation does support the prebiotic "RNA world" hypothesis for the evolution of life. According to this hypothesis, noncoding RNAs (ncRNAs) may be remnants of ancient riboorganisms, RNA-based organisms, which were later replaced by the more efficient protein system. In contrast, an alternative idea of the "modern RNA world" has also been proposed due to the expanding range of functional roles observed for ncRNAs, supporting ncRNA's continued importance in modern biology (78).

### **1.4 Current strategies for studying G4s**

Strategies for studying G4 ultimately depend on the scientific question of interest. While there are numerous ways to study G4 formation in an *in vitro* environment, it becomes much more complicated to study G4s in a more biological relevant ("*in vivo*") state. Studying G4s requires a multidisciplinary approach from the combination of chemistry, biophysics, biochemistry, pharmacology, genetics, and genomics. Here, I discuss the different methods from the lowest (*in vitro* structural studies) to the highest biological evidence (*in vivo* studies).

#### **1.4.1 *In vitro* structural validation methodologies**

*In vitro* characterization of G4 is essential for initiating the beginning of the G4 field of research. Many early breakthrough discoveries were made on chemistry-based *in vitro* experiments, which is important in defining G4 formation rules. One of the earliest evidence of folded G4 was the

observation of increased electrophoretic mobility of telomeric DNA in nondenaturing gels containing monovalent cations. Electrophoretic mobility shift assay is an easy and quick method for checking for higher-ordered structures formed by oligonucleotides (25). However, detailed structural studies require more advanced techniques.

### Circular Dichroism Spectroscopy (CD)

CD is one of the most commonly used techniques for studying nucleic acid secondary structure conformations. The principle behind CD is the differential absorption profile of left- and right-handed circularly polarized light. As mentioned earlier, the different structural G4 conformations (e.g. parallel, antiparallel, hybrids) are associated with different *syn/anti* orientations of the guanines (39). These quartet stacks would give different CD spectra signatures depending on the G4 structural conformation. CD profile typically displays a range of wavelengths ( $\lambda$ , nm) and corresponding  $\Delta\epsilon$  (molar circular dichroism value, a function of the wavelength). As a result, CD spectra usually shows a few strong peaks in either the positive or negative direction. The pattern of the peaks corresponds with a particular nucleic acid structural conformation. Table 1.2 lists the CD spectra peak patterns for different nucleic acid structures (79-80). While A-form duplexes resemble some forms of parallel G4s (both with a strong positive peak at 265 nm and negative peak at 240 nm), A-form duplexes also show a strong negative peak at 210 nm, whereas parallel G4s typically show a much weaker peak instead. Other forms, including B-form duplexes, triplexes and i-motifs, are notably different from G4 structures as seen by their different peak patterns and absence of peaks at characteristic wavelengths. Conformational changes can be tracked using CD and measuring the spectra under different conditions (e.g. use of different cations and/or addition of other chemical agents in the buffer). The stability of G4s

can also be monitored during changes in temperature. While convenient, the CD does not have the highest resolution and is insufficient as the sole determinant of a G4 structure.

**Table 1.2. CD Profile of different nucleic acid secondary structures.**

Nucleic Acid Secondary Structure	CD Profile (Peak patterns) at $\lambda$ (nm): + positive peak, – negative peak			
	210	240	265	295
A-form duplex	–	–	+	
B-form duplex		–		+ (280 nm)
Triplex	–			+
i-motif	+		–	+
G4 tetramolecular parallel	+	–	+	
G4 unimolecular and bimolecular parallel	+	–	+	+
G4 unimolecular antiparallel	+	+	–	+
G4 (3+1) Hybrid	+	–	+	+

#### Ultraviolet (UV) Spectroscopy/ Thermal Difference Spectra (TDS)

UV absorbance is considered one of the classical nucleic acid structure determination methods, commonly used for generating melting curves for duplex structures. With increased temperature, the duplex UV absorbance profile displays a hyperchromic shift in which the absorption at 260 nm increases as a function of the nucleic acid's transition from a folded to an unfolded state.

However, for G4 melting studies, a wavelength of 295 nm is used as the differential absorbance for G4 is too low at 260 nm. In contrast to duplex, the UV absorbance profile of G4 displays a hypochromic shift with increasing temperature as observed by the decreased absorption at 295 nm when G4s transition from a folded to an unfolded state (79).

Thermal difference spectra (TDS) profiles can be generated by monitoring UV absorbance values at different temperatures. In most studies, a very low temperature (to ensure a folded state) and a very high temperature (to ensure an unfolded state) are chosen. The differences in

absorbance values between the folded and unfolded states across a range of wavelengths can be used to generate unique profiles for identifying different nucleic acid structures. Generally, G4s display one negative peak (at 295 nm) and two positive peaks (at 243 nm and 273 nm) (81). To differentiate from parallel G4s, anti-parallel G4s also exhibit a stronger negative peak at 295 nm. TDS is especially useful for determining G4s due to the unique G4-specific three-peak signature, which is substantially different from that of the two-peak signature seen for other structures (i.e. duplex, triplex, i-motifs) (81). Typically, a combination of CD and TDS is sufficient for validation of *in vitro* G4 structures.

### Electrospray Ionization

A more advanced technique is the electrospray ionization combined with mass spectrometry (ESI-MS), which measures charged molecule fragments to give mass-to-charge ratios (m/z). ESI-MS is especially useful for determining the number of strands in a G4 structure (e.g. unimolecular, bimolecular, trimolecular, tetramolecular) and molecularity of specific G4-binding complexes (79).

### Nuclear Magnetic Resonance (NMR) Spectroscopy

Advanced methods, including NMR and X-ray crystallography, are methods with the highest resolution possible at atomic resolution. Since X-ray crystallography does not allow structural determination in solution, the compatibility of NMR in solution may be more representative of the physiological state. Additionally, NMR spectroscopy can also track the dynamics of G4 structures. The principle behind studying G4 using 1D NMR spectroscopy is through the properties of the guanine imino protons with chemical shifts of 10-12 ppm compared to that of Watson-Crick pairing with chemical shifts of 13-14 ppm (82). For example, a reading of twelve

peaks of guanine imino protons within the range of 10-12 ppm corresponds to a G4 with three G-quartet stacks. NMR can be used to study multiple conformations, which results in several proton peaks; however, the high spectral overlaps observed in some cases can make analysis and structural determination extremely difficult.

#### Fluorescence-based ligand-G4 interaction screening

In addition to the techniques mentioned above, some methodologies were developed more specifically for screening ligand-G4 interactions. One method is the G-quadruplex fluorescent intercalator displacement (G4-FID), based on the loss of fluorescence from triazole orange from the target DNA upon competition with a G4 ligand (83). Another technique is the Forster Resonance Energy Transfer (FRET) melting assay (84). FRET requires a donor (usually FAM) and an acceptor fluorophore (usually TAMRA) attached to each end of a G4-oligo of interest. When the G4-oligo is in a folded form, energy transfer occurs, which gives off a fluorescent signal readout. The temperature-induced unfolding of the G4-oligo can disrupt energy transfer between the fluorescent molecules on either end of the oligo, resulting in a loss of fluorescence. A melting curve using the fluorescence readout can then be generated in the absence or presence of G4 ligands to evaluate for G4-stabilizing capability. A potent G4-stabilizing ligand would shift the melting temperature ( $T_m$ ) to the right (higher temperature) due to the increasing thermostability of the G4-oligo.

#### **1.4.2 *In silico* bioinformatics methodologies**

The post-genomic era brought vast opportunities to study G4 on a genomic scale using bioinformatic tools. The first implementation of a genome-wide G4 study was through using the Quadparser algorithm to search for canonical G4 sequences in the human genome, which

resulted in 375000 hits (33), many enriched in promoter regions (34). Tools developed to predict the sequences' propensity to fold into G4s include the QGRS mapper (85) and QuadBase (86). A recently developed algorithm, G4Hunter, also considers additional factors regarding sequence G-richness/skewness, which greatly expand the number of potential G4 forming sequences by 2- to 10-fold (87). While bioinformatics tools are useful for prediction, it is important to note that predicted sequences do not necessarily fold in a biologically relevant state.

### **1.4.3 *In vitro* genomics methodologies**

In search of more substantial evidence to support the high number of pG4 sequences found by *in silico* prediction tools, a method is required to generate a genome-wide map of sites where G4s can form. Since a folded G4 poses a barrier to standard DNA polymerases used in biochemical experiments, polymerase stop assays can indirectly measure G4 formation (88). A combination of polymerase stop assay and high-throughput next-generation sequencing, the new method, G4-seq, allows genome-wide G4-DNA detection in human cells (89). G4 sites can be mapped by comparing the difference in sequencing base-calling readouts between conditions that promote or disfavour G4 formation. More than 700000 G4 sites have been discovered, surpassing earlier *in silico* estimates. While shorter loops were found more likely to form G4s, non-canonical structures with long loops and bulges were also detected. Studies of G4-RNA on a transcriptome-wide scale can exploit a similar strategy. A new approach, called rG4-Seq, is based on the principle that a folded G4-RNA in a transcript can hinder its reverse transcription to a cDNA product (90). The rG4-seq generated an *in vitro* map of thousands of potential G4-RNA sequences, including several noncanonical G4 sequences with long loops, 2-G-quartet and bulges. While indirect approaches, like G4-seq and rG4-seq, have advantages of sensitive

detection of G4 sites, a fundamental limitation is that the *in vitro* conditions in these approaches are not reflective of the complex biological environments within living cells.

#### **1.4.4 *In situ* cellular imaging methodologies**

Tools, such as antibodies and probes, are crucial for imaging and affinity-purification of G4s in a direct and more biologically relevant manner. Table 1-3 lists known tools (antibodies and probes) for studying G4s.

##### G4-specific Antibodies

The first G4-specific antibody, Sty49, single-chain fragment antibody (scFv) isolated from the antibody library using ciliate telomeric repeats, allowed visualization of G4s in ciliates (91). A few attempts afterward to develop antibodies for imaging G4s in mammalian cells proved unsuccessful, including GQ1 isolated from zinc-finger libraries (92), the scFv hf2 isolated from phage display libraries (93), and designed ankyrin repeat binding proteins (DARPin) (94). Adopting the phage display selection strategy allowed for quick screening against the c-Kit G4 sequence, resulting in the development of the scFv, hf2 (93). Unfortunately, hf2 was incapable of being used for imaging and was only suitable for binding to G4-DNA *in vitro* (95). In 2013, two G4-specific antibodies allowing for cellular imaging in mammalian cells had finally been developed. BG4 is another scFv antibody, developed using the phage display library selection strategy on a series of oligos that fold into different intramolecular and intermolecular G4-DNA structures (37). This strategy resulted in the selection of an antibody of high specificity for different types of G4 structures. BG4 was also capable of detecting G4-RNAs via cellular imaging (96). Around the same time, IH6, a monoclonal mouse antibody, was developed by immunizing mice with G4-DNAs folded by the Oxy-2 sequence, from *Oxytrichia* telomere (36).

Later, another scFv antibody, D1, was developed with high specificity for parallel G4s (97). A caveat of these antibodies is their applications in only fixed cells, but not live cells, limiting their use in studying the dynamic changes in G4s in living biological systems.

#### G4-specific Chemical Probes

Other than antibodies, G4-specific chemical probes can prove to be valuable tools for studying G4s. Compared to antibodies, chemical fluorescent probes often have the advantage of being able to be used in both fixed and live cells. A modified G4 ligand, pyridostatin- $\alpha$ , allowing fluorophore attachment in cells via click chemistry, succeeded in cellular imaging in fixed cells (98). A fluorescence probe, 3,6-bis(1-methyl-4-vinylpyridinium) carbazole diiodide (BMVC), was one of the earlier probes that allowed cellular imaging in both fixed and live cells (99). Likewise, a guanidino-anthrathiophenediones (ATPDs) was used for cellular imaging for both G4-DNAs and G4-RNAs (100).

#### **1.4.5 *In vivo* cellular imaging**

Studying G4s in live cells requires the probes' ability to enter the cells efficiently. N-TASQ, a biomimetic ligand with "turn-on" signal properties upon association with a G4, readily entered cells and showed that G4-RNAs are detectable in live cells (72). DAOTA-M2, another probe capable of visualizing G4s, more specifically G4-DNAs, showed similar ability in live cells due to its cell permeability and low toxicity (101). Using fluorescence lifetime imaging microscopy with DAOTA-M2, the differences in emissive lifetime were used to differentiate between G4s and other structures (i.e. ssDNA, dsDNA). CyT, a fluorescent cyanine dye with high specificity for G4-RNAs, also allowed live-cell imaging of G4s (102). Finally, QUMA-1, a coumarin-

hemicyanine fluorophore, was used to visualize at high resolution the dynamic folding and unfolding process of G4-RNAs within live cells (103).

**Table 1.3. List of selected tools for studying G4s**

Name of Tool	Year	Type of tool	Target(s) * preferred target	Compatibility		Applications	Ref
				Fixed Cells	Live Cells		
Sty49 (scFv)	2001	Antibody	G4-DNA	Yes	No	ICC/IF	(91)
GQ1	2001	Antibody	G4-DNA	No	No	<i>In vitro</i> only	(92)
Hf2 (scFv)	2008	Antibody	G4-DNA	No	No	Genomic DNA pull-down	(93,95)
BG4 (scFv)	2013	Antibody	G4-DNA (*) G4-RNA	Yes	No	ICC/IF, ChIP-seq	(37, 96, 104, 105)
IH6	2013	Antibody	G4-DNA	Yes	No	ICC/IF	(36)
Pyridostatin- $\alpha$	2012	Probe	G4-DNA	Yes	No	ICC/IF	(98)
BMVC	2013	Probe	G4-DNA	Yes	Yes	ICC/IF, live-cell imaging	(99)
Anthrathiophenediones (ATPDs)	2013	Probe	G4-DNA G4-RNA	Yes	No	ICC/IF	(100)
Naptho-TASQ	2015	Probe	G4-DNA G4-RNA (*)	Yes	Yes	ICC/IF, live-cell imaging	(72)
DAOTA-M2	2015	Probe	G4-DNA (*) G4-RNA	Yes	Yes	ICC/IF, live-cell imaging, fluorescent life-time imaging	(101)
CyT	2015	Probe	G4-RNA	Yes	Yes	ICC/IF, live-cell imaging	(102)
D1 (scFv)	2016	Antibody	G4-DNA	Yes	No	ICC/IF, ChIP-seq	(97)
QUMA-1	2018	Probe	G4-RNA	Yes	Yes	ICC/IF, live-cell imaging	(103)

#### 1.4.6 *In vivo* genomics methodologies

While visualization of the intracellular G4s is informative, it does not provide the exact identities of these G4s, which requires some form of affinity purification technique and identification through high-throughput sequencing. Reliable genomics techniques are invaluable for studying the biology of G4 on a large scale, but such an endeavor is challenging with many potential pitfalls. The best-characterized antibody capable of genomics study is the scFv antibody, BG4,

which is compatible with the chromatin-immunoprecipitation (ChIP) assay. Genome-wide G4 mapping using ChIP coupled with sequencing (G4-ChIP-seq) revealed approximately 10000 G4-DNA sites, much lower than the previous estimates measured using the G4-seq technique (104-105). The observed G4-ChIP enrichment in nucleosome-depleted regions emphasizes the importance of opportunities (e.g. chromatin structure dynamics) for G4 formation. The D1 antibody was also compatible with ChIP-seq, revealing 8000 peaks distributed across the human genome (97). However, for G4-RNA, there is currently no G4-specific antibody compatible with the ChIP-seq-equivalent technique for RNA, the RNA immunoprecipitation and sequencing (RIP-seq).

An alternative method for G4 mapping is to perform ChIP-seq or RIP-seq with antibodies that target G4-binding proteins to infer G4-DNA or G4-RNA sites. G4-associated proteins such as a-thalassemia mental retardation X-linked protein (ATRX) (106), and xeroderma pigmentosum group B (XPB) and group (XPD) (107), have all been shown through ChIP-seq to bind to G4-like motifs in human cells. A key issue with these indirect methods is that they do not provide definitive proof of folded G4s at these G4-binding protein sites.

Due to a lack of suitable tools for studying *in vivo* G4-RNAs on a genomic scale, an *in vivo* chemical modification step with dimethyl sulfate (DMS) was used in combination with the reverse-transcription (RT)-stall sequencing (also used in rG4-seq) (108). The principle behind this method is that live cells treated with DMS lead to the chemical modification of intracellular RNA molecules. Folded G4 blocks DMS from modifying the bases, whereas unfolded regions (presumably non-G4 regions) would be modified. During the *in vitro* refolding step (using conditions that promote G4s such as high K<sup>+</sup>), DMS-modified G4-containing sequences would not be able to refold back into G4s and thus would not lead to any RT-stall sites during the

analysis of the sequencing data. Surprisingly, using this method on mouse embryonic stem cells, the entire transcriptome appeared to be nearly devoid of folded G4-RNAs. Treatment with a SHAPE reagent, 2-methylnicotinic acid imidazolide (NAI), which leads to acylation of folded G4, also showed an unfolded G4-RNA landscape in yeast. Thus, the authors concluded that G4-RNAs are globally unfolded *in vivo*, a controversial finding given all the existing evidence from cellular imaging studies (108). This discrepancy is discussed and reexamined in detail with new findings in Chapter 4.

## **1.5 Biological Relevance of G4**

The discovery of G4s in genomic sequences of high importance, including the telomere, oncogene promoters, and high repeat-regions, provides a strong impetus to understand these structures' biological roles. Furthermore, the high number of pG4 sequences in the genome/transcriptome and the high intracellular monovalent cation (i.e.  $K^+$ ,  $Na^+$ ) concentration offer suitable G4 formation conditions. However, understanding the biological relevance of G4s is one of the most, if not the most, challenging aspects in G4 research.

### **1.5.1 Evolutionary Conservation of G4 motifs**

The advancement of next-generation sequencing techniques led to the assembly of genomes from several different organisms, which allowed pG4 motifs from these genomes of different complexity to be analyzed and compared. Large-scale computational search for pG4 motifs showed highly conserved pG4s in promoter regions when comparing humans, mouse, chimpanzee and dog genomes (109). When studying human SNPs, G4-disrupting sites in pG4 motifs display lower polymorphism levels, suggesting that these pG4 motifs are more conserved than other random sites (110). In corroboration, in a genome-wide study of several evolutionarily

distant yeast species, pG4 motifs were found to be highly conserved at promoters, rDNA, mitochondrial DNA (mtDNA) (111). These exciting discoveries prompted the development of a tool, QGRS-Conserve, used for computing conservation between pG4 motifs of different species, which also revealed the highest conservation of pG4 motifs between human and other mammalian species (112). In addition to animals, the genomes of nearly all organisms likely harbour pG4 motifs to varying degrees, including bacteria, protozoa, plants, yeasts and even viruses (112). The high pG4 evolutionary conservation suggests that G4s likely mediate key biological processes essential for survival and propagating the organisms.

## **1.5.2 G4-interacting Proteins**

Studying proteins that interact with G4s can provide clues to the biological relevance and regulation of G4s. Broadly, G4-interacting proteins can be classified into two categories: G4 helicases (specialized proteins that unfold G4s), and other non-helicase G4-interacting proteins (that can either stabilize or destabilize G4s).

### **1.5.2.1 G4 Helicases**

The characterization of G4s using biochemical assays suggests that G4s pose as barriers for the replication and transcription machinery (88). Additionally, evidence suggests that G4 may also be a barrier for the translation machinery (113). It is thus conceivable that regulatory processes are necessary to avoid global disruption of these critical biological processes. These regulation processes are mediated by helicases, specialized enzymes that can remodel nucleic acid structures. While identifications of key G4 helicases, especially early on, have stemmed from other organisms (yeast and ciliates), most of the helicases covered in this section are focused on those in humans (and human-equivalent homologs). It is important to note that these helicases often have multiple functions and can mediate the unfolding of other nucleic acid structures

(primarily duplexes). There are nearly 100 known human helicases (114), of which 18 of them have confirmed activities against G4s. Table 1-4 outlines a list of known human G4 helicases and their biological characteristics. Broadly, there are six superfamilies (SF) of human helicases based on shared sequence motifs. The majority of G4 helicases belong in the Superfamily 2 (SF2), which is also the largest superfamily of human helicases.

### SF1 G4 Helicases

There are two known G4 helicases in the SF1 superfamily: PIF1 and DNA2. Pif1 was first characterized in yeast, which functioned to prevent G4-mediated genomic instability in inserted CEB1 sequences (115). The human homolog, PIF1, was confirmed to have the same G4-DNA ATP-dependent unfolding activity (116). Furthermore, ChIP-seq showed that Pif1 binds to pG4 motifs in the genome (117), and overexpression of Pif1 was able to rescue G4-mediated DNA damage (117). DNA2 is another SF1 helicase identified in both yeast and human with the ability to unfold G4-DNA with high affinity (118). DNA2 has been implicated in reducing replication stress and telomere instability (119).

### SF2 – RecQ G4 Helicases

Some of the earliest identified G4 helicases fall under the RecQ family, including BLM helicase and WRN helicase. RecQ helicases are evolutionarily conserved with homologs in bacteria (RecQ) and yeast (Sgs1). All RecQ helicases share several similar structural features, including the RecQ-core helicase domain, RecQ C-terminal (RQC), and helicase and RNase D C-terminal (HRDC). BLM, one of the most well-characterized G4 helicase, was found to have a higher affinity for G4-DNA over duplex DNA and efficiently unfold G4s (120). Additionally, BLM has been shown to unfold both intermolecular and intramolecular G4 substrates (121). Interestingly,

BLM RecQ-core ssDNA binding near a telomeric G4 resulted in the subsequent unfolding of that G4 in the absence of ATP (122). WRN, another RecQ helicase, was similarly shown to unfold G4s, as demonstrated by its G4-unfolding activity on CGG trinucleotide repeat of Fragile X Syndrome (31) and other G4 structures as well (123). Single-molecular imaging later revealed that BLM and WRN use a common G4-unfolding mechanism by repetitive cycles of unfolding in a consecutive manner (124).

### SF2 – DEAH-box G4 Helicases

The DEAH-box subfamily of the SF2 superfamily has several well-characterized G4 helicases. These DEAH-box helicases share the Walker B motif (D-E-A-H or asp-glu-ala-his), thus the protein family's name. One of the best characterized DEAH-box G4 helicase to date is the DHX36 (also known as RNA helicase associated with AU-rich element (RHAU) or G4 Resolvase 1 (G4R1)). DHX36 is also one of the few helicases that can unfold both G4-DNAs and G4-RNAs (125). G4-DNA unfolding mechanism for DHX36 is shown to be similar to that of BLM and WRN, adopting an ATP-independent repetitive unfolding manner (125). A rare high-resolution crystal structure of DHX36 bound to a G4-DNA in a recent study revealed key structural details into the helicase-substrate interactions (126). The N-terminal motif of DHX36 appeared to form an  $\alpha$ -helix that works in conjunction with the OB-fold-like subdomain to bind parallel G4-DNAs selectively. In terms of G4-RNAs, RIP-ChIP analysis with DHX36 showed the binding of several RNAs harbouring pG4 motifs (127). Interestingly, in contrast to the DHX36-mediated unfolding of G4-DNAs, a later study showed that the activity of DHX36 on G4-RNAs proceeds through an ATP-independent unfolding followed by an ATP-dependent refolding of the RNA tail (128). This peculiar observation suggests a highly dynamic nature of DHX36 to unfold G4-RNAs but restore surrounding structures (i.e. duplexes). The BRIP (also

known as FANCI) G4-unfolding activity was first observed in its *C. elegans* ortholog, *dog-1*, in which its disruption led to the deletion of G-rich DNA regions (129). Human BRIP1 was later shown to unfold G4-DNAs in an ATP-dependent manner (130). Another important G4 helicase in this family is RTEL1, with strong connections with the telomere, as shown by the observation in which an absence of RTEL1 resulted in shortened telomere in mouse models (131). RTEL1 may counteract against G4 formation at the telomere to prevent telomere fragility and loss (132). Finally, DHX9, a relatively newly discovered G4 helicase, exhibited both G4-DNA and RNA unfolding activities (133).

#### SF2 – DEAD-box G4 Helicases

The DEAD-box family of SF2 contains many recently discovered G4 helicases. Like the DEAH-box family, these DEAD-box helicases share the Walker B motif (D-E-A-D or asp-glu-ala-glu). Several of these helicases have known functions in various RNA metabolism processes. The eIF4A (also known as DDX2), a well-known component of the translation initiation complex, has been shown through ribosome footprinting to promote the translation of oncogenes harbouring G4 motifs in their 5' untranslated region (UTR) (134). Using mass spectrometry with biotinylated G4-RNA as the guide, a new G4 helicase with G4-RNA unfolding activity, DDX21, was also discovered (135). Similarly, affinity proteomics uncovered several other new G4 helicases using the *NRAS* G4-RNA as the bait, including DDX3X, DDX5, and DDX17 (136). Additionally, an SF2 G4 helicase, DDX1, was also found to affect IgH class switching by unfolding G4-RNAs and converting them into R-loops (137).

### SF2/3/4 – Other G4 Helicases

Other G4 helicases that fall under the SF2 family include ATRX and XPD. ATRX is an SWI/SNF helicase with chromatin remodelling functions. Using ChIP, ATRX was shown to bind to tandem repeat sequences in the telomere and euchromatic sites, many of which contain pG4 motifs (106). Similarly, XPD and XPB, transcription-associated DNA helicases, were found using ChIP-seq to have a genome-wide distribution in regions with high levels of pG4 motifs (107). For the SF3 superfamily, SV40 large T-antigen, a viral helicase, showed remarkable specificity for G4 as well as G4-unfolding ability, suggesting G4-related viral-host interactions (138). For the SF4 group, twinkle, a mitochondrial helicase, was shown to be able to unfold G4 arising from mtDNA but only at a modest level (139).

#### **1.5.2.2 Other Non-helicase G4-interacting Proteins**

There are also non-helicase G4-interacting proteins. These can be divided based on their target regions: telomere and other genomic regions, promoter regions, and RNA. Table 1-5 lists some of the known non-helicase G4-interacting proteins.

#### Telomere and other genomic regions

Telomere end-binding proteins (TEBPs) in ciliates were found to regulate telomeric G4 formation (140). The human equivalents of TEBPs are the POT-TPP1 members of the six-member telomere-binding complex, shelterin. Thus, a study in POT1 G4-binding activity revealed its ability to bind and unfold G4s formed in the telomeric ssDNA G-rich overhang (141). Another shelterin member, TRF2, was also found to bind and modulate G4 formation (142). ssDNA binding protein, RPA, essential for DNA replication and repair, was found to unfold a variety of G4s, which prevents G4 formation when the DNA is single-stranded (143).

Additionally, other critical DNA damage sensors and repair proteins, BRCA1 (144) and PARP1 (145) have been shown to associate with G4s.

### Promoter regions

Several proteins also associate with G4s formed in the promoter regions. Nucleolin, a multifunction protein, was observed to interact with a folded G4 from within the *MYC* promoter (146). Nucleolin appeared to stabilize G4 upon binding (147) with a preference for long looped G4s (148). Another multifunction protein, nucleophosmin (NPM1), was also associated with G4s (149). Zinc-finger transcription factors, MAZ (150) and CNBP (151), showed G4 binding capabilities through interactions with *KRAS* and *MYC* oncogene promoters, respectively, which inhibited their transcription. Recently, DNA methyltransferase (DNMTs), known for mediating methylation of CpG islands, also showed a higher affinity for G4-DNAs than duplex DNAs (152).

### RNA-binding proteins

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a group of critical RNA-binding proteins. Some of these hnRNPs were found to disrupt G4s, including hnRNPA1 and hnRNPA2 (153). However, there have been some discrepancies on whether these hnRNPs are stabilizing or destabilizing G4s. These discrepancies were exemplified in the case for hnRNP F and hnRNP H, which were first proposed to bind directly and stabilize G4-RNAs (154) but were later reported to bind unfolded G4 sequences to prevent G4 folding (155). FMR2, related to mental retardation syndrome, were also found to bind G4-RNAs (156). Telomeric repeat-containing RNA (TERRA), RNA molecules transcribed from telomeres, contained G4s bound to telomere-binding proteins, TLS (157). Mass spectrometry analysis of G4-binding proteins, using G4-RNA

oligos from *MMP16* and *ARPC2* as baits, revealed many ribosomal proteins and splicing factors (158). It is unknown how G4-RNAs are related to these ribosomal proteins (e.g. RPL6,7,10,12,14,19,26) and splicing factors (e.g. SRSF1, SRSF9, U2AF), but these targets suggest roles of G4-RNA in several aspects of RNA metabolism.

**Table 1.4. List of G4 helicases and their G4-related functions.**

Family	Helicase Name	Target	ATP (+/-)	Effect on replication	Effect on transcription	Effect on translation	Other effects	Ref
SF1	PIF1	DNA	+	Yes	-	-		(115-117)
	DNA2	DNA	+	Yes (?)	-	-	Telomere	(118-119)
SF2-RecQ	BLM	DNA	-	Yes	Yes	-	Telomere	(30, 120-123)
	WRN	DNA	-	Yes	Yes	-	Telomere	(31, 123)
SF2-DEAH	DHX36 (RHAU)	DNA, RNA	-	-	Yes	Yes	Telomere	(124-128)
	BRIP1 (FANCF)	DNA	+	Yes	-	-	DNA repair	(129-130)
	RTEL1	DNA	+	Yes			Telomere	(131-132)
	DHX9	DNA, RNA	+	Yes (?)	Yes (?)	Yes	Telomere	(133)
SF2-DEAD	DDX21	RNA	+	-	-	Yes	-	(135)
	DDX11 (ChIR1)	DNA	+	Yes	-	-	DNA repair	(136)
	DDX5	DNA, RNA	-	-	Yes	Yes	Splicing	(136)
	DDX17	RNA	+	(?)	Yes (?)	Yes	Splicing	(136)
	DDX3X	RNA	+	-	-	Yes (?)	-	(136)
	DDX1	RNA	+	-	-	Yes (?)	IgH Class switch	(137)
	eIF4A (DDX2)	RNA	+	-	-	Yes	-	(134)
SF2-others	ATRX	DNA	+	Yes	Yes	-	Telomere	(106)
	XPD (XPB)	DNA	+	-	Yes	-	DNA repair	(107)
SF3	SV40 T-ag	DNA	+	Yes (?)	-	-	-	(138)
SF4	Twinkle	mtDNA	+	Yes	-	-	-	(139)

(?) denotes speculated roles with very limited or insufficient evidence.

**Table 1.5. List of non-helicase G4-interacting proteins**

<b>G4 Target</b>	<b>Target Regions</b>	<b>Protein Name</b>	<b>Functions</b>	<b>Ref</b>
DNA	Telomere & other genomic regions	POT1	Telomere maintenance	(140-141)
		TRF2	Telomere maintenance	(142)
		RPA	Replication and repair	(143)
		BRCA1	Replication and repair	(144)
		PARP1	Replication and repair	(145)
	Promoter regions	Nucleolin	Transcription	(146-148)
		Nucleophosmin	Transcription	(149)
		MAZ	Transcription	(150)
		CNBP	Transcription	(151)
		DNMT1	Methylation	(152)
RNA	Wide range	Nucleolin	Ribosome biogenesis (?)	(148)
		hnRNPA1/2	Translation	(153)
		hnRNP F/H	Translation	(154-155)
		FMR2	Alternative splicing	(156)
		Ribosomal proteins (several)	Ribosome biogenesis (?)	(158)
		SRSF1/9	Alternative splicing	(158)
		TLS/FUS	Telomere maintenance	(157)

### 1.5.3 Biological Functions of G4-DNAs

Much of early research on the biological roles of G4s had focused on G4-DNAs. Early biological characterizations of G4s were challenging and highly reliant on indirect approaches such as manipulation of G4-interaction proteins and/or treatment with G4 ligands. Several paradoxical observations emphasize the complexity of the biological effects of G4s. On the one hand, folded G4s appear to have global negative effects in cells by acting as barriers to key biological processes, yet evidence for the positive effects of G4s is increasing recently. The biological functions of G4-DNAs can be divided into three sections: telomere maintenance, transcription, genome instability and epigenetic control.

#### 1.5.3.1 Telomere maintenance & integrity

The interwoven research in telomere and G4s at the early stages was primarily due to ideal G4-forming motifs in telomeric G-rich repeat sequences. Telomeres require normal capping to

protect chromosomal ends. In yeast models with dysfunction of the capping protein Cdc13, upregulation of G4s was observed to rescue telomere capping defects, suggesting a positive role for G4s in the absence of natural capping (159). Whether this mechanism is translatable to humans is unknown, although telomere-binding proteins, such as TRF2, have been found to bind G4s (142). The interaction between *TERRA* and TLS led to an intriguing model in which G4-RNAs in *TERRA* could act as scaffolds for telomere binding proteins (157) and recruit other proteins for telomere maintenance. G4 ligands' ability to inhibit telomerase suggests that a persistently folded G4 can pose a problem to telomere elongation by telomerase. Folded G4 in the 3' overhang of telomere may inhibit telomerase from associating with the 3' end to mediate elongation (26). Thus, the ability of POT1 to unfold G4s at the telomeric ssDNA overhang may prevent spontaneous G4 folding that could otherwise inhibit telomerase (141). Interestingly, telomerase appeared to exhibit partial G4-unfolding capabilities despite the lack of a helicase domain (160). In surprising contrast to previous findings, it has also been recently suggested that the formation of G4 within actively extending telomerase mediates dissociation and repeat addition in conjunction with POT1-TPP1, supporting a positive role for G4s in telomerase function (161). Thus, the role of G4 concerning telomerase appears to be much more complicated than previously thought. While G4 ligand-induced telomere shortening was previously believed to be through telomerase inhibition (26), new data suggest that the primary mechanism may be more likely through the effect of G4s directly on telomere replication as telomeric repeats harbour the longest stretch of pG4 motifs. Accordingly, telomere replication stress was evident when RTEL1 helicase was dysfunctional (131), in addition to G4 ligand treatment (61), resulting in telomere fragility and loss.

### 1.5.3.2 Transcription

The relationship between G4 and transcription is also complex and context-dependent.

Computational prediction of pG4 sites in the human genome showed high saturation of pG4 in promoter regions suggesting a strong connection between G4 and transcription (35). G4s may pose as barriers to the transcriptional machinery when formed on the template strand. This notion is corroborated by studies of individual genes (*MYC*, *KRAS* and *KIT*) (29, 162, 163), which showed gene expression inhibition upon G4 ligand treatment. In support of previous studies from individual genes, genome-wide enrichment of transcription-coupled helicase, XPD/XPB, at G4 motifs suggests a dynamic unfolding of G4s during transcription (107). Separately, transcription factor interactions with G4 could also impact transcription in both negative and positive directions. An example of a G4-transcription factor interaction was seen for the interaction between nucleolin and the *MYC* promoter, leading to transcriptional inhibition (146). Using G4 ChIP-seq, several G4 sites were also found to harbour transcription factor binding motifs suggesting that other transcription factors could be involved (104-105). *In vitro* experiments suggest that G4s may act by directly binding transcription factors (e.g. nucleolin, CNBP, MAZ, SP1) (146, 150, 151, 164) but may also disrupt the binding of proteins such as transcriptional repressor, CCCTC-binding factor (CTCF) (165). Regarding G4s formed on the non-template strand, both increased and decreased transcription have been observed. A genomic study showed that gene promoters with more pG4s on the non-template strand have higher baseline transcription levels, suggesting a transcription promoting effect (166). In contrast, another study showed that G4 sequences inserted into the non-template strand instead led to inhibition of transcription due to the promotion of R-loop formation on the template strand (167). A strange phenomenon was also seen in the transcription of mitochondrial DNA in which the nascent RNA

of a GC-rich sequence formed a hybrid G4 with the non-template DNA strand (168). Overall, the effects of G4 on transcription are complex and appear to highly context-dependent.

### **1.5.3.3 Genome instability**

G4s may pose barriers to the replication machinery and thus cause replication stress (169). If not properly resolved, G4s can lead to replication-fork collapse and double-strand breaks. G4 formation leads to genome instability and is postulated as one of the driving forces for cancer genome evolution. A genome-wide study of a large number of cancer samples showed that pG4 sequences were associated with breakpoints and somatic copy number alteration (170). Studies using models with dysfunctional helicases, including FANCD1, PIF1, BLM, ATRX showed deletions, gross chromosomal rearrangement and common fragile sites in loci associated with high G-rich content (115, 130, 171, 172). A recent study has shown that BLM and WRN may function with HERC2 and RPA to suppress G4 formation and prevent genomic instability (173). Given the broad targeting mechanism of G4 ligands, a major contributor to cytotoxicity by these ligands is likely through the induction of DNA double-strand breaks (48, 169). Observation further supports this notion that loss of G4 helicases and/or dysfunction in double-strand DNA break repair mechanisms often leads to the cells becoming more sensitive to G4 ligands (173).

### **1.5.3.4 Epigenetic Control**

A new branch of study in G4-DNA functions focuses on the epigenetic aspects. Using ChIP-seq, several of these G4-DNA sites were found to be situated at DNMT1 occupancy sites and were also hypomethylated. Thus, G4 was suggested to inhibit the methylase activity of DNMT1 (174). Additionally, G4s-mediated effects on histone recycling were implicated in a study which showed that ligand-induced G4 stabilization led to alteration of local histone marks (175),

potentially mediated through G4-dependent replication fork pausing and subsequent decoupling of the replication machinery from the histone recycling system.

#### **1.5.4 Biological Functions of G4-RNAs**

Unlike the wealth of information from studies on G4-DNAs, much less is known about the biological function of G4-RNAs. Some potential biological functions of G4-RNAs include translation regulation, alternative splicing and other RNA metabolism processes.

##### **1.5.4.1 Translation**

Within the limited scope of our understanding of the effects of G4-RNAs, translation regulation is the most well studied. G4 formation in mRNA can pose as barriers to the translation machinery (i.e. ribosome). The first example of the effect of G4 on translation was observed in fragile-X mental retardation protein (FMRP), in which the interaction with a G4 in its mRNA led to translation repression, forming a negative feedback regulation (176). Later, a translation reporter assay showed clear evidence of G4-mediated inhibition of translation by the formation of a G4 in the 5' UTR of *NRAS* mRNA (113). Similar types of translation inhibition were observed for other genes, including *BCL2* (177) and *ADAM10* (178), all of which showed translational inhibitory effects of the 5' UTR-G4s. G4-RNA interacting protein hnRNP A2 was shown to alleviate the G4-mediated translation repression of *FMRI*, suggesting that a G4-unfolding mechanism is necessary for proper translation (179). A core component of the translation initiation complex, eIF4A, was proposed as the G4 helicase responsible for unfolding G4 during the translation step and was suggested as a malignant mechanism used by cancer cells to upregulate oncogene translation (134). On the other hand, G4 formation may not always lead to translation inhibition, but may instead, promote internal ribosome entry site (IRES)-mediated translation. For example, a G4-mediated increase in IRES translation was observed for FGF2

(180) and VEGF (181). Recently, a study reported that loss of the G4 helicase, DHX36, led to the accumulation of translationally incompetent target mRNAs with increased transcript abundance, increased ribosome occupancy, decreased protein output, and increased stress granule formation (182).

#### **1.5.4.2 Other potential functions of G4-RNAs**

G4-RNAs may act as a type of molecular beacon for RNA-binding proteins. For example, 3'-UTR G4s in the mRNA of *PSD-95* and *CaMKIIa* may be required for the mRNAs' dendritic localization via binding FMRP (183). Another potential function of G4-RNA is in alternative splicing, supported by the observed association of G4-RNAs with several splicing factors such as hnRNP H, hnRNP F, SRSF1 and SRFS9 (158). Individual examples of splicing alteration have been observed in *TP53* (184), *BACE1* (185) and *PAX9* (186). A recent transcriptomic study also showed that treatment with a small molecule, emetine, can disrupt G4-RNAs leading to changes in alternative splicing patterns (187). Additionally, there have been some limited pieces of evidence for other G4-RNA functions, including miRNA processing (188-190), chromatin regulation (191) and retrotransposition (192). Given the widespread distribution of pG4 motifs in the human transcriptome (34, 90), G4-RNAs may impact a wide range of RNA metabolism processes, some of which are still poorly characterized and require further research.

#### **1.5.5 G4 in human diseases & potential therapeutics**

The relevance of G4s in human diseases indicates potential therapeutic values for targeting G4s. Three categories of known G4-related human diseases include cancer, neurodegenerative diseases and helicase dysfunction-related diseases.

### 1.5.5.1 Cancer

Cancer is the most well studied in G4-related human disease due to the high number of pG4s at telomeres and oncogene promoters. As telomere maintenance is required for immortal growth in all cancers, the ability of G4 ligands to inhibit telomerase (26) and/or directly target telomeres (57-58, 61) makes the development of G4 ligands as novel anticancer therapies a logical approach. Furthermore, many of the G4-related genes studied have been well-characterized oncogenes, including *MYC*, *KRAS*, *VEGF*, *BCL2*, and *KIT*. The transcription of these oncogenes can be inhibited by G4 ligands (193). The inhibition of the translation of these oncogenes was also shown to be achievable in the case of *NRAS* (113). The consensus of the difference between G4s in cancer cells versus non-cancer cells based on cellular imaging and genomics studies is that there are typically more G4 formations in cancer cells (96, 104). This difference may be due to the overall higher level of proliferation, transcription, translation and other nucleic acid metabolism processes in cancer. Thus, like classical chemotherapeutics, this would also provide more opportunity for G4 ligands to act on cancer cells. However, developing a specific G4 ligand with high specificity, while still retaining favourable pharmacological profiles, remains a challenging endeavour and a key hurdle to clinical use (55). Nevertheless, it is encouraging that a few G4 ligands have been able to reach Phase 2 clinical trials, including quarfloxin (63) and CX-5461 (64).

### 1.5.5.2 Neurodegenerative diseases

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two of the most common types of neurodegenerative diseases. They are shown to be caused by the GGGGCC hexanucleotide repeat expansion (HRE) in the gene non-protein-coding region of *C9orf72*. Importantly, these repeats in both DNA (194) and the RNA transcript (195) were found to form

G4s, which were responsible for affecting the normal expression of the *C9orf72* and causing abortive transcripts. Additionally, G4s in these HREs may also sequester hnRNPs from their normal function (196) and cause phase separation (197). Due to the etiological roles of *C9orf72* HRE G4-RNAs in these diseases, the therapeutic value of G4-targeting approaches has been explored. G4 ligand, TmPyP4, which have potential G4-disruptive effects on some G4-RNA targets, blocked the interaction between G4s in *C9orf72* HRE and hnRNPs (198). Another study aimed at screening for new G4 ligands to stabilize G4-RNAs formed in the HRE of *C9orf72*, which resulted in three structurally related compounds. Interestingly, these new G4-stabilizing compounds, when tested against iPSC-derived motor and cortical neurons from *C9orf72* patients, were able to ameliorate disease phenotypes and improve survival of the cells (199). G4-based therapeutics may be viable options against these neurodegenerative diseases; however, future research is necessary to delineate whether the therapeutic effect is primarily through G4-disruption or G4-stabilization.

### **1.5.5.3 G4 helicase dysfunction-related diseases**

The last category of human diseases is related to helicase dysfunction (200). These diseases are often inherited and very rare. They share similar characteristics such as developmental disorders, immunodeficiencies, premature ageing, bone marrow failure, and cancer predisposition. Table 1-6 includes a list of G4 helicases and human diseases that are related to their dysfunctions.

**Table 1.6. List of G4 helicase-related human syndrome**

<b>G4 Helicase</b>	<b>Human diseases related its dysfunction</b>
PIF1	Familial breast cancer
BLM	Bloom syndrome
WRN	Werner syndrome
BRIP1 (FANCD1)	Fanconi Anaemia (FA), breast cancer
RTEL1	Dyskeratosis congenita
DDX11	Warsaw Breakage Syndrome
XPD	Xeroderma Pigmentosum, Cockayne syndrome, Trichothodystrophy, Cerebro-oculo-facio-skeletal syndrome
eIF4A	T cell leukemia (overexpressed)
ATRX	Alpha thalassemia with mental retardation syndrome

## **1.6 Rationale, Hypothesis & Aim**

### **1.6.1 Research Questions of Interest**

While the G4 field has been expanding rapidly in recent years, there are still noticeable major barriers to further advancement. Section 1.3 demonstrates that the chemical-physical laws that dictate the formation of G4s have been well defined over the last few decades. Section 1.4 outlines the tools that have already been developed, each with certain advantages and disadvantages. As reliance on a single tool/method can lead to biases, ongoing analytical methods development is necessary to advance the current state of knowledge in G4.

Furthermore, most existing methods are designed for testing the chemo-physical properties of G4 oligos and are only suitable for *in vitro* studies. Accordingly, a lack of tools and techniques for studying G4s in a biologically relevant manner remains a pressing issue. As shown in the current literature on G4-related biological roles (reviewed in Section 1.5), reports with contrasting findings suggest that our current understanding of G4 biology is inadequate. Below, I introduce three sets of unanswered questions that convey the most pressing issues in the G4 field, which have shaped my thesis's research directions.

Are G4s formed in nature? If so, where are they formed on a genome- and transcriptome-wide scale?

Despite the relatively long history of G4 research, the simple question of “*in vivo* existence” remains an “Achilles heel” for scientists in this field. It is also concerning that the belief in which G4s are *in vitro* artifacts continues to persist to this day. The main issue is that many early G4 studies have often relied on the inferred formation of G4s from sequences predicted *in silico* and confirmed *in vitro* (79). While these observations are important for evaluating the potential of certain sequences to fold into G4s, it does not necessarily prove the existence of folded G4s in nature. The recent finding that G4-RNAs are nearly completely unfolded inside living biological systems adds further complications to this issue (108).

Another limitation is that G4 functional studies to date have often been limited to the study of one gene, one protein and one ligand at a time. The issue may have stemmed from the current paradigm in biology, favouring the traditional bottom-up approach, focusing on individual components and their local associations to explain more complex observations. While it has merits in characterizing novel mechanisms in detail, it lacks a global view of all potential interactions. The problem is that the biological regulations are rarely driven by a single factor or even a few. With advancements in genomics and cellular imaging, I argue that a top-down approach can offer a more flexible and comprehensive view of G4 biology rather than a bottom-up approach. Currently, G4 research suffers from the lack of tools and methods that allow such top-down approaches. Combined with the lack of proper ways to study G4s in a biologically relevant manner, it is difficult to generate any meaningful data of high generalizability applicable to multiple biological systems.

Where are G4 ligands targeting specifically on a genome- and transcriptome-wide scale?

While several G4 ligands have already been developed (section 1.3.6), there is currently a lack of reliable tools/methods for testing G4 ligand-induced global effects on the genomic and transcriptomic G4 landscape. Most of the known screening techniques for G4 ligands are testing ligand-G4 oligo interactions under *in vitro* settings. However, given the wide variety of G4 conformations that can potentially form in nature, it is unlikely that a specific G4 ligand can target all G4s in equal efficiency. In addition to the different G4 conformations, the biological environment may present another level of restrictions in which only a subset of G4s may be accessible by G4 ligands. Thus, for G4 ligand development, a global (genome- and transcriptome-wide) screening method would be invaluable for optimizing and designing novel and specific G4-targeting compounds.

What are the biological roles of G4s? Are G4 roadblocks in nucleic acid metabolism? Can G4 provide positive regulation in any cellular function?

Early studies have suggested that G4s act as barriers or roadblocks in nucleic acid metabolism processes, including replication, transcription and translation (Section 1.5). However, the evolutionary conservation and nonrandom genomic distribution of pG4 sequences suggest G4s may provide beneficial or supportive biological roles (109-112). While some positive effects for G4 have been suggested in recent studies (161, 183, 201), the extent of these effects remains mostly unknown. A more global and comprehensive approach for studying G4s would help shed light on the possible duality nature of G4 formation.

What is the relationship between G4s and other nucleic acid structures? Are there any potential interactions with other non-canonical nucleic acid structures?

The interactions between G4s and potential protein partners have extensively been reviewed in Section 1.5.2. However, there is little to no information on the interactions between G4s and other non-canonical structures. Given the repetitive nature of pG4 motifs, such skewed sequences may fulfill the requirement to form other nucleic acid structures. For example, the importance of skewed repeats is evident in DNA/RNA hybrids (R-loops), triplexes, i-motif, hairpins and cruciform (Section 1.1). Specifically, R-loops and i-motifs are relevant to G4s. The formation of an intramolecular G4-DNA can displace the other strand, which can either interact with complementary RNA to form R-loops or fold into an i-motif due to the C-rich repeats. While the simultaneous formation of G4s with either R-loops or i-motifs has been observed *in vitro* (202-203), it is unclear if the interconnected structures are formed *in vivo*.

### **1.6.2 Hypothesis**

In response to these unanswered research questions, I aim to, in my thesis, develop top-down approaches to study G4s in a more biologically relevant manner by using both cellular imaging and genomic techniques. These approaches may help in providing evidence for the existence of biologically relevant G4s in human cells. Additionally, these methods may help in evaluating global G4 ligands-target interactions and elucidating novel G4-related biological roles. Thus, my overall thesis hypothesis is that **G-quadruplexes are widespread and can be studied on a global-level (genomic, transcriptomic, cellular imaging) in human cells through novel tools and methods.**

### 1.6.3 Aims

For my thesis hypothesis, my specific thesis aims are as follows:

1) Pilot cellular imaging investigation of G4s in multiple cell lines

Here, I aim to examine and compare the formation of G4s in multiple cell lines using various cellular imaging techniques (including antibodies and probes). Chapter 3 and Appendix A describes cellular cell imaging techniques in both fixed and live cells using the G4-specific probe, N-TASQ. Chapter 5 describes a multi-cell line study that quantified differences in G4 levels between cell lines that employ different telomere maintenance methods.

2) Genome- & transcriptome-wide mapping of G-quadruplexes

Here, I aim to study G4s on a genome- and transcriptome-wide level using newly optimized tools and methods. Chapter 4 and Appendix B describe a novel method for studying G4s on a transcriptomic level. Appendix C summarizes the preliminary results from a genome-wide G4 mapping study.

3) Global profiling of G4 ligand-induced changes

Here, I aim to study G4 ligand-induced changes on a genome- and transcriptome-wide level using the optimized tools/methods developed in Aim 2. Chapter 4, Appendix B and Appendix C demonstrate the use of genomic/transcriptomic methods to generate global induction profiles for different G4 ligands.

4) Elucidating the roles of G4s at the telomere

Here, I aim to provide evidence for novel biological roles of G4s related to the telomere using methods developed in previous aims. Cancer cells that utilize the telomerase-independent mechanism, Alternative-lengthening-of-telomere (ALT), may be especially useful models since they harbour much longer telomeres and potentially higher opportunity

for G4 formation. Due to the access of several ALT cell models in the lab, I aim to characterize G4s in the telomeres of ALT cells and to elucidate potential functions related to these G4s. Chapter 5 describes the study of G4s in ALT.

5) Studying the interactions of G4s with other non-canonical structures

In conjunction with the previous aim, ALT cell models may also provide ideal conditions to study the interactions between G4s and other non-canonical structures. Specifically, R-loops has a high potential for interacting with G4s due to the abundance of TERRA in ALT cells and recent reports of ALT-associated R-loop formation. The relationship between G4s and R-loops is explored in Chapter 5.

## Chapter 2: Detailed Protocols: Materials and Methods

This chapter contains a collection of key protocols used in this thesis that are not widely available or are in any manufacturers' recommended protocols from commercial kits. Each protocol contains step-by-step instructions.

### 2.1 Purification of the BG4 antibody

The G4-specific antibody, BG4, used in this thesis, is produced in-house by expressing the recombinant protein in bacteria and subsequent purification. This way usually provides a more reproducible and stable stock of antibodies. The procedure takes at least one week to complete depending on whether a new transformation is required.

#### 2.1.1 Reagents and Buffers

- BL21-DE3 competent *E. Coli*
- pSANG10-BG4 plasmid DNA
- TSS buffer – Made from adding 1X LB broth, 10% polyethylene glycol 3350 or 8000, 5% DMSO, 50mM Mg<sup>2+</sup>, pH 6.5 and the sterile filter
- Kanamycin (50mg/mL stock) – make 50mg/mL from powder in ddH<sub>2</sub>O and then sterile-filter and store in -20°C.
- LB-Kanamycin agar plate (standard agar recipe + 30 µg/mL final concentration of Kanamycin; store in 4°C)
- 1M IPTG
- 5M imidazole, pH 7.5
- TGI buffer: Make using 20mM Tris-Cl, 10% glycerol, 20mM imidazole, pH 7.5 and the sterile-filter.

- Polymin P (10%)
- Nickel affinity beads (Ni-NTA) (Qiagen)
- Gravity flow column
- Ultra-15 Filter system (Amicon)

### **2.1.2 Procedures**

#### Transformation (Day 1-2)

1. Grow BL21-DE3 in 5 mL of LB broth to OD<sub>600</sub> of 0.3-0.4 and then pellet 1-2 mL and then resuspend in 100 µL of TSS.
2. Make two tubes: one for negative control and one for transformation. For negative control, do not add any plasmid. For transformation tube, add 1µL of the plasmid.
3. Place both tubes on ice for 30 min.
4. Heat shock at 42°C for 30 sec.
5. Add 900 µL of LB and shake at 225 rpm for 1h at 37°C.
6. Spread on pre-warmed LB-kanamycin agar plate.
7. Grow overnight
8. Pick colony and check using restriction enzyme digestion by NcoI/HindIII. This should produce two bands, at 0.9 and 5.2 kb. The confirmed colony can be grown and stored in 50% glycerol in -70°C for future use.
9. Grow confirmed colony in 6mL overnight to use as starter for later growth. Grow more starter culture if more flasks are used.

### Bacterial growth (Day 3)

1. The next day, add the start to 500 mL of LB (LB-Kanamycin at 30 µg/mL in 2L flask) and grow at 37°C at 170 rpm. Typically, per round of purification, at least two flasks of 500 mL or more are required.
2. Wait until OD<sub>600</sub> reaches 0.5-0.6, which typically takes around 5 hours.
3. Add IPTG to a final concentration of 0.5-1 mM. Move flasks to room temperature shaker and shake at 120rpm overnight.

### Protein purification by His-tag purification (Day 4-6)

1. The next morning, spin at 4000 rpm for 30 min to collect pellets. These pellets can be stored in -20°C for later processing. Immediate processing of fresh material is recommended.
2. For frozen pellet, thaw on ice for 30 min. The following numbers are for one flask. For multiple flasks, process them in parallel in separate tubes.
3. Resuspend pellet in 5 mL, which typically gives a final volume of 7 mL. In a 50mL conical tube, sonicate the suspension with a common probe sonicator (output setting at 2) in a cold setting (i.e. cold room).
4. Sonicate for 4 cycles at 15 sec, each with 15 sec in between each cycle on ice for cool-down. The lysate should become slightly less viscous.
5. Add to the lysate the following to a final concentration of 250 mM NaCl and 0.01% NP40.
6. Spin at 12500 rpm for 15 min at 4°C. Transfer the supernatant to a 15 mL conical tube.
7. Add 35µL 10% Polymin-P four times with a quick vortex in between each addition. Rotate at 4°C for 20 min.

8. Spin at 12500 rpm (transfer to an appropriate container if necessary) for 15 min at 4°C.
9. Transfer supernatant by filtering through miracloth into a new conical tube. You can combine tubes from multiple flasks at this step. Aliquot a small fraction (~50 µL of the lysate) for running an SDS-PAGE check later.
10. Prepare Ni-NTA by washing beads 3 times in TGI Wash buffer (TGI + 250 mM NaCl). Aspirate and resuspend in an equal volume of wash buffer to make 50% slurry.
11. Approximately 7 mL of lysate can be collected from one flask of bacteria. Add 2.5 mL of 50% Ni-NTA slurry to the lysate and rotate at 4°C for 2h. This can be scaled for multiple flasks.
12. Prepare the following elutions buffer using TGI Wash buffer (TGI + 250 mM NaCl). Typically, a 5mL elution buffer is sufficient for each flask. It can be scaled up to 10mL based on the number of flasks. Make 3 elution buffer at 5mL, each with a different final concentration of imidazole (45 mM, 220 mM, 520 mM) with 1 mM PMSF in each.
13. Perform wash and elution in a cold environment (e.g. cold room). Pour bead+lysate mix into the gravity flow column. Wait until the beads are packed at the bottom of the column. Collect all flow-throughs for later analysis in case troubleshooting is required.
14. Add 2-5 mL (depending on the number of flasks) of the buffers in the following order and wait until the buffer has flowed through completely before adding the next buffer: Wash 1, Wash 2, Elute 1 (45 mM imidazole), Elute 2 (220 mM imidazole), Elute 3 (520 mM imidazole). Store at 4°C.
15. Run an SDS-PAGE gel to check the following fractions: Lysate, First Flow-through, Wash Flow-through, elute 1, elute 2, elute 3. BG4 antibody should show a band ~30kDa. Proceed with the elute fraction with the BG4 band (typically Elute 2) immediately.

(Caution: Process immediately. Do not leave in a high concentration of imidazole for more than a day!)

16. There are two ways to remove imidazole: either through dialysis cassette or buffer-exchange using a centrifugal unit.
17. For dialysis cassette, add eluted protein solution in cassette and then drop into a large flask of PBS and mix overnight at 4°C. For the Amicon Ultra-15 Centrifugal unit (10kDa cutoff), buffer exchange out the imidazole using manufacturer recommended protocol.
18. Buffer exchange using PBS at least three times or more. The final imidazole concentration should be below 5mM. Concentrate to a final volume of around 0.75-1 mL.
19. Add glycerol to a final concentration of 20-30 %. Quantify protein concentration using Bradford Assay. Typically, the final concentration is 1 mg/mL.
20. Sterile-filter and then store the final antibody solution in aliquots of 20-50  $\mu$ L and store in -20°C. The antibody should be useable for at least 1-2 years. Western blot and ICC can also be performed to test the quality of the antibody.

## **2.2 Immunocytochemistry (ICC) Staining Standard Protocol**

ICC staining is a commonly used technique in this thesis. Using antibody, designated targets can then be visualized using the confocal microscope.

### **2.2.1 Reagents and Buffers**

- Methanol (ice-cold) – store -20°C in a polypropylene conical tube
- Phosphate Buffered Saline (PBS)
- PBS-azide – made by adding sodium azide to PBS to a final concentration of 0.02%
- Tris-buffered Saline (TBS) –150 mM NaCl, 50 mM Tris-Cl, pH 7.5

- Tween-20
- Bovine serum albumin (BSA)
- Blocking Buffer – made by making 4% BSA in TBS (4 g/100 mL), then sterile-filter
- Washing Buffer (PBS-T) – add Tween-20 to PBS to a final concentration of 0.1%
- Primary antibody
- Secondary antibody
- 4',6-diamidino-2-phenylindole (DAPI)
- Fluoromount-G

### **2.2.2 Standard Staining Procedure**

#### Fixation

1. Seed cells on coverslips in a 24-well plate. Depending on the experiment, cells can also be treated.
2. On the day of harvest, aspirate media from wells and wash with PBS.
3. Fix the cells by adding pre-chilled ice-cold methanol (300-500  $\mu$ L) to each well and incubate on a rocker for 10 min. (4% paraformaldehyde, but we find methanol to be better for staining nuclear objects)
4. Rinse twice in PBS and then store in PBS-azide. Saran-wrap and store at 4°C. Typically, fixed cells can be stored for 3-6 months without problems.

#### Standard Staining

1. Transfer coverslips to a new 24-well plate and quickly wash once in PBS.
2. Block for 1h on a rocker by adding 250  $\mu$ L blocking buffer to each well.

3. Aspirate and add primary antibody diluted in blocking buffer. The concentration depends on the antibody. I typically use 2-4  $\mu\text{g}/\text{mL}$  final concentration in blocking buffer.  
Incubate either at room temp for 3h or at 4°C overnight.
4. Wash three times for 10 min each with PBS-T
5. Add secondary antibody diluted in blocking buffer (make sure the species correctly match with the secondary antibody). I typically use a concentration of 1-2  $\mu\text{g}/\text{mL}$ .
6. Wash three times for 10 min each with PBS-T
7. Counterstain with DAPI for 5 min and wash twice with PBS for 5 min each.
8. Mount using Fluoromount G. For negative control, repeat every step without the addition of primary antibodies.

### **2.2.3 G4-staining with BG4**

For G4 staining with the BG4 antibody, there are some deviation from the standard protocol.

1. Transfer coverslips to a new 24-well plate and quickly wash once in PBS.
2. Block for 1h on a rocker by adding 250  $\mu\text{L}$  blocking buffer to each well.
3. Add BG4 at 4  $\mu\text{g}/\text{mL}$  and incubated for 3h at room temp
4. Wash three times in PBS-T for 10 min each
5. Add anti-FLAG primary antibody diluted in blocking buffer at 2  $\mu\text{g}/\text{mL}$  and incubate overnight at 4°C.
6. Wash three times for 10 min each with PBS-T
7. Add secondary antibody diluted in blocking buffer at 1  $\mu\text{g}/\text{mL}$ .
8. Wash three times for 10 min each with PBS-T
9. Counterstain with DAPI for 5 min and wash twice with PBS for 5 min each.
10. Mount using Fluoromount G.

#### **2.2.4 G4-staining with N-TASQ**

For N-TASQ staining, no secondary staining agent is required.

1. Transfer coverslips to a new 24-well plate and quickly wash once in PBS.
2. Block for 1h on a rocker by adding 250  $\mu$ L blocking buffer to each well.
3. Add N-TASQ (50-100  $\mu$ M) in blocking buffer.
4. Add N-TASQ buffer on coverslips and incubate at room temp for 1-3h.
5. Wash three times with PBS for 10 min each.
6. (Optional) Counterstain with DAPI for 5 min and wash twice with PBS for 5 min each.
7. Mount using Fluoromount G.

#### **2.2.5 Imaging on Confocal Microscope**

Stained slides can be imaged on a confocal microscope (e.g. Zeiss LSM 700). These procedures are specially for the Zeiss series of microscopes.

1. Switch lens to 40X (oil immersion)
2. Place one drop of Immersol oil on the slide and load the slide on the slide holder with the oil between the lens and the slide.
3. Start by opening the blue channel (DAPI) and manually focus until the DAPI staining of the cells is clear.
4. In the acquisition tab, press on “live” and make some more fine adjustments to improve the focus.
5. Then, press the “Find focus” button. Followed by the “set exposure” to auto adjust the gain. The gain should be the same across all of your slides to be imaged for this round. Typically, the gain is adjusted to the sample with the high fluorescence signal to avoid signal saturation.

6. Open the other channels depending on the fluorophore.
7. Press “snap” to acquire an image at high resolution.

### **2.2.6 Automated foci quantification using CellProfiler**

The CellProfiler software is a convenient tool to analyze cell-imaging data. You can insert instructions for each image, and the software can automatically process a list of images through several cycles of the same instructions. Thus, the set of instructions (or “pipeline”) requires optimization.

1. The first step is to optimize the pipeline. A default pipeline for nuclear foci counting can be found on the CellProfiler website. Adjustments to the settings can be tweaked depending on the antibody or probe target.
2. Load a set of training images for pipeline optimization. Files should be named with keywords to identify the image type. The pipeline has the following steps:
  - a. Identify primary objects – these are normally the DAPI images
  - b. Mask image – this creates a mask for the DAPI which only focuses on objects inside the nuclei
  - c. Identify secondary objects – these can be nuclear foci. The diameter of the object can be set (normally around 3-50). The threshold strategy, method to extinguish between signal or background, can be changed. The threshold correction factor can be shifted up or down to make the cutoff more stringent or lenient.
  - d. Measure intensity – compute the intensity of the object
  - e. Relate object – groups the secondary objects into the appropriate primary objects (which are numbered).
3. Start the test mode to check a set of training images and make adjustments accordingly.

4. After satisfactory adjustments are made, load the entire set of images and press “analyze images” to process all the images at once. This process usually takes a long time.

### **2.3 Real-time Live-cell imaging in Incucyte ZOOM**

Real-time live-cell optical imagers, available commercially, can be used to image cell growth and monitor treatment-induced cytotoxicity. IncuCyte ZOOM is a microscope add-on to a standard cell incubator and an external controller that collects and processes images. Appendix 1 contains more details on this procedure.

1. Cells are seeded into a 96-well plate at pre-optimized density (typically 1000-10000 cells/well) for a 5-day study.
2. Place 96-well plate inside the IncuCyte plate holder and image for 1 day. Check for proper cell growth.
3. The next day, the plate can be taken out, and the desired treatment regimen can be added to the cells. It is also possible to do a factorial experiment with a range of different concentrations for two different drugs.
4. Images can be collected at 2-3 hour time intervals on three channels: phase-contrast, green and red.
5. The Phase-contrast channel can be used to calculate cell confluency.
6. Quantifiable output can be analyzed by initial normalization. The values are normalized to the maximal reachable confluency value (usually untreated or vehicle control) and the lowest reachable confluency value (usually treated with a high dose). Normalized value can then be used to generate a dose-response curve.

## 2.4 Chromatin-immunoprecipitation (ChIP) & Re-ChIP

### 2.4.1 Reagents and Buffers

- 37% Formaldehyde
- PBS
- 1M Glycine (sterile filtered)
- 10 mg/mL BSA (sterile filtered)
- Fixing Buffer (50 mM Hepes KOH, pH 7.5, 100 mM NaCl, 1 mM EDTA, pH 8.0, 0.5 mM EGTA, pH 8.0)
- Shearing Buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris, pH 8.1)
- 1:1 IP Buffer (For diluting 1:1 with sonicate) (0.1% SDS, 1 mM EDTA, 10mM Tris, pH 8.1, 2% Triton X, 300 mM NaCl)
- IP Buffer (0.1% SDS, 1 mM EDTA, 10mM Tris, pH 8.1, 1% Triton X, 150 mM NaCl)
- ChIP Wash Buffer A (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl)
- ChIP Wash Buffer B (0.1% SDS, 1% Triton X-100  
2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl)
- ChIP Wash Buffer C (0.25 M LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA,  
10 mM Tris-HCl pH 8.0)
- TE Buffer (10 mM Tris pH 8.0, 1 mM EDTA)

- ReChIP Wash Buffer (0.1% SDS, 1% NP40, 2 mM EDTA, 500 mM NaCl)
- Primary antibody (eg. BG4 antibody, anti- $\gamma$ H2AX antibody)
- Beads (eg. M2 anti-FLAG agarose, Protein G sepharose)
- Qiagen PB binding buffer
- DNA extraction kit (BioBasics)

### 2.4.2 Procedures

The ChIP procedure can typically be completed in 3 days and Re-ChIP in 4 days. However, cell culturing requires prior planning in order to fit the ChIP procedure within the same week. Cells are usually seeded in 10 cm dishes (250000-1000000 cells/dish) for a typical 3-4 day experiment.

#### Cross-linking and Sonication (Day 1)

1. Aspirate media and add 10 mL of 1% formaldehyde/1X Shearing Buffer for 5 min on a rocker in RT.
2. Quench cross-link by adding 0.125M glycine (1.5mL of 1.0M stock) and incubate 5 min on a rocker in RT.
3. Remove aldehyde waste.
4. Add 800uL of 1X PBS and scrape cells into a conical tube
5. Spin tube at 3k rpm for 3 min (You can snap freeze it here but it's not recommended for high-value samples)
6. Aspirate and resuspend in 300uL Shearing Buffer (+PIC).

7. Sonicate on Covaris m220 using 10% duty Chromatin shearing preset for 4 min. Each sonication tube can only contain 130 uL, so repeat until all lysate is sonicated. Make sure there are no bubbles forming during sonication. Use different tubes for different samples.
8. Spin sonicate at 13200 rpm for 10 min at 4°C. Move the supernatant to new tubes.
9. The sonicate can now be snap-frozen and stored at -70°C for later use. However, IP done on the same day are usually better.
10. Preclear using non-blocked beads (20uL of 50% slurry) for 1 h. Spin down at 5K rpm for 1 min.
11. Split the sonicate into two fractions (120uL). Save 5% (6ul) for 1<sup>st</sup> ChIP input control for later.
12. Dilute in 1:1 ratio (usually 120ul +120u) with 2X IP Buffer.
13. Add 0.1% BSA to each tube (approximately 26uL of 10mg/mL stock).
14. Add 5ug of 1<sup>st</sup> antibody to each tube.
15. Incubate on a rotator overnight at 4°C.

#### Bead-binding & Washing (Day 2)

1. On the following day, add 25 uL of 50% slurry of pre-blocked agarose beads and incubate at 4°C on a rotator for 2 h.
2. Centrifuge beads at 5000 rpm for 1 min.
3. Wash the beads sequentially using the same following repeated steps:
  - Centrifuge, aspirate, add wash solutions, 5 min at room temperature on a rotator
  - Sequence of wash solutions:
    - i. 1 mL ChIP Wash Buffer A (2 times)

- ii. 1 mL ChIP Wash Buffer B (2 times)
  - iii. 1 mL ChIP Wash Buffer C (2 times)
  - iv. 1 mL TE (2 times)
4. After the final wash, add 100  $\mu$ L of freshly prepared 1% SDS/TE to the samples.
  5. Incubate on rotator at room temp for 30 min then incubate sample overnight at 65°C to reverse crosslinking.

ChIP-ReChIP Procedure (This replaces the Bead-binding & Washing step of single ChIP)

*1<sup>st</sup> ChIP Wash (Day 2)*

1. Add 25  $\mu$ L of 50% slurry of pre-blocked 1<sup>st</sup> ChIP beads (e.g. Protein G) and incubate for 2h on rotator at 4C.
2. Wash 2X with ChIP Wash A, B, C, TE.
3. Add 50  $\mu$ L of 1%SDS/10mM DTT/1X PIC/TE (fresh DTT!) to each tube. Incubate for 30 min at 37C.
4. Spin down and transfer 50  $\mu$ L to a new tube.
5. Make 1X PIC in IP Buffer.
6. Add 950  $\mu$ L of IP Buffer to beads to wash remaining. Spin down and take the 950  $\mu$ L and add to the tube with 50  $\mu$ L eluted fraction.
7. Combine the two fractions for approximately 2mL of the eluate.
8. Split into two fractions (~1mL) each. One will be with the antibody, and the other will be the no antibody control. Take a 5% control (50  $\mu$ L).
9. Add 0.1% BSA (~100  $\mu$ L) to each tube.
10. Add 4-5  $\mu$ L of 2<sup>nd</sup> ChIP antibody to each tube.

11. Incubate O/N on rotator at 4C.

*2<sup>nd</sup> ChIP Wash (Day 3)*

1. Add 20uL of 50% slurry of pre-blocked 2<sup>nd</sup> ChIP beads (eg. M2 beads) and incubate for 2 hours on rotator at 4C.
2. Wash 3X with ReChIP Wash buffer.
3. Wash 1X with TE Wash buffer
4. Add 100uL of 1% SDS/15mM DTT/TE to each tube.
5. Leave on rotator at RT for 30 min before incubating at 65 O/N.
6. Extract using column purification the next day.

DNA Extraction and ChIP-qPCR (Day >3)

1. Centrifuge beads at 5000 rpm for 1 min and transfer 100 uL of the supernatant to a new tube.
2. Add 5X volume of Qiagen PB binding buffer to DNA solution.
3. Mix by pipetting up & down until homogenous and transfer to column (BioBasic).
4. Spin at 11500 rpm for 1 min. Discard flow through.
5. Wash with 500 uL Wash buffer (ethanol should be added already) and centrifuge at 11500 rpm for 1min. Discard flow through.
6. Wash again with 500 uL and centrifuge at 11500 rpm for 1 min. Discard flow through.
7. Centrifuge again at 11500 rpm for 1 min. Transfer column to low binding 1.5 mL centrifuge tubes.
8. Add 35 uL of nuclease-free elution buffer (2 mM Tris-HCl) to column filter and incubate at room temperature for 1 min. Centrifuge at 13200 rpm for 2 min to elute the DNA.

9. For qPCR strip-tubes, there should be a total of 10 uL. Run triplicates. Each tube should contain the following: DNA sample (1 uL), Forward Primer (0.5 uL), Reverse Primer (0.5 uL), 2x SYBR mix (5 uL), PCR-grade water (3 uL). Make a master mix and pipette to each tube before adding 1 uL to each.

## **2.5 G4-RNA Precipitation (G4RP)**

G4-RNA targets can be pulled down using the G4RP protocol, which uses the G4-specific probe, BioTASQ. BioTASQ contains a biotin tag that allows streptavidin-mediated affinity purification of the bound targets. Please refer to Appendix 2 for a detailed step-by-step guide to the G4RP protocol.

## **2.6 C-circle Assay (CCA) with qPCR**

This protocol is based on the method developed by Reddel's group with several modifications. I find the modifications significantly improve the CCA signal compared to the original protocol for cell-based investigations. This protocol takes advantage of the phi29 polymerase, which has a higher specificity to amplify circular DNA.

### **2.6.1 Reagents and Buffers**

- Quick C-circle Prep (QCP) Lysis Buffer (50 mM KCl, 10 mM Tris-HCL pH 8.5, 2 mM MgCl<sub>2</sub>, 0.5% IGEPAL, 0.5% Twee-20) – 10X stock can be made
- Proteinase K – 10mg/mL stock
- Phi29 polymerase and buffer
- 1M DTT
- Tween-20
- BSA

- Deoxynucleotides – dATP, dGTP, dTTP, dCTP
- Qiagen PB binding buffer
- DNA extraction kit (BioBasic)

## **2.6.2 Procedures**

### Quick C-circle Lysate Prep

1. Typically, cells seeded in a 12-well plate are enough for CCA protocol. A smaller amount may also work.
2. Either freeze down the pellet or proceed to the next step. Thaw cell pellet if using a frozen sample.
3. Thaw proteinase K on ice and mix gently. Pre-heat QCP lysis buffer to 56°C in the incubator.
4. Add proteinase K to a final concentration of 200 µg/mL into QCP buffer, flick-mix, and quick-spin.
5. Add 50 µL QPC lysis buffer to each tube with the pellet. Ensure lid is closed.
6. Vortex at 1400 rpm (speed 4-5 for most basic vortexer) for 15s.
7. Place in 56 C for 1 hour. Vortex 1400 rpm intermittently (every 15 min).
8. Vortex 1400 rpm and incubate at 85°C for 10 min.
9. Cool to room temp and spin down.
10. Vortex for 15s at 2000rpm and spin down again
11. The lysate can be stored at -20C. Aliquot to avoid the freeze-thaw cycle.

### DNA purification

1. Typically, half of the lysate (25 µL) is normally sufficient for downstream analysis.

2. Add 700  $\mu\text{L}$  Qiagen PB binding buffer to DNA solution.
3. Mix by pipetting up & down until homogenous and transfer to column (BioBasic).
4. Spin at 11500 rpm for 1 min. Discard flow through.
5. Wash with 700  $\mu\text{L}$  Wash buffer (ethanol should be added already) and centrifuge at 11500 rpm for 1min. Discard flow through.
6. Wash again with 700  $\mu\text{L}$  and centrifuge at 11500 rpm for 1 min. Discard flow through.
7. Centrifuge again at 11500 rpm for 1 min. Transfer column to low binding 1.5 mL centrifuge tubes.
8. Add 30  $\mu\text{L}$  of nuclease-free elution buffer (2 mM Tris-HCl) to column filter and incubate at room temperature for 1 min. Centrifuge at 13200 rpm for 2 min to elute the DNA.

#### C-circle rolling circle amplification

Make the following Master Mix (2.16x) for rolling circle amplification (make this and aliquot and freeze stock)

1. Make the CCA rolling-circle amplification master mix (2.16x) (this can be aliquot and made into frozen stocks) with following pre-calculated formulation and the following stocks: For 2.16x master mix (925  $\mu\text{L}$ ): Ultrapure water (577  $\mu\text{L}$ ), 1M DTT (8  $\mu\text{L}$ ), 10x phi29 Buffer (200  $\mu\text{L}$ ), 0.2 mg/mL BSA (40  $\mu\text{L}$ ), 10% Tween-20 (20  $\mu\text{L}$ ), 100 mM dATP (20  $\mu\text{L}$ ), 100 mM dGTP (20  $\mu\text{L}$ ), 100 mM dTTP (20  $\mu\text{L}$ ), 100 mM dCTP (20  $\mu\text{L}$ )
2. Dilute purified DNA by 10x in 2 mM Tris pH7.6.
3. CCA reaction mix (total 20  $\mu\text{L}$ ) – use PCR tubes:
  - a. 9.25  $\mu\text{L}$  – 2.16x master mix
  - b. 0.75  $\mu\text{L}$  – phi29 polymerase
  - c. 10  $\mu\text{L}$  – diluted purified DNA

4. Mix gently and quick-spin. Keep on ice until they are ready for the thermal cycler.
5. Put into the thermal cycler with the following protocol:
  - d. 30°C for 8h
  - e. 70°C for 20min
  - f. 11°C forever
6. Real-time qPCR-quantify using tel1b-tel2b primer set

## **Chapter 3: Real-time and quantitative fluorescent live-cell imaging methods**

*(BBA-General Subjects, 2017)*

### **3.1 Summary**

The development of quadruplex-directed molecular diagnostic and therapy rely on mechanistic insights gained at both cellular and tissue levels by fluorescence imaging. This technique is based on fluorescent reporters that label cellular DNA and RNA quadruplexes to spatiotemporally address their complex cell biology. The photophysical characteristics of quadruplex probes usually dictate the modality of cell imaging by governing the selection of the light source (lamp, LED, laser), the optical light filters and the detection modality. Here, we report the characterizations of prototype from a new generation of quadruplex dye termed G4-REP (for quadruplex-specific red-edge probe) that provides fluorescence responses regardless of the excitation wavelength and modality (owing to the versatility gained through the red-edge effect), thus allowing for diverse applications and most imaging facilities. This is demonstrated by cell images (and associated quantifications) collected through confocal and multiphoton microscopy as well as through real-time live-cell imaging system over extended period, monitoring both non-cancerous and cancerous human cell lines. Our results promote a new way of designing versatile, efficient and convenient quadruplex-reporting dyes for tracking these higher-order nucleic acid structures in living human cells. This article is part of a Special Issue entitled "G-quadruplex" Guest Editor: Dr. Concetta Giancola and Dr. Daniela Montesarchio.

### 3.2 Introduction

Ligand-DNA interaction is a subtle balance between polar (*i.e.*, electrostatic) and non-polar (*i.e.*, hydrophobic) contributions. It is now generally accepted that the coulombic attraction of DNA to cationic ligands drives binding while local interactions (including H-bonds and  $\pi$ -stacking) further stabilize the ligand-DNA complexes (201). While polar and non-polar contributions of the ligand-DNA associations are well characterized, less is known about the influence of the aqueous environment on the ligand-DNA association. Water indeed plays critical roles in molecular and cell biology, being actively involved in the folding, dynamics and ligand binding of biomolecules such as proteins and nucleic acids (202-203). Water also forms clathrate-like structures around non-polar solutes (*i.e.*, organic ligands) (204-205) and ordered networks at hydrophobic interfaces (*e.g.*, in between organic ligands and their nucleic acid targets, when they interact *via*  $\pi$ -stacking) in which individual water molecule mobility is restricted as compared to bulk water [206-208]. Local modifications of hydration shells provide invaluable information about the interactions between the solute and its environment. These modifications are of critical importance when the solute is fluorogenic; under specific conditions (*vide infra*), the contacts between surrounding constrained water molecules and fluorophores are so entwined that they can be regarded as a single quantum system, characterized by its own energy states. This new fluorogenic system is endowed with its own ground, excited and intermediate states, which offer the possibility of absorbing light at wavelengths far beyond the upper limit defined by the UV-vis spectrum of the fluorophore alone. This phenomenon, known as red-edge effect (REE) due to the absorbance wavelengths being red-shifted, opens fascinating optical opportunities (209).

We recently reported on the very first quadruplex-specific red-edge probe (G4-REP); (210) the molecular basis of our approach is based on the targeting of native G-quartets (planar arrays of four guanines that constitute the major ligand binding site of DNA/RNA quadruplexes, (211)) by fluorogenic synthetic G-quartets (a biomimetic quadruplex ligand and turn-on fluorescent probe known as NaphthoTASQ, or N-TASQ, for naphthalene template-assembled synthetic G-quartet, Fig. 3.1). N-TASQ belongs to the family of *twice-as-smart* ligands (both a smart quadruplex ligand and a smart fluorescent probe) that were already reported to allow for visualizing quadruplexes *in vitro* (71, 212) and in live human cells *via* confocal and two-photon laser scanning microscopes (72). Both the aromatic nature of G-quartets (ligands and targets) and the efficiency of their nature-inspired interactions (69, 213) result in highly ordered hydration shells around the TASQ/quadruplex assemblies; when combined with the spectroscopic properties of the N-TASQ template (a naphthalene), the resulting assemblies exhibit red-edge excitation shifts that allow them to absorb light up to the near-infrared region and emitting light correspondingly, in an excitation wavelength-dependent manner, thus deviating from the classical applications of the Kasha-Vavilov rules (*vide infra*). The REE is usually restricted to rigid or highly restrained environments (low-temperature glasses, condensed polymers, *etc.*) (209, 214-217). This very first application to a chemical biology purpose (*i.e.*, the assessment of the existence of alternative nucleic structures in cells) (5, 200, 218-220) was critical in that it provides an opportunity for new ways of designing efficient and versatile nucleic acid stains.

Here, we expand on the scope of applications for G4-REP by demonstrating that N-TASQ is suitable for quantitative real-time analysis of live-cell behavior over long incubation times, owing to the recent development of live fluorescent imaging in native cell growth environment

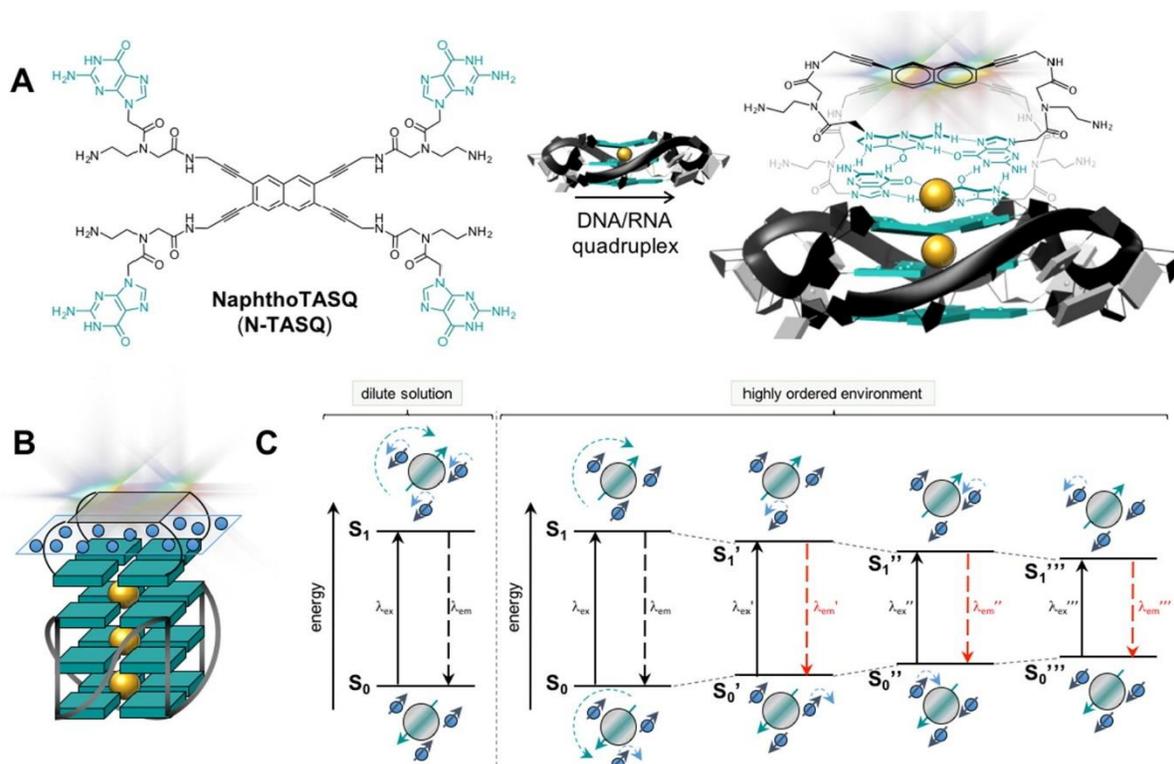
inside tissue culture incubators. This is a notable achievement, given the low-energy of the incoming light source as excitation (light-emitting diodes, LED), at specific wavelengths normally not within the range of the known photophysical properties of N-TASQ but made possible with the REE effect.

### 3.3 Results

#### 3.3.1 The red-edge effect: a mechanistic reminder

Fluorescence is mainly governed by the Kasha-Vavilov rules which stipulate that “the *quantum yield* of luminescence is independent of the wavelength of exciting radiation”. These rules explain that an irradiation at any wavelength ( $\lambda_{\text{ex}}$ ) within the absorption spectrum of a fluorophore provides an emission centered on a single wavelength ( $\lambda_{\text{em}}$ ), only the emission intensity varies according to the molar absorptivity corresponding to the selected irradiation wavelength. The Kasha-Vavilov rules apply to all fluorophores, which translate to an independence of  $\lambda_{\text{em}}$  on  $\lambda_{\text{ex}}$ . However, some exceptions exist. In dilute solution, fluorescence emission usually occurs slower than solvent relaxation (by 1 to 4 order of magnitude), explaining why solvent relaxation does not influence the emission of the fluorophore (and why solvent relaxation is usually neglected). In contrast, in highly ordered environments, the solvent mobility is restricted and solvent relaxation becomes slower than fluorescence emission (with the same order of magnitude, or less). This leads to various intermediates with partially aligned dipole moments, and thus various energies ( $S_1'$ ,  $S_0'$ ,  $S_1''$ ,  $S_0''$ , *etc.*, Fig. 3.1C) in between the ground ( $S_0$ ) and excited states ( $S_1$ ) of the dye, making dye sensitization possible with photons of lower energy (that is, with red-shifted wavelengths) than those strictly defined by the absorption spectrum of the dye. This red-shifted excitation is responsible for the REE, which is

characterized by a rather uncommon dependence of  $\lambda_{em}$  on  $\lambda_{ex}$ . Thus, while REE does not violate the Kasha-Vavilov rules *per se*, it is made possible under very specific molecular crowding conditions.



**Figure 3.1 Principles of the twice-as-smart quadruplex ligands.**

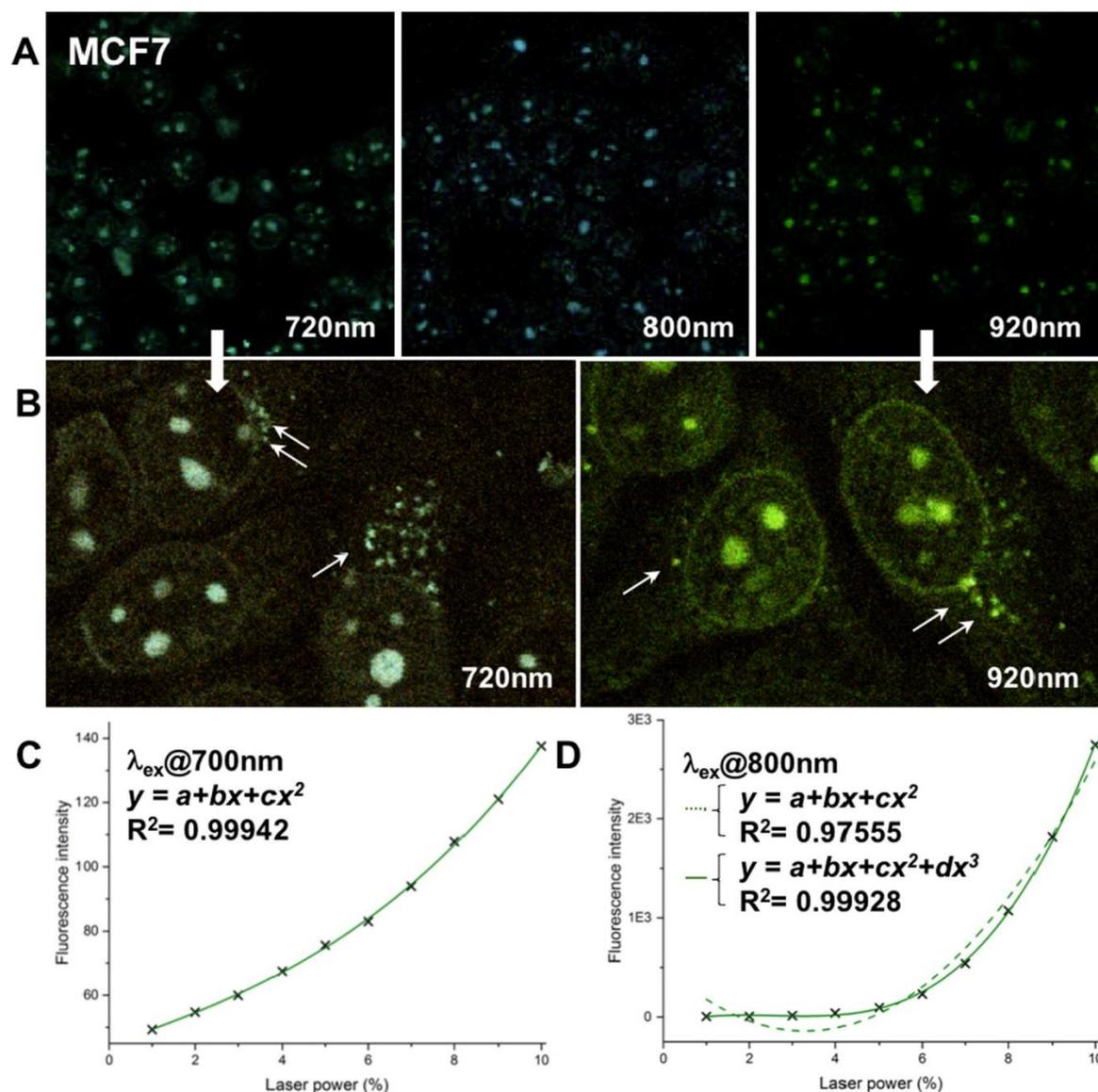
(A) structure of N-TASQ under its open (left) and closed conformations (right); the bioinspired interaction between N-TASQ and quadruplexes takes place through quartet-quartet recognition, stabilized by  $\pi$ -stacking interactions and cation chelation. B/C. Principles of the red-edge effect: (B) Schematic representation of the TASQ/quadruplex assembly, the ordered hydration shell is represented as blue circles and pale blue rectangle. C. Simplified Jablonski diagram depicting the ground ( $S_0$ ) and excited states ( $S_1$ ) of a fluorophore in dilute conditions (left), in which solvent (blue circles) relaxation occurs more rapidly than light emission, and in a constrained environment (right), in which solvent relaxation is slowed down: this condition creates a continuous model of solvent relaxation, characterized by intermediate ground ( $S_0'$ ,  $S_0''$ , etc.) and excited states ( $S_1'$ ,  $S_1''$ , etc.), leading to intermediate excitation and emission wavelengths that account for REE.

### 3.3.2 N-TASQ multimodality: multi-photon absorption and the red-edge effect

We previously demonstrated that N-TASQ is the first prototype G4-REP *via* a series of *in vitro* and cell-based fluorescence investigations performed far beyond the limits defined by its own absorbance spectrum ( $< 320$  nm). Briefly, we first demonstrated the dependence of the wavelength of the emission maximum on the excitation wavelength,  $\lambda_{em}^{max} = f(\lambda_{ex})$  (with  $\lambda_{ex}$  every 10 nm between 488 and 588 nm and  $\lambda_{em}^{max}$  from 593 to 673 nm) when N-TASQ (10  $\mu$ M) interacts with the quadruplex-forming sequence found in the human telomeres, 22AG (5  $\mu$ M, d[<sup>5'</sup>AG<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>']) *in vitro*. We subsequently demonstrated that N-TASQ/quadruplex assemblies provide fluorescence signals when excited at wavelengths corresponding to standard confocal laser excitation (*i.e.*, 408, 488 and 555 nm); stepwise titrations carried out for each excitation wavelength with N-TASQ (10  $\mu$ M) interacts with 22AG (from 0 to 5  $\mu$ M), providing emission patterns (with maxima at 474, 519, 592, 607 and 647 nm) that fit within the emission filter sets routinely used with confocal microscopes (DAPI:  $< 495$  nm, FITC: 495-590 nm, and Alexa:  $> 585$  nm), and thus demonstrating the suitability of N-TASQ for confocal microscopy analyses. We then exploited further these properties to visualize both DNA and RNA quadruplexes in cells. MCF7 cells were either live-incubated for 24 h with N-TASQ (2.5  $\mu$ M) before fixation and mounting steps required for confocal analyses, or labeled with N-TASQ (100  $\mu$ M) after cell fixation but prior to the mounting step, respectively. These series of experiments demonstrated the suitability of N-TASQ as REE nucleic acid stain for confocal analyses and emphasized its versatility for cell-based imaging experiments since it uniquely allows for *i*- tracking both DNA and RNA quadruplexes, *ii*- in both live and fixed human cells, *iii*- with both two-photon and confocal microscopes (72, 210).

An open question is whether the images collected with a two-photon microscope actually originate in two-photon absorption by N-TASQ bound to its nucleic acid targets or that REE may also take place under two-photon microscopy conditions; the latter option is certainly possible given the use of high-energy lasers as light source. To differentiate between the two possibilities, MCF7 breast cancer cells were incubated with N-TASQ both as a live-cell treatment (2.5  $\mu$ M for 24 h) and as a post-staining agent (100  $\mu$ M after MeOH fixation) and images were collected at selected wavelengths (*i.e.*, 720, 800 and 920 nm) with a two-photon microscope. Double N-TASQ treatments aim at maximizing the detection of its cellular targets, with live-cell incubation promoting cytoplasmic quadruplex-based ribonucleoprotein assemblies and the post-fixation labelling favoring nucleolar staining. As seen in Fig. 3.2A, high-quality images were collected at each wavelength, highlighting the expected strong nucleolar and cytoplasmic foci (white arrows in Fig. 3.2B). These results demonstrate that cellular N-TASQ/quadruplex assemblies are responsive to multiphoton excitation far beyond double the maximum of their one-photon absorption ( $< 320$  nm), which could represent a signature of the REE. To further investigate this phenomenon, the fluorescence intensity of a 2  $\mu$ M N-TASQ/quadruplex assembly (1:1 ratio; here, we used the quadruplex forming sequence found in the promoter region of the c-myc oncogene, d[<sup>5'</sup>GAG<sub>3</sub>TG<sub>4</sub>AG<sub>3</sub>TG<sub>4</sub>A<sub>2</sub>G<sup>3'</sup>]) was quantified as a function of the laser intensity (0–10%, *i.e.*, up to 3000 mW) under excitation wavelengths at 700 and 800 nm, respectively, to assess whether the resulting fluorescence emission varies with the square of the excitation intensity. As seen in Fig. 3.2C, we observed a good quadratic relationship between excitation at 700 nm and emission collected through the FITC channel ( $525 \pm 30$  nm), with a strong regression ( $R^2 = 0.99942$ ), which indicates that N-TASQ acts as a two-photon probe. More surprisingly, we found an excellent cubic relationship between excitation at 800 nm and

fluorescence emission (Fig. 3.2D), with a very strong regression ( $R^2 = 0.99928$  versus 0.97555 for the quadratic fit), which indicates that N-TASQ acts also as a three-photon probe. These results show that both two-photon and three-photon excitations can activate the fluorescence properties of N-TASQ/quadruplex assemblies. Despite the under-utilization and reporting of three-photon excitation in the literature, it occurs in the same way as the two-photon process except that three photons must interact with the fluorophore simultaneously to activate emission. In light of the quantum-mechanical properties involved, the photon density required for three-photon excitation is only about 5-to-10-times greater than the density needed for two-photon absorption, (221-223) making it useful for cell imaging, as was demonstrated here. Two- and three-photon excitations have both advantages (including deeper cell/tissue penetration, less photobleaching issues, *etc.*) and drawbacks (including lower absorption rates, requiring the use of high-intensity lasers that may damage cells, and lower brightness that requires the use of high-sensitivity detectors) as compared to conventional one-photon excitation. This highlights the interest of using multivalent probes, such as N-TASQ that is both a multi-photon and a REE probe, which could be visualized regardless of the microscopy facilities and optical setups, therefore allowing for the selection of the experimental conditions most adapted to the intended application.

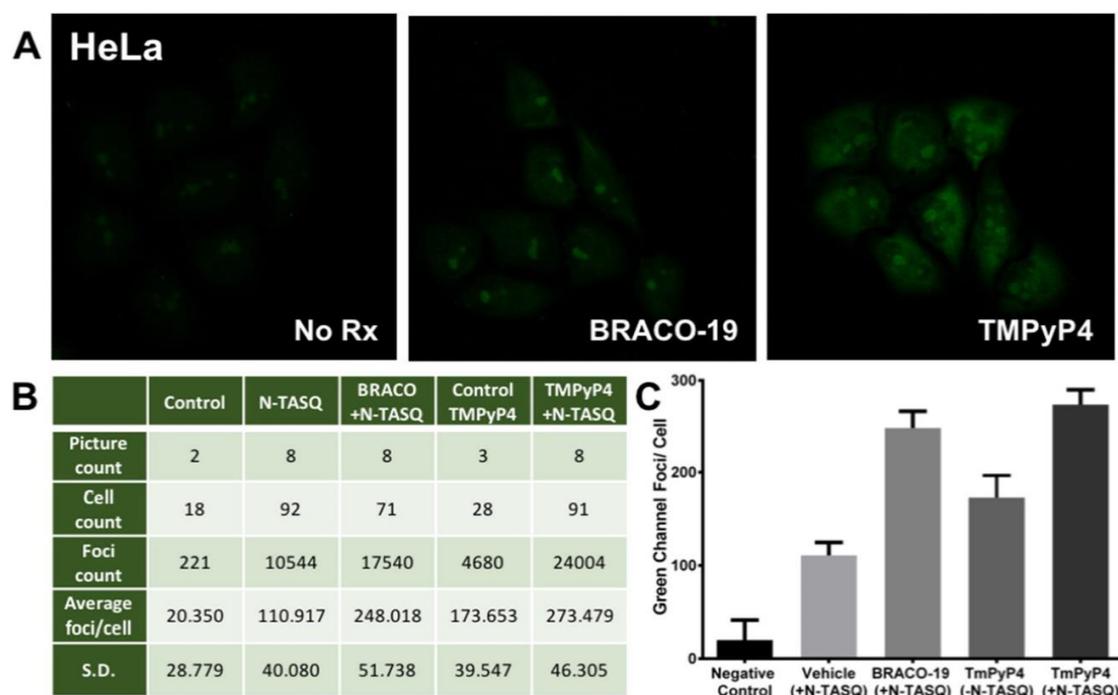


**Figure 3.2 Multiphoton images with N-TASQ.**

(A-B) Fluorescence cell images collected after both live-cell incubation and post-fixation labelling of MCF7 cells with N-TASQ (2.5 and 100  $\mu$ M, respectively) upon multi-photon excitation at either 720, 800 and 920 nm. White arrows indicate cytoplasmic quadruplex-based ribonucleoprotein assemblies. C,D. Analyses of the multiphoton properties of N-TASQ. Quantification of the fluorescence intensity of a N-TASQ/c-myc assembly (1:1 ratio, 2  $\mu$ M) through the FITC channel ( $525 \pm 30$  nm) as a function of the laser intensity (0–10%, *i.e.*, up to 3000 mW) under excitation wavelength at 700 nm (C) and 800 nm (D).

### 3.3.3 Disclosing the quadruplex entity of N-TASQ cellular targets: a quantitative analysis

In our previous studies, the quadruplex nature of the N-TASQ cellular targets was mainly demonstrated by co-labelling studies performed with N-TASQ and the quadruplex-specific antibody BG4 (37). Inspired by a benchmark study from Rodriguez *et al.* in which the authors performed live cell incubation with pyridostatin PDS- $\alpha$  to promote the formation of quadruplexes and visualize the resulting ligand-quadruplex assemblies *via* fluorescence measurements, (98) we performed similar live cell incubation experiments by treating human transformed cells with two established quadruplex-promoting agents, (225) *i.e.*, BRACO-19 (57) and TMPyP4 (226), and quantify the resulting quadruplexes through measuring the fluorescence signal intensity of N-TASQ, applied as a post-staining agent. HeLa cells were treated with BRACO-19 (1.8  $\mu$ M) or TMPyP4 (2.5  $\mu$ M) for 48 h before paraformaldehyde (PFA)/Triton X-100 fixation, followed by post-fixation staining with N-TASQ (10  $\mu$ M) for 3 h. Collected confocal images (Fig. 3.3) were subsequently quantified using the cell imaging analysis software Cell Profiler. Our results clearly demonstrated that N-TASQ was able to identify the formation of quadruplexes in cells by the two ligands in a statistically significant manner, with an average of  $248.0 \pm 51.7$  foci after BRACO-19 incubation (*versus*  $110.9 \pm 40.0$  foci in untreated cells) and  $273.5 \pm 46.3$  foci after TMPyP4 incubation (*versus*  $173.6 \pm 39.5$  foci in cells treated with TMPyP4 but without N-TASQ post-labelling, due to the intrinsic fluorescence of the porphyrin core under these experimental conditions). These results not only further support the quadruplex nature of the N-TASQ labelling targets but also clearly show that N-TASQ is an efficient G4-REP, labelling quadruplexes in human cells and offering a convenient way to quantify them computationally in an unbiased manner.



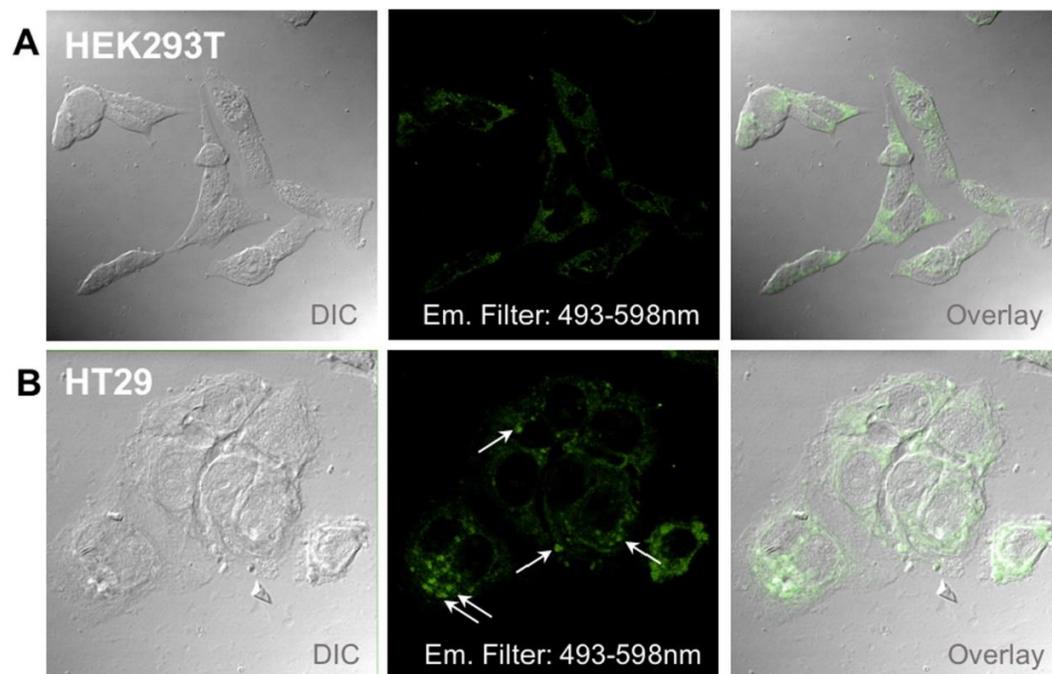
**Figure 3.3 Quantitative analysis of the N-TASQ cellular targets.**

(A) Confocal images of HeLa cells without treatment (upper panel, left) or treated with BRACO-19 (1.8  $\mu$ M, upper panel, center) or TMPyP4 (2.5  $\mu$ M, upper panel, right) for 48 h before fixation (PFA/triton) and post-staining with N-TASQ (10  $\mu$ M). (B-C) Quantification of N-TASQ cellular foci under different conditions through the FITC filter (495–590 nm) using Cell Profiler software.

### 3.3.4 N-TASQ, quadruplexes and the red-edge effect: what is next?

We next decided to further study the G4-REP properties of N-TASQ in cancerous (colorectal adenocarcinoma HT29) *versus* non-cancerous (embryonic kidney HEK293T) human cell models. These investigations were focused on the property of N-TASQ live-cell incubation and on whether N-TASQ can enter and label human cells differently depending on their transformation status. Both HT29 and HEK293T cell lines were live-incubated with N-TASQ (10  $\mu$ M) for 24 h and directly mounted (Fluoromount-G) for confocal imaging, without fixation step. As seen in Fig. 3.4, the high resolution images collected showed that N-TASQ labels both cell lines efficiently. However, a notable difference in fluorescence labelling pattern can be observed;

HEK293T cells displayed homogeneous cytoplasmic staining (Fig. 3.4A), whereas HT29 cells showed bright cytoplasmic foci along with a diffuse cytoplasmic staining (Fig. 3.4B). Our results indicated that N-TASQ can enter human cells efficiently regardless of their transformation status but highlighted different cellular responses upon N-TASQ exposure. The discrete foci seen in HT29 cells (white arrows, Fig. 3.4) likely correspond to quadruplex-based ribonucleoprotein assemblies that accumulate in cytoplasmic granules (227) in response to N-TASQ treatments. Notably, these foci are not seen in HEK293T under the same treatment conditions and dose of N-TASQ. We then decided to further explore this difference by monitoring the real-time labelling capabilities of N-TASQ in transformed *versus* immortalized cells over longer-term N-TASQ incubations.

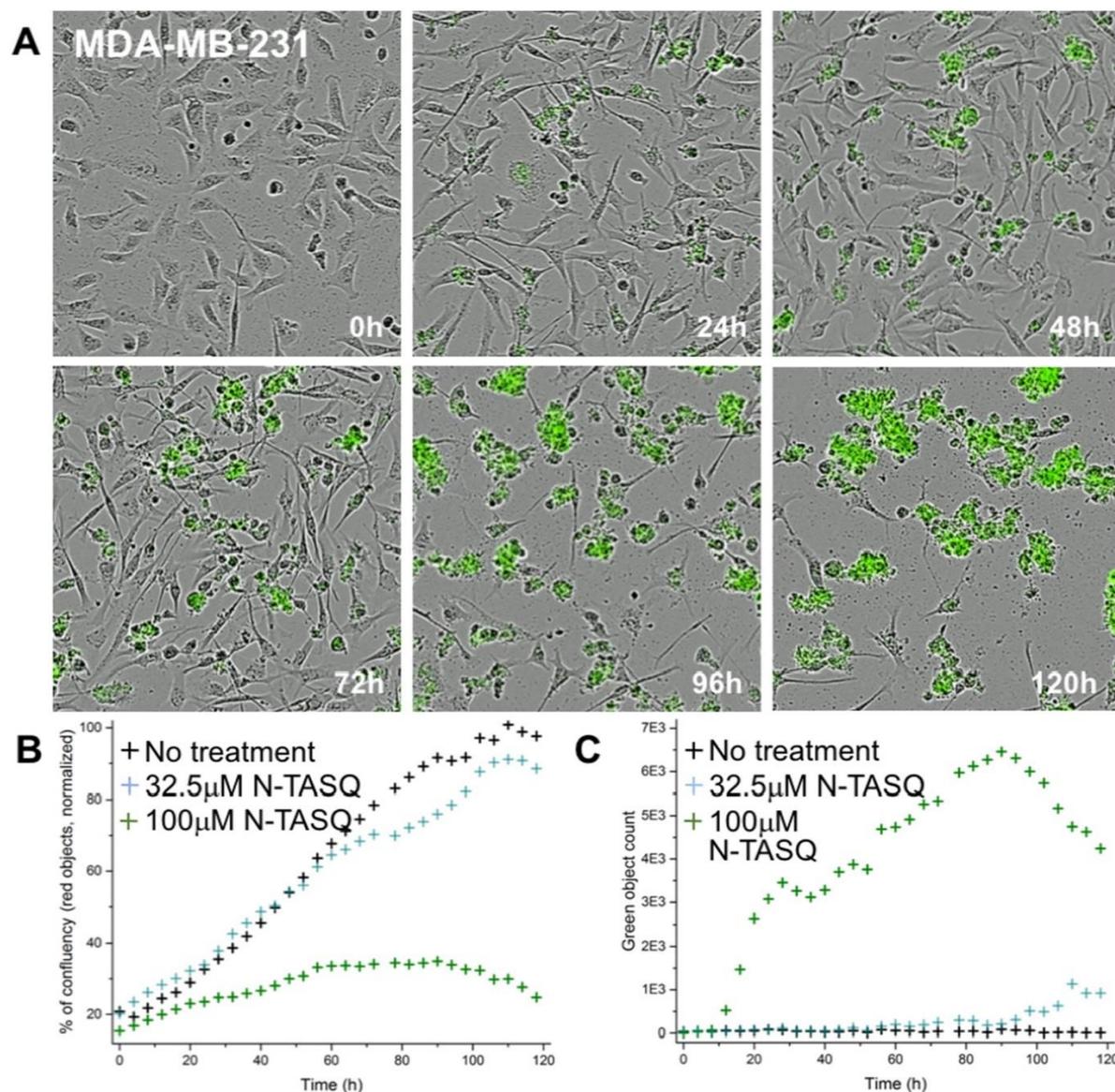


**Figure 3.4 Implementation of the red-edge effect for fluorescence cell imaging.**

Live cell incubation of HEK293T (A) and HT29 (B) cells with N-TASQ (10  $\mu$ M) for 24 h before the mounting step (Fluoromount-G); confocal images were collected after an excitation at 488 nm and visualized through differential interference contrast (DIC) and FITC channels (493–598 nm) or both (overlay). White arrows indicate cytoplasmic quadruplex-based ribonucleoprotein assemblies.

### 3.3.5 A spatiotemporal step beyond

To assess whether N-TASQ would be suitable for real-time analyses over extended incubation period, we employed the IncuCyte ZOOM live cell imaging system that relies on a microplate-compatible LED-based microscope housed inside a cell incubator. We first verified that N-TASQ can label cells and be detected through IncuCyte technology; to this end, we used breast cancer MDA-MB-231 cells transduced with NucLight Red lentivirus to produce red fluorescent proteins which label the breast cancer cell nucleus and can be visualized through the red channel of the microscope ( $\lambda_{\text{ex}} = 585$  nm, emission filter: 625-705 nm), leaving N-TASQ to be detected *via* the green channel ( $\lambda_{\text{ex}} = 460$  nm, emission filter: 504–544 nm) where its fluorescence signal is highest (*vide infra*). MDA-MB-231 NucLight Red cells were incubated with increasing concentration of N-TASQ (2-100  $\mu\text{M}$ ) for 120 h. The images collected through the green channel, seen in Fig. 3.5A for 100  $\mu\text{M}$  N-TASQ treatment (the corresponding movie is available as Supplementary material), clearly demonstrate that N-TASQ is suitable for real-time IncuCyte analysis. This achievement is notable in light of the low-energy excitation through a LED light source, at a wavelength ( $\lambda_{\text{ex}} = 460$  nm) normally not able to activate N-TASQ fluorescence. Interestingly, with real-time imaging, we observe N-TASQ accumulation inside and on the cytoplasmic membranes of the cells in a dose-response manner since higher concentrations lead to higher cell uptake, as seen in Fig. 3.5C (cyan and green crosses for 32.5 and 100  $\mu\text{M}$  N-TASQ concentration, respectively). We determine an  $\text{EC}_{50} = 58$   $\mu\text{M}$  at 120 h, calculated *via* NucLight Red fluorescence quantification (see the Supplementary material, Fig. 3.S1).



**Figure 3.5 Implementation of the red-edge effect for live cell imaging (part 1).**

(A) MDA-MB-231 cells were incubated with N-TASQ (100  $\mu\text{M}$ ) for 120 h; cells were continuously imaged (every 4 h) using the IncuCyte ZOOM Live-Cell Analysis System upon irradiation at  $\lambda_{\text{ex}} = 460 \text{ nm}$  and through the emission filter: 504–544 nm. (B) Normalized quantification of the MDA-MB-231 growth through the NucLight red object count for cells either untreated (black crosses) or treated with either 32.5  $\mu\text{M}$  (cyan crosses) or 100  $\mu\text{M}$  N-TASQ (green crosses). (C) Quantification of the MDA-MB-231 growth through the green object count (*i.e.*, N-TASQ fluorescence) for cells either untreated (black crosses) or treated with either 32.5  $\mu\text{M}$  (cyan crosses) or 100  $\mu\text{M}$  N-TASQ (green crosses).

It is noteworthy that given the G4-REP nature of N-TASQ, an overlap between its emission and the NucLight Red fluorescence occurs; however, we previously demonstrated (210) that antibody-based and N-TASQ labeling co-localization experiments can be performed reliably since the red-edge nature of the N-TASQ signals makes the quantum output much lower at the red channel, *a fortiori* far weaker than a dedicated dye with emission at dedicated wavelengths. Consequently, the two reporters can be reliably monitored through the red channel (where NucLight Red fluorescence is dominant, Fig. 3.5B) and the green channel (where only N-TASQ fluorescence is detected, Fig. 3.5C), making the cell growth quantification easier.

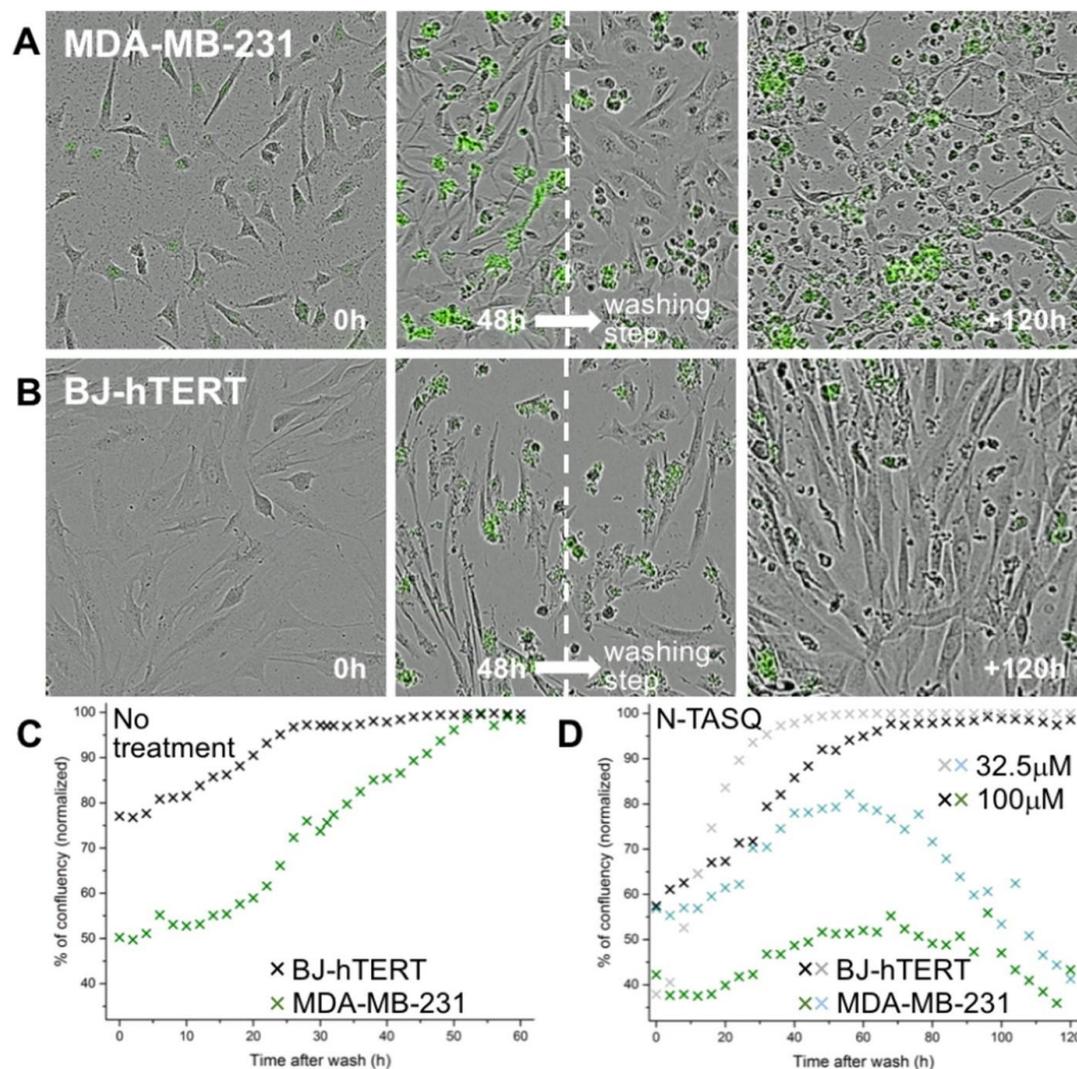
To further evaluate the anticancer activity of internalized N-TASQ following cellular uptake, we modified our experimental setup by introducing a washing step after 48 h N-TASQ incubation (at both 32.5 or 100  $\mu$ M concentration) to remove extracellular TASQ. We performed these investigations with both cancerous (breast cancer MDA-MB-231) and non-cancerous (immortalized foreskin fibroblast BJ-hTERT) human cell lines. Images collected through the green channel (for N-TASQ signals) can be seen in Fig. 3.6A-B for 100  $\mu$ M N-TASQ treatment (the corresponding movies are available as Supplementary material). With MDA-MB-231 cells (Fig. 3.6A), we observed accumulation of N-TASQ in both intra- and extra-cellular sites in the first 48 h. The washing step (which does not affect the growth of untreated cells, see Fig. 3.6C) clearly removed extracellular N-TASQ, while the internalized ligand keeps on acting as an anti-proliferative agent inducing massive cell necrosis by 120 h (Fig. 3.6A and the related quantification in Fig. 3.6D for 32.5 and 100  $\mu$ M, cyan and green crosses, respectively; see also the Supplementary material). In contrast, N-TASQ appeared to accumulate within a subpopulation of BJ-hTERT cells only (Fig. 3.6B for 100  $\mu$ M N-TASQ treatment), leading to rapid cell death. Again, the washing step neither affect the growth of untreated cells (Fig. 3.6C),

nor modify the labelling patterns (Fig. 3.6B) nor impede cell growth afterwards (Fig. 3.6D, grey and black crosses for 32.5 and 100  $\mu$ M N-TASQ concentration, respectively). Together, these results again demonstrate the versatility of N-TASQ as a G4-REP, allowing for *i-* the visualization of quadruplexes at the cellular level on real-time, *ii-* acting as an anti-proliferative agent at high doses in a transformation-status dependent manner, inducing cancer cell death effectively over non-transformed human cells, and *iii-* the quantification of induced quadruplex formation through automatic fluorescent foci counting in a straightforward manner. Our results provide the impetus to assess the specific toxicity of N-TASQ against cancer *versus* healthy normal cells in comparative experiments (as shown here) and also co-culture experiments in the near future to properly visualize and quantify the therapeutic window of such treatments.

### 3.4 Discussion

The art of fluorescence imaging with chemical sensors (228) has led to the parallel development of molecular tools and microscope facilities, with improved resolutions, as well as complete automation. These recent development have made live cell imaging readily accessible thanks to the arrays of commercial markers that allow for the specific labelling of subcellular compartments. For nucleic acid staining, the downside of employing these chemical probes often lies in their poor spectroscopic properties that most-often require excitation with high-energy blue light (mainly DAPI but also Hoechst dyes), which are damaging to nucleic acids as well as other cellular structures. A way to tackle this issue is to design probes excitable by red/far-red light, but the chemical modifications responsible for a significant red-shift of absorbance wavelengths often alter their water-solubility, cell permeability and target specificity. When a dye has addressed these concerns satisfactorily, other parameters must be taken into account,

such as its overall toxicity, compatibility with available imaging facilities, *etc.* All these constraints make the designing of a useful nucleic acid stain an involute task (229). A way to circumvent many of these issues is to design a new type of fluorophore.



**Figure 3.6 Implementation of the red-edge effect for live cell imaging (part 2).**

(A-B) MDA-MB-231 (A) and BJ-hTERT cells (B) were incubated with N-TASQ (100 μM) and continuously imaged (every 2 h) using the IncuCyte ZOOM Live-Cell Analysis System upon irradiation at  $\lambda_{ex} = 460$  nm and through the emission filter: 504–544 nm. After 48 h, a washing step removes the excess of N-TASQ and the imaging was continued for 120 h. (C-D) Normalized quantification of the cell growth (BJ-hTERT, black and grey crosses; MDA-MB-231, green and cyan crosses), either untreated (C) or treated with 32.5 μM (D, grey and cyan crosses) 100 μM N-TASQ (D, black and green crosses) after the washing step through the NuLight red object count/phase contrast analysis (MDA-MB-231 and BJ-hTERT cells, respectively).

The conventional discovery process is focused on the nucleic acids dye's spectroscopic properties *per se*, either alone in solution or in interaction with its target. UV–vis titrations allow for the determination of useful wavelengths (*e.g.*, absorption maxima, isosbestic points), which are usually restricted over a few dozen nanometers. An alternative to this development standard is to consider the protagonists (dyes and targets) in their actual solvation envelope. Upon strong interaction between the two partners, the water molecule motions in the hydration layer of the dye/target assembly are slowed down, giving rise to various intermediate relaxed states that could be excited by photons of lower energy and emit photons of even lower energy. This leads to the possibility of sensitizing the dye on its long wavelength edge (*i.e.*, the red edge) over hundreds of nanometers, *via* the so-called red-edge effect (REE).

Herein, we demonstrate the applications of REE probes in the field of cell biology and chemical genetics, focusing on their use to visualize quadruplexes in human cells by fluorescence cell imaging. The first prototype of quadruplex-selective red-edge probe (G4-REP) studied here, N-TASQ, provides good fluorescence responses through confocal analysis both as a post-fixation staining agent and *via* live-cell incubation. Even though N-TASQ labels both cancerous and non-cancerous cells, its binding patterns and cellular effects differ between these cell models, as demonstrated by long-term incubation experiments carried out with the breast cancer MDA-MB-231 and the fibroblast BJ-hTERT models. While N-TASQ enter the two human cell lines at low concentrations and for short incubation times (10  $\mu$ M, 24 h), it harbored a significant propensity to adhere to the surface of cancer cells at higher doses and for extended period of time (100  $\mu$ M, 120 h), subsequently triggered massive cell death by necrosis. This observation implies the existence of cancer-specific membrane receptors, which must also promote and/or stabilize the closed conformation of N-TASQ to make it fluorescent, as seen in Fig. 3.5 and Fig. 3.6. With all

these clues, we contend that cell surface nucleolin may be involved: (230) *i*- nucleolin is over-expressed at the surface of cancer cells (MDA-MB-231 strongly express nucleolin while confluent BJ cells do so to a lesser extent; (231) more generally, it is known that the presence of surface nucleolin is higher in cancerous cells than in non-transformed, normal cells), and *ii*- nucleolin displays a high-affinity for quadruplex-forming oligonucleotides, that is, for G-quartet-containing binding agents (232). We hypothesize that N-TASQ may interact with cell surface nucleolin through its closed conformation (its fluorescent conformation), thereby triggering cell death in a way that is similar to the guanine-rich aptamer AS1411 developed by P. Bates and colleagues (232). To investigate this, we performed co-staining experiments with TAMRA-labeled AS1411 (AS1411-T) and N-TASQ (seen in the Supplementary material). Concomitant or sequential cell incubations with both AS1411-T and N-TASQ provide evidences that both molecular tools share common cellular targets (likely to be nucleolin), albeit with different affinity and specificity (further discussed in the Supplementary material, see Fig. 3.S2 to 3.S6). Further efforts will be necessary to assess the validity of this hypothesis, but undoubtedly, these results will provide the experimental tools and incentives for future investigations.

### **3.5 Methods**

#### Materials

N-TASQ was prepared according to the protocol described in (72); BRACO-19 and TMPyP4 were purchased from Sigma-Aldrich and AS1411-T from Eurogentec and used without further purification.

### Cell culture

HeLa, MCF7, HT29, HEK293T and U2OS cell lines were obtained from the American Type Culture Collection (ATCC). MDA-MB-231 NuLight Red cell line was obtained from Essen Bioscience. BJ fibroblast cells were obtained from ATCC and immortalized by transducing in hTERT. Cells were cultured in 75 and 175 cm<sup>2</sup> tissue culture flasks (Corning) in DMEM (Life Technologies) supplemented with 5% synthetic fetal bovine serum (FetalClone III, GE Life Sciences) and 100 U penicillin-streptomycin mixture (1.0 U · mL<sup>-1</sup> Pen/1.0 mg · mL<sup>-1</sup> Strep) at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere-controlled incubator (HERAcell). The standard protocols were used for subculturing the cells: aspiration of medium, PBS (Gibco) wash, trypsinization in Trypsin-EDTA (0.25%) and reseeding in appropriate density. All cell counting was performed using the Coulter Counter (Beckman Coulter).

### Fluorescence microscopy

Cells were seeded on glass coverslips. After 24 h, the cells were treated with either vehicle (ddH<sub>2</sub>O), BRACO 1.8 μM or TMPyP4 2.5 μM for 48 h. The cells were then washed with PBS and fixed in 3.7% paraformaldehyde, solubilized in 0.1% Triton X-100/PBS. The cells were then blocked in 4% bovine serum albumin in PBS for 1 h then incubate in 10 μM N-TASQ in PBS for 3 h before mounting onto the microscope slides. Confocal microscopy was performed either on a Zeiss LSM700 or on a Leica DMi8 microscope with the appropriate filters using the 63 × objective. Foci quantification is done using Cell Profiler with custom pipeline.

### Live-cell imaging

Cells were washed once with PBS, trypsinized and resuspended in fresh media. The cells were counted and seeded into the wells of the plates at appropriate densities. Cells were continuously imaged over a repeated schedule (every 2 h) in the IncuCyte ZOOM Live-Cell Analysis System

(Essen Bioscience) using 10 × objective. Analysis of live-cell images was performed using the IncuCyte ZOOM software. For cytotoxicity profiling, MDA MB-231 NucLight Red were seeded at a density of 3000 cells per well respectively in 96 well plate (Corning). After 24 h, the cells were treated with varying doses (2-100 μM) of N-TASQ for 5 days. Each treatment was performed with three repeats. The growth curve was plotted using the IncuCyte ZOOM software by analyzing phase-contrast or red object count (for NucLight Red). For the single treatment plus washing step, the media was removed and the cells were washed with PBS on day 3 before adding fresh media.

### **3.6 Supplementary Materials**

#### Cell culture

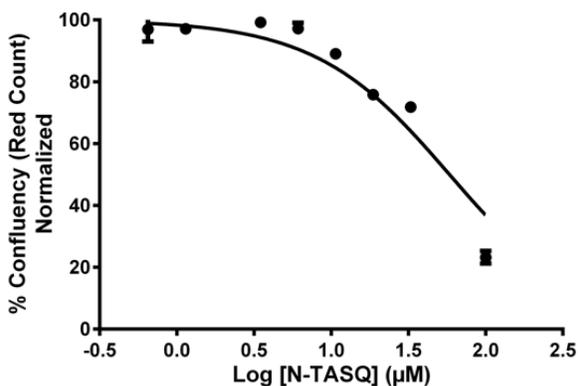
Cells were obtained from the American Type Culture Collection (ATCC).

MDA-MB-231 NucLight Red cell line was obtained from Essen Bioscience. Cells were cultured in 75 and 175cm<sup>2</sup> tissue culture flasks (Corning) in DMEM (Life Technologies) supplemented with 5% synthetic fetal bovine serum (FetalClone III, GE Life Sciences) and 100u penicillinstreptomycin mixture (1.0u·mL<sup>-1</sup> Pen/ 1.0mg·mL<sup>-1</sup> Strep) at 37°C in a humidified, 5% CO<sub>2</sub> atmosphere-controlled incubator (HERAcell). The standard protocols were used for subculturing the cells: aspiration of medium, PBS (Gibco) wash, trypsinization in Trypsin-EDTA (0.25%) and reseeded in appropriate density.

#### Cytotoxicity assessment

Cells were continuously imaged over a repeated schedule (every 2 hours) in the IncuCyte ZOOM Live-Cell Analysis System (Essen Bioscience) using 10X objective. Analysis of live-cell images was performed using the IncuCyte ZOOM software. For cytotoxicity profiling, MDA MB-231

Nuclight Red were seeded at a density of 3000 cells per well respectively in 96 well plate (Corning). After 24h, the cells were treated with varying doses 2-100 $\mu$ M) of N-TASQ for 5 days. Dose-response is generated (Fig. 3.S1).



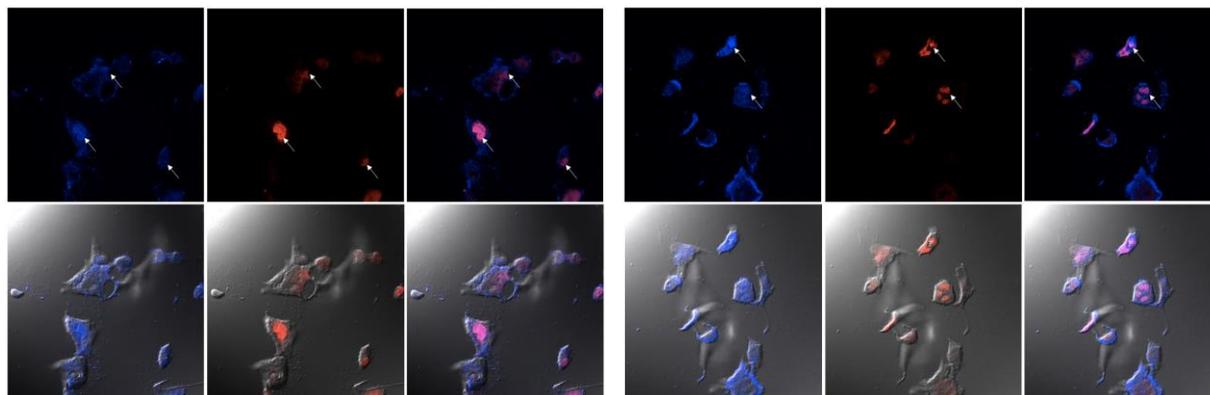
**Figure 3.S1.** Dose-response curve for N-TASQ in MDA MB-231

#### Nucleolin-related investigations

TAMRA-labeled AS1411, or AS1411-T (5'-G2TG2TG2TG2T2GTG2TG2TG2-3'-TAMRA) was purchased from Eurogentec (Seraing, Belgium). Both HEK293T and MCF7 cells were seeded onto glass coverslips in 8-well plates (Falcon), allowed to recover for 24h and treated as follows:

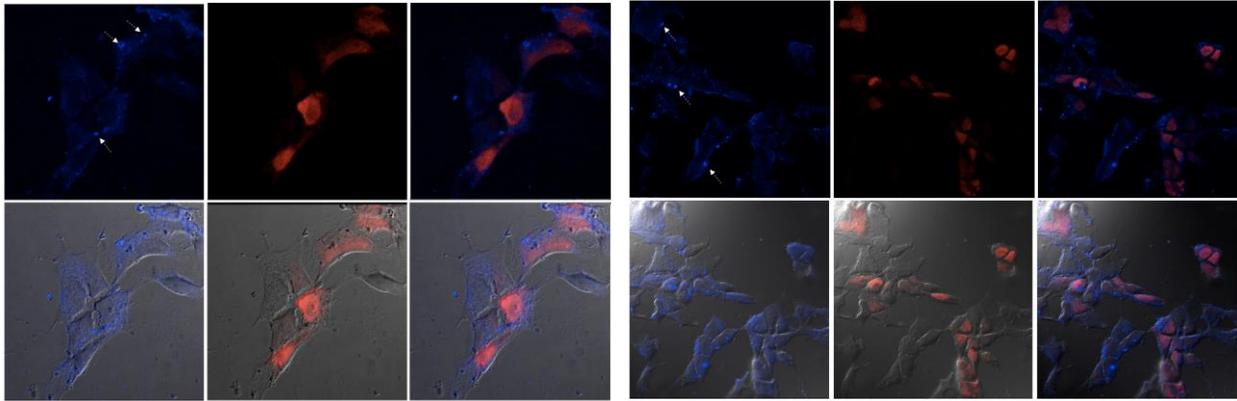
- Direct competition (Fig. 3.S2): MCF7 (Fig. 3.S2; left) or HEK293T (Fig. 3.S2; right) cells were fixed with PFA (4% in PBS) for 5min at room temperature and then incubated concomitantly with N-TASQ (5 $\mu$ M) and AS1411-T (5 $\mu$ M) for 15min at room temperature. Cells were then washed twice with PBS, mounted with Fluoromount-G, and images were collected through the “DAPI channel” (lex = 405nm, lem = 410-580nm) for N-TASQ labeling, and the “TAMRA channel” (lex = 552nm, lem = 559-705nm) for AS1411-T labeling, along with differential interference contrast. Cells were here fixed with PFA only (not supplemented with triton) in order not to completely disrupt membrane structure. Images seen in Fig. 3.S2 highlight that N-

TASQ (blue labeling) labels primarily cytoplasmic sites while AS1411-T (red labeling) is primarily located in the nucleus (where nucleolin is the most abundant). More interestingly, some foci (white arrows) clearly indicate co-labeling sites, where both N-TASQ and AS1411-T are located simultaneously.



**Figure 3.S2.** Direction competition experiments in MCF7 (left) or HEK293T (right)

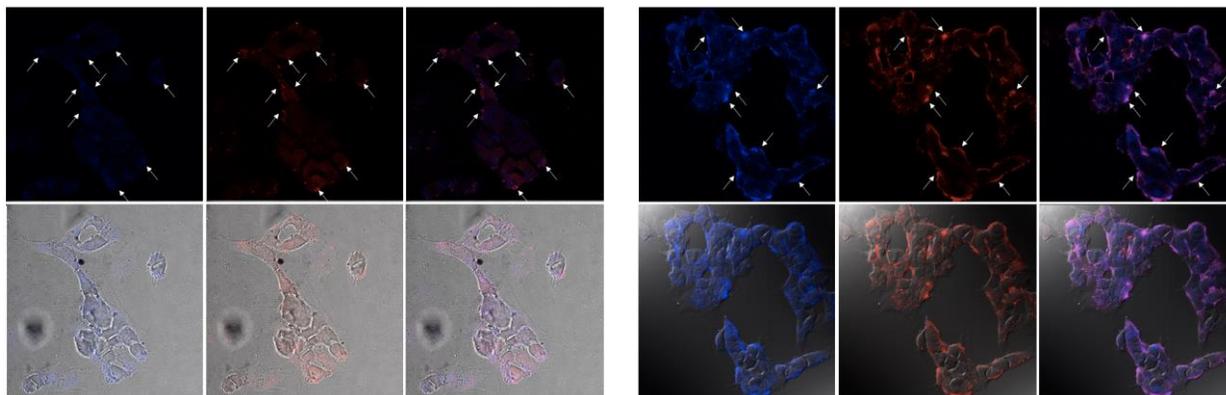
- Indirect competition (Fig. 3.S3 and 3.S4): MCF7 (Fig. 3.S3 and 3.S4; left) or HEK293T (Fig. 3.S3 and S4; right) cells were incubated with either N-TASQ ( $5\mu\text{M}$ , Fig. 3.S3) or AS1411-T ( $5\mu\text{M}$ , Fig. 3.S4) for 2h at  $37^\circ\text{C}$  prior to fixation with PFA (4% in PBS, 5min at room temperature) and then incubated with the other compounds, that is either AS1411-T ( $5\mu\text{M}$ , Fig. 3.S3) or N-TASQ ( $5\mu\text{M}$ , Fig 3.S4) for 15min at room temperature. Cells were then washed twice with PBS, mounted with Fluoromount-G, and images were collected through the “DAPI channel” (lex = 405nm, lem = 410-580nm) for N-TASQ labeling, and the “TAMRA channel” (lex = 552nm, lem = 559-705nm) for AS1411-T labeling, along with differential interference contrast.



**Figure 3.S3.** Indirection competition with N-TASQ in MCF7 (left) or HEK293T (right)

Images seen in Fig. 3.S3 enlightens that, as expected, N-TASQ live incubation triggers the appearance of cytoplasmic granules (dashed arrows) while post-fixation labeling with AS1411-T leads to exclusive nuclear staining (as above, Fig. 3.S2). This stepwise protocol makes colabeling foci unlikely since it leads to a distinct compartmentalization of the two probes. The opposite protocol in which cells were live-incubated with AS1411-T and post labelled with NTASQ leads to a completely different labeling pattern. As seen in Fig. 3.S4, live cell incubation with AS1411-T provides an intense membrane staining (along with a diffuse cytoplasm labeling), which strongly differs from the pattern obtained after post-fixation labeling (mostly located in the nucleus, *vide supra*). This might originate in a complete redistribution of the AS1411-T targets upon live incubation, although cytotoxicity issues cannot be firmly ruled out (even if incubation time is shortened to 2h here). This possible new target relocalization also strongly impacts N-TASQ staining since a cytoplasm labeling is observed as above along with an intense membrane staining (Fig. 3.S4); as seen in merged images (right panels), numerous co-staining foci (white arrows) are observed under this experimental setup, indicating again that both molecular tools AS1411-T and N-TASQ share common targets. Altogether, this series of images substantiates the hypothesis that AS1411-T and N-TASQ target –at least partially– the

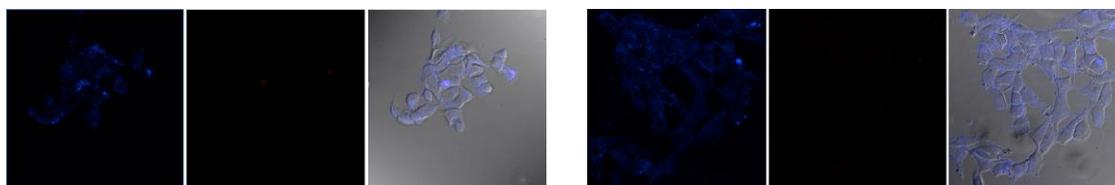
same cellular component (likely to be nucleolin), without identifying it in a definitive manner. Subsequent experiments (chiefly immunodetection studies) are needed to firmly assess the validity of this hypothesis.



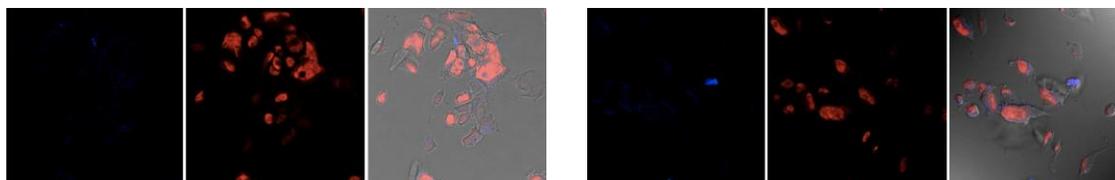
**Figure 3.S4.** Indirection competition with AS1411-T in MCF7 (left) or HEK293T (right)

#### Control experiments (Figures S5 and S6)

Two control experiments were performed, with a single post-labeling agent, either N-TASQ (Fig. 3.S5) or AS1411-T (Fig. 3.S6). MCF7 (Fig. 3.S5 and S6; left) or HEK293T (Fig. 3.S5 and 3.S6; right) cells were fixed with PFA (4% in PBS) for 5min at room temperature and then incubated with either N-TASQ (5 $\mu$ M, Fig. 3.S5) or AS1411-T (5 $\mu$ M, Fig. 3.S6) for 15min at room temperature. Cells were then washed twice with PBS, mounted with Fluoromount-G, and images were collected through the “DAPI channel” (lex = 405nm, lem = 410-580nm) and the “TAMRA channel” (lex = 552nm, lem = 559-705nm), along with differential interference contrast. These control images confirm that N-TASQ signal is seen *via* the “DAPI channel” only (Fig. 3.S5) and that of ASA1411-T *via* the “TAMRA channel” only (Fig. 3.S6), thus supporting the suitability of our experimental setup for co-staining investigations.



**Figure 3.S5.** Control experiment for MCF7 (left) or HEK293T (right) post-labelled with N-TASQ alone



**Figure 3.S6.** Control experiment for MCF7 (left) or HEK293T (right) post-labelled with AS1411-T alone

## **Chapter 4: Transcriptome-wide identification of transient RNA G-quadruplexes in human cells (*Nature Communications*, 2018)**

### **4.1 Summary**

Guanine-rich RNA sequences can fold into four-stranded structures, termed G-quadruplexes (G4-RNAs), whose biological roles are poorly understood, and *in vivo* existence is debated. To profile biologically relevant G4-RNA in the human transcriptome, we report here on G4RP-seq, which combines G4-RNA-specific precipitation (G4RP) with sequencing. This protocol comprises a chemical crosslinking step, followed by affinity capture with the G4-specific small-molecule ligand/probe BioTASQ, and target identification by sequencing, allowing for capturing global snapshots of transiently folded G4-RNAs. We detect widespread G4-RNA targets within the transcriptome, indicative of transient G4 formation in living human cells. Using G4RP-seq, we also demonstrate that G4-stabilizing ligands (BRACO-19 and RHPS4) can change the G4 transcriptomic landscape, most notably in long non-coding RNAs. G4RP-seq thus provides a method for studying the G4-RNA landscape, as well as ways of considering the mechanisms underlying G4-RNA formation, and the activity of G4-stabilizing ligands.

### **4.2 Introduction**

The biological functions and cellular regulations of RNAs are dependent on their secondary and tertiary structures (233-234). RNAs can adopt intricate bulged, stem-loop structures involving duplex-, triplex-, and quadruplex-RNA motifs (235-236). G-quadruplexes (G4s) are structures formed by Hoogsteen bonding of four guanines to form planar guanine quartet (G-quartet) units, which  $\pi$ -stack on each other, to assemble into columnar four-stranded structures with the central

cavity stabilized by monovalent cations (i.e.,  $K^+$ ,  $Na^+$ ). G4 folding is spontaneous *in vitro* and results in a highly stable structure. While both single-stranded DNA and RNA can fold into G4s, the latter is less studied (5, 218), even though G4 formation in RNA molecules is generally more stable, and RNA molecules can fold more readily due to their predominant single-stranded nature *in vivo*.

Formation of G4-RNA has recently been implicated in key RNA metabolism events, including the regulation of RNA processing and translation (237-238). To better understand the roles of G4-RNA in cell biology, there has been a strong interest in mapping the distribution of G4-RNAs within the human transcriptome. However, to date, *in vitro* and *in vivo* evidence appear to be contradictory, limiting interpretation of the relevance of G4-RNAs (72, 90, 96, 108). *In vitro* transcription experiments suggest that opportunities exist for G4-RNA formation during co-transcriptional folding of nascent RNA (239). Kwok et al. (90) used reverse transcription (RT)-stalling coupled with next-generation sequencing to map thousands of G4-RNA sites *in vitro*, showing widespread potential G4-forming sites within the human transcriptome. In contrast, Guo and Bartel (108) showed through *in vivo* DMS-mediated RNA modification with RT-stall-sequencing that G4s are nearly entirely in an unfolded state *in vivo* in mammalian cells. This observation is surprising since it has widely been assumed that G4-RNAs are formed *in vivo*, at least transiently. Conversely, evidence for *in vivo* G4 formation has been provided by cellular imaging studies using G4-specific antibodies and probes (72, 96). To reconcile these results, we hypothesized that G4-RNAs must be able to form, at least transiently, in live human cells, and that the identity of these G4-RNAs may provide valuable insights into their regulatory

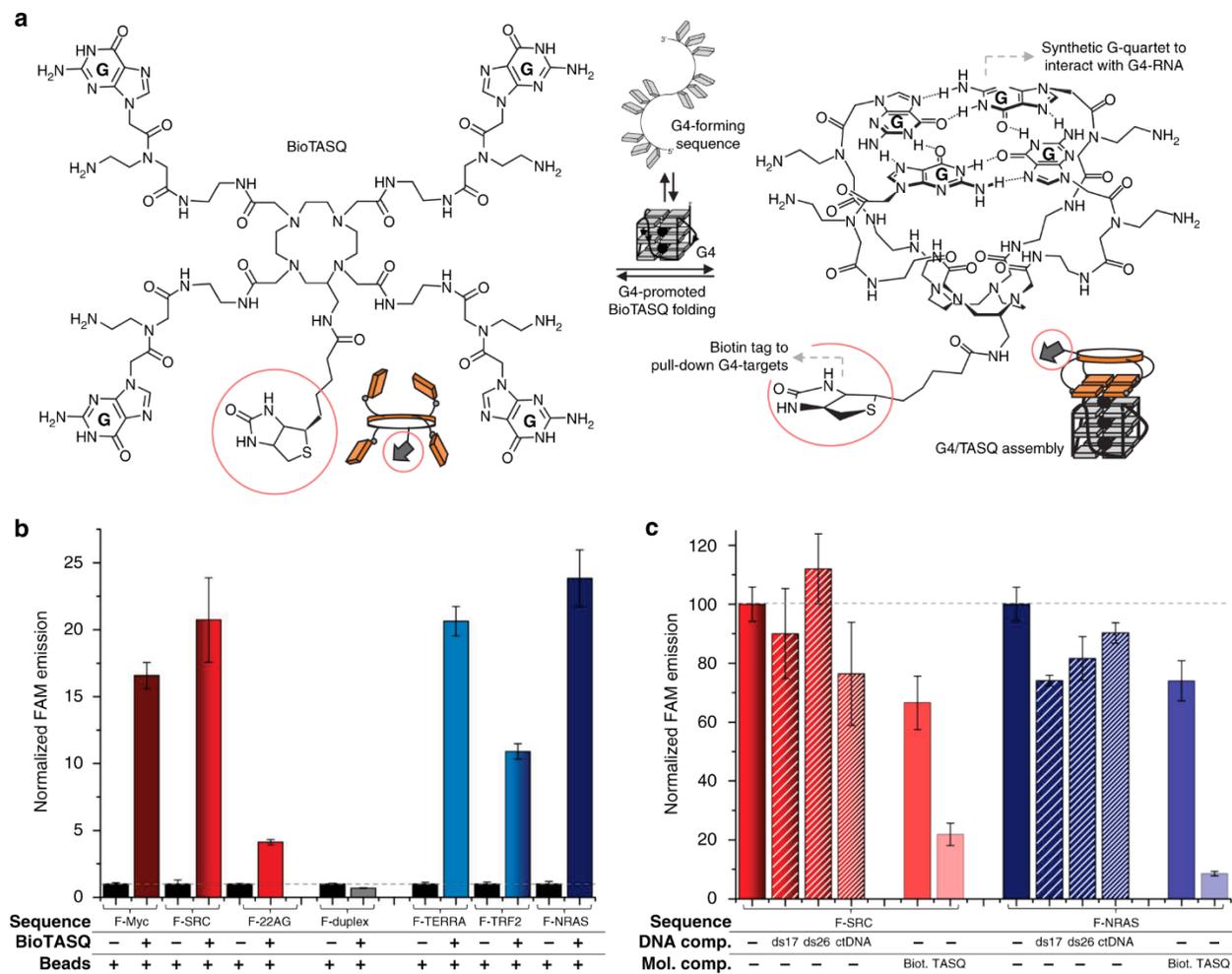
mechanisms and functions. However, because none of the above methods are suitable for capturing transient G4s, an alternative approach was needed.

Here, we report on a small-molecule-based approach to assessing the existence and the identity of transient G4-RNAs in the human transcriptome. We designed a biotinylated version of a previously characterized G4-specific ligand/probe, template-assembled synthetic G-quartet (TASQ), which self-assembles into a synthetic G-quartet upon association with a G4 target through end-quartet stacking. We developed a protocol, G4RP-seq (G4-RNA-specific precipitation and sequencing) using BioTASQ to capture G4-RNAs from human breast cancer cells in log-phase growth. Using this protocol to characterize *in vivo* transcriptomic landscapes, we showed that more G4s are present in gene transcripts that are GC-rich and have higher densities of predicted G4 motifs. We also evaluated G4 ligand-induced changes to the G4-RNA landscape following treatments with the G4 ligand, BRACO-19 or RHPS4, showing both similarities and differences in their respective induction profiles. Our data show that G4-RNAs can be ligand-induced in diverse RNA entities that include long non-coding RNAs; further, differential G4-RNAs induced by G4 ligands suggest that specific G4 structure–ligand interactions could be exploited.

## 4.3 Results

### 4.3.1 BioTASQ selectively captures G4 targets *in vitro*

To support affinity purification and identification of functional transcriptomic G4-RNA targets, we added a biotin tag to the biomimetic quadruplex ligands known as TASQ (Fig. 4.1A) (71, 213), known for their high G4-selectivity and, for some of them, their ability to track G4-RNAs in live cells (N-TASQ)(72, 210). Detailed synthesis and characterization of the biotinylated TASQ, or BioTASQ, can be found in Supplementary Figure 4.S1 and Methods. We first evaluated the G4-interacting properties of BioTASQ via a fluorescence resonance energy transfer (FRET)-melting assay (Supplementary Figure 4.S2) (84) and electrospray ionization mass spectrometry (ESI-MS) analyses (Supplementary Figure 4.S3) (240). FRET-melting experiments were performed with BioTASQ against a panel of dual-labeled nucleic acid sequences that included: (a) three G4-DNAs (F-Myc-T and F-kit-T, found in the promoter regions of *MYC* and *KIT* gene, respectively, and F21T, the human telomeric sequence);(63, 241) (b) one G4-RNA (F-TERRA-T, the human telomeric transcript); (242-243) and (c) one duplex-DNA as a control (F-DS-T) (Supplementary Figure 4.S2). Introduction of a biotin tag negatively impacted G4-affinity of BioTASQ (further confirmed by ESI-MS measurements) compared to the non-biotinylated parent <sup>PNA</sup>DOTASQ. However, it did not affect the ability of BioTASQ to discriminate G4s over duplexes (213, 244).



**Figure 4.1 Characterization of G4-specific affinity of BioTASQ.**

**A)** Structure of BioTASQ displaying a biotin affinity tag (red circles), and schematic representation of its open (left) and closed, quadruplex-associated conformation (right), in which the intramolecular G-quartet is formed. Schematic representation of a guanine-rich RNA sequence (guanines as gray squares) in its unfolded, random-coil and folded G4 structure. **B)** Fluorescence analysis of pull-down experiments carried out with (1) FAM-labeled oligonucleotides (1  $\mu$ M): either G4-DNA (F-MYC, F-SRC, and F-22AG), duplex-DNA (F-duplex) or G4-RNA (F-TERRA, F-TRF2, and F-RAS); (2) BioTASQ (20  $\mu$ M); and (3) streptavidin-coated magnetic beads. **C)** Competitive pull-down experiments performed with F-SRC and F-NRAS (1  $\mu$ M), BioTASQ (20  $\mu$ M) in the absence or presence of duplex-DNA competitors (ds17 or ds26, 20  $\mu$ M) or DNA extracts (calf thymus DNA, ctDNA, 100  $\mu$ M), or of molecular competitors (biotin (or Biot.), 80  $\mu$ M or <sup>PNA</sup>DOTASQ (or TASQ), 10  $\mu$ M). All experiments were done in triplicates. Error bars represent SD.

Next, we assessed whether BioTASQ could efficiently capture G4s from solutions *in vitro* (24). We used fluorescein (F)-labeled nucleic acids since fluorescence signal measurements allow for convenient and sensitive detection of the ligand/G4 assemblies. We performed these experiments with: (a) three G4-DNAs (F-Myc and F-SRC, two sequences found in the promoter regions of *MYC* and *SRC* gene, respectively (29, 98), and F-22AG, the human telomeric sequence); (b) three G4-RNAs (F-TRF2 and F-NRAS, two sequences found in the mRNA of *TRF2* and *NRAS* gene, respectively (246) and F-TERRA (the human telomeric transcripts); and (c) one duplex-DNA as control (F-Duplex). Labeled oligonucleotides (1  $\mu$ M) were incubated with BioTASQ (20  $\mu$ M) and streptavidin-coated magnetic beads (MagneSphere®, 25  $\mu$ L). After overnight incubation at 25 °C, streptavidin beads were precipitated, the supernatant removed, and the beads resuspended in denaturing buffer before thermal denaturation (8 min at 90 °C). After separation from the beads (via centrifugation and magnet immobilization), fluorescent signals from the supernatant solutions were measured to quantify nucleic acid capture. Our results confirmed the efficiency of BioTASQ-mediated G4 capture (Fig. 4.1B) and showed that the level of G4s recovered was dependent on both the G4 nucleic acid type and topology. G4-DNA was enriched 4.1–20.7-fold, whereas G4-RNA was enriched between 10.9- and 23.8-fold when compared to controls, while duplex-DNA was not enriched. BioTASQ appeared to have stronger preferences for certain G4 topologies, as type I (or ‘parallel’) (247) G4s displayed better enrichment than type II (or “mixed-hybrid”) G4s (with 16.6-, 20.7-, 20.6-, 10.9-, and 23.8-fold enhancement for F-Myc, F-SRC, F-TERRA, F-TRF2, and F-NRAS, respectively, versus 4.1-fold enhancement for F-22AG). The preference of BioTASQ for type I G4, which displays accessible external G-quartets, is expected, given that TASQs are sterically demanding ligands that require accessible, loopless G-quartets for binding G4 targets efficiently. This property represents a

limitation to the use of BioTASQ for G4 detection, especially for G4-DNAs, which have higher conformational diversity than G4-RNAs. We have not yet tested the affinity of BioTASQ on the recently reported antiparallel G4-RNA (51). While the cellular prevalence of G4-RNA with antiparallel topology remains to be established, they may provide key insights into the topological preference of TASQ ligands. We confirmed that the streptavidin bead/BioTASQ system did not extract duplex-DNA (0.7-fold). We further confirmed G4 selectivity of BioTASQ via competitive pull-down experiments, which we performed with F-SRC and F-NRAS (1  $\mu$ M) in the presence of an excess of duplex-DNA (either ds17 or ds26, 20  $\mu$ M) or DNA extracts (calf thymus DNA, ctDNA, 100  $\mu$ M, expressed in base pairs). The capture efficiencies of the fluorescently labeled G4-RNA were not significantly affected by an excess of synthetic duplexes (90–112% with F-SRC, 74–82% with F-NRAS) or with DNA extracts (76% and 90% with F-SRC and F-NRAS, respectively). We also tested BioTASQ/streptavidin association with an excess of either biotin (80  $\mu$ M, to compete with streptavidin interaction) or <sup>PNA</sup>DOTASQ (10  $\mu$ M, to compete with G4 interaction) (Fig. 4.1C). Together, these results show the strong ternary interaction between G4s, BioTASQ, and streptavidin (beads), which provided the basis for the development of our G4RP protocol (described below).

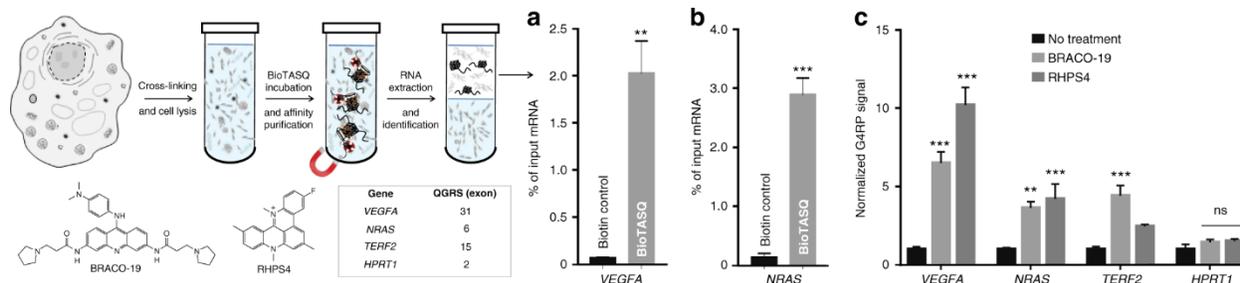
#### **4.3.2 G4RP isolates G4-RNA targets from human cell extracts**

After confirming that BioTASQ could interact with and capture G4s *in vitro*, we then assessed whether it could capture G4 targets from human cell extracts. For this, we developed the G4-RNA-specific precipitation (G4RP) protocol, a modified version of the commonly used RNA-immunoprecipitation (RIP) protocol (248). MCF7 cells were first crosslinked with formaldehyde to halt biological processes and stabilize transient structural interactions. Harvested cells were then sonicated briefly to release cellular content. Cell lysates were incubated with a high

concentration of BioTASQ (100  $\mu$ M) overnight (Supplementary Figure 4.S3C) before affinity purification with magnetic streptavidin beads.

We first used RT quantitative PCR (RT-qPCR) with gene-specific primers to confirm the efficiency of the G4RP protocol (249). G4RP-qPCR analysis of RNAs extracted from the BioTASQ-enriched fractions showed that non-specific binding was negligible (black bars) (Fig. 4.2A-B) while demonstrating the enrichment of two known G4-forming mRNAs, i.e., *VEGFA* and *NRAS* (gray bars) (Fig. 4.2A-B). To confirm bona fide G4 formation in target mRNA sequences, we collected MCF7 cell extracts following treatments with two well-established G4-stabilizing ligands, BRACO-19 and RHPS4 (61, 250). We chose the treatment ligand doses to be between the respective IC<sub>15</sub> and IC<sub>25</sub> (doses that are growth inhibitory in 15–25% of the cells; BRACO-19: 5  $\mu$ g/mL and RHPS4: 1.5  $\mu$ M), as determined by growth kinetics profiling of MCF7 in the presence of G4 ligands, compared with vehicle controls (Supplementary Figure 4.S4). Treatment with G4 ligands significantly increased G4RP-qPCR signals by 6.5-fold, 3.6-fold, and 4.4-fold in BRACO-19-treated cells and by 10.2-fold, 4.2-fold, and 2.5-fold in RHPS4-treated samples in three selected G4-rich regions, *VEGFA*, *NRAS*, and *TERF2*, respectively, compared to the untreated control (Fig. 4.2C; Supplementary Figure 4.S5), demonstrating that BioTASQ can specifically enrich for G4-containing RNA sequences. *HPRT1* was selected as an unstructured RNA control, as this housekeeping mRNA is expected to have low G4-forming potential (251). Neither BRACO-19 nor RHPS4 treatment induced a significant change in BioTASQ-captured *HPRT1* signals, confirming that these G4 ligands were selective for G4-rich targets. Collectively, these results indicate that the G4RP protocol is suitable for the purification and identification of G4-

containing RNAs from human cell extracts, as well as the quantification of the G4 ligand-induced changes.



**Figure 4.2 Isolation of G4 targets from human cell extracts using G4RP.**

(left) Schematic representation of G4RP protocol. G4RP signals of biotin control versus BioTASQ through RT-qPCR quantification of **A) VEGFA** and **B) NRAS** mRNA levels in untreated MCF7 cells. **C)** Changes induced by BRACO-19 and RHPS4 measured by the fold change of G4RP signal in *VEGFA*, *NRAS*, *TERF2*, and *HPRT1* mRNA. Values are normalized to their individual untreated controls. Three biological replicates were used for the quantifications. Student's *t*-test and two-way ANOVA were performed. *p*-Values: \**p* < 0.05, \*\* *p* < 0.01, and \*\*\**p* < 0.001. Error bars represent SEM

We performed G4RP in samples that were crosslinked to preserve the transiently formed cellular G4s while preventing the induction of G4 formation *in vitro*, as crosslinked nucleic acids would be in an immobilized state. We used the reported formaldehyde concentration and crosslinking conditions where over 90% of nuclear DNA are immobilized (252) and anticipated that the crosslink efficiency for cellular RNA to be similar. As controls and to illustrate the importance of this crosslinking step, G4RP was performed in non-crosslinked samples. We selected three targets from the top- and bottom-ranked transcripts obtained with our G4RP-seq results (see below). BioTASQ enrichment of these targets was quantified using RT-qPCR and compared between the non-crosslinked and crosslinked samples (Supplementary Figure 4.S6). We observed a loss of difference in BioTASQ enrichment in the non-crosslinked samples, despite the overall higher signals. These increased signals in non-crosslinked samples were likely G4s that

were formed *in vitro*, due to the high concentration of BioTASQ, arguing that the crosslinking step is necessary to immobilize cellular RNA structures and to minimize the effects of *in vitro* G4 formation and destabilization through the biochemical evaluation steps. As naked, non-crosslinked RNA targets are susceptible to the *in vitro* G4 stabilization effects of high concentration of BioTASQ during the incubation steps, the true differences between the *in vivo* transient levels of the top and bottom ranking targets would be masked.

### **4.3.3 G4RP-Seq identifies transcriptome-wide transient G4-RNAs**

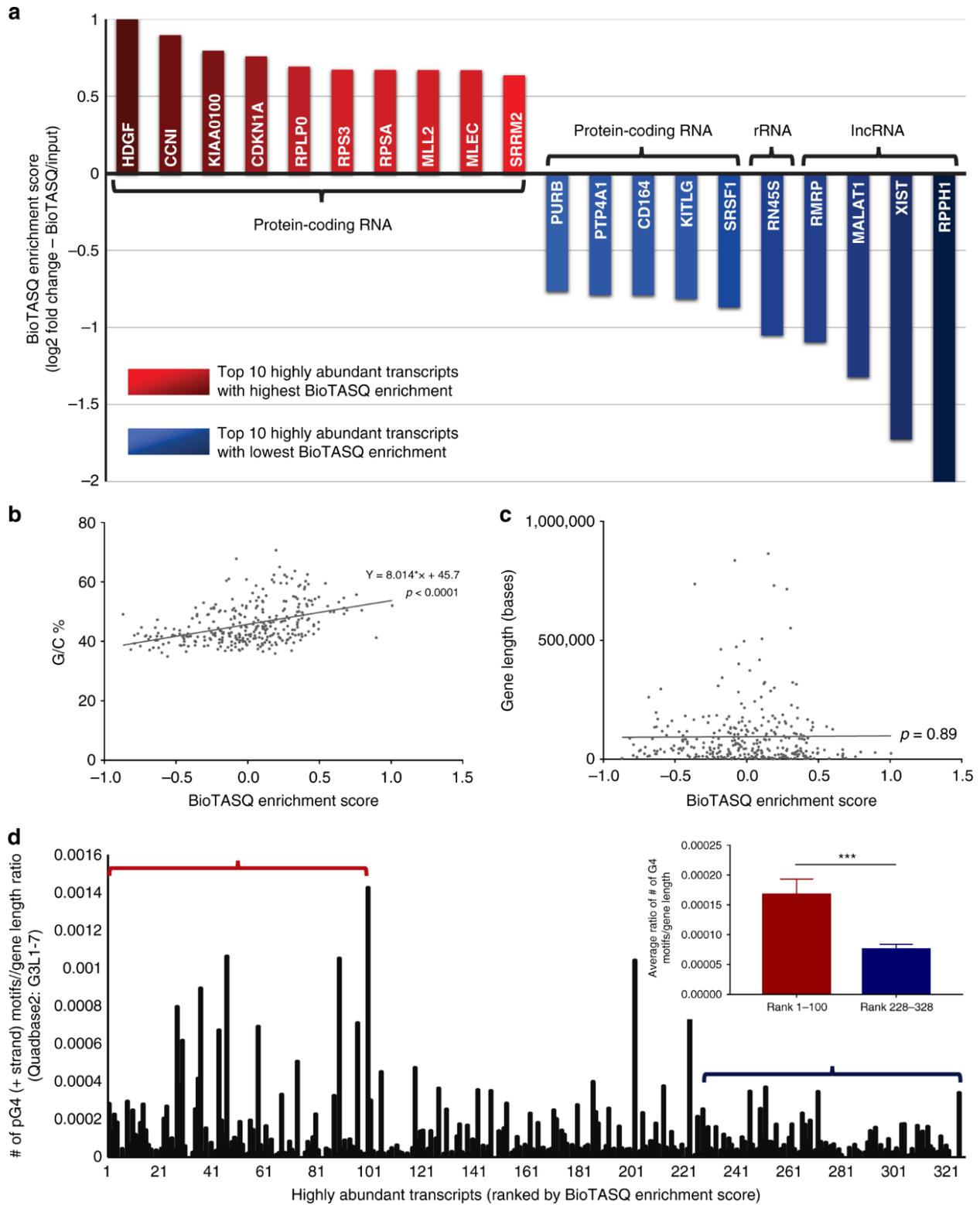
To survey the baseline *in vivo* G4 transcriptomic landscape, we performed G4RP followed by sequencing (G4RP-seq) in human breast cancer cells harvested at log-phase growth. Due to the crosslinking step in the G4RP protocol, we expected to capture global levels of transient G4s in the transcriptome. Notably, as G4 ligand treatment resulted in gene expression changes in human cells (253), we needed to account for these input differences; therefore, an internal input control was included for each treatment condition. To enrich for diverse G4-forming RNAs, we elected to remove ribosomal RNA targets at the cDNA library preparation step, as rRNA constitute over 80% of total cellular RNAs. Ribodepletion instead of poly-A selection was used, as we anticipated that non-poly-A RNAs, including many non-coding RNAs (ncRNAs), not only harbor but contribute to a substantial proportion of cellular G4s (254). For the sequencing analysis, two comparisons were required: BioTASQ versus input (which gives normalized global levels of transient G4s for a specific transcript as Enrichment Score (ES)), and G4 ligand-treated versus untreated cells (given as Enrichment Score Change ( $\Delta$ ES)) (Supplementary Figure 4.S7). After mapping reads to the hg19 reference human genome assembly with HISAT2 (255), normalization and differential gene expression analyses were performed with HTSeq/DESeq2 pipeline (256-257), by comparing the BioTASQ-enriched samples with the corresponding inputs

(258) (Supplementary Data 1). Of note, the relative enrichment levels by BioTASQ (given as ES) were not direct quantitative readouts of G4 formation but instead indicate the relative propensity of a specific transcript to fold into G4. ES values allow ranking of transcripts by relative G4-folding status under specific experimental conditions.

We observed BioTASQ enrichment of many gene transcripts, suggesting the existence of widespread G4-RNAs, at least transiently, in live human cells. This observation was expected since we were capturing a snapshot of the G4-RNA landscape in which some G4-rich sequences were in folded states, while others were in unfolded states. While the sequencing depth was not high enough to detect subtle changes in individual G4-forming sequences, we were able to confidently determine gene-level changes by focusing on a subset of highly abundant transcripts, filtered by high normalized mean read counts (as an estimate of relative expression level). We compared ES for each condition to generate initial lists of gene transcripts filtered by a minimal abundance threshold ( $>50$  normalized mean read counts) (Supplementary Data 1). To ensure high-confidence hits with substantive changes, we further filtered the list and included only those with high transcript read abundance ( $\geq 500$  mean read counts; approximately the top 5% of the list), which we used for downstream analyses. Gene transcripts in this filtered list were then ES-ranked for downstream analyses (Supplementary Data 1D).

ES, a gene-specific ratio of the BioTASQ signal normalized to the corresponding input signal, moderately positively correlated with their respective G/C content (Pearson correlation = 0.43,  $p < 0.0001$ ) (Fig. 4.3B). ES was uncorrelated to gene length (Pearson correlation = 0.02,  $p = 0.89$ ) (Fig. 4.3c). To evaluate whether the ES is related to the density of

potential G4 sequences, Quadbase2 (259) was then used to assess the number of predicted G4 (pG4) motifs (i.e. sequences with the canonical G3L1–7 G4 sequence) in each gene transcript. The ratios of the number of pG4s to gene length were calculated and then plotted against the transcripts ranked by their ES (Fig. 4.3D). Comparison of the ratios between the top and bottom 100 ES-ranked transcripts showed significantly higher values (2.2-fold difference) for top-ranked transcripts. Results from gene ontology analysis of the top and bottom 100 ES-ranked transcripts are summarized in Supplementary Table 4.S1 (260). Together, our bioinformatics analyses confirmed that transcripts with higher ES tend to have higher G/C content and higher pG4 density. We also found that highly expressed lncRNAs had some of the lowest relative G4 levels, as shown by their low ES (Fig. 4.3A). These bottom-ranked transcripts included well-known lncRNAs such as *MALAT1*, *RPPH1*, *RMRP*, and *XIST*. The presence of *MALAT1* on this list was unexpected, as it has been previously reported as a G4-forming lncRNA *in vitro* (90, 108). Interestingly, residual rRNAs (that escaped from the ribodepletion step during library preparation) were also among the most depleted within this list.



**Figure 4.3** Characterization of the baseline level of G4-RNA landscape using G4RP-seq.

**A)** Top 10 highly abundant transcripts (filtered by at least 500 base-read counts) with the lowest BioTASQ enrichment (blue bars, normalized to the input in the untreated sample) or with the highest BioTASQ enrichment (red bars) ranked by Enrichment Scores. **B)** Regression plot of BioTASQ-Enrichment Score for each transcript versus its corresponding G/C content ( $R^2 = 0.187$ ,  $p < 0.001$ , significant non-zero relationship). **C)** Regression plot of BioTASQ-Enrichment Score for each transcript versus its corresponding gene length ( $R^2 = 0.00005$ ,  $p = 0.89$ , non-significant relationship). **D)** Number of pG4 motifs (calculated by Quadbase2 using mid stringency G3L1–7) to gene length ratio plotted against the subset of highly abundant transcripts ranked by their BioTASQ enrichment. The bar graph (inset) shows the average pG4 motif/gene length ratio between the top 100 ranked transcripts versus the bottom 100 ranked transcripts ( $p < 0.001$ , significant difference, Student's *t*-test)

#### 4.3.4 G4RP-seq identifies ligand-induced changes in G4-RNA profiles

We evaluated ligand-induced changes in the G4-RNA landscapes by applying G4RP-seq to samples treated with G4 ligands and calculating an Enrichment Score Change ( $\Delta ES$ ) from the ratio of treated versus non-treated samples. By filtering the initial list  $\Delta ES > 1.75$ , we found BioTASQ enrichment in 251 and 463 gene transcripts to be highly induced by BRACO-19 and RHPS4, respectively (Supplementary Data 1B–C). Results from gene ontology analysis performed on the lists of top genes are summarized in Supplementary Table 4.S2 (260).

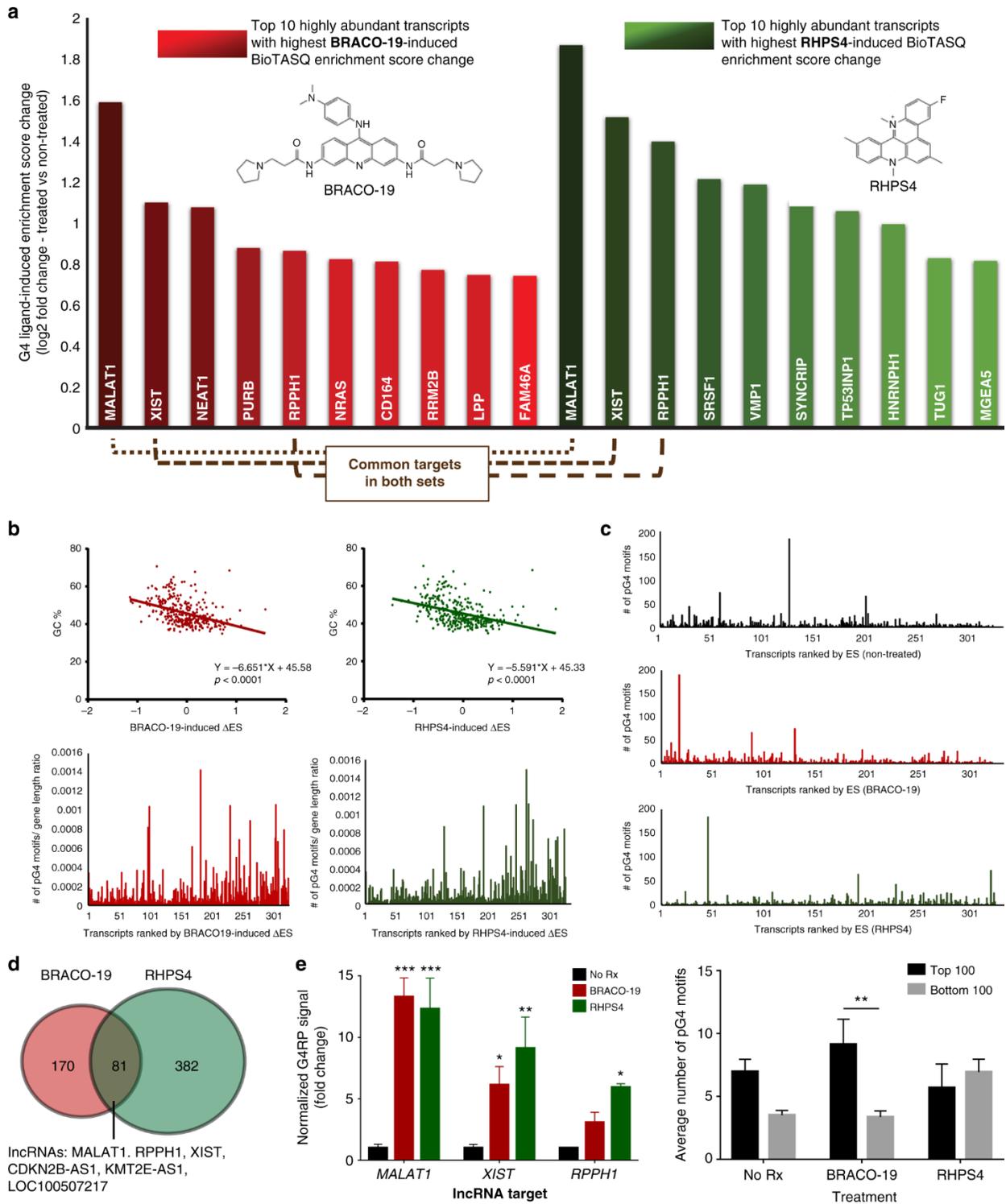
Among the list of gene transcripts with a high ligand-induced increase ( $\Delta ES > 1.75$ ) and at least 500 mean read counts, the lncRNAs *MALAT1*, *RPPH1*, and *XIST* were highly ranked in both BRACO-19- and RHPS4-induced gene lists (Fig. 4.4A). Some RNAs previously reported to harbor G4s *in vitro* were also identified, including *NEAT1* and *NRAS* (90, 246). We found no correlation between the ligand-induced  $\Delta ES$  and the read counts of the transcripts (Pearson correlation = 0.02,  $p = 0.69$ ) (Supplementary Figure 4.S8). The ligand-induced  $\Delta ES$  of each transcript was compared to the corresponding G/C content of the transcripts, which interestingly showed a negative correlation (Fig. 4.4B). When the transcripts were ordered by  $\Delta ES$ , the pG4 density appeared to be distributed toward lower scores (1.7-fold and 2.7-fold difference between

the average ratio of top and bottom 100 ranked transcripts for BRACO-19 and RHPS4, respectively) (bottom panel) (Fig. 4.4B). Overall, our observations suggest that transcripts with higher pG4 density were more likely to be captured in a folded state in the absence of ligands, resulting in  $\Delta$ ES being lower due to the higher baseline level of G4 formation. In contrast, transcripts with lower pG4 density were more likely to be unfolded in the absence of ligands and to have their G4 structures stabilized in the presence of ligands, leading to a higher  $\Delta$ ES.

When we compared the absolute number of pG4 motifs (i.e. without normalization to gene length) between the three treatment conditions ranked by their respective ES, we observed differential changes in pG4 profiles between the two G4 ligand treatments (Fig. 4.4C). pG4 scores generated using different stringency of searches (G2L1–10, G3L1–5, and G3L1–7) showed similar trends (Supplementary Figure 4.S9). BioTASQ-captured targets generated from BRACO-19-treated samples exhibited higher levels of pG4-dependent enrichment regardless of the search stringency, conceivably due to the broader range of intramolecular G4s (longer loops, 2-quartet G4s, etc.) stabilized by this ligand. On the other hand, targets generated from RHPS4-treated samples showed lower levels of pG4-dependent enrichment and a pG4-dependency could only be observed when the plots were obtained using G4 motif searches with the lower stringency. We reasoned that RHPS4-binding preference could be selective towards sequences with lower numbers of G4 motifs or highly specific G4 sequences that are less prevalent within the transcriptome. However, we cannot rule out the possibility of intermolecular G4s, as computational algorithms are currently unable to predict these structures. Given this caveat, the lack of pG4-dependent enrichment in samples treated with RHPS4 could be partially explained by a preferential ligand-induced stabilization of intermolecular G4s. The overlap between the

gene lists for the two G4 ligand treatments demonstrated that they have differential G4-induction profiles, in agreement with their differential *in vitro* G4-structure-specific binding profiles (Fig. 4.4D) (261).

We further validated our findings from G4RP-seq by performing qPCR on separate G4RP samples obtained with biological repeat experiments, using primers specific for the top three common lncRNA targets: *MALAT1* (262), *XIST* (263), and *RPPH1* (264). Fold change differences between G4RP-qPCR and G4RP-seq were seen and expected due to the small qPCR region amplified. Nevertheless, the qPCR results were consistent with those obtained from sequencing, confirming that these lncRNAs were targets of G4 ligands (Fig. 4.4E; Supplementary Figure 4.S10). Circular dichroism (CD) and thermal differential spectra (TDS) analyses (81) of the three selected pG4 regions of *MALAT1*, *XIST*, and *RPPH1* were consistent with the formation of parallel-type G4 structures *in vitro* in the selected pG4 motif sequences extracted from these genes (Supplementary Figure 4.S11). In summary, we observed that the BRACO-19 and RHPS4 treatments in MCF7 cells similarly induced G4 stabilization in several highly expressed lncRNA targets, but the treatments also displayed distinct ligand specificity towards other RNA targets.



**Figure 4.4 Characterization of the ligand-induced changes in the G4-RNA landscape.**  
**A)** Top 10 highly abundant transcripts with highest fold increase in BioTASQ enrichment (ranked by Enrichment Score Change) for BRACO-19 (red) and RHPS4 (green). Common targets that were ranked highly for both ligands are indicated in brown dashed lines.

**B)** (Top) Regression plot of BioTASQ-Enrichment Score Change for each transcript versus its corresponding G/C content (BRACO-19:  $R^2 = 0.16$ ,  $p < 0.0001$ ; RHPS4:  $R^2 = 0.16$ ,  $p < 0.0001$ , significant non-zero relationship). (Bottom) Number of G4 motifs (calculated by Quadbase2 using mid stringency G3L1–7) to gene length ratio plotted against the subset of highly abundant transcripts ranked by their BioTASQ-Enrichment Score Change. Left panel is BRACO-19-induced changes and right panel is RHPS4-induced changes. **C)** Absolute number of pG4 motifs of highly abundant gene transcripts ranked by BioTASQ-Enrichment Score for the three conditions: non-treated (black), BRACO-19 (red), and RHPS4 (green). (Bottom) Quantification of average number of pG4 motifs for top 100 and bottom 100 ranked genes for the three sets of conditions. **D)** Venn diagram comparing top filtered BioTASQ-enriched gene lists for BRACO-19 and RHPS4. **e** G4RP-qPCR controls of the top lncRNA hits MALAT1, RPPH1, and XIST, in G4 ligand-treated (BRACO-19 or RHPS4) samples normalized to corresponding untreated controls. Three biological replicates were used for quantification. Two-way ANOVA was performed.  $p$ -Values:  $*p < 0.05$ ,  $**p < 0.01$ , and  $***p < 0.001$ . Error bars represent SEM

#### 4.4 Discussion

It has been widely assumed that G4s must be formed *in vivo*, at least transiently, due to the high structural stability of G4 nucleic acids and the favorable intracellular potassium concentration.

The extent to which G4-RNA formation occurs *in vivo* in human cells is debated (265), since it has been recently reported that G4s are globally unfolded in the mammalian transcriptome (108).

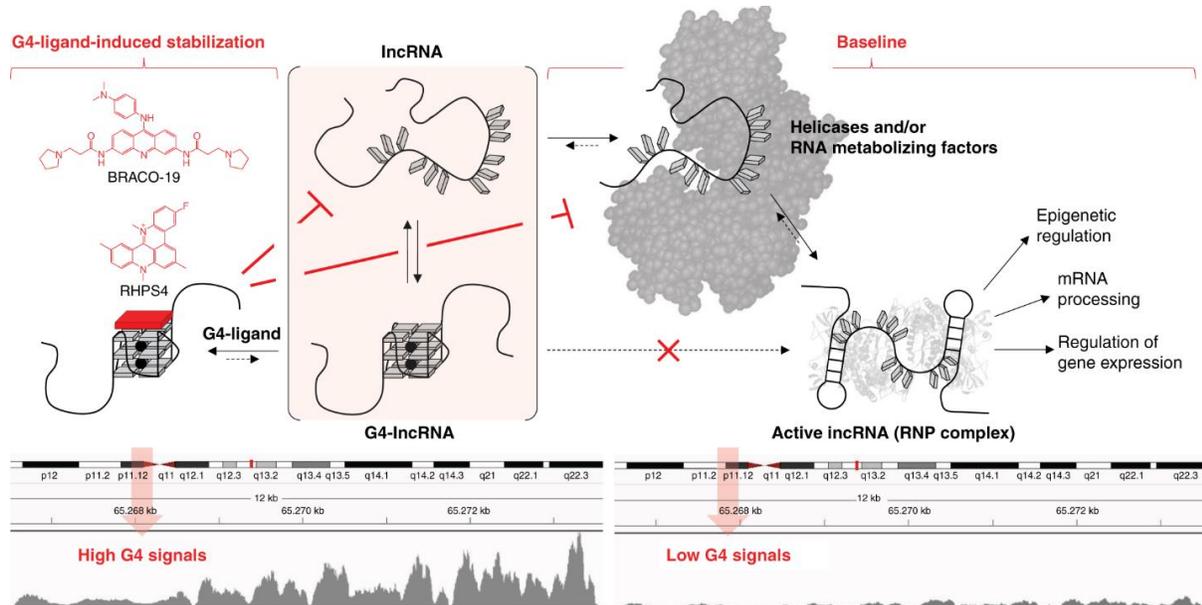
To address this conundrum, we developed and reported here on G4RP-seq, a protocol that provides evidence supporting the existence of transient G4-RNAs in the *in vivo* human transcriptome. We also used G4RP-seq to generate readouts of the changes in transcriptome-wide G4-RNA landscape upon treatments with G4 ligands.

Our first key finding of global transient G4-RNA formation suggests an alternative perspective on Guo and Bartel's (108) lack of detected G4-RNAs by DMS/RT-profiling, an observation that can be interpreted in two ways: either G4s are seldomly formed *in vivo*, or G4s are formed but are unfolded quickly by destabilizing mechanisms such as through the actions of helicases.

DMS/RT profiling may be reporting on an unfolded G4 landscape since it measures the biological endpoint in which all G4s are eventually unfolded. A recent study corroborated this by

demonstrating that dynamic folding and unfolding of some G4-RNAs in live cells could occur within seconds (103). We postulated that even if the G4 equilibrium heavily favors an unfolded G4 state, global snapshots of transiently folded G4s could still be captured by a chemical crosslinking step. Our observation of widespread G4-RNAs using a snapshot approach, combined with Bartel's observation of endpoint globally unfolded G4-RNAs, suggest that *in vivo* G4-RNAs can form continuously and are rapidly resolved to an unfolded state. Further, we anticipate that the equilibrium between an RNA's transient folded and unfolded states can be influenced by competing factors such as the transcription rate of the RNA, the availability of helicases (266), the chaperone activities of RNA-binding proteins (267), and the structural stabilization by G4 ligands (268) (Fig. 4.5). Importantly, the effectiveness of capturing these transient G4-RNAs within transcripts, as measured by BioTASQ enrichment, is correlated with their G/C content and pG4 density. This observation suggests that gene transcripts with a higher density of pG4 motifs would have a higher probability of having at least one of these motifs captured in a folded state with our crosslinking snapshot approach. It is currently undetermined whether such transient G4s have any biological roles or are simply stochastic passenger events that arise from the chemical properties of guanines. Strikingly, lncRNAs appeared to avoid G4-formation and were found mainly in an unfolded state, in the absence of G4 ligands. The functions and interactions of lncRNAs with other molecules (DNA, RNA, and protein) are dependent on their folding into higher-order structures (269), and the formation of G4 may interfere with such folding. Consistent with this, a study on hnRNP F binding and G4 formation showed that hnRNP F ribonucleoprotein formation prevents G4 formation/reformation by sequestering the G-tracts in a single-stranded RNA state (155). As the rate of protein-RNA

complex formation is faster than that for G4 formation (154), the shift of cellular G4-RNAs toward an unfolded state could be partially due to interactions with RNA-binding proteins.



**Figure 4.5 Proposed model of G4 structural equilibrium of lncRNAs in the presence of G4-stabilizing and G4-destabilizing factors.**

Guanine-rich RNAs in lncRNAs exist in equilibrium between single-strand and folded G4 state (center panel). At steady state (normal cell biology or untreated state), the nascent lncRNAs avoid G4-formation by the actions of G4-destabilizing factors such as helicases and RNA metabolizing factors (i.e., RNA-binding proteins) (right panel). G4-destabilizing factors may contribute to proper folding (including duplex structures) into the active form of lncRNA to mediate biological functions such as epigenetic regulation, mRNA processing, and gene expression regulation. The lack of G4 formation at steady state is shown by a lower relative level of G4RP-seq signal. However, in the presence of G4-stabilizing ligand (such as BRACO-19 and RHPS4), the nascent G-rich lncRNAs can be trapped in a folded G4 state due to G4-stabilizing ligands outcompetes the G4-destabilizing factors (left panel). This increase in folded G4 state is shown by an increase in the relative level of G4RP-seq signal

The second key finding from our report is the characterization of G4 ligand treatment effects on the G4-RNA folding state within the *in vivo* human transcriptome. While previous studies have characterized G4 ligand-induced changes for human genome-wide G4-DNAs and transcriptome-wide G4-RNAs *in vitro* (89-90), the effects of G4 ligands on transcriptome-wide G4-RNAs *in*

*in vivo* have not been reported. Here, we showed by the comparison of BioTASQ-enrichment profiles of BRACO-19 and RHPS4 that these two G4 ligands have both shared and distinct targets. Following ligand treatments, we found, paradoxically, that transcripts with lower pG4 density were relatively more enriched than those with a higher number of G4 motifs. There are two possible explanations for this. First, the relative increase in BioTASQ enrichment is less significant in high-pG4 transcripts than that found in low-pG4 transcripts, which have, by definition, a lower probability of being crosslinked in a folded state in the absence of ligands. Upon ligand treatment, stabilization of these previously unfolded pG4s within the low-pG4 transcripts significantly increased their BioTASQ enrichment. A second possibility is that rather than the density of G4 motifs, ligand-induced G4 stabilization may be more dependent on the ligand-binding affinity of individual G4 sequences, taking into account other confounding factors such as topology, flanking or loop sequences, and the presence of modified nucleosides. This second possibility is supported by our observation that the absolute number of pG4s within a transcript seemed to be more predictive of its ligand-induced G4-structure formation for BRACO-19, a broad spectrum, pan-G4-specific ligand, but not for RHPS4, a ligand with higher structural and sequence specificity. Given that the G4-binding modes of BRACO-19 and RHPS4 are different, their pG4 profiles are expected to be different.

While BRACO-19 and RHPS4 G4-profiles are distinct in several aspects, they also share common targets within the *in vivo* MCF7 transcriptome, most notably in the abundant lncRNAs. We anticipate that these targets are of interest, as they contain G4 targets that can be accessed and stabilized by both tested G4 ligands, suggesting that these G4s may be druggable. We propose that lncRNAs can spontaneously form G4s (e.g. *MALATI*) (90, 108), but that G4-

formation is actively counteracted by the actions of helicases and/or other RNA-binding proteins (267-268). We anticipate that there may be a window of opportunity for ligand-induced G4-formation/stabilization between the initial creation of nascent RNA and the RNA–RNA/RNA–protein interactions that are part of the co-transcriptional maturation stage of the ribonucleoprotein complex (239) (Fig. 4.5). It is important to consider the folding/unfolding kinetics, since the balance may be shifted as ligand-mediated G4 stabilization outcompetes the G4-destabilizing factors (i.e. duplex structures (270), helicases (266), and RNA-binding proteins (267)). As dysregulation of lncRNAs has been implicated in various human diseases, including cancers, cardiovascular, and neurodegenerative diseases, we anticipate that targeting G4s within lncRNAs may present a valuable therapeutic strategy to alter the functions of these RNA entities (271).

The G4RP-seq protocol has certain limitations and, we anticipate, can be further improved. One concern is that BioTASQ is itself a G4 ligand and thus may alter the G4 landscape. However, as demonstrated, the chemical crosslinking step before BioTASQ binding serves to minimize the effects of BioTASQ-induced stabilization (Supplementary Figure 4.S6), we contend that the G4RP protocol should provide a relatively unbiased readout of cellular G4-RNA. On the other hand, as we have stated earlier, BioTASQ preference toward parallel G4s could limit its capability in capturing rarer forms of G4-RNAs (i.e. antiparallel G4s). Additionally, the G4RP protocol uses formaldehyde as a crosslinking agent, and this is known to capture both direct and indirect RNA–RNA interactions (272); it will be informative in future work to also include the use of other crosslinking agents to better characterize the G4-RNA interactome. Additionally, the sequencing depth in our work was insufficient to differentiate small changes in BioTASQ

enrichment in low abundance transcripts and individual G4-forming sequences, and we anticipate that high-resolution sequencing in future studies will be necessary for a more complete mapping of potential transcriptomic G4 sites. Despite these limitations, we have shown here that G4RP-seq is useful for the broad identification of transient G4 structures and offers a snapshot view of the G4 landscape in live human cells. While using G4RP-seq alone cannot distinguish which competing factors (i.e. RNA-binding proteins or competing secondary structures) played more important roles at specific sites, data from G4RP-seq could be studied in combination with other functional genomic strategies (i.e. G4-ChIP-seq (105), RIP-seq (248), rG4-seq (90), DMS-seq (108), LIGR-seq (273)) to better characterize the interactions between G4 genome and transcriptome (collectively the G4ome).

In summary, our work provides a proof-of-principle for studying the mammalian G4-RNA landscape, and a method for studying the dynamics of *in vivo* transient G4-RNA under various biological conditions. Importantly, through G4RP-seq, we have also evaluated the mechanisms underlying the biological activity of G4 ligands. This opens exciting opportunities in which G4RP-seq, by providing transcriptome-wide views of G4 level changes, could be used as a quantitative method in the iterative chemical design of new G4 ligands.

## 4.5 Methods

### FRET-melting experiments

The sequences of the oligonucleotides used herein are listed in Supplementary Table 4.S3. The preparation of these sequences is described in the Supplementary Methods. Experiments were performed in a 96-well format using a Mx3005P qPCR machine (Agilent) equipped with FAM filters ( $\lambda_{\text{ex}} = 492 \text{ nm}$ ;  $\lambda_{\text{em}} = 516 \text{ nm}$ ) in 100  $\mu\text{L}$  (final volume) of 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (F21T, F-DS-T) or plus 1 mM KCl/99 mM LiCl (F-Myc-T, F-Kit-T, F-TERRA-T) with 0.2  $\mu\text{M}$  of labeled oligonucleotide and 0, 1, 5, or 10  $\mu\text{M}$  of BioTASQ. Competitive experiments were carried out with labeled oligonucleotide (0.2  $\mu\text{M}$ ), 10  $\mu\text{M}$  BioTASQ, and increasing amounts (0, 15 and 50 equiv.) of the unlabeled competitor ds26. After a first equilibration step (25  $^{\circ}\text{C}$ , 30 s), a stepwise increase of 1  $^{\circ}\text{C}$  every 30 s for 65 cycles to reach 90  $^{\circ}\text{C}$  was performed, and measurements were made after each cycle. Final data were analyzed with Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.). The emission of FAM was normalized (0 to 1), and  $T_{1/2}$  was defined as the temperature for which the normalized emission is 0.5;  $\Delta T_{1/2}$  values are means of triplicates.

### *In vitro* pull-down assay

The sequences of the oligonucleotides used herein are listed in Supplementary Table 4.S3. The preparation of these sequences is described in the Supplementary Methods. The *in vitro* quadruplex capture experiments were performed in 250  $\mu\text{L}$  final volume as follows: first, the streptavidin MagneSphere® beads (Promega) were washed three times with 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (Caco.K) buffer. To this end, 200  $\mu\text{L}$  of the commercial solution of beads were centrifuged, taken up in 200  $\mu\text{L}$  of Caco.K, and this washing step is repeated three times. After the original solutions being reconstituted (as 250- $\mu\text{L}$

mixtures in Caco.K), 25  $\mu$ L of MagneSphere<sup>®</sup> beads were added to solutions comprising BioTASQ (20  $\mu$ M) and FAM-labeled oligonucleotides (1  $\mu$ M). The mixtures were stirred overnight at 25 °C; then, the mixtures were subsequently centrifuged (60 s at 8900 r.p.m.), the beads are immobilized (attracted by a magnet) and the supernatant was removed. The solid residue was resuspended in 240  $\mu$ L of TBS 1 $\times$  buffer, heated for 8 min at 90 °C (under gentle stirring 800 r.p.m.), and then centrifuged for 2 min (8900 r.p.m.). The supernatant was taken up for analysis (the beads being immobilized by a magnet), after being splitted in three wells (80  $\mu$ L each) of a 96-well plate using a ClarioStar machine (BMG Labtech) equipped with FAM filters ( $\lambda_{\text{ex}} = 492$  nm;  $\lambda_{\text{em}} = 516$  nm). Data were analyzed with Excel and OriginPro<sup>®</sup>9.1. FAM emission was normalized as follows: the FAM emission of the three control wells (without BioTASQ) was collected and normalized to 1; then, the FAM emission of the three wells comprising labeled oligonucleotides, BioTASQ, and beads were collected and compared to the control experiments. This allowed for a direct quantification of the BioTASQ capture efficiency. Competitive experiments were performed with BioTASQ (20  $\mu$ M), labeled oligonucleotides (1  $\mu$ M), and MagneSphere<sup>®</sup> beads (25  $\mu$ L) along with the either nucleic acid competitors, ds17 (20  $\mu$ M), ds26 (20  $\mu$ M) or calf thymus DNA (CT-DNA, 100  $\mu$ M in base pairs), or small-molecule competitors, biotin (80  $\mu$ M) or <sup>PNA</sup>DOTASQ (10  $\mu$ M). All experiments were performed in triplicates.

### Cell line and culture

Human breast cancer cells MCF7 was obtained from American Type Culture Collection (ATCC). The cells were culture in Dulbecco's modified Eagle's medium (Gibco) supplemented with 5% synthetic fetal bovine serum (FetalClone III; GE Life Sciences) and 100 U penicillin–streptomycin mixture (Gibco). Cells were incubated at 37 °C in a humidified, 5%

CO<sub>2</sub> atmosphere-controlled incubator (HERAcell). Standard cell-culturing procedures were employed including aspiration and washing with phosphate-buffered saline (PBS, Gibco). Cells were trypsinized using Trypsin-EDTA (0.25%) (Gibco). Cell counting was performed using a Coulter Counter (Beckman Coulter).

#### Dose–response profiling of G4 ligands

MCF7 cells were seeded at 3000/well in a 96-well flat bottom plate. The cells were treated with a series of BRACO-19 or RHPS4 concentrations, made from serial dilutions. The cells were then monitored in the Essen Bioscience IncuCyte ZOOM live-cell monitoring system (274). Phase confluency was used to measure cellular proliferation under a range of doses. The maximum and minimum achievable confluency values under these conditions were used for normalization.

Three biological replicates were used to produce the dose–response curve. Doses between LD<sub>15</sub> to LD<sub>25</sub> were calculated from dose–response curve profiles (Supplementary Figure 4.S4) and used for subsequent experiments.

#### G4 RNA-specific precipitation (G4RP)

MCF7 cells were seeded at  $3.5 \times 10^5$  cells per 10-cm dish before treatment with either vehicle (PBS), BRACO 5 µg/mL (LD<sub>15</sub>), or RHPS4 1.5 µM (LD<sub>25</sub>) for 72h. Cells were then crosslinked using 1% formaldehyde/PBS for 5 min at 25 °C and the crosslinking was then quenched with 0.125 M glycine for 5 min. Cells were scraped and resuspended in G4RP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, RNase inhibitor (Roche), homebrew protease inhibitor cocktail). Cells were then sonicated using a Covaris m220 Ultrasonicator using default settings at 10% duty for 2 min. The sonicated fractions were then incubated with 100 µM BioTASQ (or 100 µM biotin for negative controls) overnight at 4 °C.

Five percent of the sonicate was collected as input control. Ten micrograms of streptavidin-magnetic beads (Promega) was added and the extract was incubated for 2 h at 4 °C. Magnetic beads were then washed four times in G4RP buffer for 5 min. The beads were then incubated at 70 °C for 1 h to reverse crosslink. TRIZOL was then used to extract the RNA from the beads using the manufacturer's instructions.

### RT-qPCR

The primer sets used for RT-qPCR are listed in Supplementary Table 4.S4. Extracted RNA was reverse transcribed with Superscript III (Thermos) and random hexamer primers using the manufacturer's standard protocol to generate cDNA. cDNAs were quantified using 2× SYBR green mix (Bimake) with three technical replicates. *C(t)* values of pull-down samples were normalized to the input control. Three biological replicates were used for all qPCR-based quantifications. Exon-spanning primers for quantifying mRNA levels were derived from Primerbank (249).

### G4RP sequencing (G4RP-seq) and analysis

G4RP samples were DNAase-treated, briefly thermally fragmented and ribo-depleted using the Illumina Ribo-Zero rRNA removal kit. RNA library preparation was performed using the Illumina TruSeq RNA Library Prep kit by following the manufacturer's instructions. Two replicates (non-treated, BRACO-19 treated and RHPS4) along with an input for each condition were paired end sequenced at 2 × 75 bp using NextSeq 500. The sequenced reads were mapped to the human reference genome hg19 assembly using HISAT237. Exon feature count and annotation to genes were performed using HTSeq38. DESeq239 was used for normalization and differential gene expression analyses. Further filtering and analyses were performed on Excel.

The analysis workflow is outlined in Supplementary Figure 4.S7. Initial gene lists were filtered by mean read count of >50 and fold change of >1.75 (log value of 0.8). The list of abundantly expressed genes were filtered by a mean read count of >500. GO analyses were performed using Enrichr42. Detailed list of gene transcripts can be found in Supplementary Data 1.

### Statistics

Graphs were produced by Microsoft Excel or GraphPad Prism. Statistical testing for multiple groups of dataset were performed using one-way or two-way ANOVA, and multiple comparisons corrected by the Bonferroni's method. Statistical comparisons of the average of two groups were performed using two-tailed Student's *t*-test. Linear regression goodness of fit was determined from coefficient of correlation ( $R^2$ ) or Pearson correlation ( $R$ ), and non-zero slope significance was given as *p*-value <0.05. All *p*-values <0.05 were considered significant unless specified otherwise.

## **4.6 Supplementary Materials**

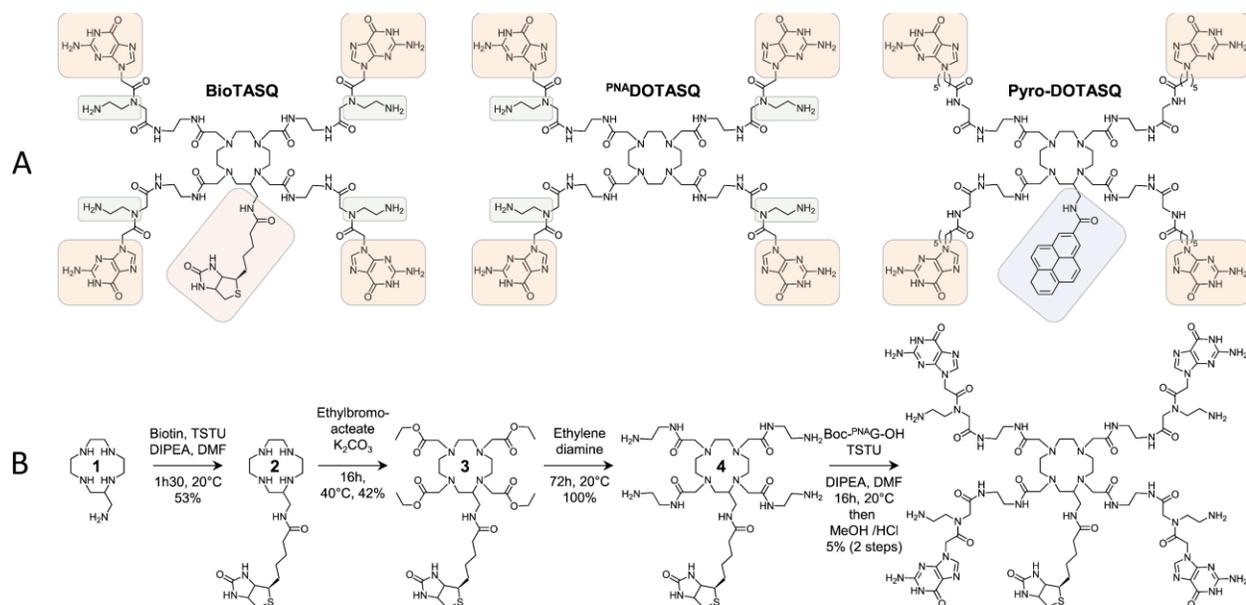
### I. Chemistry.

*Material and methods.* All chemicals were purchased from Sigma-Aldrich, except for Boc-P<sup>NA</sup>G(Z)-OH which was purchased from ASM Research Chemicals and used without further purification. NMR spectra were recorded with a Bruker 300 Avance III NanoBay spectrometer (300 MHz for <sup>1</sup>H and 75 MHz for <sup>13</sup>C). Chemical shifts are reported in δ ppm, using [D] solvents as standards. The following abbreviations are used: s: singlet, d: doublet, t: triplet, q: quartet, m: multiplet, br: broad. MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization - Time of Flight) mass spectrometry was carried out using a Bruker Ultraflex II LRF 2000 spectrometer and ESI (ElectroSpray Ionization) mass spectrometry was carried out using a Bruker Daltonics

AmaZon SL spectrometer. Purification: Flash chromatography was performed using the Automatic Reveleris® Flash Chromatography System (GRACE) equipped with multiple channel detection (UV (201 and/or 214 nm) and ELSD (Evaporative Light Scattering Detection)). A Reveleris® C18 RP 80 g cartridge was used, with CH<sub>3</sub>CN/0.1 % TFA and H<sub>2</sub>O/0.1 % TFA as eluents (linear gradient at 0% CH<sub>3</sub>CN/0.1 % TFA (12 min), linear gradient from 0% to 10% CH<sub>3</sub>CN/0.1 % TFA (10 min), linear gradient from 10% to 40% CH<sub>3</sub>CN/0.1 % TFA (5 min), linear gradient from 40% to 55% CH<sub>3</sub>CN/0.1 % TFA (5 min)) at a flow rate of 60 mL/min. Semi-preparative RP-HPLC was performed using a Thermo Beta Basic C18 column, 5 μm, 30 x 150 mm with CH<sub>3</sub>CN/0.1 % HCO<sub>2</sub>H and H<sub>2</sub>O/0.1 % HCO<sub>2</sub>H as eluents (linear gradient from 5 to 50% CH<sub>3</sub>CN/0.1 % HCO<sub>2</sub>H (45 min)) at a flow rate of 15 mL/min with a UV-Vis detection with an Ultimate 3000 diode array detector at 201, 214, 222, 260 nm.

The design of BioTASQ (Figure 4.S1A, left) is based on two previously reported TASQs, *i.e.*, <sup>PNA</sup>DOTASQ (Figure 4.S1A, center) (213) which is considered here as the parent compound, and of Pyro-DOTASQ (Figure 4.S1A, right) (275), in which a pyrene moiety was grafted on the DOTA template. Its synthesis (Figure 4.S1B) is performed as follows: the Pyro-DOTASQ key intermediate aminomethylcyclen 1, obtained after a described 5-step protocol from the commercially available triethylenetetramine (TETA), is coupled with biotin after activation with 2-succinimido-1,1,3,3-tetramethyluronium tetrafluoroborate (TSTU), to provide the biotin-cyclen conjugate 2; the TASQ core is subsequently constructed according to the <sup>PNA</sup>DOTASQ strategy, involving introduction of four ethyl bromoacetate arms (intermediate 3) subsequently elongated upon reaction with ethylene diamine to provide the intermediate 4 that finally reacts with *N*-Boc-protected <sup>PNA</sup>G monomers (Boc-<sup>PNA</sup>G-OH) to lead, after a final acidic deprotonation step, to BioTASQ. Of note, each compound is purified by semi-prep HPLC, which proves

difficult for the final compound BioTASQ, thus explaining the very low chemical yield (5%) for the very final steps.



**Figure 4.S1. A)** Chemical structures of BioTASQ, PNA-DOTASQ and Pyro-DOTASQ. **B)** Chemical synthesis of BioTASQ.

Step 1: TSTU (1.61 g, 5.36 mmol), biotin (1.19 g, 4.87 mmol) and DIPEA (0.93 mL, 5.36 mmol) were dissolved in DMF (50 mL). The mixture was stirred at room temperature for 30 min. A solution of **1** (cf. Note 1, 0.98 g, 4.87 mmol) and DIPEA (1.695 mL, 9.74 mmol) in DMF (25 mL) was slowly added and the mixture was stirred at room temperature for 1 hr. The solvent was evaporated under reduced pressure and the crude mixture was purified by flash chromatography to afford **2** as a white solid (1.11 g, 53%).  $^1\text{H NMR}$  (500 MHz,  $\text{d}_6\text{-DMSO}$ ):  $\delta$  7.50-7.37 (dd, 1H); 4.47-4.25 (m, 2H); 3.32-2.65 (m, 24H), 2.20 (m, 2H); 1.67-1.25 (m, 8H). HR-MS ( $m/z$ ):  $[\text{M}+\text{H}]^+$  calcd. for  $\text{C}_{19}\text{H}_{38}\text{N}_7\text{O}_2\text{S}$ , 428.27; found, 428.27 (and 450.26  $[\text{M}+\text{Na}]^+$ ).

Step 2: To a solution of **2** (200 mg, 0.31 mmol) and  $\text{K}_2\text{CO}_3$  (337 mg, 2.44 mmol) in acetonitrile (3 mL) was added ethylbromoacetate (138  $\mu\text{L}$ , 1.25 mmol). The mixture was stirred overnight at  $40^\circ\text{C}$ . It was filtered through dicalite and washed with acetonitrile. The filtrate was

evaporated under reduced pressure. The crude mixture was purified by semi-preparative HPLC to afford **3** as a white solid (100 mg, 42 %). <sup>1</sup>H NMR (500 MHz, d<sub>6</sub>-DMSO): δ 7.30 (s broad, 1H); 5.50-5.09 (m, 2H); 4.26-4.07 (m, 8H); 3.46-2.04 (m, 27H), 1.65-1.42 (m, 2H); 1.29-1.18 (m, 18H); 0.88 (m, 3H). HR-MS (m/z): [M+H]<sup>+</sup> calcd. for C<sub>35</sub>H<sub>62</sub>N<sub>7</sub>O<sub>10</sub>S, 772.42; found, 772.43 (and 794.42 [M+Na]<sup>+</sup>).

Steps 3 and 4: A solution of **3** (30 mg, 0.039 mmol) in ethylene diamine (1 mL) was stirred at room temperature for 3 days. Ethylene diamine was evaporated under reduced pressure. The mixture was triturated successively in acetonitrile and diethyl ether to afford **4** as a white solid, which is directly engaged in the following step: Boc-<sup>PNA</sup>G-OH (cf. note 2, 238 mg, 0.58 mmol), TSTU (175 mg, 0.58 mmol) and DIPEA (101 μL, 0.58 mmol) were dissolved in DMF (1.6 mL). The mixture was stirred at room temperature for 30 min. A solution of **4** (96 mg, 0.12 mmol) and DIPEA (40 μL, 0.23 mmol) in DMF (2 mL) was then added dropwise and the mixture was stirred overnight at room temperature. The solvent was removed under reduced pressure. The crude mixture was purified by semi-preparative HPLC to afford **5** as a white solid. <sup>1</sup>H NMR (500 MHz, d<sub>6</sub>-DMSO): δ 10.77 (s broad, 2H); 7.9-7.42 (m, 14H); 7.05 (s broad, 2H); 6.90-6.62 (m, 5H); 6.17 (dd, 5H); 5.32-4.84 (m, 9H); 4.26-4.03 (m, 10H); 3.52-3.07 (m, 43H), 1.82 (m, 5H); 1.39-0.99 (m, 58H). MS (MALDI-TOF, m/z): [M+H]<sup>+</sup> calcd for 2393.17, found 2393.22.

Step 5: 5 mL of a solution of methanol saturated with HCl were added to **5**. The mixture was stirred for 30 min at room temperature. Diethyl ether was added gently to precipitate the compound. The supernatant was removed and the residue was dried under reduced pressure to afford BioTASQ as a white powder (11 mg, 4.6%). <sup>1</sup>H NMR (500 MHz, d<sub>6</sub>-DMSO): δ 11.49 (s broad, 4H); 8.78-8.42 (m, 5H); 8.42-8.04 (2 s broad, 9H); 7.12 (s broad, 8H); 5.31-5.04 (m, 8H);

4.26-2.92 (m, water signals precluded proper interpretation, expected: 85H, 18H detectable), 1.30-1.12 (m, 2H). MS (MALDI-TOF, m/z):  $[M+H]^+$  calcd for  $C_{79}H_{122}N_{43}O_{18}S$ , 1991.96; found 1993.09. HR-MS (m/z):  $[M+2H+H_2O]^{2+}$ , calcd. for 1004.98451; found 1004.98318 ( $\Delta = -1.323$  ppm).

**Note 1:** Synthesis of aminomethylcyclen **1** (276). **Step 1:** A solution of 2,3-butanedione (88.3 g, 1.03 mol) was added dropwise to a solution of triethylenetetraamine (TETA, 150 g, 1.03 mol) and calcium hydroxide (151.8 g, 2.05 mol) in acetonitrile (6 L) at 0°C. The mixture was stirred at this temperature for 2hrs. It was then filtered over dicalite and washed with acetonitrile to afford a yellow solution. This solution was cooled down to 0°C and  $K_2CO_3$  (283.2 g, 2.05 mol) and benzotriazole (122.1 g, 1.03 mol) were added. A solution of chloroacetaldehyde in acetonitrile (1 L) was slowly added at 0°C and the resulting mixture was stirred for 2 h. NaCN (50.3 g, 1.03 mol) was added and the mixture was stirred overnight at room temperature. The solution was filtered through dicalite and washed with acetonitrile. The filtrate was evaporated. **Step 2:** A solution of the previously prepared compound (126.9 g, 0.51 mol) in dry THF (500 mL) was slowly added to a suspension of  $LiAlH_4$  (23.4 g, 0.62 mol) in dry THF (1 L) under nitrogen at -78°C. The resulting mixture was stirred overnight. Water (250 mL) was carefully added to the mixture at -78°C to neutralize the excess  $LiAlH_4$ . The mixture was evaporated and the residue was taken up in chloroform (1.5 L) and stirred overnight at room temperature. The mixture was filtered through dicalite and washed with chloroform. The filtrate was concentrated under reduced pressure. **Step 3:** A solution of the previously prepared compound (77.3 g, 0.30 mol) in 37% hydrochloric acid (250 mL, 3 mol, 10 equiv.) was stirred overnight at 80°C. The mixture was evaporated and triturated in  $Et_2O$ . It was filtered and washed successively with 37% HCl and

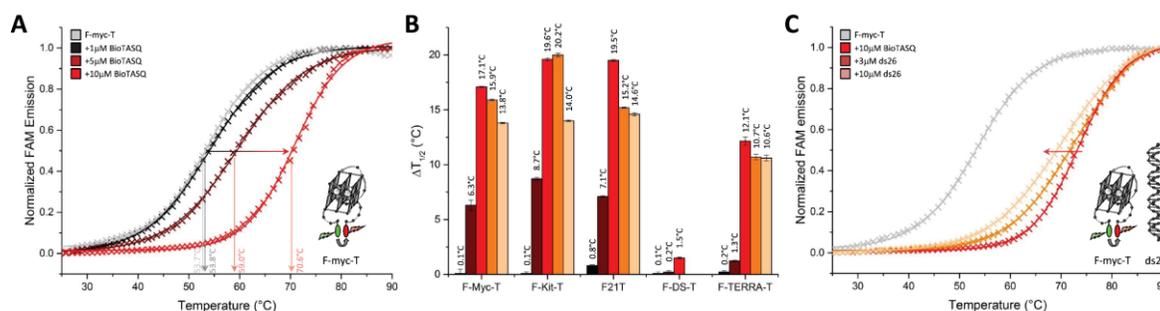
diethyl ether to afford aminomethylcyclen **1** as a white powder (21.73 g, 19%). <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 300K): Δ (ppm): 2.01 (br. s., 6H), 2.40-2.80 (m, 17H).

**Note 2:** Synthesis of Boc-<sup>PNA</sup>G-OH (213): To a solution of Boc-<sup>PNA</sup>G(Z)-OH (1 g, 1.84 mmol) in methanol (50 mL) was added Pd/C (39.1 mg, 20 mol%). The suspension was stirred overnight at room temperature under H<sub>2</sub>. The solid was then filtered over dicalite and washed with methanol. The residue was concentrated under reduced pressure to afford Boc-<sup>PNA</sup>G-OH as a white solid (540 mg, 72%). <sup>1</sup>H NMR (300 MHz, d<sub>6</sub>-DMSO, 300K): δ (ppm): 1.37 (s, 9H), 3.20-3.80 (m, 4H), 3.99 (s, 2H), 4.85 (s, 2H), 6.46 (s, 2H), 6.95 (m, 1H), 7.51 (s, 1H), 10.59 (s, 1H).

## II. FRET-melting experiments.

The nucleic acid binding capabilities of BioTASQ were investigated with the fluorescence resonance energy transfer (FRET)-melting assay (277): experiments were implemented with increasing concentrations of BioTASQ (1-10 μM) against a panel of biologically relevant doubly labeled quadruplex-forming sequences (Figure 4.S2), namely F21T, F-myc-T and F-kit-T, which are DNA sequences found in the human telomeres and in the promoter regions of myc and kit genes, respectively, F-TERRA-T, which is a RNA sequence found in the human telomeric transcripts, along with F-DS-T, a duplex-DNA control. As seen in Figures 4.S2, BioTASQ stabilizes the quadruplexes in a dose-response manner, with ΔT<sub>1/2</sub> values up to 19.6°C (for DNA) and 12.1°C (for RNA) at 10 μM concentration, albeit modestly since very low stabilizations are obtained at 1 μM concentration (ΔT<sub>1/2</sub> = 0.1-0.8°C) that is considered as the standard ligand concentration in the FRET-melting assay. Of note, in these conditions (i.e., 1 μM ligand), the parent compound <sup>PNA</sup>DOTASQ stabilizes quadruplexes more efficiently with ΔT<sub>1/2</sub> values

between 12.5-21.1°C; these results indicate that the introduction of the biotin tag reduced ligand affinity. We next investigated the quadruplex selectivity of BioTASQ, which is concluded to be similar to that of the parent compound <sup>PNA</sup>DOTASQ, as shown by a low F-DS-T stabilization ( $\Delta T_{1/2}$  up to 1.5°C, at 10  $\mu$ M concentration) and further confirmed *via* competitive FRET-melting experiments carried out with labeled quadruplexes in the presence of 15 and 50 mol. equiv. of unlabeled 26-bp duplex-DNA ds26. As seen in Figure 4.S2, the stabilization is maintained at 78-100% and 71-87% in the presence of 15 and 50 mol. equiv. ds26, respectively, with overall better affinity for quadruplex-RNA *vs.* quadruplex-DNA. Collectively, these two series of data highlighted the fair G4 affinity but high selectivity of BioTASQ for quadruplexes (*vs* duplex-DNA).



**Figure 4.S2. A.** Dose-response FRET-melting curves obtained with F-myc-T (0.2mM) and increasing amounts of BioTASQ (1-10mM). **B.** Summary of  $\Delta T_{1/2}$  values with F-myc-T, F-kit-T, F21T, F-DS-T and F-TERRA-T (0.2mM) and increasing amounts of BioTASQ (1-10mM), without or with the duplex ds26 as competitor (3-10mM). **C.** Competitive FRET-melting curves obtained with F-myc-T (0.2mM), BioTASQ (10mM) and increasing amounts of ds26 (3-10mM).

The sequences of the oligonucleotides used herein are: F-21-T: FAM-

d[<sup>5'</sup>G<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>T<sub>2</sub>AG<sub>3</sub>']-TAMRA; F-myc-T: FAM-d[<sup>5'</sup>GAG<sub>3</sub>TG<sub>4</sub>AG<sub>3</sub>TG<sub>4</sub>A<sub>2</sub>G<sup>3'</sup>]-TAMRA;

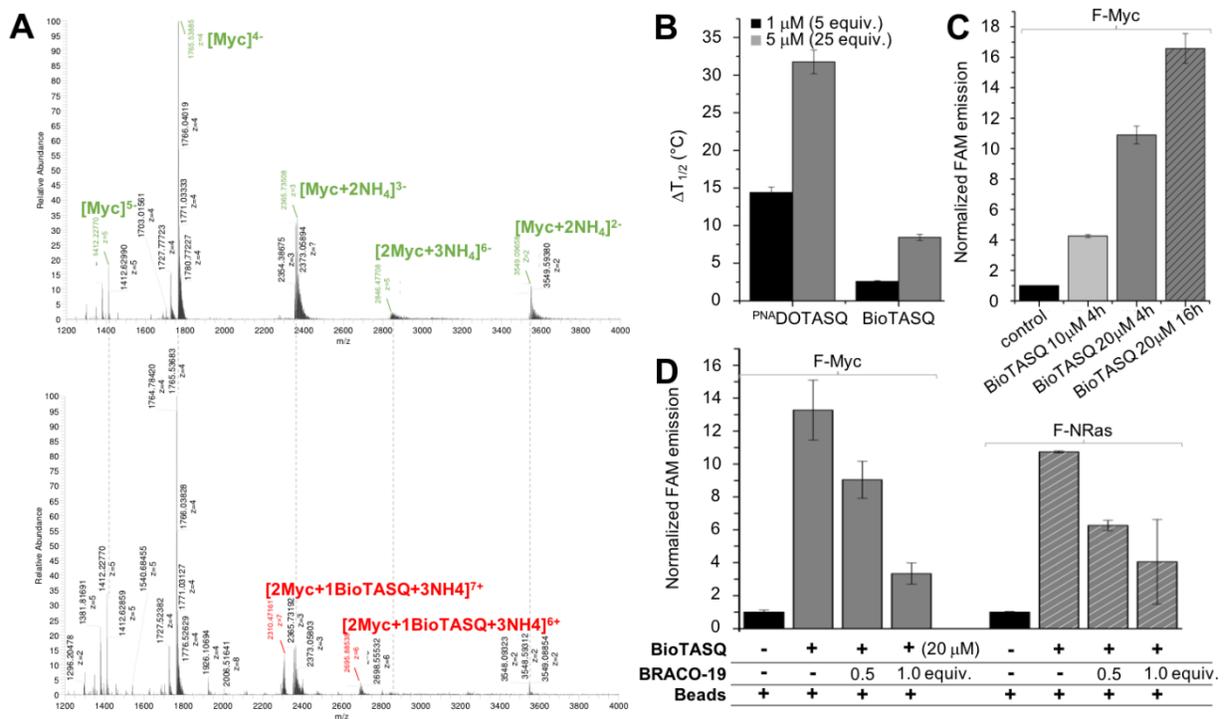
F-kit-T: FAM-d[<sup>5'</sup>CG<sub>3</sub>CG<sub>3</sub>CGCGAG<sub>3</sub>AG<sup>3'</sup>]-TAMRA); F-DS-T: FAM-

d[<sup>5'</sup>TATAGCTATAT<sub>6</sub>TATAGCTATA<sup>3'</sup>]-TAMRA; F-TERRA-T: FAM-5[<sup>5'</sup>

G<sub>3</sub>U<sub>2</sub>AG<sub>3</sub>U<sub>2</sub>AG<sub>3</sub>U<sub>2</sub>AG<sub>3</sub>']-TAMRA; and the self-complementary ds26: d[<sup>5'</sup>CA<sub>2</sub>TCG<sub>2</sub>ATCGA<sub>2</sub>T<sub>2</sub>CGATC<sub>2</sub>GAT<sub>2</sub>G<sup>3'</sup>]. Lyophilized DNA strands (purchased from Eurogentec, Seraing, Belgium) were first diluted to 500 μM in deionized water (18.2 MΩ.cm resistivity). All DNA structures were prepared in Caco.K buffer, comprised of 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (F21T, F-DS-T) or plus 1 mM KCl/99 mM LiCl (F-MYC-T, F-KIT-T, F-TERRA-T). Quadruplex structures were prepared by mixing 40 μL of the constitutive strand (500 μM) with 8 μL of a lithium cacodylate buffer solution (100 mM, pH 7.2), plus 8 μL of a KCl/LiCl solution (100 mM/900 mM) and 24 μL of water. The duplex structure (ds26) was prepared by mixing 40 μL of each constitutive strand (500 μM) with 16 μL of a lithium cacodylate buffer solution (100 mM, pH 7.2), plus 16 μL of a KCl/LiCl solution (100 mM/900 mM) and 48 μL of water. The final concentrations were theoretically 250 and 125 μM, for mono- and bimolecular DNA structures, respectively. Actual concentration of each DNA was determined through dilution to 1 μM theoretical concentration and confirmed with UV spectral analysis at 260 nm (after 5 min at 90 °C) with the following molar extinction coefficient values: 268300 (F21T), 232000 (F-myc-T), 205600 (F-kit-T), 258900 (F-DS-T), 276700 (F-TERRA-T) and 506400 M<sup>-1</sup>.cm<sup>-1</sup> (ds26). Higher-order DNA structures were folded according to two procedures: (a) for the monomolecular architectures, solutions were heated (90°C, 5 min), cooled on ice (7 h) and then stored at least overnight (4°C); (b) for the folding of all other structures, the solutions were heated (90°C, 5 min), gradually cooled (65, 60, 55, 50, 40 and 30°C (30 min/step), 25°C (2 h)) and then stored at least overnight (4°C).

FRET-melting experiments were performed in 96-well format using a Mx3005P qPCR machine (Agilent) equipped with FAM filters ( $\lambda_{ex} = 492$  nm;  $\lambda_{em} = 516$  nm) in 100 μL (final volume) 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (F21T, F-DS-

T) or plus 1 mM KCl/99 mM LiCl (F-MYC-T, F-KIT-T, F-TERRA-T) with 0.2  $\mu\text{M}$  of labeled oligonucleotide and 0, 1, 5 or 10  $\mu\text{M}$  of BioTASQ. Competitive experiments were carried out with labeled oligonucleotide (0.2  $\mu\text{M}$ ), 10  $\mu\text{M}$  BioTASQ and increasing amounts (0, 15 and 50 mol. equiv.) of the unlabeled competitor ds26. After a first equilibration step (25  $^{\circ}\text{C}$ , 30 s), a stepwise increase of 1  $^{\circ}\text{C}$  every 30 s for 65 cycles to reach 90  $^{\circ}\text{C}$  was performed, and measurements were made after each cycle. Final data were analyzed with Excel (Microsoft Corp.) and OriginPro<sup>®</sup>9.1 (OriginLab Corp.). The emission of FAM was normalized (0 to 1), and  $T_{1/2}$  was defined as the temperature for which the normalized emission is 0.5;  $\Delta T_{1/2}$  values are means of 2-4 experiments.



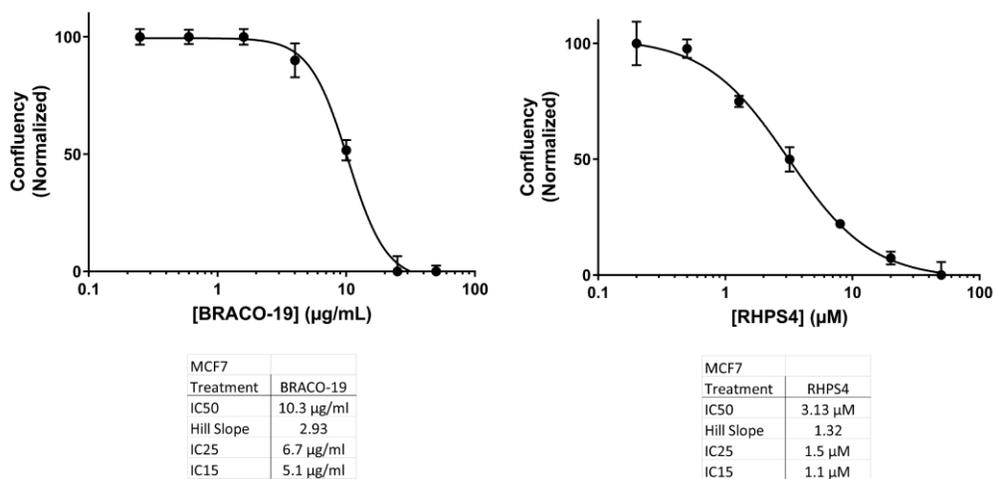
**Figure 4.S3.** A) Electrospray ionization mass spectra (ESI-MS) analysis demonstrated the low affinity and stoichiometry (1:1) of association between Myc quadruplex and BioTASQ. B) Confirmation of the lower quadruplex-affinity of BioTASQ as compared to the parent molecule PNA/DOTASQ. FRET-melting experiments were performed with F-Myc-T (0.2  $\mu\text{M}$ ) with increasing amounts of ligands (1 and 5  $\mu\text{M}$ , *i.e.*, 5 and 25 mol. equiv.). C) Pull-down experiments were performed with F-Myc (1  $\mu\text{M}$ ) and increasing BioTASQ concentrations (10

and 20  $\mu\text{M}$ ) and incubation times (4 and 16h). Experiments were done with triplicates D) Competition BioTASQ (20  $\mu\text{M}$ ) pull-down experiments were performed with F-Myc and F-NRas (1  $\mu\text{M}$ ), and increasing amounts of BRACO-19 (0.5 and 1  $\mu\text{M}$ ). Experiments were done with triplicates. Error bars represented SD.

Electrospray mass spectrometry experiments were performed on a LTQ Orbitrap XL (Thermo Scientific) spectrometer equipped with Ion Max source and HESI-II probe in the negative ion mode, according to the previously described protocol (278). Myc ( $d[{}^5\text{GAG}_3\text{TG}_4\text{AG}_3\text{TG}_4\text{A}_2\text{G}^3]$ , 10  $\mu\text{M}$ ) and the corresponding Myc:BioTASQ mixture (1:1 ratio) were prepared in 100 mM ammonium acetate buffer and equilibrated at 25°C for 1 hour. To obtain a stable electrospray signal, 20% of methanol were added to the solution just before injection. The solutions were injected with syringe pump at a flow rate of 5  $\mu\text{L}/\text{min}$ . The full scan mass was recorded in 600-4000 m/z range. The following tuning parameters were used: heater temperature = 50°C, spray voltage = 4.0 kV, capillary temperature = 275°C, Tube lens = -160.00 (negative ion mode) and the capillary voltage varied between -35.00 V and -60.00V.

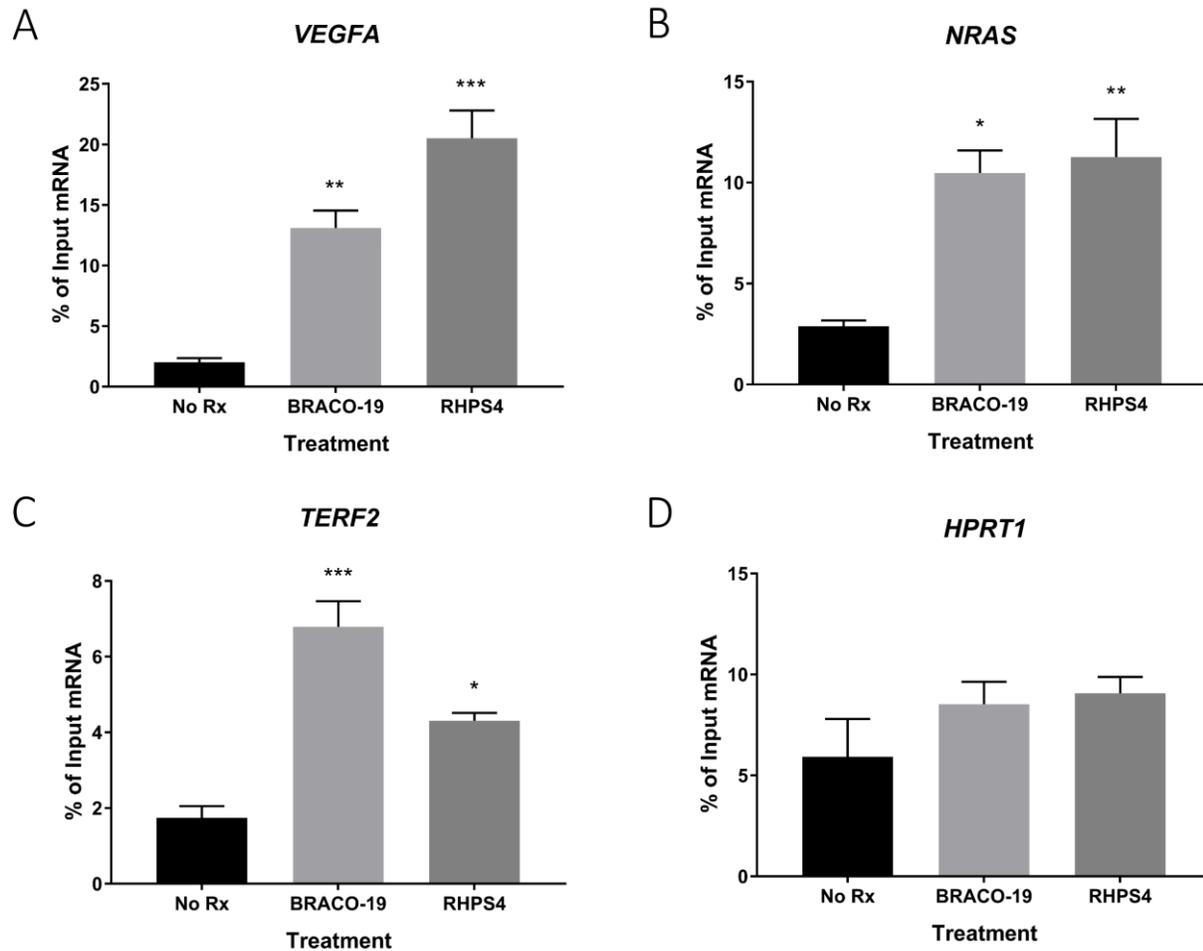
### III. Dose-response Profiling of G4 ligands

MCF7 cells were cultured with 5% FBS in DMEM (Hyclone), washed with PBS (Gibco) and trypsinized with 0.025% Trypsin/EDTA (Gibco). MCF7 cells were seeded at 3000/well in a 96-well flat bottom plate (Corning). The cells were treated with serial dilutions of indicated G4 ligands. Growth kinetics and cell morphology were recorded using the Essenbio IncuCyte ZOOM live-cell monitoring system (274). Dose response curves were generated using Graphpad Prism. The inhibitory dose 50 (or lethal dose 50, or LD<sub>50</sub>) at which 50% growth inhibition occurred was calculated from the generated dose response curves. Doses between IC<sub>15</sub> (or LD<sub>15</sub>) to IC<sub>25</sub> (or LD<sub>25</sub>) were used for subsequent experiments.

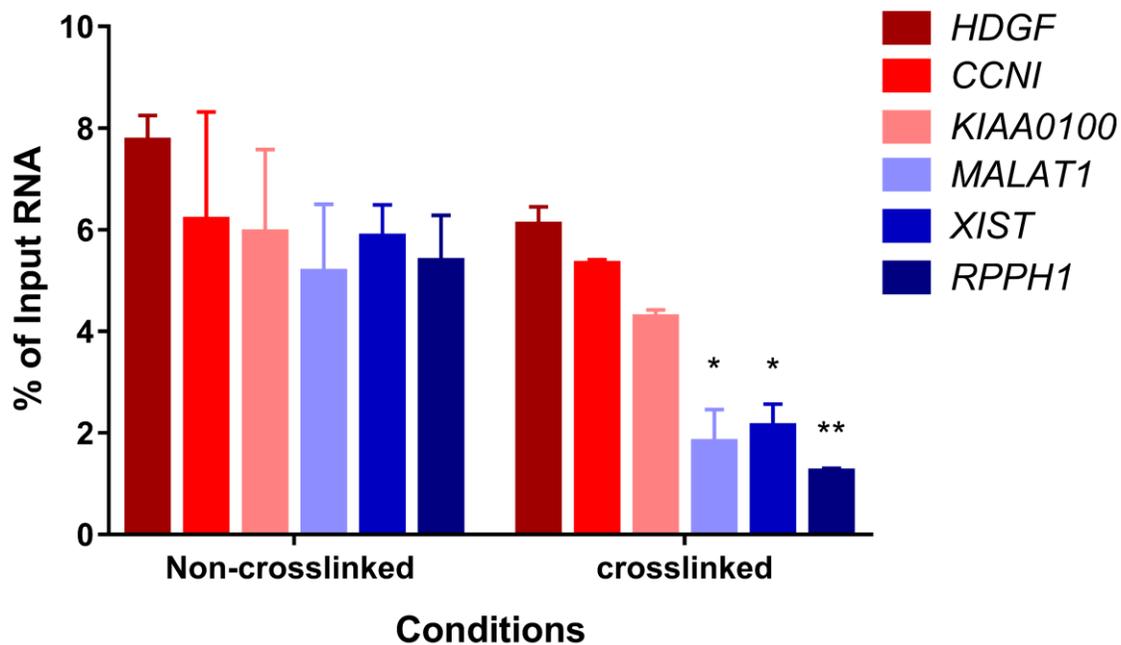


**Figure 4.S4.** Dose response profiles for BRACO-19 and RHPS4 in MCF7 cells generated from IncuCyte ZOOM live-cell monitoring system. Experiments were conducted with three biological replicates. The IC50, hill slope, IC25 and IC15 are listed in the tables below. Error bars represent SEM.

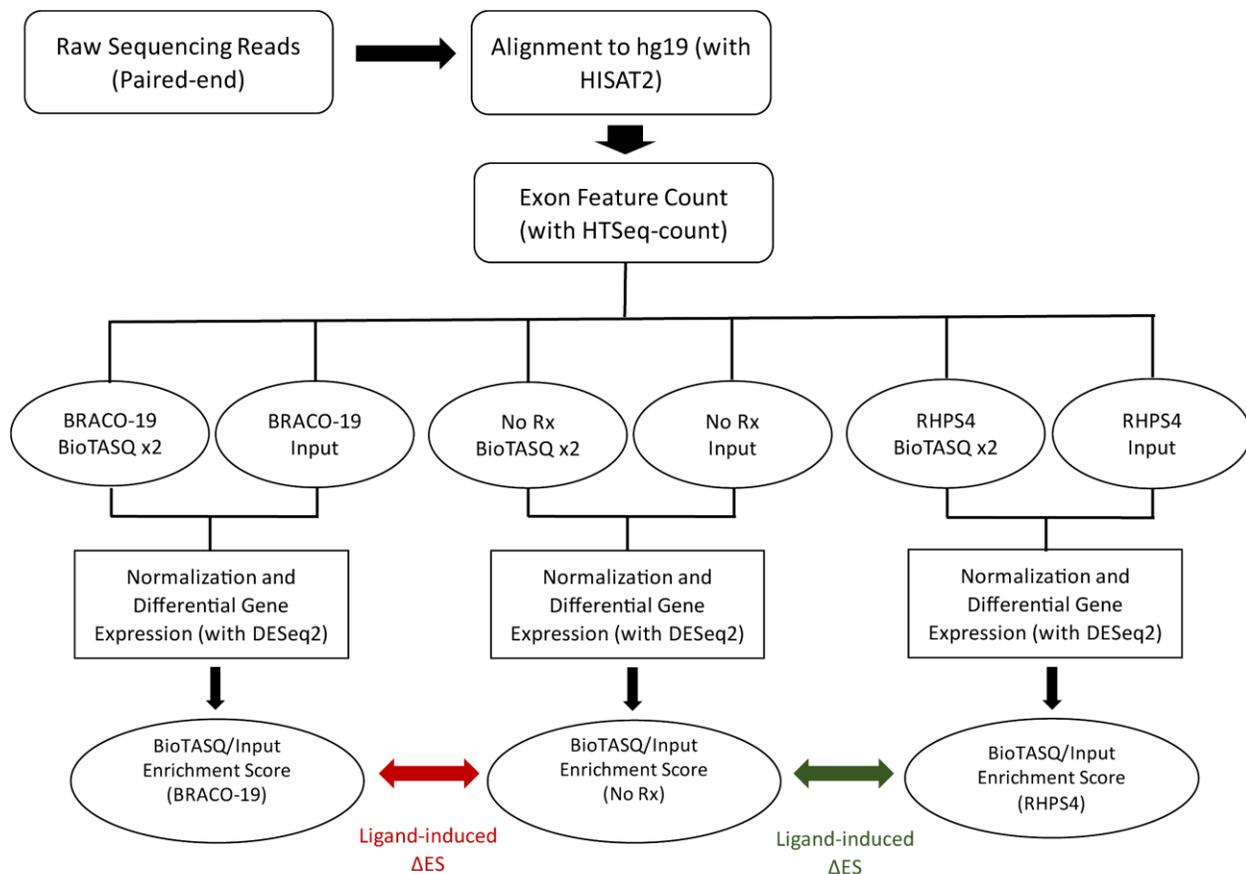
IV. Additional G4RP & G4RP-seq results.



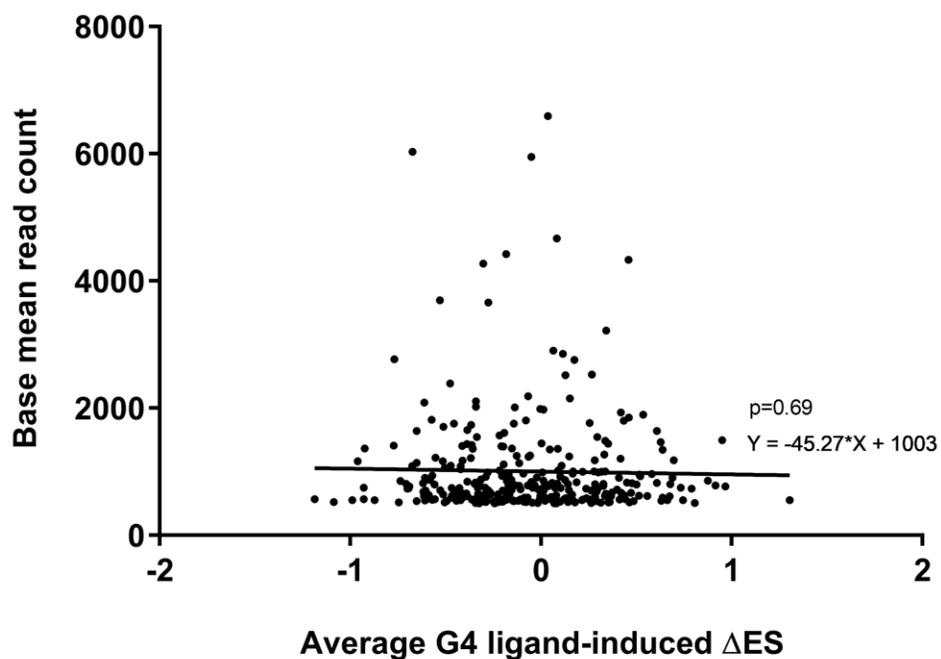
**Figure 4.S5.** BioTASQ G4RP signals were quantified with RT-qPCR measurements of (A) VEGFA, (B) NRAS, (C) TERF2 or (D) HPRT1 mRNA levels in untreated (No Rx), BRACO-19-treated or RHPS4-treated MCF7 cells. Values are normalized to their individual input control used for the BioTASQ pull down. One-way ANOVA was performed against No Rx control. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001. Experiments were conducted with three biological replicates. Error bars represent SEM.



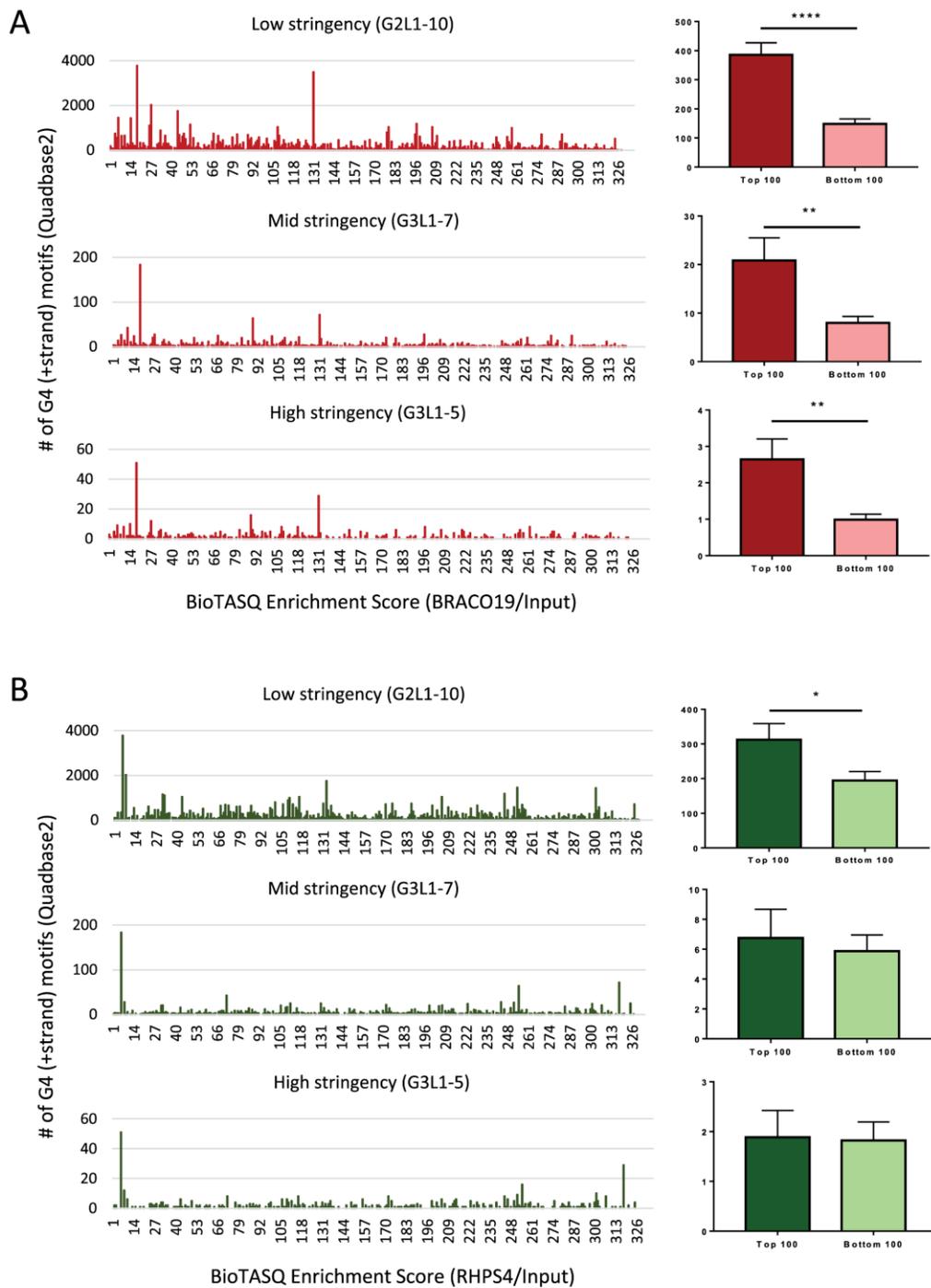
**Figure 4.S6.** BioTASQ enrichment of G4-RNA using crosslinked versus non-crosslinked samples. RT-qPCR measurements of top ranked genes (HDGF, CCNI, KIAA0100) and lowest-ranked gene (MALAT, XIST, RPPH1) from the G4RP-seq dataset at baseline (with no G4 ligand treatments) condition were tested. Two-way ANOVA was performed against HDGF (top ranked gene) for each gene under each condition. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Experiments were performed with three biological replicates. Error bars represent SEM.



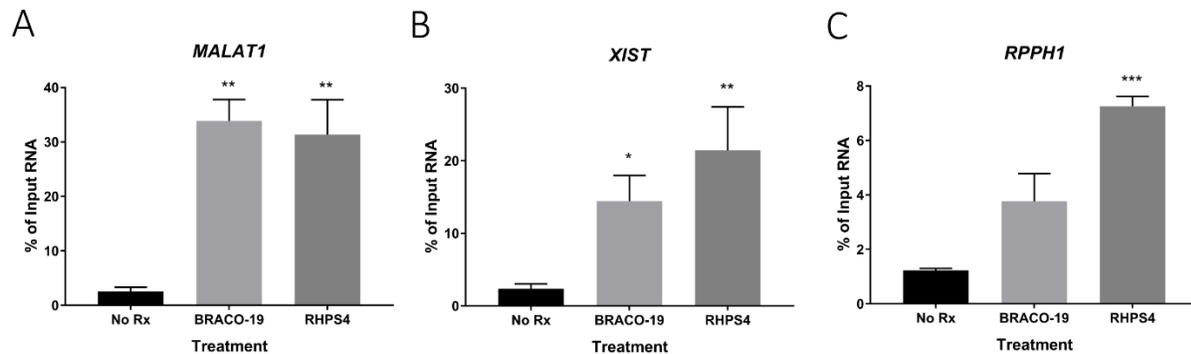
**Figure 4.S7.** G4RP-seq analysis workflow. Pair-end raw sequencing reads were first aligned to human genome assembly hg19 using HISAT2. The aligned reads within exons were then counted and annotated to genes using HTSeq. The counts were then normalized using DESeq2 before calculating the differential gene expression for each BioTASQ and input pair to obtain the Enrichment Score (ES). BioTASQ/input values (from DESeq2) were compared between conditions to calculate ligand-induced Enrichment Score Change ( $\Delta$ ES) using Excel.



**Figure 4.S8.** Regression plot of average G4 ligand-induced BioTASQ  $\Delta$ ES versus transcript abundance (as measured by baseline mean read count) ( $p=0.69$ , non-significant relationship). Genes with read count of over 8000 were removed to avoid regression being heavily weighted and thus driven by a few points (3 genes out of the 329 removed due to values beyond 35000).



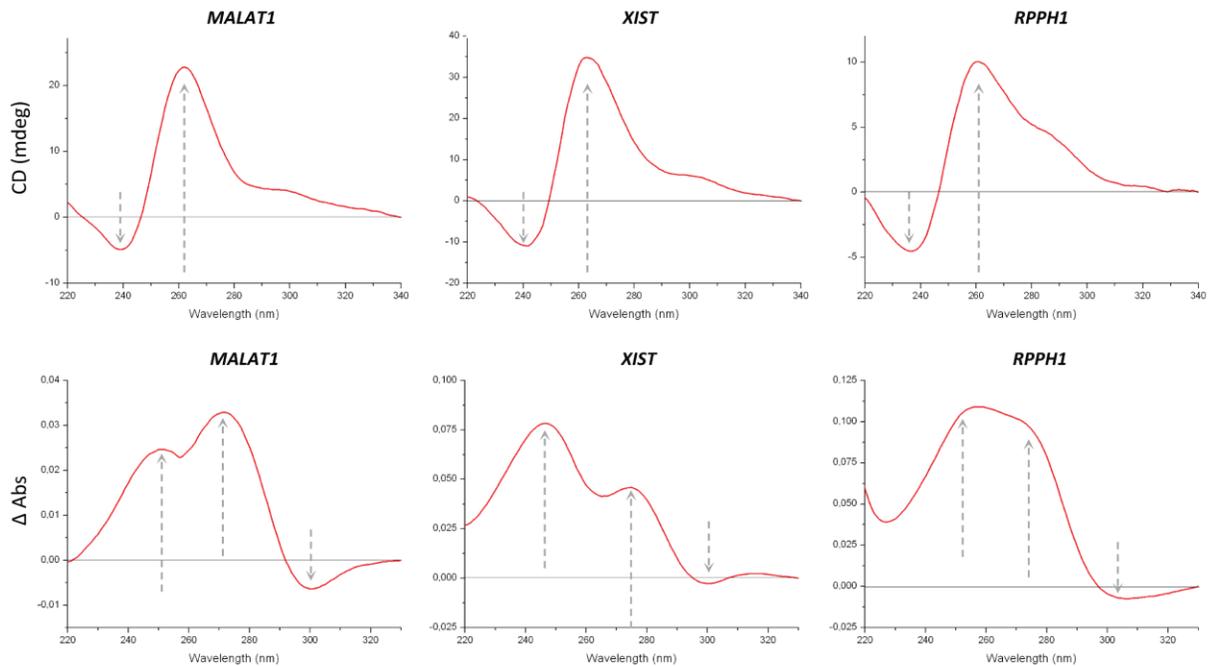
**Figure 4.S9.** Number of non-overlapping G4 motifs (calculated by Quadbase2) in genes ranked by ES of a) BRACO-19 treated sample (red) and b) RHPS4 treated sample (green). Three different stringency settings are shown: low (G2L1-10), mid (G3L1-7), and high (G3L1-5). Right panel shows the quantification for the top and bottom 100 ranked transcripts. Two-tailed student t-test was performed. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent SEM.



**Figure 4.S10.** G4RP signal of BioTASQ by RT-qPCR quantification of (A) *MALAT1*, (B) *XIST*, (C) *RPPH1* RNA levels in untreated (No Rx), BRACO-19-treated or RHPS4-treated MCF7 cells. Values are normalized to their individual input control used for RIP. Experiments were performed with three biological replicates. One-way ANOVA was performed against non-treated control. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent SEM.

#### -- V. Circular Dichroism results

Circular dichroism (CD) spectra were recorded on a JASCO J-815 spectropolarimeter in a 10mm path-length quartz semi-micro cuvette (Starna). CD spectra were recorded over a range of 220-400nm (bandwidth = 0.5nm, 1nm pitch, 1s response, scan speed = 500nm.mn<sup>-1</sup>, averaged over 3 scans, zeroed at 340nm) with MALAT1: r[G<sub>3</sub>ATG<sub>3</sub>AG<sub>2</sub>AG<sub>5</sub>TG<sub>3</sub>], XIST: r[G<sub>2</sub>AAG<sub>2</sub>AAG<sub>2</sub>TTG<sub>2</sub>] and RPPH1: r[G<sub>2</sub>AG<sub>4</sub>CCCG<sub>2</sub>CG<sub>2</sub>] (10 μM) in 1mL (final volume) of 10mM lithium cacodylate buffer (pH 7.2) + 90mM LiCl/10mM KCl. Thermal difference spectra (TDS) spectra were collected Spectra were recorded on a JASCO V630Bio spectrophotometer and calculated using the same cuvettes (81).



**Figure 4.S11.** Confirmation of quadruplex topology by circular dichroism (CD) and thermal differential spectra (TDS) investigations of selected quadruplex-forming sequences (QFS) of top ranked gene (MALAT1: r[G<sub>3</sub>ATG<sub>3</sub>AG<sub>2</sub>AG<sub>5</sub>TG<sub>3</sub>] (G-score: 40); XIST: r[G<sub>2</sub>AAG<sub>2</sub>AAG<sub>2</sub>TTG<sub>2</sub>] (G-score: 21); and RPPH1: r[G<sub>2</sub>AG<sub>4</sub>CCCG<sub>2</sub>CG<sub>2</sub>] (G-score: 21)). Typical G4-RNA parallel quadruplex signatures were obtained by both CD (positive (260-265nm) and negative peaks (240nm)) and TDS (positive (250-270nm) and negative peaks (295-300nm), dashed arrows).

-- VI. Gene Ontology analysis results

**Table 4.S1.** Gene Ontology (GO) geneset analysis on baseline G4 levels using the top 100 and bottom 100 highly abundant transcripts ranked by BioTASQ Enrichment Score. The top 5 ranked terms are listed for each category (Cellular component, Biological Process and Molecular Function).

<b>Top 100</b>		<b>Bottom 100</b>		
<b>Rank</b>	<b>Cellular Component GO Term</b>	<b>Score</b>	<b>Cellular Component GO Term</b>	<b>Score</b>
1	cytosolic part	76.46	cytoplasmic stress granule	22.24
2	cytosolic ribosome	69.62	ribonucleoprotein granule	15.79
3	ribosome	55.69	cytoplasmic ribonucleoprotein granule	12.76
4	polysome	50.09	nuclear matrix	12.07
5	cytosolic large ribosomal subunit	49.36	nucleolus	11.3
<b>Rank</b>	<b>Biological Process GO Term</b>	<b>Score</b>	<b>Biological Process GO Term</b>	<b>Score</b>
1	protein targeting to ER	110.01	ribosomal small subunit export from nucleus	31.5
2	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	107.93	RNA metabolic process	28.88
3	peptide biosynthetic process	107.5	alternative mRNA splicing, via spliceosome	25.48
4	SRP-dependent cotranslational protein targeting to membrane	100.23	rRNA-containing ribonucleoprotein complex export from nucleus	22.19
5	viral transcription	89.05	ribosomal large subunit export from nucleus	21.62
<b>Rank</b>	<b>Molecular Function GO Term</b>	<b>Score</b>	<b>Molecular Function GO Term</b>	<b>Score</b>
1	RNA binding	79.86	RNA binding	31.18
2	cadherin binding	43.5	DNA helicase activity	17.57
3	eukaryotic initiation factor 4E binding	31.76	RNA helicase activity	16.9
4	small ribosomal subunit rRNA binding	31.08	protein kinase activator activity	16.79
5	histone methyltransferase activity	28.59	RNA stem-loop binding	15.65

**Table 4.S2.** Gene Ontology (GO) geneset analysis on filtered gene list (50 mean base count and >1.75 fold change) with BRACO-19- (left) or RHPS4- (right) induced BioTASQ-enrichment. BRACO-19, n=251; RHPS4, n=463. The top 5 ranked terms are listed for each category (Cellular component, Biological Process and Molecular Function).

<b>BRACO-19</b>		<b>RHPS4</b>	
<b>Rank</b>	<b>Cellular Component GO Term</b>	<b>Cellular Component GO Term</b>	<b>Score</b>
1	trans-Golgi network membrane	nucleolar part	32.64
2	Golgi membrane	nucleolus	32.41
3	cytoplasmic side of Golgi membrane	Golgi membrane	22.58
4	extrinsic component of Golgi membrane	cytoplasmic side of Golgi membrane	22.49
5	Golgi cisterna membrane	extrinsic component of Golgi membrane	22.46
<b>Rank</b>	<b>Biological Process GO Term</b>	<b>Biological Process GO Term</b>	<b>Score</b>
1	negative regulation of transcription from RNA polymerase II promoter	G1/S transition of mitotic cell cycle	33.61
2	negative regulation of mitotic cell cycle	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	33.59
3	vesicle targeting, rough ER to cis-Golgi	RNA splicing	24.25
4	endoplasmic reticulum-Golgi intermediate compartment (ERGIC)	mRNA splice site selection	22.96
5	COPII vesicle uncoating	RNA export from nucleus	22.44
<b>Rank</b>	<b>Molecular Function GO Term</b>	<b>Molecular Function GO Term</b>	<b>Score</b>
1	single-stranded RNA binding	siRNA binding	46.15
2	double-stranded RNA binding	pre-miRNA binding	46.11
3	misfolded RNA binding	tRNA binding	45.48
4	RNA binding	double-stranded RNA binding	44.75
5	base pairing with RNA	mRNA binding	44.36

-- VII. Sequences used in current study

Table 4.S3. Sequences of oligonucleotides used in the <i>in vitro</i> assays			
FRET-melting Assay	Dual-labeled DNA	F21T	FAM-d[ <sup>5'</sup> GGGTTAGGGTTAGGGTTAGGG <sup>3'</sup> ]-TAMRA
		F-myc-T	FAM-d[ <sup>5'</sup> GAGGGTGGGGAGGGTGGGGAAG <sup>3'</sup> ]-TAMRA
		F-kit-T	FAM-d[ <sup>5'</sup> CGGGCGGGCGCGAGGGAGGGG <sup>3'</sup> ]-TAMRA
		F-DS-T	FAM-d[ <sup>5'</sup> TATAGCTATATTTTTTTATAGCTATA <sup>3'</sup> ]-TAMRA
	Dual-labeled RNA	F-TERRA-T	FAM-r[ <sup>5'</sup> GGGUUAGGGUUAGGGUUAGGG <sup>3'</sup> ]-TAMRA
	Unlabeled DNA	ds26	d[ <sup>5'</sup> CAATCGGATCGAATTCGATCCGATTG <sup>3'</sup> ]
<i>In vitro</i> Pull-down Assay	Labeled DNA	F-myc	FAM-d[ <sup>5'</sup> GAGGGTGGGGAGGGTGGGGAAG <sup>3'</sup> ]
		F-SRC	FAM-d[ <sup>5'</sup> GGGAGGGAGGGCTGGGGG <sup>3'</sup> ]
		F-22AG	FAM-d[ <sup>5'</sup> AGGGTTAGGGTTAGGGTTAGGG <sup>3'</sup> ]
		F-DS	FAM-d[ <sup>5'</sup> TATAGCTATATTTTTTTATAGCTATA <sup>3'</sup> ]
	Labeled RNA	F-TERRA	FAM-r[ <sup>5'</sup> GGGUUAGGGUUAGGGUUAGGG <sup>3'</sup> ]
		F-TRF2	FAM-r[ <sup>5'</sup> CGGGAGGGCGGGGAGGGC <sup>3'</sup> ]
		F-RAS	FAM-r[ <sup>5'</sup> GGGAGGGGCGGGUCUGGG <sup>3'</sup> ]
	Unlabeled DNA	ds17-1	d[ <sup>5'</sup> CCAGTTCGTAGTAACCC <sup>3'</sup> ]
		ds17-2	d[ <sup>5'</sup> GGGTACTACGAACTGG <sup>3'</sup> ]
		ds26	d[ <sup>5'</sup> CAATCGGATCGAATTCGATCCGATTG <sup>3'</sup> ]
ESI-MS		Myc	d[ <sup>5'</sup> GAGGGTGGGGAGGGTGGGGAAG <sup>3'</sup> ]
CD, TDS		MALAT1	r[G <sub>3</sub> ATG <sub>3</sub> AG <sub>2</sub> AG <sub>5</sub> TG <sub>3</sub> ]
		XIST	r[G <sub>2</sub> AAG <sub>2</sub> AAG <sub>2</sub> TTG <sub>2</sub> ]
		RPPH1	r[G <sub>2</sub> AG <sub>4</sub> CCCG <sub>2</sub> CG <sub>2</sub> ]

<b>Table 4.S4. Primer set used for RT-qPCR</b>			
<b>mRNA</b>	<b>Forward</b>	<b>Reverse</b>	<b>Source</b>
<i>VEGFA</i>	CCTTGCCTTGCTGCTCTACC	AGATGTCCACCAGGGTCTCG	Ref. 249
<i>NRAS</i>	ATGACTGAGTACAACTGGTG GT	CATGTATTGGTCTCTCATGGCA C	Ref. 249
<i>TERF2</i>	GTACGGGGACTTCAGACAGAT	CGCGACAGACACTGCATAAC	Ref. 249
<i>HPRT1</i>	ACCAGTCAACAGGGGACATAA	CTTCGTGGGGTCCTTTTCACC	Ref. 249
<i>MALAT1</i>	AAAGCAAGGTCTCCCCACAAG	GGTCTGTGCTAGATCAAAAGG CA	Ref. 262
<i>XIST</i>	GGTCTGTGCTAGATCAAAAGG CA	AGCTCCTCGGACAGCTGTAA	Ref. 263
<i>RPPH1</i>	GAGCTGAGTGCGTCCTGTC	TCAGGGAGAGCCCTGTTAGG	Ref. 264
<i>HDGF</i>	AACAACCCTACTGTCAAGGCT	TCTTCAACGCTCCTTTCTCGT	Ref. 249
<i>CCNI</i>	AAGTGAATGTGCGGAAAATGC	GGCCAGCCATTGAATTACTTC AT	Ref. 249
<i>KIAA0100</i>	GGTGGGCATTAGCAGTCGG	ATACCACGCTTGTGTTCTCCA	Ref. 249

## **Chapter 5: Telomeric G-loops stimulate Alternative-lengthening-of-telomere cancers**

### **5.1 Summary**

10-15% of all human cancer cells employ a telomerase-independent recombination-based telomere maintenance method, known as Alternative-lengthening-of-telomere (ALT), of which the full mechanism remains incompletely understood. Here, we examine the connections between ALT and two non-canonical nucleic acid structures, G-quadruplexes (G4s) and DNA/RNA hybrids (R-loops). We observe higher levels of telomeric G4s (TelG4s) and R-loops (TelR-loops) in ALT-positive (ALT+) compared to telomerase-positive (TERT+) cells, which are broadly associated with telomere length and accompanied by increases in DNA damage signals. ALT TelG4s and TelR-loops are found spatially linked into unique structures, termed telomeric G-loops (TelG-loops). Genetic and pharmacological manipulations reveal that G4s and R-loops are interdependent. Additionally, stabilizations of both G4s and R-loops in ALT+ cells result in functional cooperation in enhancing ALT activity at sub-lethal doses, with higher doses leading to the synergism of cytotoxicity. We present these nucleic acid structures as targetable ALT cancer hallmarks with potential applications in novel biomarkers and therapeutics.

### **5.2 Introduction**

Telomeres are essential protective structures that cap the ends of chromosomes to maintain the termini's integrity and stability. All cancers require a functional telomere maintenance mechanism to sustain replicative immortality (279). While the majority (~85%) of cancers utilize a telomerase-dependent (TERT+) method of telomere maintenance, a subset (~10-

15%) of cancers adopt telomerase-independent alternative-lengthening of telomeres (ALT or ALT+) mechanisms. Some key features of ALT+ cells include long mean telomere lengths, a higher number of degenerate variant repeats, and elevated levels of ALT-associated Promyelocytic leukemia (PML) bodies and extrachromosomal telomeric DNA (especially c-circles, single-stranded extrachromosomal C-rich telomeric circular DNA) (279-280). Clinically, the prevalence of ALT in several types of cancers, including pancreatic neuroendocrine tumor (281-282) and soft tissue sarcoma (283), has been associated with unfavourable prognosis and increased mortality. Despite being discovered more than two decades ago, the mechanisms behind ALT are still not fully understood, necessitating the need to further study the biology of ALT in hopes of spurring therapeutic development.

It is important to note that the telomeric chromatin environments are different between ALT+ and TERT+ cancer cells. Among the differences, a reduction in the compaction state of ALT telomeres is thought to facilitate the use of recombination mechanisms (284-285). While multiple factors may be involved in establishing the unique state of ALT telomeres, two factors are the most well-characterized. First, over 80% of all ALT cancers are observed with genetic dysfunction of the Alpha Thalassemia/Mental Retardation Syndrome X-linked Chromatin Remodeler (ATRX) and its partner Death Domain Associated Protein (DAXX), components of the chromatin remodelling complex functioning in histone H3.3 deposition and heterochromatin maintenance (286-287). Second, ALT+ cancers carry higher numbers of degenerate variant telomeric repeats that cannot associate properly with shelterin, the six-membered telomeric DNA binding complex, and instead these variant repeats recognized by orphan nuclear receptors that promote chromatin decompaction (288-290). Telomere elongation in ALT undergoes a type of homology-directed repair (HDR) process related to that of the break-induced replication (BIR)

pathway, specifically called break-induced telomere synthesis (BITS) (280). Two different BITS pathways have been identified, including the slower RAD51/HOP2-MND1-mediated long-range homology search mechanism (291) and the faster PCNA DNA damage sensor-mediated mechanism (292). Additionally, replication defects at an ALT telomere can trigger a RAD52-dependent mitotic DNA synthesis (MiDAS) at the telomere (293). However, all these ALT pathways share a common prerequisite in that initiation of the recombination process requires some form of persistent DNA damage, yet the full mechanism of the induction of this damage signal is unclear.

Given the "open" telomere environment in ALT, we speculate that there is a higher chance of forming non-canonical structures, and these structures may be the source of this persistent DNA damage. Recently, telomeric R-loops (TelR-loops), DNA/RNA hybrid structures resulting from the association of Telomeric repeat-containing RNA (TERRA), telomeric RNA molecules transcribed from telomeres, with telomeric DNA, have been reported in ALT (284, 294, 295). It is an intriguing possibility that R-loop formation displaces the G-rich strand, thus increasing the opportunity to form another non-canonical nucleic acid structure of interest, G-quadruplexes (G4s). G4s are formed by Hoogsteen bonding of adjacent guanine bases to form planar units, G-quartets, which then stack on each other through  $\pi$ -stacking to form the overall columnar structure. While potential G4 forming sequences (pG4s) are widespread in the genome (105) and transcriptome (296), telomeric regions hold the largest pG4 reservoir due to their long consecutive G-rich repeats. We speculate that this is especially true in the case of ALT<sup>+</sup> cells due to their long telomere tracks in addition to their more permissive telomeric chromatin environment. Whether G4s and R-loops are commonly found at higher levels at ALT telomeres is unknown due to the lack of direct comparisons between multiple ALT<sup>+</sup> and TERT<sup>+</sup> cell lines.

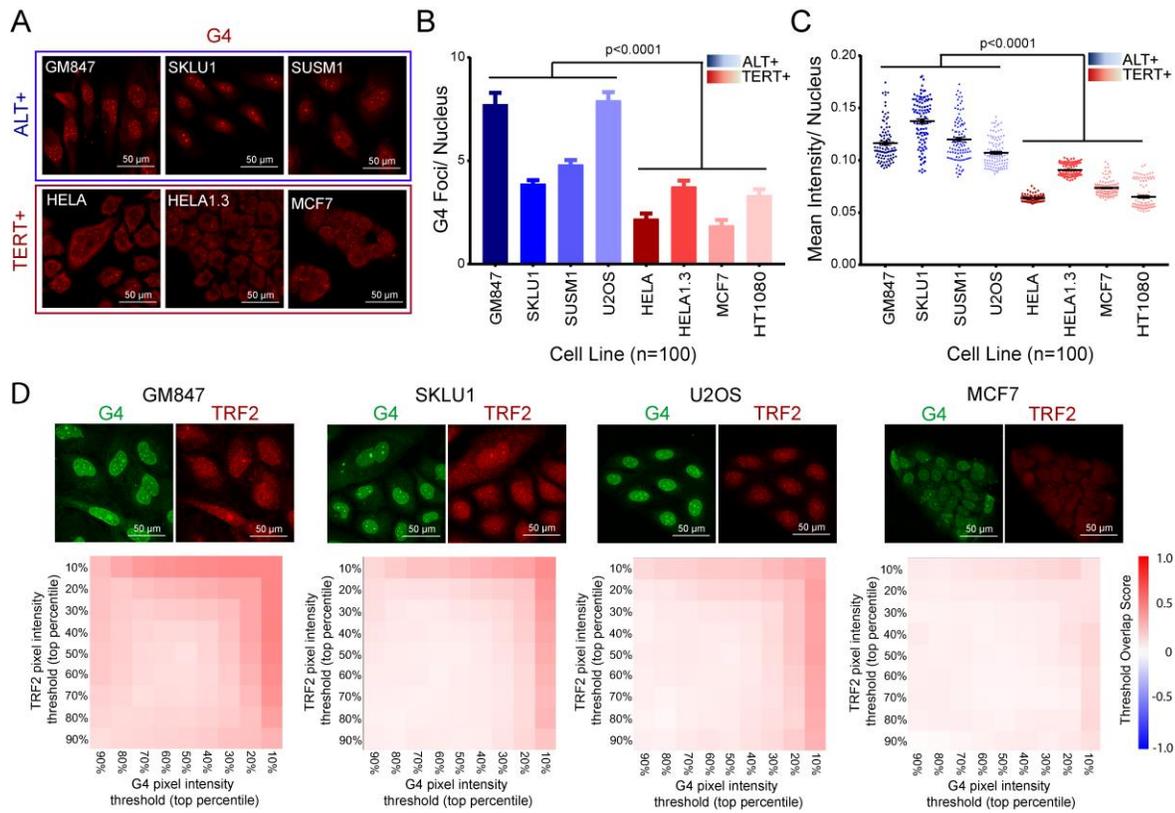
To expand our understanding of the biology of ALT, we set to examine the relationship between ALT and the two non-canonical nucleic acid structures of interest, G4s and R-loops.

### **5.3 Results**

#### **5.3.1 G4 and R-loop levels are elevated in ALT+ cells compared to TERT+ cells**

The ALT status of eight different cell lines was confirmed by measuring the mean telomere length with two established assays, telomere restriction fragment analysis and quantitative PCR and examining the levels of PML bodies by immunocytochemistry (ICC) (Supplementary Figure 1). These analyses classified four ALT+ (i.e. exhibiting longer telomeres and more PML bodies) and four TERT+ cell lines for further experiments. We next performed ICC staining of a panel of confirmed ALT+ and TERT+ cell lines using the G4-specific antibody, BG4 (37). While we observed punctate foci within cellular nuclei, there were qualitative and quantitative differences between ALT+ and TERT+ cells (Fig. 1A). Notably, with an automated image processing software, we determined that a higher number of foci and an increase in foci intensity per nucleus were associated with ALT+ cells compared to TERT+ cells (Fig. 1B-C; Supplementary Figure 2). In parallel, we investigated R-loop distribution in a panel of ALT+ and TERT+ cell lines by performing ICC staining using an R-loop-specific antibody, S9.6. Similar to our observations with G4s, the number of R-loop foci was also markedly higher in ALT+ cells than TERT+ cells (Supplementary Figures 3-4); however, foci intensity remained similar between ALT+ and TERT+ cells. We observed that these G4 and R-loop foci patterns appeared to be very similar to large telomere foci widely seen in ALT+ cells in multiple previous studies (297-299). Therefore, we reasoned that the observed increases in foci number and intensity could represent clusters of telomeric G4s and/or stretches of telomeric R-loops, a function of the long ALT

telomere tracks. To that end, we performed co-staining ICC experiments for G4 or R-loop with an antibody against Telomeric Repeat Binding Factor 2 (TRF2), a member of the shelterin telomere-binding protein complex. We found a high degree of nuclear colocalization, as evidenced by the increase in threshold overlap score (TOS) when shifting the threshold toward the top 10 percentile of pixels by intensity (Fig. 1D).



**Figure 5.1 ALT+ cells harbour higher level of telomeric G4s than TERT+ cells.**

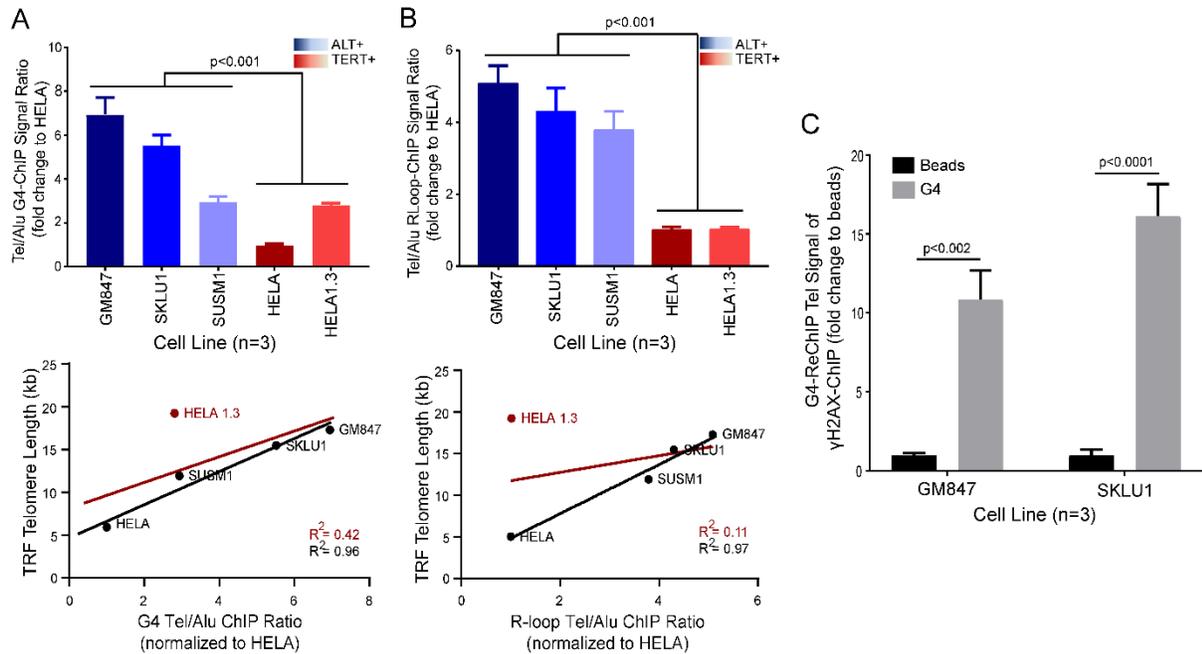
(A) Immunocytochemistry (ICC) straining experiments for G4 in ALT+ cells lines (within blue frame) versus TERT+ (within red frame) (B) Average number of G4 foci per nucleus. Automated quantification of 100 cells per cell line was performed in parallel. (C) Mean G4 foci intensity per nucleus as quantified. (D) ICC co-staining for G4 (green) and TRF2 (red) for ALT+ cell lines (GM847, SKLU1, U2OS) and TERT+ cell line (MCF7). Below is the metric matrix of the threshold overlap score (TOS) for the top percentile of pixel (by intensity) of nuclear G4 signal (x-axis) and TRF2 signal (y-axis). Darker red colour corresponds with higher TOS.

### **5.3.2 Telomeric G4 and R-loop levels are correlated with telomere length in ALT+ cells and associated with DNA damage signals**

While ICC provides useful visualization of the targeted molecules, it cannot differentiate between genomic and telomeric sites. For more robust quantification against telomeric G4s and R-loops, we performed chromatin-immunoprecipitation coupled with real-time quantitative polymerase chain reaction (ChIP-qPCR) assay to quantify for telomeric DNA. We used Alu repeats as a genome background control between different cell lines and calculated Tel/Alu ratios to measure the level of telomeric localization. Both G4-ChIP and R-loop-ChIP showed higher Tel/Alu ChIP ratios in ALT+ compared to TERT+ cell lines. Additionally, we separately plotted the Tel/Alu ratios derived from G4-ChIP or R-loop-ChIP experiments against each cell line's corresponding telomere length. Linear regression plot of G4-ChIP signals against telomere length showed a strong positive correlation between the two variables ( $R^2=0.96$ ) (Fig. 2A-B). Likewise, R-loop-ChIP displayed a similar pattern ( $R^2=0.97$ ). We repeated the linear regression analyses with the addition of data from HELA1.3, a TERT+ cell line with telomere length comparable to those observed in ALT+ cell lines. Notably, HELA1.3 was a clear outlier in the linear regression plots (Fig. 2A-B, red line). Accordingly, after the addition of HELA1.3 to the analyses, the coefficients of determination ( $R^2$ ) were significantly decreased for both TelG4 ( $R^2=0.42$ ) and TelR-loop ( $R^2=0.11$ ). We concluded that while TelG4 and TelR-loop levels are highly correlative with telomere length, a longer telomere length does not fully explain the magnitude of TelG4 and TelR-loop signals in ALT+ cells.

ALT+ cells harbour elevated basal levels of DNA damage signals (286), and these damage signals may present initiation signals for recombination-based repair. We speculated that the increases in TelG4s and TelR-loops could lead to increases in DNA damage signals. We first

performed ChIP against the DNA damage response (DDR) marker,  $\gamma$ H2AX, and confirmed higher telomere-enriched signals of DNA damage in ALT+ cells compared to TERT+ cells, as well as a strong positive correlation with telomere length ( $R^2=0.91$ ) (Supplementary Figures 5). To demonstrate that the increases in telomeric DNA damage signals are closely associated with G4 formation, we performed the ChIP-ReChIP assay, a sequential ChIP protocol using  $\gamma$ H2AX antibody in the first ChIP followed by Re-ChIP of the eluate with the BG4 antibody in the second ChIP reaction, and subsequent quantification using qPCR (Fig. 2C). The antibody order was chosen based on ReChIP compatibility screening of the antibodies in which BG4 was found to be more tolerant of the rigorous ReChIP procedure and thus applied for the second ChIP. ReChIP enrichment of telomeric DNA signals over beads-only control in both ALT+ cell lines, GM847 and SKLU1, indicated close proximity (within approximately 500bp, the median ChIP sonicated chromatin length) between telomeric G4s and DNA damage signals.

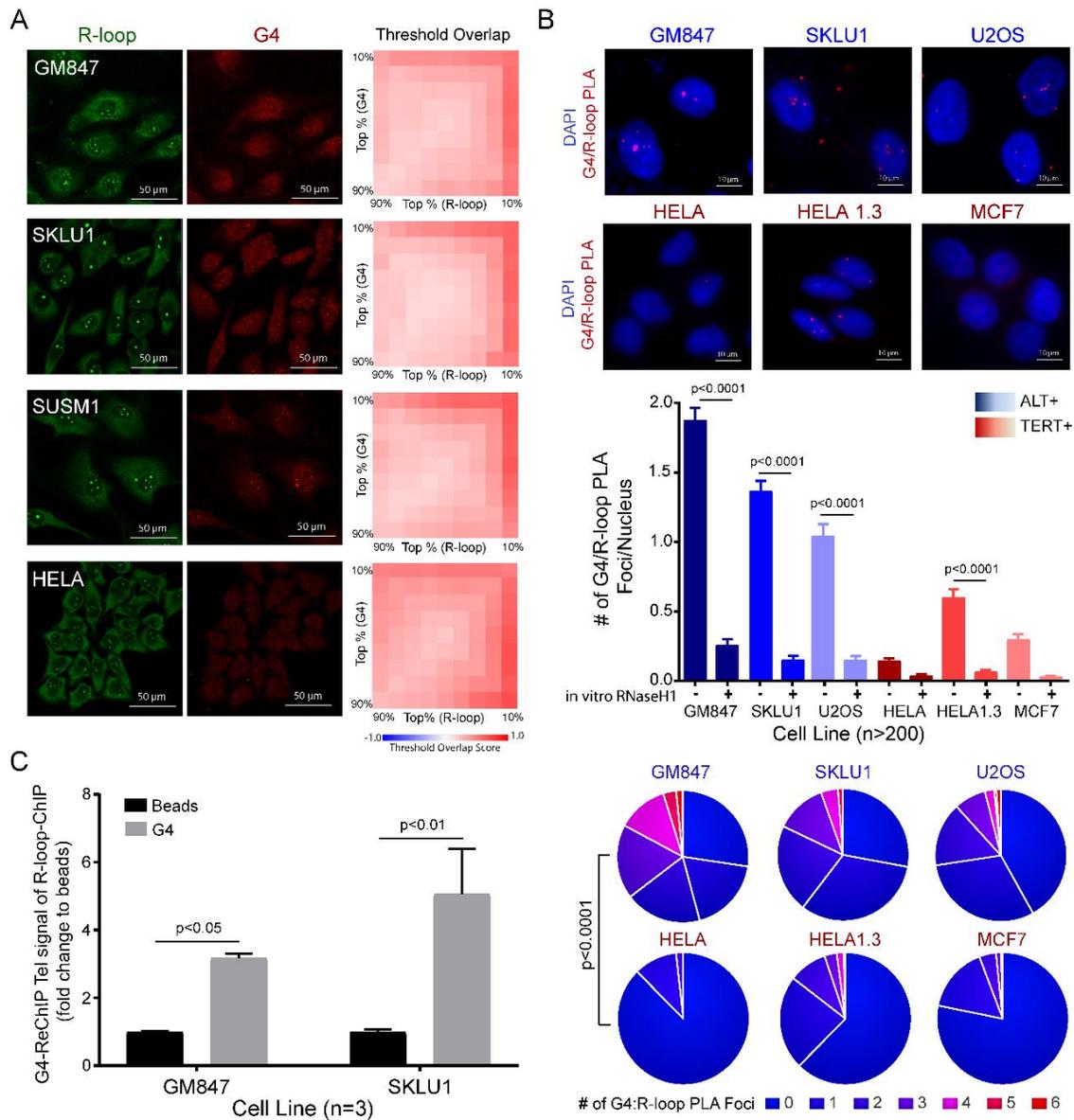


**Figure 5.2 Levels of telomeric G4s and R-loops are higher in ALT+ cells and are broadly associated with telomere length and coupled with DNA damage signals.**

(A-B) Top; (A) G4-ChIP assay or (B) R-loop-ChIP assay in ALT+ cell lines (blue shades) versus TERT+ cell lines (red shades). Data shown are the means from 3 biological repeat experiments. T-test (two-tailed) was performed for the combined ALT+ versus TERT+ datasets. Bottom; linear regression plot of Tel/Alu ChIP signal ratio against telomere length (as measured by telomere restriction fragment (TRF) analysis). Line of best fit is plotted along with R<sup>2</sup> values with HELA 1.3 (red) or without HELA 1.3 (black). (C) ChIP-Re-ChIP (1<sup>st</sup> ChIP for  $\gamma$ H2AX followed by 2<sup>nd</sup> ChIP for G4) and quantified for Tel DNA in ALT+ cells, GM847 and SKLU1. ReChIP signal for each sample was then normalized to its respective beads only control. Two-way ANOVA with Sidak multiple correction was performed. Data shown are the means from 3 biological repeat experiments. Error bars represent the SEM.

### 5.3.3 Telomeric G4s and R-loops are spatially linked in ALT+ cells

Given the elevated levels of both TelG4s and TelR-loops in ALT+ cells and the mechanistic plausibility of their simultaneous formation, we examined whether these two structures are linked to each other spatially. We performed ICC co-staining experiments for G4 and R-loop and observed highly colocalized nuclear staining, as illustrated by the TOS increase when examining the top 10 percentile of pixels by intensity (Fig. 3A). To increase our imaging experiments' resolution, we employed the proximity ligation assay (PLA) with BG4 and S9.6 antibodies to probe for colocalization of G4s and R-loops within nanometer ranges. We observed a significantly higher number of G4:R-loop PLA signals in ALT+ cells compared with TERT+ cells (Fig. 3B). As controls, these PLA signals were efficiently removed by *in vitro* RNaseH1 treatment. Finally, we performed a ChIP-ReChIP experiment with two ALT+ cell lines, GM847 and SKLU1, using the S9.6 (for R-loop) for the 1<sup>st</sup> ChIP and BG4 (for G4) antibody for 2<sup>nd</sup> ChIP, and then followed by qPCR-quantification for telomeric DNA. ChIP-ReChIP data confirmed that G4 and R-loop structures are co-enriched at the telomere (Fig. 3C). Together, our data demonstrated that TelG4 and TelR-loop are in close proximity within ALT telomeres, suggesting that their formations are tightly spatially linked.

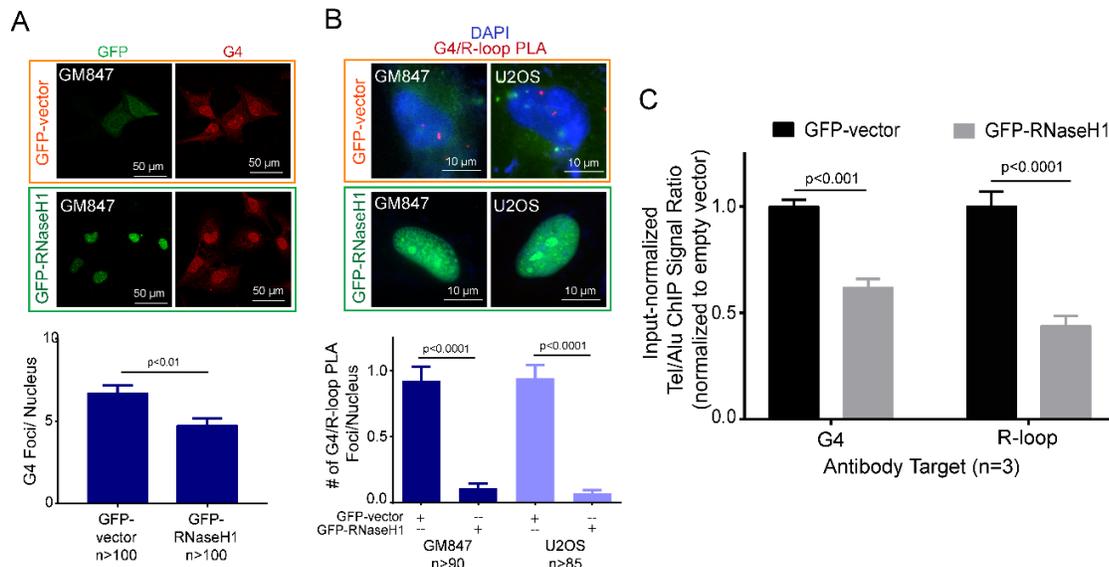


### Figure 5.3 TelG4 and TelR-loops are spatially linked.

(A) ICC co-staining for R-loop (green) and G4 (red) in ALT+ (GM847, SKLU1, SUSM1) and TERT+ (HELA). Right is the corresponding metric matrix of the threshold overlap score (TOS) for the top percentile of pixel (by intensity) of nuclear R-loop signal (x-axis) and G4 signal (y-axis). (B) Top; Proximity ligation assay for G4/R-loop colocalization in a panel of ALT+ and TERT+ cell lines. Middle; Quantification of the mean PLA foci per nucleus. 200 cells per condition were quantified. *In vitro* RNaseH1 digestion was used as a control for each cell line. One-way ANOVA with Dunnett's multiple test correction was performed comparing non-treated samples and RNaseH1-treated samples. Bottom; Pie charts depicting the distribution of PLA foci per nucleus. T-test (two-tailed) was performed for the combined set of ALT+ versus TERT+ cell lines. (C) ChIP-ReChIP (1<sup>st</sup> ChIP for R-loop followed by 2<sup>nd</sup> ChIP for G4) and quantified for Tel DNA in ALT+ cells, GM847 and SKLU1. ReChIP signal for each sample was then normalized to its respective beads only control. Two-way ANOVA with Sidak multiple correction was performed. Data shown are the means from 3 biological repeat experiments. Error bars represent the SEM.

### **5.3.4 Genetic and pharmacological manipulation reveals interdependency between G4s and R-loops**

Given the close association of G4 and R-loop at the ALT telomere, we speculated whether these two structures could affect each other's formation or stability. To study the effect of R-loop changes on G4 abundance, we targeted biological pathways responsible for R-loop resolutions. Telomeric R-loops are resolved mainly in cells by RNaseH1 (294) and FANCM(295). We first overexpressed recombinant GFP-tagged RNaseH1 in GM847 ALT+ cells to reduce the cellular levels of R-loops. ICC staining for G4s in GM847 overexpressing RNaseH1 showed a decrease in G4 levels compared to control GFP-overexpressing cells (Fig. 4A). Then, we evaluated G4:R-loop PLA in GM847 and U2OS ALT+ cells overexpressing RNaseH1 and showed significant reductions of PLA foci compared to that of the vector control (Fig. 4B). Using ChIP, we observed parallel reductions of TelG4 and TelR-loops upon overexpression with RNaseH1 (Fig. 4C). Reciprocally, we used the siRNA-mediated knockdown of FANCM to increase the cellular accumulation of R-loops. We demonstrated significant increases in G4:R-loop PLA foci in GM847 and U2OS cells treated with FANCM-siRNA compared to control-siRNA knockdown cells (Supplementary Figure 6). Taken together, we showed that both positive and negative changes in TelR-loop levels induced corresponding changes in TelG4 stability, in the same direction.

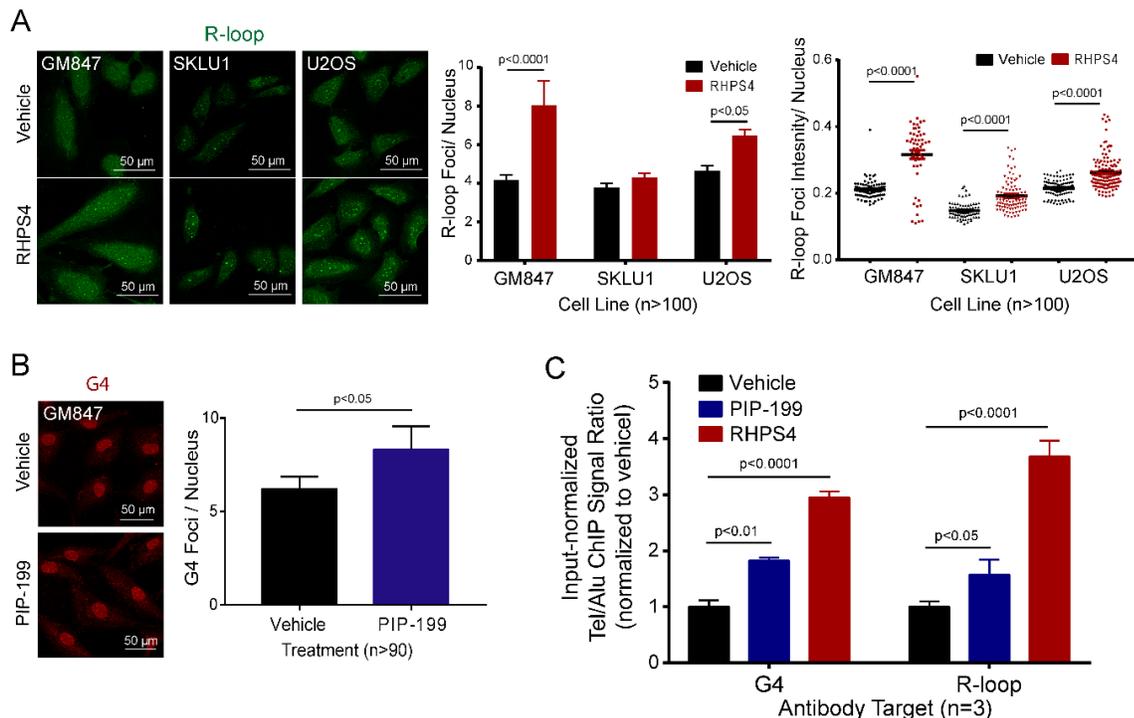


**Figure 5.4 RNaseH-mediated reduction of R-loops decreases TelG4 levels.**

(A) Top; ICC staining for GFP (green) and G4 (red) in GM847 transfected with GFP-vector control (within orange frame) or GFP-RNaseH1 (within green frame). Bottom; Quantification for mean BG4 foci count per nucleus. (B) Top; Proximity ligation assay (PLA) for G4 and R-loop (red foci) in GM847 and U2OS transfected with either GFP-vector (within orange frame) or GFP-RNaseH1 (within green frame). Bottom; Quantification for mean PLA foci count per nucleus (C) G4-ChIP or R-loop-ChIP results from GM847 transfected with either GFP-vector or GFP-RNaseH1. Tel/Alu ChIP signals were then normalized against GFP-vector control. Data shown are the mean from 3 biological repeat experiments. Two-way ANOVA with Sidak multiple comparisons was performed.

We also tested the effects of changes in G4 abundance on R-loop accumulation. Treatment with G4 ligands, small molecules that can bind and increase the stability of G4s, is a well-established method to study G4-related biological changes. In this study, we used sub-lethal doses of RHPS4, a G4 ligand that targets the telomere and causes telomere dysfunction (61). ICC staining experiments revealed a significant increase of R-loops in ALT+ cells treated with RHPS4 (Fig. 5A). In addition, we tested the effects of a newly identified small-molecule R-loop stabilizer, PIP-199, which increases R-loop accumulations through the inhibition of MM2-RMI1/2 interactions that disrupts the formation of the FANCM-BTR (BLM-TOP3-RMI1) complex (300). ICC staining experiments revealed that treatment with PIP-199 increases G4 levels in ALT+ cells (Fig. 5B). We performed parallel ChIP experiments to quantify the levels of TelG4 and

TelR-loop in the vehicle-, PIP-199- and RHPS4-treated GM847 ALT+ cells. Treatments with both chemicals showed significant and parallel increases in G4 and R-loop levels, with PIP-199 (R-loop stabilizer) displaying a parallel 1.5-2 fold increases in TelR-loop and corresponding TelG4; and RHPS4 (G4 ligand) displaying a parallel 3-4 fold increases in TelG4 and corresponding TelR-loop (Fig. 5C). Together, our ICC and ChIP experiments quantitatively showed that the two telomeric structures are interdependent.



**Figure 5.5 Small molecule-mediated inductions of TelG4s and TelR-loops show a reciprocal relationship.**

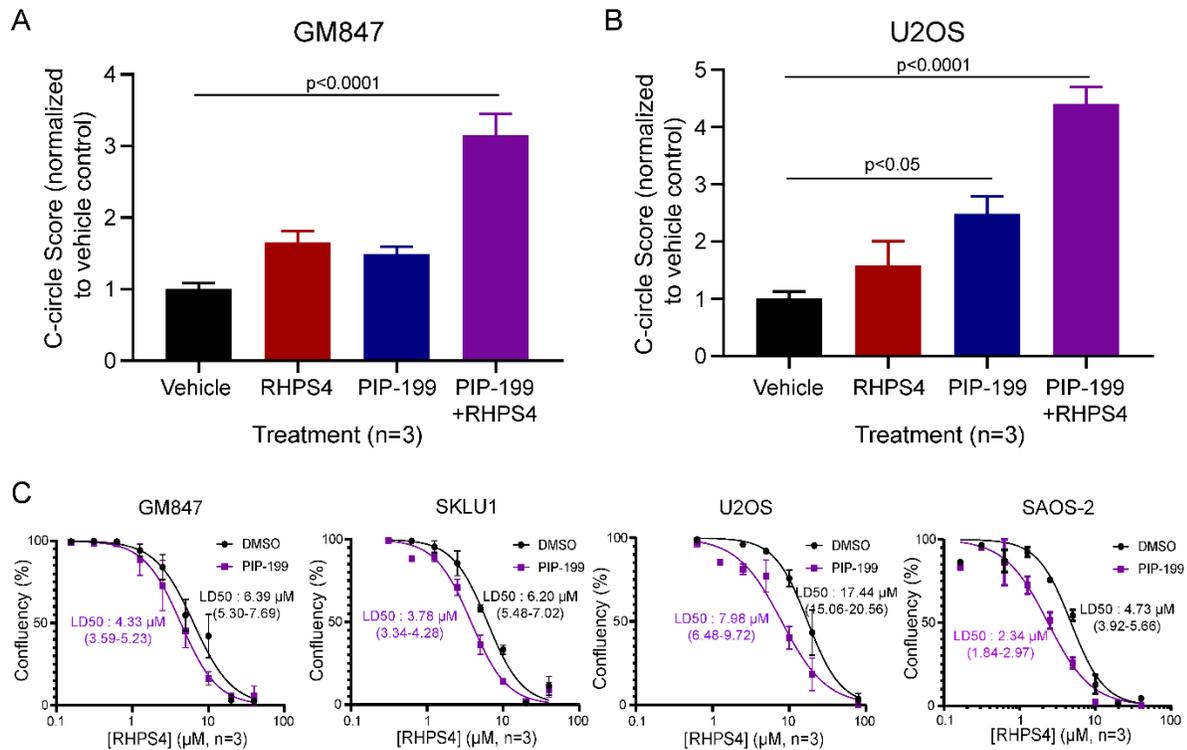
(A) Left; R-loop ICC staining (green) in ALT+ cell lines (GM847, SKLU1, U2OS) treated with vehicle or G4 ligand, RHPS4 (1  $\mu$ M, 1.5  $\mu$ M, 5  $\mu$ M respectively). Mean R-loop foci count (center) and intensity (right) per nucleus were quantified. 100 cells per treatment were quantified. Two-way ANOVA with Sidak-corrected multiple comparisons was performed against the untreated controls. (B) G4 ICC staining (red) in GM847 with vehicle or R-loop stabilizer, PIP-199 0.75  $\mu$ M. 50 cells per treatment were quantified. T-test (two-tailed) was performed. (C) G4-ChIP or R-loop-ChIP of GM847 treated with vehicle, 0.75  $\mu$ M PIP-199 or 1  $\mu$ M RHPS4. ChIP signal was first normalized to input then the Tel/Alu ratio calculated for each condition was then normalized against that of the vehicle control. Data shown are the mean from 3 biological repeat experiments. Two-way ANOVA with Sidak-corrected multiple comparisons was performed against the vehicle control.

### 5.3.5 Telomeric G4 and R-loop cooperatively stimulate ALT activity.

To test the hypothesis that TelG4 and TelR-loop cooperate in providing the initiation signals for ALT recombination, we co-treated GM847 and U2OS cells with a previously determined sub-lethal dose of RHPS4 and PIP-199 (Supplementary Figure 7). These ALT+ cells were treated for 24h prior to DNA extraction. We measured changes in c-circle levels, the most widely used and sensitive ALT activity measure, using the rolling circle amplification and qPCR method (301). The sensitivity of the C-circle Assay (CCA)-qPCR method was demonstrated by the significant increase of Tel DNA signals with the addition of phi29 ( $\phi$ ) polymerase (+Phi) to that of a mock reaction control (-Phi) in ALT+ cells (~10 fold in GM847 and ~6.5 fold in U2OS) compared to TERT+ cells (~1.4 fold in HELA 1.3) (Supplementary Figure 8). Single-agent treatments with G4 ligand or R-loop stabilizer induced a slight but reproducible increase in CCA score for GM847 ALT+ cells, while the co-treatment significantly increased the CCA score to ~3.2 fold (Fig. 6A). U2OS ALT+ cells showed a similar CCA score pattern, with a reproducible increase in telomeric c-circles following RHPS4 single-agent treatments, a higher increase after PIP-199 single-agent treatments, and the co-treatment with both agents resulted in the highest increase in CCA scores (Fig. 6B). For these data, we concluded that an increase in TelG4 and TelR-loop levels positively correlate with ALT activity, and that stabilization of the two linked structures induced ALT beyond a simple additive effect (GM847: +70% vs. +220%; U2OS: +200% vs. 350%).

### **5.3.6 Co-treatment with a G4 ligand and R-loop stabilizer leads to the synergism of cytotoxicity in ALT+ cells**

We observed compound increase in c-circle accumulation, a surrogate measure for ALT activity, when Tel-G4 and Tel-R-loop levels increased. However, over-accumulations of TelG4 and TelR-loop are expected to be cytotoxic, and we anticipated that targeting both structures at the same time could create synergism. We performed a curve-shift analysis (302) on the dose-response curve for cytotoxicity of RHPS4 in four ALT+ cell lines, with or without co-treatment with a previously determined sub-lethal dose (lethal dose (LD) of 5-15%) of PIP-199. Dose-response curves showed a leftward shift of the normalized RHPS4 cytotoxicity dose-response curve after the addition of a sub-lethal dose of PIP-199 in all four ALT+ cell lines: GM847 (LD50: 4.33 vs 6.39  $\mu$ M), SKLU1 (LD50: 3.78 vs 6.20  $\mu$ M), U2OS (LD50: 7.98 vs 17.44  $\mu$ M) and SAOS-2 (LD50: 2.34 vs 4.73  $\mu$ M) (Fig. 6C). These results suggest that the drug-drug interaction is of synergistic nature (leftward shift) rather than additive (expected minimal or no shift) or antagonistic (rightward shift) nature. Together, the observations support the view that combination treatments of G4 and R-loop stabilizing chemicals may have therapeutic potential in the clinical management of ALT+ cancers.



**Figure 5.6 Induction of G4 and R-loops cooperatively stimulate ALT activity and leads to synergism of cytotoxicity.**

(A-B) C-circle assay (CCA) followed by qPCR in (A) GM847 treated with either vehicle (black), RHPS4 1.25  $\mu$ M (red), PIP-199 0.37  $\mu$ M (blue) or cotreated with both RHPS4 and PIP-199 (purple) (B) U2OS treated with either vehicle (black), RHPS4 5  $\mu$ M (red), PIP-199 0.37  $\mu$ M (blue) or cotreated with both RHPS4 and PIP-199 (purple). C-circle score from quantifying Tel DNA ratio between the presence and absence of phi polymerase was first normalized to single copy gene then normalized to the vehicle control. Data was collected from 3 biological repeat experiments. (C) Curve-shift analysis in ALT+ cells, GM847, SKLU1, U2OS and SAOS-2. RHPS4 cytotoxicity dose response curve was generated with treatment of either DMSO vehicle control or a nontoxic dose of PIP199 (0.75  $\mu$ M, 3  $\mu$ M, 0.75  $\mu$ M, 1.5  $\mu$ M respectively) for each cell line. The LD50, dose kills half of the cell population, was calculated with 95% confidence intervals for each condition per cell line, from data collected in 3 biological repeat experiments.

## 5.4 Discussion

Persistent DNA damage is obligatory for the initiation of ALT telomere maintenance.

Here, we explore the possibility that non-canonical nucleic acid structures, G4s and R-loops, may provide the source of this damage. We observe a strikingly parallel elevation of both TelG4s and TelR-loops in ALT+ cells compared to TERT+ cells (Fig. 1). TelG4s and TelR-loops levels

are broadly associated with telomere length and are found localized in the vicinity of DNA double-strand break (DSB) damage signals at the telomere (Fig. 2). We also demonstrate that TelG4s and TelR-loops are spatially linked and positively influence each other's stability (Fig. 3-5). Additionally, we show that small molecule-mediated stabilization of both G4 and R-loop can cooperatively enhance ALT activity at sub-lethal doses and that higher doses of co-treatment led to synergistic cytotoxicity (Fig. 6). Together, these observations support the formation of special structures at ALT telomeres, comprising co-existing G4s and R-loops on opposing strands, referred to here as "G-loops" (303), with a role in stimulating ALT recombination (Fig. 7). We present these non-canonical nucleic acid structures as potential ALT biomarkers and therapeutic targets.

Our study here provides key missing links, which helps broaden the perspective on the full ALT mechanism. The unique ALT telomere environment is established by multiple factors, including dysfunction of ATRX/DAXX and increased number of degenerate telomere repeats (286-290). Recently, TelR-loops have also been implicated in ALT (284, 294, 295). While we have confirmed that an elevated level of TelR-loops is a general feature of ALT, we have also observed parallel elevation of TelG4 levels in ALT+ cell lines. Recent studies also suggest that ATRX may be involved in suppressing G4s (304) and R-loops (305); therefore, the dysfunction of ATRX may be one of the selected ways in which ALT+ cells upregulate these non-canonical nucleic acid structures. Based on our observation of the strong interaction between G4 and R-loop, we contend that G-loops are more effective than either one of the two structures on their own to promote the contextual environment required for ALT maintenance by keeping a larger and more stable opened chromatin conformation (Fig. 7). Given that the average size of TERRA ranges from 100 to 9000nt (306) (which can be longer in ALT+ cells), it is thus likely that our

mean ChIP chromatin size of 500bp contains a single continuous R-loop and clusters of G4 on the opposite strand. G4 clusters within the G-loops would likely be structurally different due to the length of single-stranded telomere G-rich sequence available for quartet formation and  $\pi$  stacking. The long telomeric G-rich sequence displaced by R-loop could be assembled into a higher-order structure, analogous to the "stacked beads-on-string" model, in which "bead" (or monomers) of individual G4s can stack on each other to form a multimeric structure (307). Furthermore, using the c-circle level as a readout of ALT activity, we observed significantly greater ALT activity in the presence of both G4 and R-loop structures than either one alone (Fig. 6). The coexistence of both structures may reduce the energy threshold required to keep an "open" and more favourable chromatin state for the formation of spontaneous DSB DNA damage, perhaps through the interaction with ALT-associated and/or G4-associated endonucleases, such as MUS81 (308) and XPF (309) (Fig. 7). However, we expect that TelG4s and TelR-loops are formed only transiently and highly regulated by protein complexes composed of helicases to prevent genome instability. In this context, the FANCM-BTR complex (300), a key regulator of ALT homeostasis, may have evolved to process G4s (by BLM in the BTR complex) (122) and R-loops (by FANCM) (295) simultaneously within G-loops. There is also an intriguing possibility that large G-loops can promote enhanced accessibility, facilitating telomere clustering in the intertelomeric recombination ALT pathway (291).

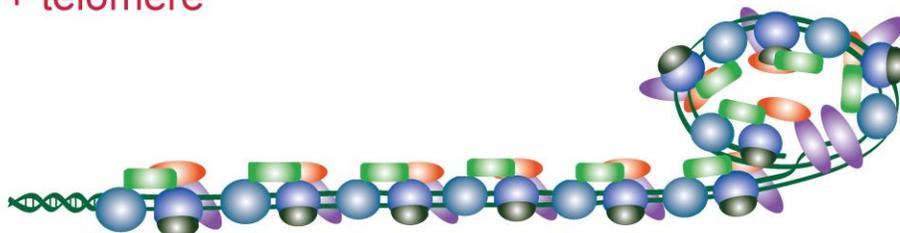
We have shown here that higher levels of TelG4s, TelR-loops and their combined structures of TelG-loops are general features of ALT and, therefore, may be viable candidates as novel biomarkers and therapeutic targets for ALT cancers. Currently, one of the most significant predictors of ALT is the dysfunction of ATRX/DAXX; however, some ALT+ tumours (~20%) do not harbour defective ATRX/DAXX and have normal DNA repair capabilities (286). One

such example in our study is the SKLU1 cell line, derived from lung epithelial carcinoma, which has functional ATRX/DAXX, yet displayed higher levels of TelG-loops. Thus, these special nucleic acid structures could potentially present more specific biomarkers for ALT+ cancers. Current molecular diagnostics for testing ALT include TRF southern blotting and c-circle assays, both of which are relatively labour-intensive and/or require large amount of tumour materials. Future studies to comprehensively evaluate different ALT biomarkers should expand to include G4s and G-loops as diagnostic and/or prognostic markers. Based on our cytotoxicity screens using a combination of G4 ligand and R-loop inducer, we propose that TelG-loops may also be valuable therapeutic targets for ALT+ cancers (Fig. 5). As co-stabilization of G4 and R-loop leads to significantly enhanced ALT activity (Fig. 6A-B), the observed cytotoxicity is likely due to hyper-ALT phenotypes, manifested in excessive increases in ALT features and accumulation of toxic recombination intermediates. Here, we provide a proof-of-principle for exploiting these non-canonical nucleic acid structures in future clinical applications for ALT+ cancers.

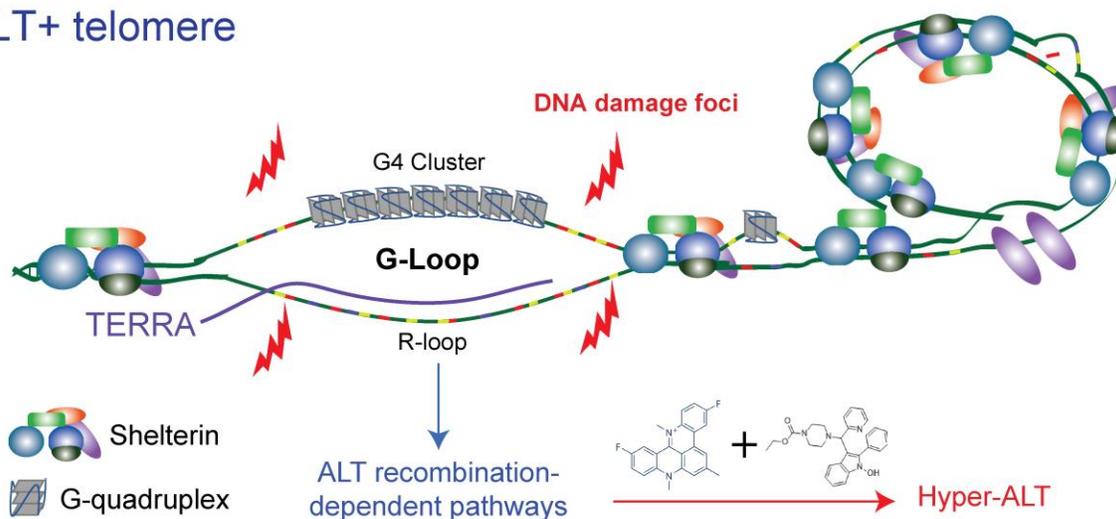
While the connection between G4s and R-loops are studied here within the context of ALT telomeres, our findings can be extended to G-rich regions in the rest of the genome and impact other important homeostatic processes in human biology. Biological processes, including transcription, replication, and chromatin structure remodelling, are potentially conducive to transient G-loops formation. Our findings thus provide incentives to study similar phenomena in other biological pathways and inform future studies investigating the functional genomics of these nucleic acid structures to consider the potential confounding formation of the other counterpart. In summary, we have shown that elevated TelG4s and TelR-loops are ALT-specific hallmarks and that these G4s and R-loops are interdependent and can co-exist in G-loops, which plays a role in stimulating ALT recombination. Co-targeting TelG4s and TelR-loop can also

synergistically induce cytotoxicity in ALT+ cells. Our study provides new insights into the ALT biology and G4/R-loop interplay as well as evidence supporting the future development of novel biomarkers and therapeutics against ALT+ cancers.

## TERT+ telomere



## ALT+ telomere



**Figure 5.7 Proposed model of telomeric G-loop formation in ALT.**

At the TERT+ telomere, there are proper shelterin binding and proper telomere structure. Telomere capping function is intact. At the ALT+ telomere, chromatin remodeler dysfunction (i.e. ATRX), long telomere tract with degenerate repeats, reduced shelterin binding, and other ALT-promoting events contribute to a less compacted chromatin and more dysfunctional telomere structure. This provides the opportunity for the formation of G4s and/or R-loops. TERRA (purple) associate with the C-rich telomere strand to form R-loop. The other displaced strand harbours long G-rich regions which associate with each other to form a G4 cluster. The cooperative actions of G4 and R-loop leads to further opening with increased stability of the R-loop and G4 cluster. The actions of endonuclease at the opening of G-loop leads to double-strand break which facilitates the initiation of ALT-associated recombination-based repair mechanisms. Stabilization of G-loops through treatments with G4-ligands and R-loop inducers drive the cancer cell to a hyper-ALT state.

## 5.5 Methods

### Cell line and culture

Cell lines (SKLU1, U2OS, HELA, MCF7, HT1080) were obtained from American Type Culture Collection (ATCC). GM847 was obtained from Coriell Institute cell repository. SUSM1 and HELA1.3 were obtained from Collins Lab (UC Berkeley, USA). The cells were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 5% synthetic fetal bovine serum (FetalClone III; GE Life Sciences) and 1% penicillin-streptomycin mixture (Gibco). Incubation was done at 37°C in a humidified, 5% CO<sub>2</sub> atmosphere-controlled incubator (HERAcell). All cell culturing was performed using standard procedures including aspiration and washing with phosphate-buffered saline (PBS, Gibco), trypsinizing using Trypsin-EDTA (Gibco). Coulter Counter (Beckman Coulter) was used for cell counting.

### Immunocytochemistry (ICC) Staining

Cell lines were seeded on coverslips in a 24-well plate at density of 5000-15000 cells per well. In experiments with treatments, the cells were left to grow for 24h in the incubator before dosing with a predetermined dose. The cells were then incubated for another 48h. The coverslips were then rinsed in PBS. For fixation, 100% ice cold methanol was added to each well for 10 min. Then, the coverslips were washed twice with PBS before storage in PBS-azide. For staining, the fixed cells on coverslips were blocked in blocking solution (4% BSA/TBS) for 1h. Coverslips were incubated with 2 µg/mL BG4 in blocking solution at room temperature for 3h. The coverslips were then washed 3 times with PBS-T (0.02% Tween) for 5 min each. Next, the primary antibody was added. For BG4, anti-FLAG antibody was added at 2 µg/mL and incubated at 4°C overnight. For any other primary antibodies used, 2 µg/mL of antibody was

added and incubated at 4°C overnight. The coverslips were then washed three times with PBS-T for 10 min each. Appropriate secondary antibodies were added at 1 µg/mL and incubated at room temperature for 2h. The coverslips were again washed three times with PBS-T for 10 min each, counterstained in DAPI and mounted using Fluoromount G (Southern Biotech). A detailed list of antibodies used in this study are listed (Supplementary Table 1). Images were taken using Zeiss LSM700 confocal microscope. The raw images were then processed and quantified using CellProfiler (310) using custom pipeline. First, the DAPI signal was used to generate masks to identify the nuclei as the regions of interest. Foci, either red or green, passing the intensity and size threshold were then counted within these regions. The number of foci count in each nucleus was exported out for further analysis. For the overlap analysis, EZColocalization (311), a plugin in ImageJ, was used to generate a metric matrix for threshold overlap score (TOS). The DAPI signals were used to create masks for nuclear regions. Different threshold of the top percentile (in 10% increments) based on pixel intensity were used to generate the matrix. A score of 1 (red colour) represent colocalization whereas a score of -1 (blue colour) represent complete anti-colocalization.

#### Chromatin-immunoprecipitation (ChIP) assay

Cells were seeded at a density of 250000-750000 (depending on the cell line) cells per 10-cm dish in DMEM containing 5% FBS and 1% Pen/Strep. Treatments with ligands were performed 24h after seeding and incubated for 48h. In experiments without ligand treatments, the cells were incubated for 72h after seeding. Cells were washed once with PBS and cross-linked in 1% formaldehyde/fixation buffer (50 mM HEPES KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA pH8) for 5 min at room temperature on rocking platform. 0.125 M glycine was added to each dish and incubated for 5 min to quench crosslinking reaction. The fixing solution

was removed to waste. Cells were quickly rinsed with PBS and scraped into a 1 mL Eppendorf tube. Tubes were spun at 3k rpm for 3 min then resuspended in shearing buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris pH 8) supplemented with 1XPIC. Resuspended cells were then sonicated using Covaris m220 at 10% duty and chromatin shearing preset setting for 4 min. Sonicate was then spun down at 13200 rpm for 10 min at 4°C and the supernatant is moved to a new tube. The sonicate was then precleared using non-blocked agarose beads. Sonicate was then split into two fractions for immunoprecipitation and diluted 1:1 with 2X IP Buffer (0.1% SDS, 1 mM DTA, 10 mM Tris pH 8.1, 2% Triton-X, 300 mM NaCl) and 0.1% BSA was added to each IP. A small aliquot of sonicate was saved for the input. 5 µg of antibody was added to each ChIP tube and no antibody was added to the control tube. The IP was performed overnight at 4°C on rotator. On the following day, 25 µL of 50% slurry of pre-blocked agarose beads (M2 anti-FLAG (Sigma, A2220) for G4-ChIP or Protein G Sepharose beads (GE Life Sciences, 17061802) for other ChIPs) and incubated at 4°C on rotator for 2h. Beads were centrifuged at 5000 rpm and washed sequentially twice for each buffer for 5 min each: ChIP Wash Buffer A (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl), Wash Buffer B (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 500 mM NaCl, Wash Buffer C (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8) and TE (10 mM Tris-HCl pH 8, 1 mM EDTA). After final wash, beads were resuspended in 1% SDS/TE buffer and incubated for 30 min at room temperature on rotator before incubating overnight at 65°C. Next day, the beads were spun down, and the supernatant was collected for DNA purification using DNA purification kit (BioBasics) using the manufacturer recommended protocol. The ChIP DNA was eluted in 35 µL elution buffer.



Each ChIP sample was then quantified by running three technical repeats. Tel/Alu ChIP signal ratio was then computed, and the ratio of each cell line was then normalized to that of HELA. For comparisons between conditions, the ChIP signal was first normalized to the input then the Tel/Alu ratios were calculated from the normalized values. The input-normalized Tel/Alu ratios were then normalized against the appropriate controls (either GFP-empty vector control or non-treated control).

#### Transfection for ICC and ChIP

Cells were transfected using JetPrime transfection reagent (Polyplus). Cells were seeded 20000-60000 cells per coverslips in 24-well plate and incubated for 24h at 37°C. Next day, the cells were transfected with 250 ng of vector DNA using ratio according to the manufacturer recommended protocol. Cells were then incubated for 6h at 37°C. Then, the media were replaced with fresh media and incubate for another 24h. Cells were then harvested for ICC or ChIP as mentioned above.

#### Proximity Ligation Assay (PLA)

Cells were seeded on coverslips in 12-well plate at a density of  $1 \times 10^5$  cells per well. One day post-seeding, cells were washed with PBS, fixed with ice-cold methanol for 10 min, and permeabilized with ice-cold acetone for 1 min. After blocking in 3% BSA, 0.1% Tween-2- in 4XSSC for 1h at room temperature, cells were then incubated with or without 2  $\mu$ g/mL BG4 in blocking solution for 3h at room temperature on a rocker. Following 3X PBST wash for 5 minutes, cells were incubated with primary antibodies [1:250 rabbit anti-S9.6 antibody (Ab01137-23.0, Absolute Antibody) and 1:500 mouse anti-Flag-M2 antibody (F1804, Sigma)] at 4°C overnight. The next day, subsequent steps in proximity ligation assay were carried out with

Duolink In Situ Kit (Sigma) in accordance with the manufacturer's instructions. Briefly, after washing twice with PBS and once with PLA wash buffer A for 5 min each, cells were incubated with pre-mixed PLA probe anti-mouse minus and PLA probe anti-rabbit plus (Sigma) for 1h at 37°C. After 5 min wash with Buffer A 3 times, ligation reaction was carried out for 30 minutes at 37°C, followed by additional wash with Buffer A for 2 min 2 times, and amplification reaction for 100 minutes at 37°C. After washing with Buffer B for 10 min 3 times and 0.01% Buffer B for 1 min, cells were stained with DAPI, mounted and imaged on LeicaDM18 microscope at 100X. Negative controls were included in each experiment, treated identically but either without BG4 antibody incubation for 3h or without anti-S9.6 antibody overnight.

#### Transfection for PLA

For experiments with *in vitro* RNaseH1 treatment, RNaseH1 treatment was done after permeabilization for 2h at 37°C before blocking. For experiments with GFP or GFP-RNH1 overexpression, transfection was done 24h post seeding, and cells were fixed 24h post-transfection. For RNA interference, cells were transfected with siGENOME-SMARTpool siRNAs from Dharmacon: Non-targeting siRNA Pool #1 as si-Control and si-FANCM.

Transfections were done with Lipofectamine RNAiMAX transfection reagent (Invitrogen) according to manufacturer's protocol and cells were fixed 48h after the siRNA administration.

#### C-circle Assay

C-circle assay (CCA) coupled with qPCR quantification was performed using a modified version of a published protocol (301). 75000 cells per well in 12-well plate were seeded and grown for 24h. Cells were then treated with appropriate ligands and/or vehicle controls and incubated for another 24h. Cells were then harvested by scrapping and centrifugation. The cell pellets were

then incubated at 56°C in 50 µL of Quick C-circle Preparation (QCP) buffer (50 mM KCl, 10 mM Tris-HCl, 2 mM MgCl<sub>2</sub>, 0.5% IGEPAL CA-630 detergent, 0.5% Tween-20) supplemented with fresh 1 mg/mL Proteinase K (Sigma) and vortex intermittently for 1h. DNA from the lysate was then purified using column purification (BioBasic). The purified DNA is then diluted 10x then 10 µL of the diluted material is added to 9.25 µL 2.16X CCA buffer (final reaction concentration of 4 mM DTT, 4 µg/mL BSA, 0.1% Tween-20, 1 mM dATP, 1 mM dGTP, 1 mM dTTP, 1 mM dCTP, 1X phi29 Buffer) and 0.75 µL phi29 polymerase (NEB). Two reaction tubes per sample was made: one with phi29 polymerase added (Phi+) and one without (Phi-). Samples were incubated at 30°C for 8h then 70°C for 20 min. Real-time qPCR was performed using the Tel1b/Tel2b primer pair. Calculate the ratio of Phi+/Phi- signal for the raw CCA signal. Real-time qPCR was performed on a single copy gene, *ACTB* with the primer pair (forward: CTGGAACGGTGAAGGTGACA, reverse: AAGGGACTTCCTGTAACAACGCA), and used quantified for input differences. The raw CCA signal is normalized between samples using *ACTB*.

### Cytotoxicity Profiling

Cytotoxicity screening of ligands were performed as previously described (274). Cells were seeded at 3000-6000 cells per well in 96-well plate and were grown for 24h. Cells were then treated with varying doses of ligands. The plate was then inserted in the Incucyte ZOOM Live-cell Imaging System (Essen Bioscience) and imaged at 3h interval for 72h. Phase confluency was measured for each well over the period and exported into numerical values. The confluency values were then normalized per experiment. Dose response curves were then plotted using GraphPad Prism.

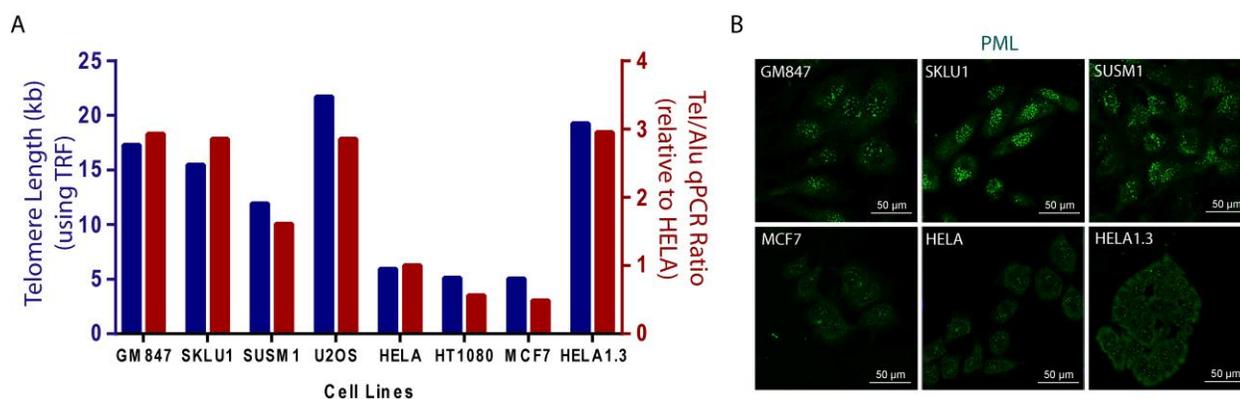
## Statistical Analysis

Statistical analysis of the comparison between ALT+ and TERT+ cell lines were performed using student's t-test on aggregated datasets from multiple cell lines. One-way or two-way ANOVA with multiple-test corrections was performed for comparison between conditions. All statistical analyses were performed using GraphPad Prism.

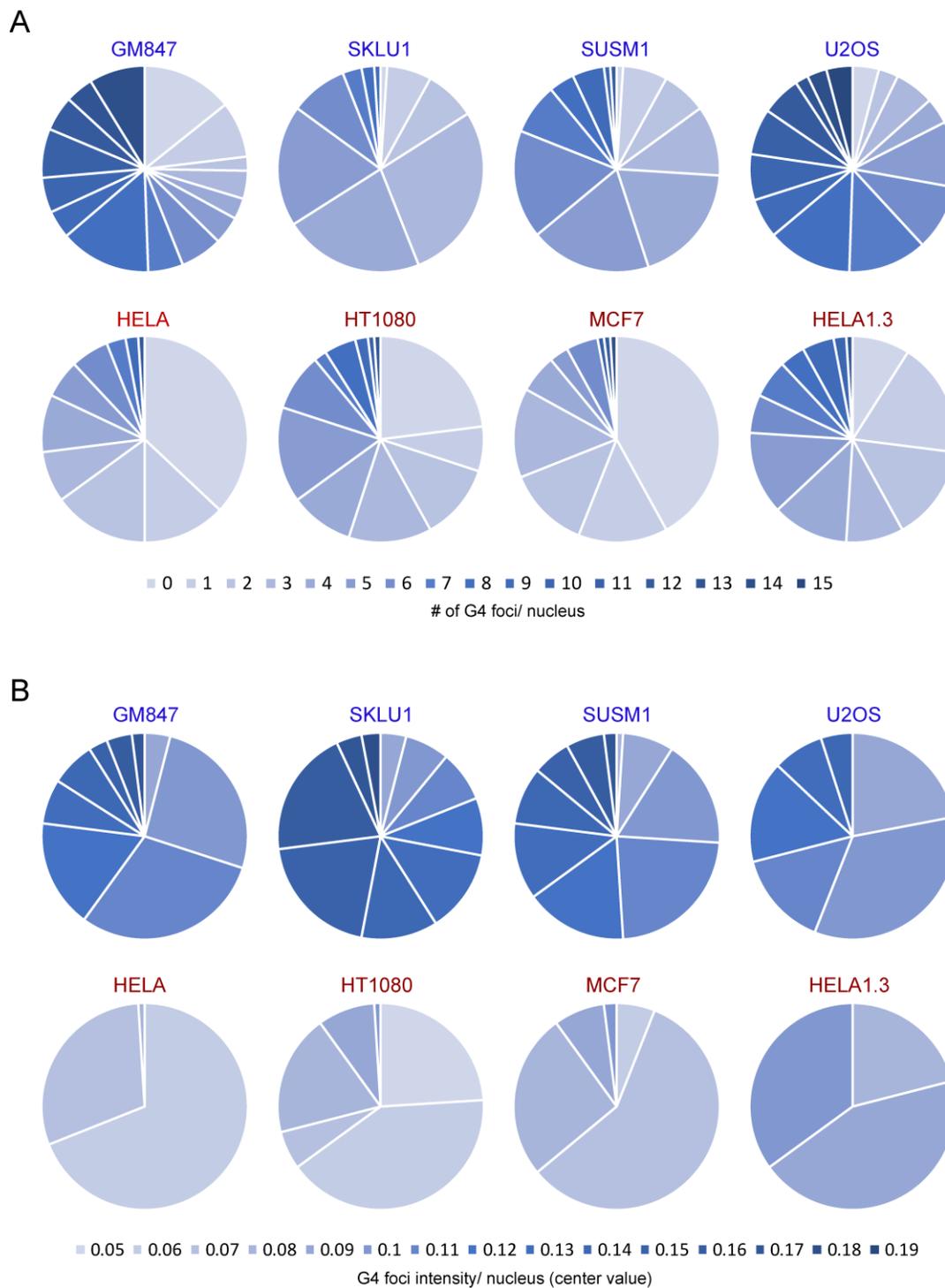
## 5.6 Supplementary Materials

### ---Validation of ALT cell lines---

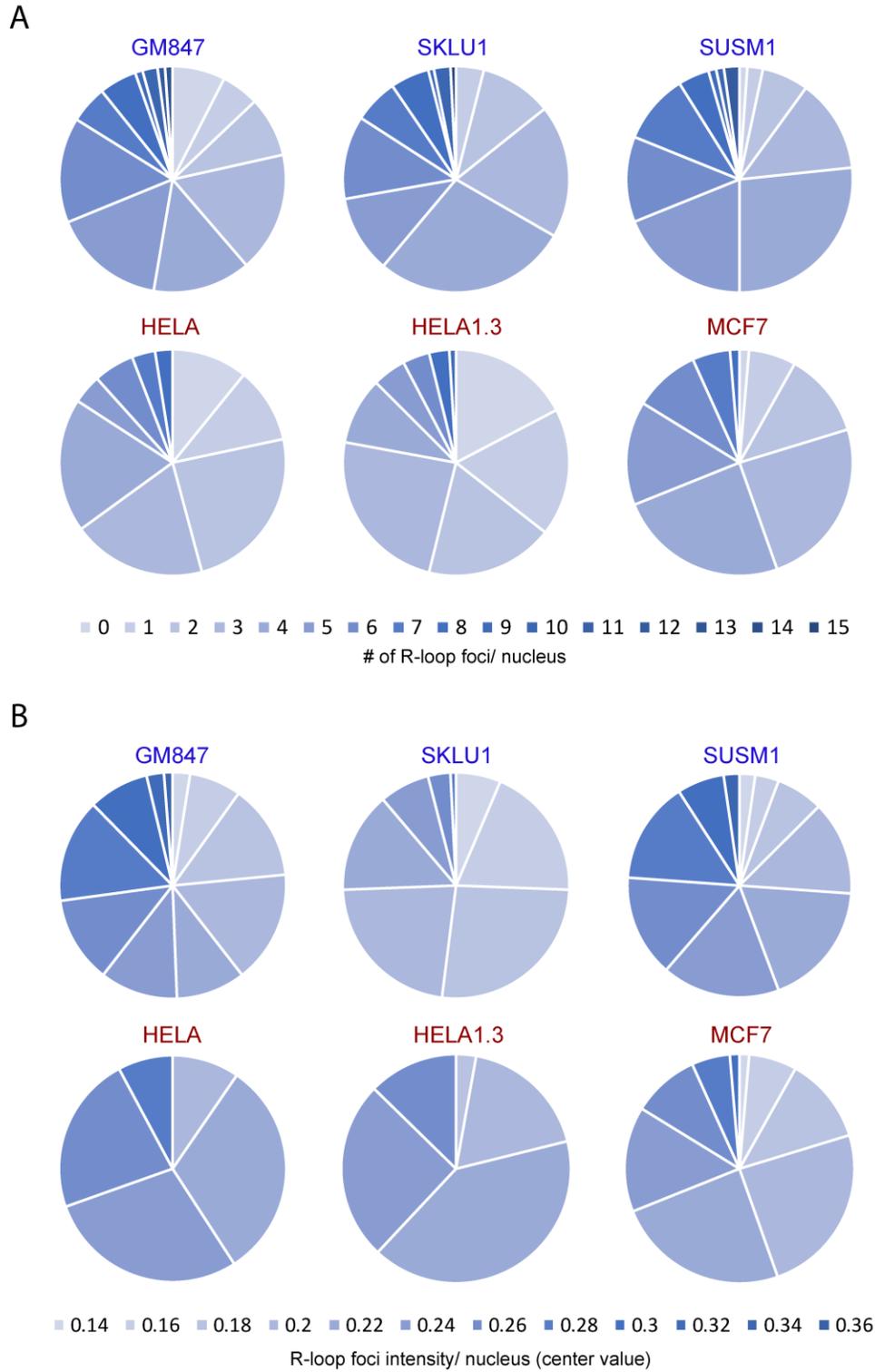
To validate the ALT status in a panel of human cell lines used in the study, we first performed telomere restriction fragment (TRF) assay and Tel-qPCR to measure the mean telomere length, as well as immunocytochemistry (ICC) staining using an antibody against promyelocytic leukemia (PML) bodies. As a control, we also included a TERT+ cell line HELA1.3, a clone of HELA harbouring exceptionally long mean telomere length, comparable to those seen in ALT+ cells.



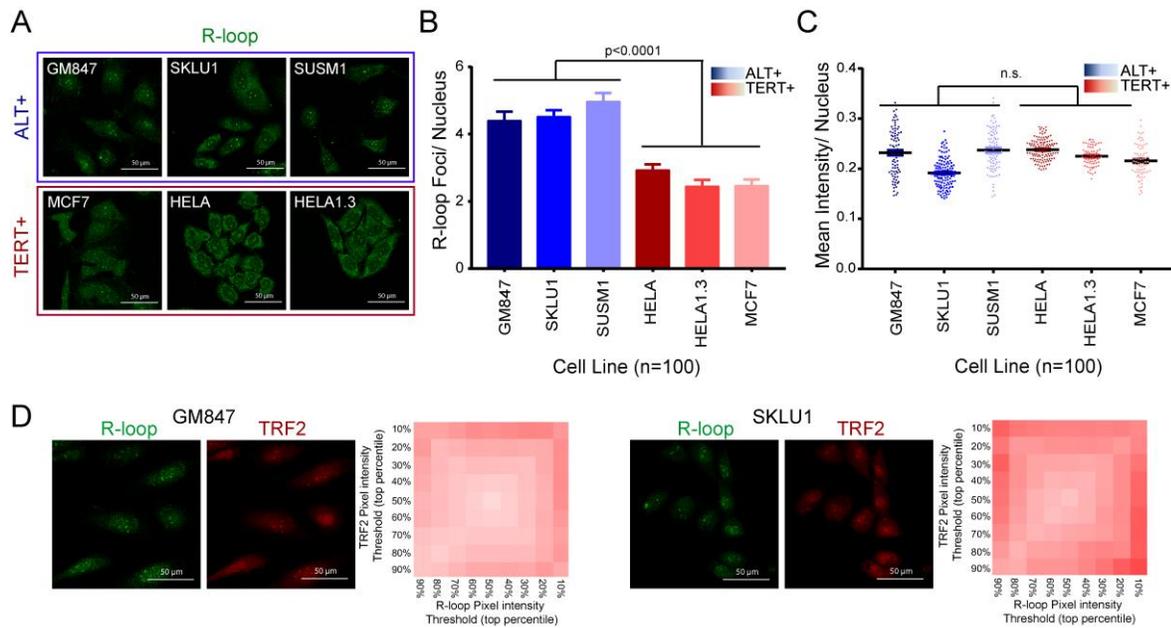
**Figure 5.S1. Confirmation of ALT status in a panel of ALT+ and TERT+ cell lines. (A)** Mean telomere length as measured by TRF (blue) or Tel-qPCR (red) in ALT+ cell lines (GM847, SKLU1, SUSM1, U2OS) and TERT+ cell lines (HELA, HT1080, MCF7, HELA1.3) **(B)** ICC staining for PML in a panel of ALT+ (top) and TERT+ (bottom) cell lines.



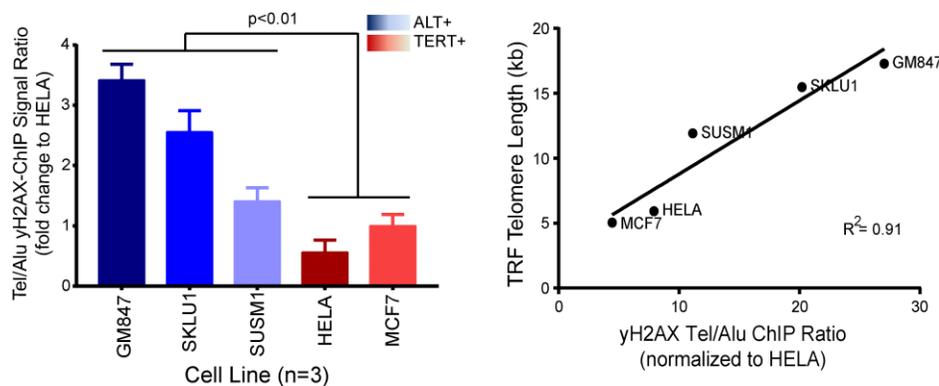
**Figure 5.S2. Distribution profiles of G4 foci in a panel of ALT+ (blue) and TERT+ (red) cell lines. (A) G4 foci count and (B) G4 foci intensity. Distribution is calculated from 100 foci count. Increases in value are depicted by increasing darkness of colour.**



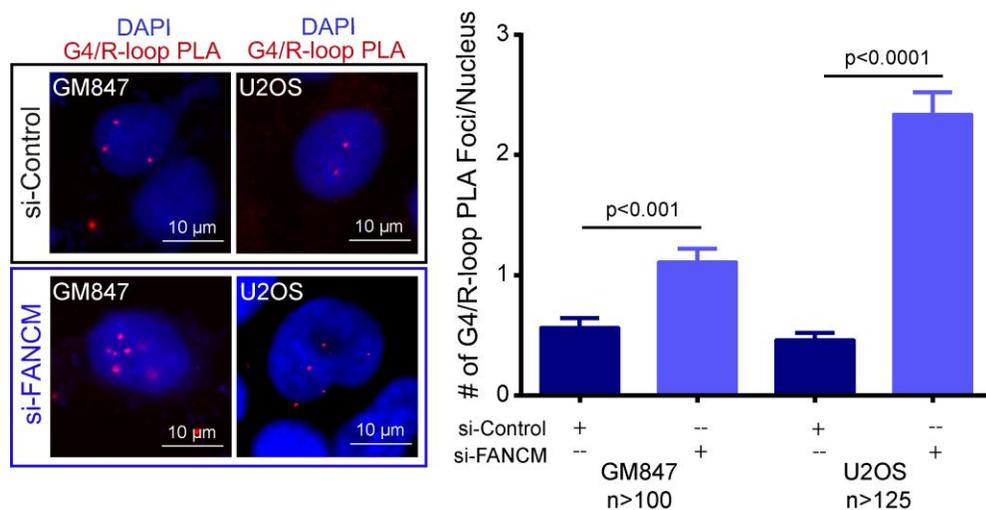
**Figure 5.S3. Distribution profiles of R-loop foci in a panel of ALT+ (blue) and TERT+ (red) cell lines. (A) R-loop foci count and (B) R-loop foci intensity. Distribution is calculated from 100 foci count. Increases in value are depicted by increasing darkness of colour.**



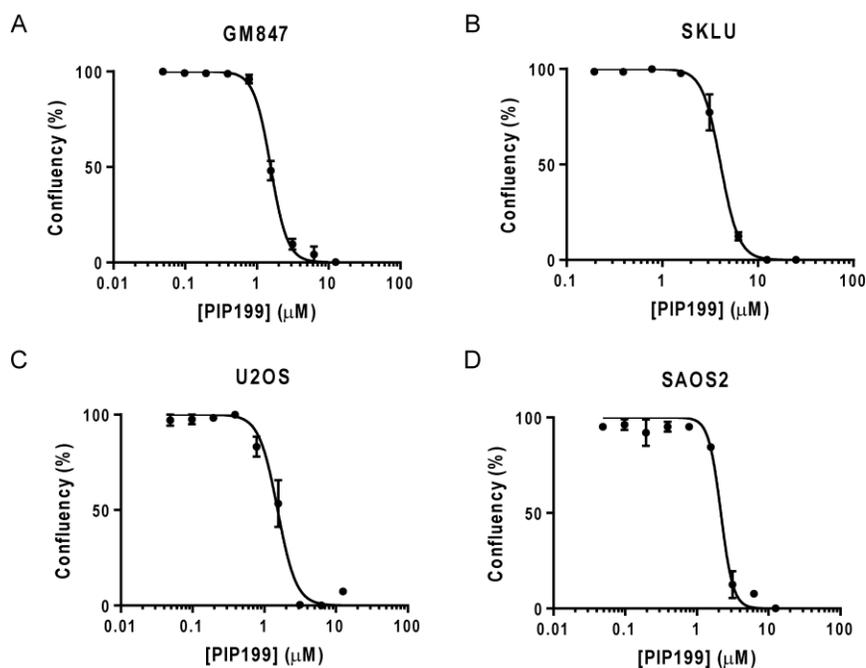
**Figure 5.S4. ALT+ cells harbour higher level of telomeric R-loops than TERT+ cells.** (A) Immunocytochemistry (ICC) staining experiments for R-loop in ALT+ cells lines (within blue frame) versus TERT+ (within red frame) (B) Average number of R-loop foci per nucleus. Automated quantification of 100 cells per cell line was performed in parallel (C) Mean R-loop foci intensity per nucleus as quantified (D) ICC co-staining for R-loop (green) and TRF2 (red) for ALT+ cell lines (GM847, SKLU1). To the right of the ICC images is the metric matrix of the threshold overlap score (TOS) for the top percentile of pixel (by intensity) of R-loop signal (x-axis) and TRF2 signal (y-axis). Darker red colour corresponds with higher TOS.



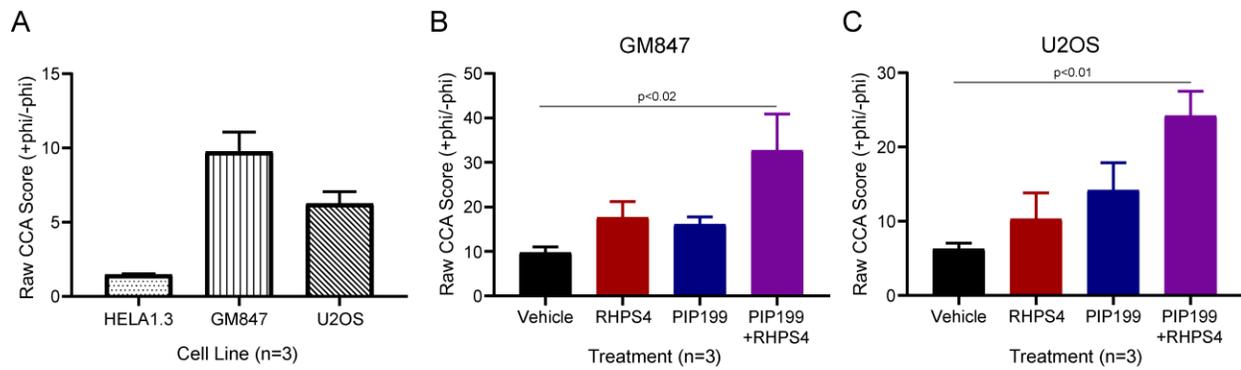
**Figure 5.S5. Level of DNA damage signal,  $\gamma$ H2AX, is higher in ALT+ cells and are broadly associated with telomere length.**  $\gamma$ H2AX-ChIP assay in ALT+ cell lines (blue shades) versus TERT+ cell lines (red shades). Data shown are the mean from 3 biological repeat experiments. T-test (two-tailed) was performed for the combined ALT+ versus TERT+ datasets. Bottom; linear regression plot of Tel/Alu ChIP signal ratio against telomere length (as measured by telomere restriction fragment (TRF) analysis). Line of best fit is plotted along with  $R^2$  value.



**Figure 5.S6. FANCM knockdown-mediated induction of R-loops increases G-loop levels.** Left; Proximity ligation assay for G4 and R-loop (red foci) in GM847 and U2OS transfected with either si-Control (within orange frame) or si-FANCM (within green frame). Right; Quantification for mean PLA foci count per nucleus.



**Figure 5.S7. Cytotoxicity profiles of PIP-199 in ALT cell lines.** Dose response curves after treatment with PIP-199 for 72h was generated for (A) GM847 (B) SKLU1 (C) U2OS (D) SAOS-2. Curves were generated by 3 biological replicates.



**Figure 5.S8. Raw C-circle Assay (CCA) score.** A) Baseline CCA score in TERT+ (HELA 1.3) and ALT+ (GM847 and U2OS) cells B) GM847 treated with either vehicle, RHPS4, PIP-199 or RHPS4+PIP-199. C) U2OS treated with either vehicle, RHPS4, PIP-199 or RHPS4+PIP-199. Raw CCA scores were derived from the ratio of Tel DNA in sample with phi polymerase (+phi) to sample without phi polymerase (-phi). Data was collected from 3 biological repeat experiments.

---Other Tables---

<b>Table 5.S1. Specifications of antibodies used in ICC, CHIP and PLA</b>				
<b>Antibody Name</b>	<b>Vendor (Catalog #)</b>	<b>Type</b>	<b>Target of antibody</b>	<b>Application in this study</b>
BG4	Homemade (purified from BL21 DE3)	Single-chain fragment with FLAG tag	G-quadruplex	ICC, ChIP, ChIP-ReChIP, PLA
Anti-FLAG M2	Sigma (F3165)	Mouse monoclonal primary antibody	FLAG tag	ICC
Anti-FLAG M2	Sigma (F1804)	Mouse monoclonal primary antibody	FLAG tag	PLA
DYKDDDK Tag Rabbit mAB	Cell Signaling (14793)	Rabbit monoclonal primary antibody	FLAG tag	ICC
Anti-DNA-RNA Hybrid, clone S9.6	Millipore (MABE1095)	Mouse monoclonal primary antibody	DNA/RNA-hybrid (R-loop)	ICC, ChIP, ChIP-ReChIP
Anti-S9.6 antibody	Absolute Antibody (Ab01137-23.0)	Rabbit monoclonal primary antibody	DNA/RNA-hybrid (R-loop)	PLA
TRF2 Antibody	Novus Biologicals (NB110-57130)	Rabbit polyclonal primary antibody	TRF2	ICC
Anti-phospho-H2A.X, clone JBW301	Millipore-Sigma (05-636)	Mouse monoclonal primary antibody	$\gamma$ H2AX	ICC, ChIP, ChIP-ReChIP
PML antibody	Santa Cruz (sc-966)	Mouse monoclonal primary antibody	PML	ICC
Goat anti-mouse IgG, Alexa Fluor 488	Invitrogen (A11001)	Goat polyclonal secondary antibody	Mouse IgG	ICC
Goat anti-mouse IgG, Alexa Fluor 594	Invitrogen (A11032)	Goat polyclonal secondary antibody	Mouse IgG	ICC
Goat anti-rabbit IgG, Texas Red	Invitrogen (T2767)	Goat polyclonal secondary antibody	Rabbit IgG	ICC

## Chapter 6: Summary and Conclusions

### 6.1 The optimization of G4 methods through comparison of different tools/methods

Aims 1-3 of my thesis are related to characterizing G4s globally using cellular imaging and genomics techniques. To efficiently optimize methods for downstream experiments, I have performed parallel comparisons using different combinations of antibodies/probes and methods (summarized in Table 6.1). Interestingly, I have discovered that our collection of G4 tools (antibodies and biomimetic chemical probes) are highly complementary. Each confirmed combination is selectively suitable for analyzing a different subset of G4 structures, providing coverage for a broad spectrum of G4 targets.

Through these comparisons, I have found that, while the G4-specific antibody, BG4, is a valuable tool for studying G4-DNAs, it is less compatible with the detection of G4-RNAs. Furthermore, BG4 has currently only been shown to work with ChIP and not RIP. Along with other groups, I have demonstrated the compatibility of ChIP with BG4 (Appendix C; 104-105). It is not immediately clear why BG4 has a stronger preference for G4-DNAs over G4-RNAs. It is possible that since the antibody is a single-chain fragment antibody isolated from the phage display library, the use of G4-DNA forming oligos for the screening process may have resulted in a G4-DNA bias. Unfortunately, there is no high-resolution structural study that can help shed light on the details behind the BG4-target binding mechanism. It is highly probable that the BG4 antibody may have the ability to recognize G4s through multiple modes in addition to the end G-quartet interaction, including interactions with the loop and groove regions. These other interactions may explain some of the potential off-target binding events for BG4.

On the other hand, compared to BG4, the biomimetic TASQ molecules (specifically N-TASQ and BioTASQ used in my studies) have strong affinities for G4-RNAs over G4-DNAs. This preference may be due to the strictly defined binding mode of the TASQ molecules. G-quartet stacking as the target recognition method provides a highly reliable G4 binding specificity; however, this property also restricts the types of G4 conformations that are compatible with these probes. An accessible end G-quartet requirement indicates that any obstruction by an intervening loop region would impede access and inhibit binding by a TASQ probe. Due to this reason, the TASQ probes are limited to the detection of parallel G4s since the antiparallel conformations, especially the "basket" conformation, have diagonal and edgewise loops that are likely to block TASQ-binding. In this context, since G4-RNAs are restricted to fold into different parallel conformations, which effectively bypasses this TASQ binding limitation, these probes have successfully been used for the global detection of G4-RNAs.

Another level of optimization is related to that of the measurement of G4 ligand-induced changes. Compared to the measurement of baseline G4 levels, the presence of G4 ligands can lead to complications due to the following reasons. While, in general, both BG4 and BioTASQ show G4 ligand-induced increases in binding signals, the G4 ligand may compete for binding against an antibody/probe for a subset of G4 targets. Treatment with high doses of G4 ligands can block target access by the antibody/probe, as seen in *in vitro* studies (210, 312). Despite the potential competition, the observation of a ternary complex of antibody/target/ligand suggests that antibody and ligand are not entirely mutually exclusive (313). Instead, these types of interference are dependent on the binding mode compatibilities between the specific ligand and the individual G4 target, in a potentially competitive manner. Based on these observations, the level of G4-ligand-related interference may primarily be dependent on the stoichiometry, the

difference in the number of molecules between the antibody/probe and the G4 ligand at the site of interaction. Thus, a way to circumvent this *in vitro* interference is to treat live cells with a lower dose of G4 ligands, which establishes a new steady-state of the *in vivo* G4 landscape. At later steps after the washing of excess G4 ligands followed by fixation, the use of a higher concentration of antibody or probe at the affinity binding step would provide favourable stoichiometry to out-compete any interference posed by the low amount of residual G4 ligands.

**Table 6.1 Optimization of G4 techniques through parallel comparisons of different combinations of tools with methods**

Techniques	ChIP (BG4)	ChIP (BioTASQ)	RIP (BioTASQ)	RIP (BG4)
Cell lines tested	HELA, MCF7, GM847, KMST1, MDA MB-231, U2OS, SKLU1, SUSM1, VA13, HELA1.3	HELA, MDA MB-231	HELA, MCF7, MDA MB-231, HT-29, HEK293	HELA
Antibody/ tools	BG4	BioTASQ	BioTASQ	BG4
Negative Control	Beads Only	Biotin	Biotin	Beads Only
Signal over background	~10-50 fold	~2-5 fold	>20-100 fold	No signal
IP efficiency/ input	0.2-10%	~0.03-0.05%	1-20%	N/A
Gene regions tested	MYC, VEGFA promoter, Tel1b (telomere), gene desert	MYC, Tel1b	VEGFA, NRAS, TRF2, HPRT1 (control), Tel1b/2b (TERRA), MALAT1, XIST, RPPH1, HDGF, CCNI, KIAA0100	VEGFA, NRAS
Treatment (fold increase over non-treated)	~2-5 fold	~ 1-2 fold	~2-10 fold	N/A
Extraction methods	Spin column	Spin column or TRIZOL	TRIZOL	TRIZOL

Overall, I have successfully performed the optimization of different combinations of tools and methods, which allowed quick determination of suitable methods for genome- and transcriptome-wide mapping of G4s. As a result, I have found that the use of BG4 with ChIP is more suited for genome-wide G4 studies (Chapter 5 and Appendix C), whereas the use of

BioTASQ (a biotinylated version of TASQ) with RIP is more suited for transcriptome-wide G4 studies (Chapter 4). Table 6.1 summarizes the optimization details.

## **6.2 G-quadruplexes exist in nature and are widespread in the human genome and transcriptome**

In my doctoral work, I have provided multiple levels of evidence to support the *in vivo* existence of G4s. First, I have successfully demonstrated, through cellular imaging, that widespread G4s are detectable in fixed cells. Additionally, for both the BG4 antibody and N-TASQ probe, I have shown that G4 ligand-treated cells harbour a significantly increased number of G4 foci, suggesting the existence of targets that can respond to G4 ligand-mediated stabilization. A key limitation is that cellular imaging in fixed cells offers only a snapshot view of the G4 distribution. Since N-TASQ is a smart ligand/probe, which acts as a turn-on probe upon binding to a G4 target and can also readily enter live cells, I have successfully implemented the use of N-TASQ with a live-cell imaging system to monitor the dynamic G4 binding events in real-time (Chapter 3). Interestingly, N-TASQ signals appear to be concurrent with increased cytotoxicity, suggesting that the G4-binding and stabilization events may cause cellular dysfunction.

While imaging techniques are useful for visualization, identifying specific G4 targets requires global mapping methods. To this end, the newest version of TASQ, BioTASQ, has been designed by adding a biotin tag for affinity purification. BG4 and BioTASQ are compatible with different global G4 mapping protocols, ChIP (G4-ChIP-seq) and RIP (G4RP-seq), respectively. They are both performed in sonicated cell lysates that have been chemically crosslinked to preserve biological interactions. Both methods are successful in identifying G-rich sequences

using high-throughput sequencing. Notably, G4RP-seq has shown a highly pG4-dependent enrichment of RNA transcripts, demonstrating its strong G4 specificity (Chapter 4).

Results from G4 genome- and transcriptome-wide mapping studies complement with those from cellular imaging experiments in suggesting widespread G4 formations. It is important to note that these are global measurements from pooled collections of cells (typically in the range of 1-10 million) that do not represent G4 formation events at a single-cell level. However, it does provide an overview of the propensity of particular genomic or transcriptomic regions for forming G4s within living cells in log-phase growth. Results from Chapter 4, together with observations from other studies (90, 108), have established the transient nature of G4s. In the case of G4-RNAs, our results suggest that, rather than being unfolded at all times, RNAs with strong G4 motifs have a higher probability of folding into G4s, but the folded state only lasted for a short time. While this observation provides strong evidence for the *in vivo* existence of G4-RNAs, the biological roles of these global G4-RNAs remain a mystery. The most accepted view to date is that G4 formations are merely stochastic events that occur in G-rich regions during single-stranded states and that G4 helicases are evolutionarily evolved to resolve such events. Such occurrences may explain the high redundancy of functions found for many of these helicases as the accumulation of these G4 folding events can be detrimental to the cells.

From an evolutionary standpoint, it seems counterintuitive for cells to allow transiently folded G4s, then expend extra resources (i.e. through G4 helicases) in unfolding them. While speculative, G4-RNAs may act as transient signals to recruit proteins and become unfolded immediately after the protein has successfully associated with the loci. Separately, using gene ontology analyses, a detailed examination of the identities of gene transcripts enriched by G4RP-seq (Chapter 4) has shown several of them to be RNA-binding proteins. This observation

provides an intriguing hypothesis that G4-RNAs formed in the gene transcript of RNA-binding proteins can act as negative-feedback regulators by binding to its protein product and subsequently inhibiting its translation. An old study has shown evidence of such interactions in the FMRP protein. However, this well-designed feedback system would suggest that every G4-RNA is unique and can differentiate between protein targets or a subset of proteins, which remained to be proven. A deeper understanding of the unique properties of individual G4s would be imperative for delineating the interactome of G4s.

### **6.3 Limitations of tools and methods for G4 studies**

A major limitation for using any G4 detection tools has always been the degree of specificity as there are no perfect antibodies or probes. Improvement of specificity through continual development of these tools can help mitigate most of the off-target biases. In the case of antibodies, batch-to-batch variations exist. We have tried to produce the BG4 antibodies to minimize such variations, but some differences in signals between batches are still expected and observed. For G4 probes, the complex chemical synthesis procedures of some probes (e.g. BioTASQ) limit the availability of these tools to a broader network of research labs. As mentioned previously, the use of antibodies and probes for studying G4 ligand interactions with G4 targets may also be confounding due to the competition of G4 ligand with these antibodies/probes. Careful considerations in the stoichiometry between antibody/probes and G4 ligands can help minimize potential interference.

Several issues remain for global G4 profiling methods (i.e. genome- and transcriptome-wide profiling). During the chemical crosslinking step, biological interactions are "frozen" in their current state. Consequently, biomolecules that are nearby each other would be effectively

crosslinked to each other. While this is essential for preventing probe-mediated *de novo* G4 stabilization *in vitro* during sample processing, other closeby molecules may also be crosslinked to the primary target. Potential indirect interactions via other biomolecules, including proteins, other nucleic acid structures, or even a mixture of both, can be detected but are often weaker in signal than those of the primary targets. These indirect interactions may be biologically interesting, especially for determining the formation of potential G4-related complexes. Future studies by supplementing other compatible methods (e.g. proteomics and sequential immunoprecipitations) can provide insight into this interactome.

Using sequencing to locate the exact G4-forming sites at the nucleotide level may also be challenging due to the low resolutions and co-enrichment of peripheral regions (often ranging in hundreds of nucleotides). Further optimization through fine-tuning the sonication level and increasing sequencing depth may help in improving this deficiency. Additionally, the use of a cross-sectional measurement for G4 mapping in a cell population offers only snapshot views of the G4-folding events and does not provide any information on the dynamic capabilities of these sequences. Finally, none of these techniques can effectively differentiate between intermolecular or intramolecular G4s. For example, a region may have no predicted pG4 motif while harbouring one or two long stretches of guanines. These regions are not necessarily false positives as intermolecular G4 formation is theoretically possible between distant sites arranged close together intracellularly. The addition of methods, including RNA-RNA interactome mapping and advanced bioinformatics strategies, may help elucidate such complex interactions.

#### **6.4 The biological role of G4s – a matter of balance?**

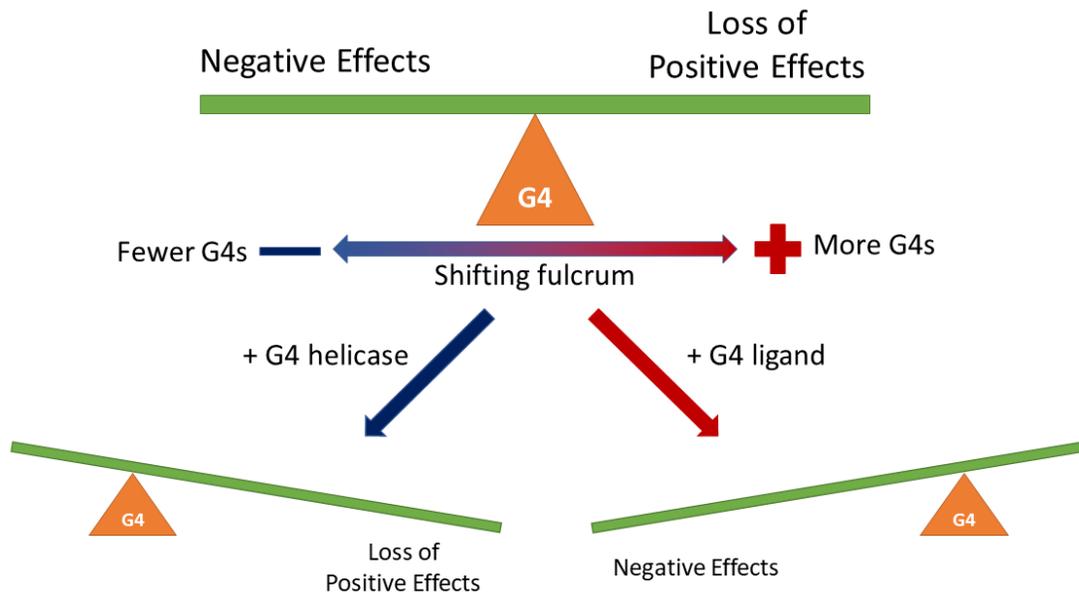
As our understanding of the G4 biology continues to evolve, we become more aware of its complexity. Early studies in the G4 field overwhelmingly support the paradigm in which G4 formations invariably result in negative consequences to normal cell physiology. This deeply ingrained view stemmed from early evidence that has suggested that G4s are barriers to key biological processes, including replication, transcription and translation. In this context, the formation of G4s would be significantly disadvantageous for cells, especially in cancer, as blockade of these key events in the flow of genetic information would inhibit immortal growth. Cancers may compensate for this by overexpressing G4 helicases, corroborated by observations that several common cancer cell lines (e.g. HELA, MCF7, U2OS) have significantly higher expression of well-characterized G4 helicases, including PIF1 and FANCI (derived from Human Protein Atlas). Likewise, leukemia cells overexpress G4 helicase, eIF4A, to unfold G4-RNAs arising during translation (134). Accordingly, overwhelming the G4-unfolding mechanisms of cancer cells through chemical stabilization of G4 is a potentially useful strategy, as demonstrated in Chapter 3, using the N-TASQ ligand/probe. The negative consequences of accumulated G4 formation are also evident in other G4-related diseases, including several developmental disorders linked with G4 helicase dysfunction and neurodegenerative diseases associated with G4 repeat expansions.

Recent studies, including my work presented in Chapter 5, offer an alternate perspective on G4 biology based on a balanced or "G4 homeostasis" model. Unfavourable consequences arise due to disruptions of this G4 homeostasis in both directions. Early G4 studies may have only uncovered one side of the balance (negative effects on global cell physiology), while new evidence is starting to emerge for the other side of the balance (positive effects on selective

biological mechanisms). The evolutionary conservation and the non-random genomic distribution of pG4 motifs are highly suggestive of potential beneficial functions for G4s (109-112). In the case of replication, in contrast to the popular belief that G4s are a barrier to replication, a new study has suggested that G4s may promote origin firing (314). While G4s can pose as roadblocks for transcription, they can potentially promote transcription by keeping the chromatin in a more "open" state or through interactions with proteins (165-166). Similarly, observations of G4s as barriers to translation are contrasted by findings of translation upregulation by IRES (180-181). Despite the earlier reports that folded G4 inhibits telomerase activity, in sharp contrast, a recent report suggested that folded telomeric G4s within actively extending telomerase play a role in the kinetics of telomerase catalysis and dissociation (161). My doctoral work also provides another example (presented in Chapter 5) in which ALT+ cancers harbour increased telomeric G4s, which functions in stimulating the telomere recombination process. Thus, the observations of both positive and negative effects of G4s supports a model in which maintaining a balance of G4 formation capability is essential for optimal cellular functions. Figure 6.1 visually depicts this G4 balance as a lever system in which the fulcrum of this balance represents the level of G4s. G4-interacting factors, such as G4 helicases and G4 ligands, can shift this fulcrum. For example, G4 ligand treatment would shift the fulcrum toward the side with more G4s, which tips the balance of lever favouring negative effects, manifesting in the excess formation of G4s and disruption of key biological processes by forming roadblocks. Shifting the fulcrum toward the side with fewer G4s may lead to the loss of beneficial effects mediated by G4s. However, as our knowledge in the positive biological effects of G4 is still limited, contributed by the lack of genetic and pharmacological tools to specifically destabilize G4 formation, future studies aimed at uncovering the full spectrum of positive

biological roles within G4 homeostasis will still necessarily be through a target-specific approach.

Targeting the G4 homeostasis may be a particularly useful strategy against cancers, as a proper balance of G4s is necessary for cancer cells to thrive. The G4 homeostasis setpoint between normal and cancer cells may also be different, where cancer cells may be functioning at a much higher setpoint, as evident in the general observation of higher intracellular G4 levels in cancer. The high adaptability of cancers may lead to leverage in exploiting G4s' positive biological effects. An interesting observation in the whole genome sequencing showed enrichment of pG4 sequences near breakpoints associated with copy-number alterations (170). This observation suggests an intriguing hypothesis that G4s are drivers of cancer evolution. The association of G4s with hypomethylation suggests that cancers may be able to promote a favourable environment through G4-mediated epigenetic controls (170, 174). Additionally, my work in Chapter 5 also provides another piece of evidence in which elevated levels of telomeric G4s in ALT+ cancers function in stimulating the telomere recombination process. Together, these observations suggest specific vulnerabilities in cancer types that are reliant on G4-mediated beneficial effects. Treatments with G4 ligands may cause higher lethality, as any resistance mechanism through compensatory upregulation of G4-unfolding machinery, will also abolish G4-specific beneficial effects. This notion is supported by evidence from my work in ALT+ cancers (Chapter 5) and others in HR-deficient cancers (62, 64).



**Figure 6.1 Balance of G4 formation (G4 homeostasis), depicted as a lever, for optimal cellular functions.**

The fulcrum represents the G4 levels, which can be shifted toward fewer (blue) or more (red) G4s by factors, such as G4 helicases and G4 ligands. The fulcrum movement shifts the lever's balance to tip in favour of one side, as seen in the bottom two scenarios.

## 6.5 Relationship between G4s and other nucleic acid structures

Interactions between nucleic acids can be functionally important. The G-rich repeat sequence prerequisite for G4 formation may also be permissive for forming other nucleic acid structures. Chapter 5 demonstrates that G4-DNAs can form in parallel with R-loops at the telomeres of ALT<sup>+</sup> cancer cells. Additionally, we have shown that G4s and R-loops can affect each other's stability and cooperate in stimulating ALT recombination. The effect on G4 stability by forming R-loops on the opposing strand may present another G4-DNA regulation mechanism.

For G4-RNAs, G4 formation may compete with other canonical structures, such as hairpin and stem-loops. Interestingly, while highly expressed long non-coding RNAs (lncRNAs), such as *MALAT1* and *NEAT1*, have been found to contain strong pG4 sites using rG4-seq, G4RP-

seq has shown that they are also some of the least enriched targets (with lower propensity for G4 formation in a biologically relevant state). Notably, G4RP-seq in cells treated with G4 ligands has shown significantly increased enrichment of these lncRNAs, suggesting that they have the potential to fold into G4s but that their folded states are highly suppressed in cells in log-phase growth. LncRNAs are known for folding into various RNA structures (e.g. hairpins, loops, knots) to properly associate with RNA-binding proteins to form RNPs. As a result, the interactions involved in RNP formation may restrict G4 formation in lncRNAs. Alternatively, lncRNAs may also be priority targets for G4 helicases, and thus any G4 formation would be short-lived. G4 helicases, like DHX36, may help prevent illegitimate G4 folding from arising during the RNP assembly or remodelling processes.

## **6.6 Future considerations and applications**

Despite significant advances in recent years in the G4 field, several questions remain unanswered. Current techniques for genome- and transcriptome-wide G4 mapping are not without limitations. A more complete and high-resolution mapping may require a combination of strategies and powerful bioinformatics tools to integrate a large amount of data. For example, to improve G4 transcriptomic mapping resolution, G4RP-seq (Chapter 5 and Appendix B) can be combined with rG4-seq. While G4RP-seq lacks resolution, rG4-seq lacks the biologically relevant information by showing only the *in vitro* potential for G4 formation (90). Thus, an intersection of both methods would mitigate these limitations. A separate issue with rG4-seq is its limitation for detecting intramolecular G4s and not intermolecular G4s. Thus, the addition of RNA-RNA interactome mapping would be valuable for mapping out potential intermolecular RNA interactions. Theoretically, the combination of all three techniques, G4RP-seq, rG4-seq and

RNA-RNA interactome mapping, should provide the highest resolution of G4-RNA mapping with the ability to differentiate between intramolecular and intermolecular G4s. The use of chemical probing and time-course experiments may also provide insights into the dynamic nature of G4s on a global scale. Likewise, for G4-ChIP, integration with other chromatin interaction mapping techniques, such as Hi-C, may be useful in providing a more comprehensive view of intermolecular G4-DNA interactions.

Despite the listed shortcomings, these global G4 profiling methods can still be immediately useful. As mentioned earlier, a potential use of these methods is for identifying the types of genomic and transcriptomic G4 targets bound by G4 ligands, with potential application in the design and development of novel G4 ligands. Another possible application is for identifying G4-binding proteins. The current strategy for identifying G4-binding proteins is by using biotinylated G4-oligo as bait followed by proteomics, which may only identify a subset of G4-binding proteins due to the high dependence on a single specific G4-oligo. The G-quartet recognition of BioTASQ may allow the identification of a much broader spectrum of G4-binding proteins. These global profiling methods would also allow broad comparisons across different cell lines and tissues. Global G4 profiling in organisms of different complexity, including bacteria, plants and even viruses, can provide critical insights into the G4 evolutionary biology. Such broad characterization would be essential for studying the relationship between G4s and human diseases (including cancer, neurodegenerative diseases, and developmental diseases). For example, large scale comparisons between normal and cancer cells can reveal potential new therapeutic targets that can be complemented by the parallel development of specific G4-targeting ligands. These tools can also be useful as biomarkers for G4-related diseases (i.e. cancers and neurodegenerative diseases). I have also provided the proof-of-principle for the use

of G4s as biomarkers for ALT+ cancers (Chapter 5). Early detection and classification of these diseases would be essential for guiding therapeutic options and disease management.

While future research in G4s may seem daunting, the rapid advancement in tools and methods are encouraging as they allow previously unapproachable research questions to be empirically tested. My doctoral work contributes to the methodological advancements and provides evidence to support a shift in the current understanding of G4 biology, including previously under-appreciated, context-dependent positive effects of G4. My work here also provides the proof-of-principle and framework for future global-scale G4 studies to develop novel G4 ligands and study the G4 interactome. Thus, the new strategies described in this thesis, combined with the parallel development of complementary approaches in related research fields, will provide the means for unraveling the complexity of G4.

## Bibliography

1. Saenger, W. (2013). *Principles of nucleic acid structure*. Springer Science & Business Media.
2. Kool, E. T. (1997). Preorganization of DNA: design principles for improving nucleic acid recognition by synthetic oligonucleotides. *Chemical reviews*, 97(5), 1473-1488.
3. Vinogradov, S. N., & Linnell, R. H. (1971). *Hydrogen bonding*. Van Nostrand Reinhold.
4. Hoogsteen, K. (1959). The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta crystallographica*, 12(10), 822-823.
5. Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews Genetics*, 13(11), 770-780.
6. Varani, G., & McClain, W. H. (2000). The G·U wobble base pair. *EMBO reports*, 1(1), 18-23.
7. Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., ... & Turner, D. H. (1986). Stability of XGCGCp, GCGCYp, and XGCGCYp helices: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry*, 25(11), 3214-3219.
8. Mignon, P., Loverix, S., Steyaert, J., & Geerlings, P. (2005). Influence of the  $\pi$ - $\pi$  interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic acids research*, 33(6), 1779-1789.
9. Dickerson, R. E., & Klug, A. (1983). Base sequence and helix structure variation in B and A DNA. *Journal of Molecular Biology*, 166(3), 419-441.

10. Grdadolnik, J., Merzel, F., & Avbelj, F. (2017). Origin of hydrophobicity and enhanced water hydrogen bond strength near purely hydrophobic solutes. *Proceedings of the National Academy of Sciences*, 114(2), 322-327.
11. Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.
12. Arnott, S. (2006). Historical article: DNA polymorphism and the early history of the double helix. *Trends in biochemical sciences*, 31(6), 349-354.
13. Leslie, A. G. W., Arnott, S., Chandrasekaran, R., & Ratliff, R. L. (1980). Polymorphism of DNA double helices. *Journal of molecular biology*, 143(1), 49-72.
14. Seol, Y., Skinner, G. M., Visscher, K., Buhot, A., & Halperin, A. (2007). Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Physical review letters*, 98(15), 158103.
15. Jain, A., Wang, G., & Vasquez, K. M. (2008). DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, 90(8), 1117-1130.
16. Abou Assi, H., Garavís, M., González, C., & Damha, M. J. (2018). i-Motif DNA: structural features and significance to cell biology. *Nucleic acids research*, 46(16), 8038-8056.
17. Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences*, 73(7), 2294-2298.
18. Li, Y., Syed, J., & Sugiyama, H. (2016). RNA-DNA triplex formation by long noncoding RNAs. *Cell chemical biology*, 23(11), 1325-1333.
19. Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and molecular life sciences*, 67(1), 43-62.

20. Butcher, S. E., & Pyle, A. M. (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of chemical research*, 44(12), 1302-1311.
21. Lagnado, J. (2013). The story of quadruplex DNA—it started with a Bang!. *The Biochemist*, 35(2), 44-46.
22. Gellert, M., Lipsett, M. N., & Davies, D. R. (1962). Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 48(12), 2013.
23. Sen, D., & Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *nature*, 334(6180), 364-366.
24. Sundquist, W. I., & Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *nature*, 342(6251), 825-829.
25. Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59(5), 871-880.
26. Zahler, A. M., Williamson, J. R., Cech, T. R., & Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DNA structures. *Nature*, 350(6320), 718-720.
27. Balagurumorthy, P., & Brahmachari, S. K. (1994). Structure and stability of human telomeric sequence. *Journal of Biological Chemistry*, 269(34), 21858-21869.
28. Miura, T., Benevides, J. M., & Thomas Jr, G. J. (1995). A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *Journal of molecular biology*, 248(2), 233-238.
29. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & 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. *Proceedings of the National Academy of Sciences*, 99(18), 11593-11598.
30. Huber, M. D., Lee, D. C., & Maizels, N. (2002). G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic acids research*, 30(18), 3954-3961.
  31. Fry, M., & Loeb, L. A. (1999). Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d (CGG) n. *Journal of Biological Chemistry*, 274(18), 12797-12802.
  32. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., ... & Funke, R. (2001). Initial sequencing and analysis of the human genome.
  33. Todd, A. K., Johnston, M., & Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic acids research*, 33(9), 2901-2907.
  34. Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, 33(9), 2908-2916.
  35. Huppert, J. L., & Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic acids research*, 35(2), 406-413.
  36. Henderson, A., Wu, Y., Huang, Y. C., Chavez, E. A., Platt, J., Johnson, F. B., ... & Lansdorp, P. M. (2013). Detection of G-quadruplex DNA in mammalian cells. *Nucleic acids research*, 42(2), 860-869.
  37. Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry*, 5(3), 182-186.

38. Haider, S. M., Parkinson, G. N., & Neidle, S. (2003). Structure of a G-quadruplex–ligand complex. *Journal of molecular biology*, 326(1), 117-125.
39. Winnerdy, F. R., & Phan, A. T. (2020). Quadruplex structure and diversity.
40. Lech, C. J., Heddi, B., & Phan, A. T. (2013). Guanine base stacking in G-quadruplex nucleic acids. *Nucleic acids research*, 41(3), 2034-2046.
41. Kogut, M., Kleist, C., & Czub, J. (2019). Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail. *PLoS computational biology*, 15(9), e1007383.
42. Qin, M., Chen, Z., Luo, Q., Wen, Y., Zhang, N., Jiang, H., & Yang, H. (2015). Two-quartet G-quadruplexes formed by DNA sequences containing four contiguous GG runs. *The Journal of Physical Chemistry B*, 119(9), 3706-3713.
43. Mukundan, V. T., & Phan, A. T. (2013). Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *Journal of the American Chemical Society*, 135(13), 5017-5028.
44. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic acids research*, 34(19), 5402-5415.
45. Bhattacharyya, D., Mirihana Arachchilage, G., & Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Frontiers in chemistry*, 4, 38.
46. Hud, N. V., Smith, F. W., Anet, F. A., & Feigon, J. (1996). The selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by <sup>1</sup>H NMR. *Biochemistry*, 35(48), 15383-15390.
47. Joachimi, A., Benz, A., & Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorganic & medicinal chemistry*, 17(19), 6811-6815.

48. Agarwala, P., Pandey, S., & Maiti, S. (2015). The tale of RNA G-quadruplex. *Organic & biomolecular chemistry*, 13(20), 5570-5585. Olsen, C. M., & Marky, L. A. (2009).
49. Energetic and hydration contributions of the removal of methyl groups from thymine to form uracil in G-quadruplexes. *The Journal of Physical Chemistry B*, 113(1), 9-11.
50. Mullen, M. A., Assmann, S. M., & Bevilacqua, P. C. (2012). Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity. *Journal of the American Chemical Society*, 134(2), 812-815.
51. Xiao, C. D., Ishizuka, T., & Xu, Y. (2017). Antiparallel RNA G-quadruplex formed by human telomere RNA containing 8-bromoguanosine. *Scientific reports*, 7(1), 1-8.
52. Huang, H., Suslov, N. B., Li, N. S., Shelke, S. A., Evans, M. E., Koldobskaya, Y., ... & Piccirilli, J. A. (2014). A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature chemical biology*, 10(8), 686-691.
53. Li, Q., Xiang, J. F., Yang, Q. F., Sun, H. X., Guan, A. J., & Tang, Y. L. (2013). G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic acids research*, 41(D1), D1115-D1123.
54. Zhang, S., Wu, Y., & Zhang, W. (2014). G-Quadruplex structures and their interaction diversity with ligands. *ChemMedChem*, 9(5), 899-911.
55. Monchaud, D., & Teulade-Fichou, M. P. (2008). A hitchhiker's guide to G-quadruplex ligands. *Organic & biomolecular chemistry*, 6(4), 627-636.
56. Asamitsu, S., Bando, T., & Sugiyama, H. (2019). Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chemistry—A European Journal*, 25(2), 417-430.
57. Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., & Neidle, S. (2005). The G-quadruplex-interactive molecule BRACO-19 inhibits tumor

- growth, consistent with telomere targeting and interference with telomerase function. *Cancer research*, 65(4), 1489-1496.
58. Rodriguez, R., Müller, S., Yeoman, J. A., Trentesaux, C., Riou, J. F., & Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *Journal of the American Chemical Society*, 130(47), 15758-15759.
59. Di Antonio, M., Biffi, G., Mariani, A., Raiber, E. A., Rodriguez, R., & Balasubramanian, S. (2012). Inside Cover: Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry (*Angew. Chem. Int. Ed.* 44/2012). *Angewandte Chemie International Edition*, 51(44), 10906-10906.
60. Wheelhouse, R. T., Sun, D., Han, H., Han, F. X., & Hurley, L. H. (1998). Cationic porphyrins as telomerase inhibitors: the interaction of tetra-(N-methyl-4-pyridyl) porphine with quadruplex DNA. *Journal of the American Chemical Society*, 120(13), 3261-3262.
61. Salvati, E., Leonetti, C., Rizzo, A., Scarsella, M., Mottolese, M., Galati, R., ... & Chiorino, G. (2007). Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *The Journal of clinical investigation*, 117(11), 3236-3247.
62. Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., ... & Lim, J. K. (2009). Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer research*, 69(19), 7653-7661.
63. Balasubramanian, S., Hurley, L. H., & Neidle, S. (2011). Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?. *Nature reviews Drug discovery*, 10(4), 261-275.

64. Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N. J., ... & Kabeer, F. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nature communications*, 8(1), 1-18.
65. Kong, D. M., Ma, Y. E., Guo, J. H., Yang, W., & Shen, H. X. (2009). Fluorescent sensor for monitoring structural changes of G-quadruplexes and detection of potassium ion. *Analytical chemistry*, 81(7), 2678-2684.
66. Scaglioni, L., Mondelli, R., Artali, R., Sirtori, F. R., & Mazzini, S. (2016). Nemorubicin and doxorubicin bind the G-quadruplex sequences of the human telomeres and of the c-MYC promoter element Pu22. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(6), 1129-1138.
67. Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., ... & Neidle, S. (2006). Natural and synthetic G-quadruplex interactive berberine derivatives. *Bioorganic & Medicinal Chemistry Letters*, 16(6), 1707-1711.
68. Nikan, M., & Sherman, J. C. (2008). Template-Assembled Synthetic G-Quartets (TASQs). *Angewandte Chemie*, 120(26), 4978-4980.
69. Stefan, L., Guedin, A., Amrane, S., Smith, N., Denat, F., Mergny, J. L., & Monchaud, D. (2011). DOTASQ as a prototype of nature-inspired G-quadruplex ligand. *Chemical Communications*, 47(17), 4992-4994.
70. Xu, H. J., Stefan, L., Haudecoeur, R., Vuong, S., Richard, P., Denat, F., ... & Monchaud, D. (2012). Porphyrin-templated synthetic G-quartet (PorphySQ): a second prototype of G-quartet-based G-quadruplex ligand. *Organic & Biomolecular Chemistry*, 10(27), 5212-5218.

71. Laguerre, A., Stefan, L., Larrouy, M., Genest, D., Novotna, J., Pirrotta, M., & Monchaud, D. (2014). A twice-as-smart synthetic G-quartet: PyroTASQ is both a smart quadruplex ligand and a smart fluorescent probe. *Journal of the American Chemical Society*, 136(35), 12406-12414.
72. Laguerre, A., Hukezalie, K., Winckler, P., Katranji, F., Chanteloup, G., Pirrotta, M., ... & Monchaud, D. (2015). Visualization of RNA-quadruplexes in live cells. *Journal of the American Chemical Society*, 137(26), 8521-8525.
73. Bates, P. J., Reyes-Reyes, E. M., Malik, M. T., Murphy, E. M., O'toole, M. G., & Trent, J. O. (2017). G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1861(5), 1414-1428.
74. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C., & Basu, S. (2012). The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic acids research*, 40(9), 4137-4145.
75. Zamiri, B., Reddy, K., Macgregor, R. B., & Pearson, C. E. (2014). TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r (GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *Journal of Biological Chemistry*, 289(8), 4653-4659.
76. O'Hagan, M. P., Haldar, S., Duchi, M., Oliver, T. A., Mulholland, A. J., Morales, J. C., & Galan, M. C. (2019). A Photoresponsive Stiff-Stilbene Ligand Fuels the Reversible Unfolding of G-Quadruplex DNA. *Angewandte Chemie International Edition*, 58(13), 4334-4338.

77. Poon, L. C. H., Methot, S. P., Morabi-Pazooki, W., Pio, F., Bennet, A. J., & Sen, D. (2011). Guanine-rich RNAs and DNAs that bind heme robustly catalyze oxygen transfer reactions. *Journal of the American Chemical Society*, 133(6), 1877-1884.
78. Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2(12), 919-929.
79. Malgowska, M., Gudanis, D., Teubert, A., Dominiak, G., & Gdaniec, Z. (2012). How to study G-quadruplex structures. *BioTechnologia. Journal of Biotechnology Computational Biology and Bionanotechnology*, 93(4).
80. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčiuk, D., & Kypr, J. (2012). Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*, 24(9), 691-698.
81. Mergny, J. L., Li, J., Lacroix, L., Amrane, S., & Chaires, J. B. (2005). Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic acids research*, 33(16), e138-e138.
82. Adrian, M., Heddi, B., & Phan, A. T. (2012). NMR spectroscopy of G-quadruplexes. *Methods*, 57(1), 11-24.
83. Monchaud, D., & Teulade-Fichou, M. P. (2010). G4-FID: a fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands. In *G-Quadruplex DNA* (pp. 257-271). Humana Press.
84. De Rache, A., & Mergny, J. L. (2015). Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie*, 115, 194-202.
85. Kikin, O., D'Antonio, L., & Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic acids research*, 34(suppl\_2), W676-W682.

86. Yadav, V. K., Abraham, J. K., Mani, P., Kulshrestha, R., & Chowdhury, S. (2007). QuadBase: genome-wide database of G4 DNA—occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic acids research*, 36(suppl\_1), D381-D385.
87. Bedrat, A., Lacroix, L., & Mergny, J. L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic acids research*, 44(4), 1746-1759.
88. Han, H., Hurley, L. H., & Salazar, M. (1999). A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic acids research*, 27(2), 537-542.
89. Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., & Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature biotechnology*, 33(8), 877-881.
90. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S., & Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature methods*, 13(10), 841-844.
91. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., & Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proceedings of the National Academy of Sciences*, 98(15), 8572-8577.
92. Isalan, M., Patel, S. D., Balasubramanian, S., & Choo, Y. (2001). Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry*, 40(3), 830-836.
93. Fernando, H., Rodriguez, R., & Balasubramanian, S. (2008). Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry*, 47(36), 9365-9371.

94. Scholz, O., Hansen, S., & Plückthun, A. (2014). G-quadruplexes are specifically recognized and distinguished by selected designed ankyrin repeat proteins. *Nucleic acids research*, 42(14), 9182-9194.
95. Lam, E. Y. N., Beraldi, D., Tannahill, D., & Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. *Nature communications*, 4(1), 1-8.
96. Biffi, G., Di Antonio, M., Tannahill, D., & Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nature chemistry*, 6(1), 75-80.
97. Liu, H. Y., Zhao, Q., Zhang, T. P., Wu, Y., Xiong, Y. X., Wang, S. K., ... & Tan, J. H. (2016). Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. *Cell chemical biology*, 23(10), 1261-1270.
98. Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., ... & Jackson, S. P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature chemical biology*, 8(3), 301.
99. Tseng, T. Y., Wang, Z. F., Chien, C. H., & Chang, T. C. (2013). In-cell optical imaging of exogenous G-quadruplex DNA by fluorogenic ligands. *Nucleic acids research*, 41(22), 10605-10618.
100. Cogoi, S., Shchekotikhin, A. E., Membrino, A., Sinkevich, Y. B., & Xodo, L. E. (2013). Guanidino anthrathiophenediones as G-quadruplex binders: uptake, intracellular localization, and anti-Harvey-Ras gene activity in bladder cancer cells. *Journal of medicinal chemistry*, 56(7), 2764-2778.

101. Shivalingam, A., Izquierdo, M. A., Le Marois, A., Vyšniauskas, A., Suhling, K., Kuimova, M. K., & Vilar, R. (2015). The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nature communications*, 6(1), 1-10.
102. Xu, S., Li, Q., Xiang, J., Yang, Q., Sun, H., Guan, A., ... & Chen, H. (2015). Directly lighting up RNA G-quadruplexes from test tubes to living human cells. *Nucleic acids research*, 43(20), 9575-9586.
103. Chen, X. C., Chen, S. B., Dai, J., Yuan, J. H., Ou, T. M., Huang, Z. S., & Tan, J. H. (2018). Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. *Angewandte Chemie*, 130(17), 4792-4796.
104. Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., ... & Tannahill, D. (2016). G-quadruplex structures mark human regulatory chromatin. *Nature genetics*, 48(10), 1267-1272.
105. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., & Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature protocols*, 13(3), 551.
106. Law, M. J., Lower, K. M., Voon, H. P., Hughes, J. R., Garrick, D., Viprakit, V., ... & Abbott, A. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*, 143(3), 367-378.
107. Gray, L. T., Vallur, A. C., Eddy, J., & Maizels, N. (2014). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nature chemical biology*, 10(4), 313.
108. Guo, J. U., & Bartel, D. P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, 353(6306).

109. Verma, A., Halder, K., Halder, R., Yadav, V. K., Rawal, P., Thakur, R. K., ... & Chowdhury, S. (2008). Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *Journal of medicinal chemistry*, 51(18), 5641-5649.
110. Nakken, S., Rognes, T., & Hovig, E. (2009). The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic acids research*, 37(17), 5749-5756.
111. Capra, J. A., Paeschke, K., Singh, M., & Zakian, V. A. (2010). G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput Biol*, 6(7), e1000861.
112. Frees, S., Menendez, C., Crum, M., & Bagga, P. S. (2014). QGRS-Conserve: a computational method for discovering evolutionarily conserved G-quadruplex motifs. *Human genomics*, 8(1), 8.
113. Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nature chemical biology*, 3(4), 218-221.
114. Umate, P., Tuteja, N., & Tuteja, R. (2011). Genome-wide comprehensive analysis of human helicases. *Communicative & integrative biology*, 4(1), 118-137.
115. Ribeyre, C., Lopes, J., Boulé, J. B., Piazza, A., Guédin, A., Zakian, V. A., ... & Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet*, 5(5), e1000475.
116. Sanders, C. M. (2010). Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochemical Journal*, 430(1), 119-128.

117. Paeschke, K., Capra, J. A., & Zakian, V. A. (2011). DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell*, 145(5), 678-691.
118. Masuda-Sasa, T., Polaczek, P., Peng, X. P., Chen, L., & Campbell, J. L. (2008). Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *Journal of Biological Chemistry*, 283(36), 24359-24373.
119. Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., ... & Chai, W. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO journal*, 32(10), 1425-1439.
120. Sun, H., Karow, J. K., Hickson, I. D., & Maizels, N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. *Journal of Biological Chemistry*, 273(42), 27587-27592.
121. Liu, J. Q., Chen, C. Y., Xue, Y., Hao, Y. H., & Tan, Z. (2010). G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *Journal of the American Chemical Society*, 132(30), 10521-10527.
122. Budhathoki, J. B., Ray, S., Urban, V., Janscak, P., Yodh, J. G., & Balci, H. (2014). RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic acids research*, 42(18), 11528-11545.
123. Mohaghegh, P., Karow, J. K., Brosh Jr, R. M., Bohr, V. A., & Hickson, I. D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic acids research*, 29(13), 2843-2849.

124. Creacy, S. D., Routh, E. D., Iwamoto, F., Nagamine, Y., Akman, S. A., & Vaughn, J. P. (2008). G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *Journal of Biological Chemistry*, 283(50), 34626-34634.
125. Tippana, R., Hwang, H., Opresko, P. L., Bohr, V. A., & Myong, S. (2016). Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proceedings of the National Academy of Sciences*, 113(30), 8448-8453.
126. Chen, M. C., Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., & Ferré-D'Amaré, A. R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, 558(7710), 465-469.
127. Lattmann, S., Stadler, M. B., Vaughn, J. P., Akman, S. A., & Nagamine, Y. (2011). The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic acids research*, 39(21), 9390-9404.
128. Tippana, R., Chen, M. C., Demeshkina, N. A., Ferré-D'Amaré, A. R., & Myong, S. (2019). RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nature communications*, 10(1), 1-10.
129. Cheung, I., Schertzer, M., Rose, A., & Lansdorp, P. M. (2002). Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nature genetics*, 31(4), 405-409.
130. Wu, Y., Shin-ya, K., & Brosh, R. M. (2008). FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Molecular and cellular biology*, 28(12), 4116-4128.

131. Ding, H., Schertzer, M., Wu, X., Gertsenstein, M., Selig, S., Kammori, M., ... & Tam, P. P. (2004). Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell*, 117(7), 873-886.
132. Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H., & Boulton, S. J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*, 149(4), 795-806.
133. Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair*, 10(6), 654-665.
134. Wolfe, A. L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V. K., Sanghvi, V. R., ... & Schatz, J. H. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature*, 513(7516), 65-70.
135. McRae, E. K., Booy, E. P., Moya-Torres, A., Ezzati, P., Stetefeld, J., & McKenna, S. A. (2017). Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic acids research*, 45(11), 6656-6668.
136. Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., ... & Balasubramanian, S. (2018). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic acids research*, 46(21), 11592-11604.
137. de Almeida, C. R., Dhir, S., Dhir, A., Moghaddam, A. E., Sattentau, Q., Meinhart, A., & Proudfoot, N. J. (2018). RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Molecular cell*, 70(4), 650-662.

138. Tuesuwan, B., Kern, J. T., Thomas, P. W., Rodriguez, M., Li, J., David, W. M., & Kerwin, S. M. (2008). Simian virus 40 large T-antigen G-quadruplex DNA helicase inhibition by G-quadruplex DNA-interactive agents. *Biochemistry*, 47(7), 1896-1909.
139. Bharti, S. K., Sommers, J. A., Zhou, J., Kaplan, D. L., Spelbrink, J. N., Mergny, J. L., & Brosh, R. M. (2014). DNA sequences proximal to human mitochondrial DNA deletion breakpoints prevalent in human disease form G-quadruplexes, a class of DNA structures inefficiently unwound by the mitochondrial replicative Twinkle helicase. *Journal of Biological Chemistry*, 289(43), 29975-29993.
140. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., & Lipps, H. J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature structural & molecular biology*, 12(10), 847-854.
141. Zaug, A. J., Podell, E. R., & Cech, T. R. (2005). Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences*, 102(31), 10864-10869.
142. Pedroso, I. M., Hayward, W., & Fletcher, T. M. (2009). The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures. *Nucleic acids research*, 37(5), 1541-1554.
143. Qureshi, M. H., Ray, S., Sewell, A. L., Basu, S., & Balci, H. (2012). Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency. *The Journal of Physical Chemistry B*, 116(19), 5588-5594.
144. Brázda, V., Hároníková, L., Liao, J. C., Fridrichová, H., & Jagelská, E. B. (2016). Strong preference of BRCA1 protein to topologically constrained non-B DNA structures. *BMC molecular biology*, 17(1), 1-9.

145. Soldatenkov, V. A., Vetcher, A. A., Duka, T., & Ladame, S. (2008). First evidence of a functional interaction between DNA quadruplexes and poly (ADP-ribose) polymerase-1. *ACS chemical biology*, 3(4), 214-219.
146. González, V., Guo, K., Hurley, L., & Sun, D. (2009). Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *Journal of Biological Chemistry*, 284(35), 23622-23635.
147. González, V., & Hurley, L. H. (2010). The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry*, 49(45), 9706-9714.
148. Lago, S., Tosoni, E., Nadai, M., Palumbo, M., & Richter, S. N. (2017). The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1861(5), 1371-1381.
149. Scognamiglio, P. L., Di Natale, C., Leone, M., Poletto, M., Vitagliano, L., Tell, G., & Marasco, D. (2014). G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1840(6), 2050-2059.
150. Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K., & 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. *Journal of Biological Chemistry*, 285(29), 22003-22016.
151. Chen, S., Su, L., Qiu, J., Xiao, N., Lin, J., Tan, J. H., ... & Li, D. (2013). Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc

- transcription. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1830(10), 4769-4777.
152. Cree, S. L., Fredericks, R., Miller, A., Pearce, F. G., Filichev, V., Fee, C., & Kennedy, M. A. (2016). DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS letters*, 590(17), 2870-2883.
153. Khateb, S., Weisman-Shomer, P., Hershcó, I., Loeb, L. A., & Fry, M. (2004). Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)<sub>n</sub> is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic acids research*, 32(14), 4145-4154.
154. Decorsière, A., Cayrel, A., Vagner, S., & Millevoi, S. (2011). Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes & development*, 25(3), 220-225.
155. Samatanga, B., Dominguez, C., Jelesarov, I., & Allain, F. H. T. (2013). The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic acids research*, 41(4), 2505-2516.
156. Bensaid, M., Melko, M., Bechara, E. G., Davidovic, L., Berretta, A., Catania, M. V., ... & Bardoni, B. (2009). FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic acids research*, 37(4), 1269-1279.
157. Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., & Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA-and TERRA-binding protein TLS/FUS. *Chemistry & biology*, 20(3), 341-350.

158. Hacht, A. V., Seifert, O., Menger, M., Schütze, T., Arora, A., Konthur, Z., ... & Kurreck, J. (2014). Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic acids research*, 42(10), 6630-6644.
159. Smith, J. S., Chen, Q., Yatsunyk, L. A., Nicoludis, J. M., Garcia, M. S., Kranaster, R., ... & Schultz, D. C. (2011). Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nature structural & molecular biology*, 18(4), 478.
160. Moye, A. L., Porter, K. C., Cohen, S. B., Phan, T., Zyner, K. G., Sasaki, N., ... & Bryan, T. M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature communications*, 6(1), 1-12.
161. Jansson, L. I., Hentschel, J., Parks, J. W., Chang, T. R., Lu, C., Baral, R., ... & Stone, M. D. (2019). Telomere DNA G-quadruplex folding within actively extending human telomerase. *Proceedings of the National Academy of Sciences*, 116(19), 9350-9359.
162. Cogoi, S., & Xodo, L. E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic acids research*, 34(9), 2536-2549.
163. Bejugam, M., Sewitz, S., Shirude, P. S., Rodriguez, R., Shahid, R., & Balasubramanian, S. (2007). Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. *Journal of the American Chemical Society*, 129(43), 12926-12927.
164. Raiber, E. A., Kranaster, R., Lam, E., Nikan, M., & Balasubramanian, S. (2012). A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic acids research*, 40(4), 1499-1508.

165. Li, P. T., Wang, Z. F., Kuan, Y. M., Li, M. H., Huang, M. C., Chiang, P. C., ... & Chen, C. T. (2017). Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation. *Journal of Biological Chemistry*, 292(51), 20859-20870.
166. Du, Z., Zhao, Y., & Li, N. (2008). Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome research*, 18(2), 233-241.
167. Belotserkovskii, B. P., Soo Shin, J. H., & Hanawalt, P. C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine–homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic acids research*, 45(11), 6589-6599.
168. Wanrooij, P. H., Uhler, J. P., Shi, Y., Westerlund, F., Falkenberg, M., & Gustafsson, C. M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic acids research*, 40(20), 10334-10344.
169. Lerner, L. K., & Sale, J. E. (2019). Replication of G quadruplex DNA. *Genes*, 10(2), 95.
170. De, S., & Michor, F. (2011). DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nature structural & molecular biology*, 18(8), 950.
171. van Wietmarschen, N., Merzouk, S., Halsema, N., Spierings, D. C., Guryev, V., & Lansdorp, P. M. (2018). BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nature communications*, 9(1), 1-12.
172. Pladevall-Morera, D., Munk, S., Ingham, A., Garribba, L., Albers, E., Liu, Y., ... & Lopez-Contreras, A. J. (2019). Proteomic characterization of chromosomal common fragile site

- (CFS)-associated proteins uncovers ATRX as a regulator of CFS stability. *Nucleic acids research*, 47(15), 8004-8018.
173. Wu, W., Rokutanda, N., Takeuchi, J., Lai, Y., Maruyama, R., Togashi, Y., ... & Saeki, Y. (2018). HERC2 facilitates BLM and WRN helicase complex interaction with RPA to suppress G-quadruplex DNA. *Cancer Research*, 78(22), 6371-6385.
174. Mao, S. Q., Ghanbarian, A. T., Spiegel, J., Cuesta, S. M., Beraldi, D., Di Antonio, M., ... & Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. *Nature structural & molecular biology*, 25(10), 951-957.
175. Guilbaud, G., Murat, P., Recolin, B., Campbell, B. C., Maiter, A., Sale, J. E., & Balasubramanian, S. (2017). Local epigenetic reprogramming induced by G-quadruplex ligands. *Nature chemistry*, 9(11), 1110.
176. Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., & Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *The EMBO journal*, 20(17), 4803-4813.
177. Shahid, R., Bugaut, A., & Balasubramanian, S. (2010). The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry*, 49(38), 8300-8306.
178. Lammich, S., Kamp, F., Wagner, J., Nuscher, B., Zilow, S., Ludwig, A. K., ... & Haass, C. (2011). Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region. *Journal of Biological Chemistry*, 286(52), 45063-45072.

179. Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A. L., & Fry, M. (2007). The tetraplex (CGG)<sub>n</sub> destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic acids research*, 35(17), 5775-5788.
180. Bonnal, S., Schaeffer, C., Créancier, L., Clamens, S., Moine, H., Prats, A. C., & Vagner, S. (2003). A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *Journal of Biological Chemistry*, 278(41), 39330-39336.
181. Morris, M. J., Negishi, Y., Pázsint, C., Schonhofs, J. D., & Basu, S. (2010). An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES. *Journal of the American Chemical Society*, 132(50), 17831-17839.
182. Sauer, M., Juranek, S. A., Marks, J., De Magis, A., Kazemier, H. G., Hilbig, D., ... & Paeschke, K. (2019). DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nature communications*, 10(1), 1-15.
183. Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J. L., & Moine, H. (2011). G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO reports*, 12(7), 697-704.
184. Marcel, V., Tran, P. L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M. P., ... & Van Dyck, E. (2011). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis*, 32(3), 271-278.
185. Fisette, J. F., Montagna, D. R., Mihailescu, M. R., & Wolfe, M. S. (2012). AG-Rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *Journal of neurochemistry*, 121(5), 763-773.

186. Ribeiro, M. M., Teixeira, G. S., Martins, L., Marques, M. R., de Souza, A. P., & Line, S. R. P. (2015). G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Human genetics*, 134(1), 37-44.
187. Zhang, J., Harvey, S. E., & Cheng, C. (2019). A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic acids research*, 47(7), 3667-3679.
188. Arachchilage, G. M., Dassanayake, A. C., & Basu, S. (2015). A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chemistry & biology*, 22(2), 262-272.
189. Pandey, S., Agarwala, P., Jayaraj, G. G., Gargallo, R., & Maiti, S. (2015). The RNA stem-loop to G-quadruplex equilibrium controls mature MicroRNA production inside the cell. *Biochemistry*, 54(48), 7067-7078.
190. Chan, K. L., Peng, B., Umar, M. I., Chan, C. Y., Sahakyan, A. B., Le, M. T., & Kwok, C. K. (2018). Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chemical communications*, 54(77), 10878-10881.
191. Wang, X., Goodrich, K. J., Gooding, A. R., Naeem, H., Archer, S., Paucek, R. D., ... & Davidovich, C. (2017). Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Molecular cell*, 65(6), 1056-1067.
192. Sahakyan, A. B., Murat, P., Mayer, C., & Balasubramanian, S. (2017). G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nature Structural & Molecular Biology*, 24(3), 243-247.

193. Rigo, R., Palumbo, M., & Sissi, C. (2017). G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1861(5), 1399-1413.
194. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., ... & Rothstein, J. D. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, 507(7491), 195-200.
195. Fratta, P., Mizielińska, S., Nicoll, A. J., Zloh, M., Fisher, E. M., Parkinson, G., & Isaacs, A. M. (2012). C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Scientific reports*, 2, 1016.
196. Conlon, E. G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N. A., & Manley, J. L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife*, 5, e17820.
197. Fay, M. M., Anderson, P. J., & Ivanov, P. (2017). ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell reports*, 21(12), 3573-3584.
198. Zamiri, B., Reddy, K., Macgregor, R. B., & Pearson, C. E. (2014). TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated (GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *Journal of Biological Chemistry*, 289(8), 4653-4659.
199. Simone, R., Balendra, R., Moens, T. G., Preza, E., Wilson, K. M., Heslegrave, A., ... & Clayton, E. L. (2018). G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO molecular medicine*, 10(1), 22-31.
200. Maizels, N. (2015). G4-associated human diseases. *EMBO reports*, 16(8), 910-922.

201. Misra, V. K., & Honig, B. (1995). On the magnitude of the electrostatic contribution to ligand-DNA interactions. *Proceedings of the National Academy of Sciences*, 92(10), 4691-4695.
202. Ball, P. (2008). Water as an active constituent in cell biology. *Chemical reviews*, 108(1), 74-108.
203. Chaplin, M. (2006). Do we underestimate the importance of water in cell biology?. *Nature Reviews Molecular Cell Biology*, 7(11), 861-866.
204. Bagchi, B. (2005). Water dynamics in the hydration layer around proteins and micelles. *Chemical Reviews*, 105(9), 3197-3219.
205. Pal, S. K., & Zewail, A. H. (2004). Dynamics of water in biological recognition. *Chemical Reviews*, 104(4), 2099-2124.
206. Davis, J. G., Rankin, B. M., Gierszal, K. P., & Ben-Amotz, D. (2013). On the cooperative formation of non-hydrogen-bonded water at molecular hydrophobic interfaces. *Nature chemistry*, 5(9), 796-802.
207. Davis, J. G., Gierszal, K. P., Wang, P., & Ben-Amotz, D. (2012). Water structural transformation at molecular hydrophobic interfaces. *Nature*, 491(7425), 582-585.
208. Perera, P. N., Fega, K. R., Lawrence, C., Sundstrom, E. J., Tomlinson-Phillips, J., & Ben-Amotz, D. (2009). Observation of water dangling OH bonds around dissolved nonpolar groups. *Proceedings of the National Academy of Sciences*, 106(30), 12230-12234.
209. Demchenko, A. P. (2002). The red-edge effects: 30 years of exploration. *Luminescence: the journal of biological and chemical luminescence*, 17(1), 19-42.

210. Laguerre, A., Wong, J. M., & Monchaud, D. (2016). Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Scientific reports*, 6, 32141. (205)
211. Collie, G. W., & Parkinson, G. N. (2011). The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chemical Society Reviews*, 40(12), 5867-5892.
212. Zhou, J., Roembke, B. T., Paragi, G., Laguerre, A., Sintim, H. O., Guerra, C. F., & Monchaud, D. (2016). Computational understanding and experimental characterization of twice-as-smart quadruplex ligands as chemical sensors of bacterial nucleotide second messengers. *Scientific reports*, 6, 33888.
213. Haudecoeur, R., Stefan, L., Denat, F., & Monchaud, D. (2013). A model of smart G-quadruplex ligand. *Journal of the American Chemical Society*, 135(2), 550-553.
214. Chattopadhyay, A., & Haldar, S. (2014). Dynamic insight into protein structure utilizing red edge excitation shift. *Accounts of chemical research*, 47(1), 12-19.
215. Haldar, S., Chaudhuri, A., & Chattopadhyay, A. (2011). Organization and dynamics of membrane probes and proteins utilizing the red edge excitation shift. *The Journal of Physical Chemistry B*, 115(19), 5693-5706.
216. Demchenko, A. P. (2008). Site-selective red-edge effects. *Methods in enzymology*, 450, 59-78.
217. Berger, O., Adler-Abramovich, L., Levy-Sakin, M., Grunwald, A., Liebes-Peer, Y., Bachar, M., ... & Ebenstein, Y. (2015). Light-emitting self-assembled peptide nucleic acids exhibit both stacking interactions and Watson–Crick base pairing. *Nature Nanotechnology*, 10(4), 353-360.

218. Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, 43(18), 8627-8637.
219. Tarsounas, M., & Tijsterman, M. (2013). Genomes and G-quadruplexes: for better or for worse. *Journal of molecular biology*, 425(23), 4782-4789.
220. Rodriguez, R., & Miller, K. M. (2014). Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nature Reviews Genetics*, 15(12), 783-796.
221. Xu, C., Zipfel, W., Shear, J. B., Williams, R. M., & Webb, W. W. (1996). Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy. *Proceedings of the National Academy of Sciences*, 93(20), 10763-10768.
222. König, K. (2000). Multiphoton microscopy in life sciences. *Journal of microscopy*, 200(2), 83-104.
223. Ustione, A., & Piston, D. W. (2011). A simple introduction to multiphoton microscopy. *Journal of microscopy*, 243(3), 221-226.
224. Neidle, S. (2016). Quadruplex nucleic acids as novel therapeutic targets. *Journal of medicinal chemistry*, 59(13), 5987-6011.
225. Müller, S., & Rodriguez, R. (2014). G-quadruplex interacting small molecules and drugs: from bench toward bedside. *Expert review of clinical pharmacology*, 7(5), 663-679.
226. Han, H., Langley, D. R., Rangan, A., & Hurley, L. H. (2001). Selective interactions of cationic porphyrins with G-quadruplex structures. *Journal of the American Chemical Society*, 123(37), 8902-8913.
227. Anderson, P., Kedersha, N., & Ivanov, P. (2015). Stress granules, P-bodies and cancer. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1849(7), 861-870.

228. Schäferling, M. (2012). The art of fluorescence imaging with chemical sensors. *Angewandte Chemie International Edition*, 51(15), 3532-3554.
229. Largy, E., Granzhan, A., Hamon, F., Verga, D., & Teulade-Fichou, M. P. (2012). Visualizing the quadruplex: from fluorescent ligands to light-up probes. In *Quadruplex Nucleic Acids* (pp. 111-177). Springer, Berlin, Heidelberg.
230. Mongelard, F., & Bouvet, P. (2007). Nucleolin: a multiFACeTed protein. *Trends in cell biology*, 17(2), 80-86.
231. Hovanessian, A. G., Soundaramourty, C., El Khoury, D., Nondier, I., Svab, J., & Krust, B. (2010). Surface expressed nucleolin is constantly induced in tumor cells to mediate calcium-dependent ligand internalization. *PloS one*, 5(12), e15787.
232. Bates, P. J., Laber, D. A., Miller, D. M., Thomas, S. D., & Trent, J. O. (2009). Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Experimental and molecular pathology*, 86(3), 151-164.
233. Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., & Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, 505(7485), 696-700.
234. Morris, K. V., & Mattick, J. S. (2014). The rise of regulatory RNA. *Nature Reviews Genetics*, 15(6), 423-437.
235. Leppek, K., Das, R., & Barna, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature reviews Molecular cell biology*, 19(3), 158.
236. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E., & Chang, H. Y. (2011). Understanding the transcriptome through RNA structure. *Nature Reviews Genetics*, 12(9), 641-655.

237. Cammas, A., & Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic acids research*, 45(4), 1584-1595.
238. Fay, M. M., Lyons, S. M., & Ivanov, P. (2017). RNA G-quadruplexes in biology: principles and molecular mechanisms. *Journal of molecular biology*, 429(14), 2127-2147.
239. Endoh, T., Rode, A. B., Takahashi, S., Kataoka, Y., Kuwahara, M., & Sugimoto, N. (2016). Real-time monitoring of G-quadruplex formation during transcription. *Analytical chemistry*, 88(4), 1984-1989.
240. Gabelica, V., Rosu, F., & De Pauw, E. (2009). A simple method to determine electrospray response factors of noncovalent complexes. *Analytical chemistry*, 81(16), 6708-6715.
241. Hänsel-Hertsch, R., Di Antonio, M., & Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature reviews Molecular cell biology*, 18(5), 279-284.
242. Azzalin, C. M., Reichenbach, P., Khoraiuli, L., Giulotto, E., & Lingner, J. (2007). Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 318(5851), 798-801.
243. Montero, J. J., De Silanes, I. L., Grana, O., & Blasco, M. A. (2016). Telomeric RNAs are essential to maintain telomeres. *Nature communications*, 7, 12534.
244. Haudecoeur, R., Stefan, L., & Monchaud, D. (2013). Multitasking Water-Soluble Synthetic G-Quartets: From Preferential RNA–Quadruplex Interaction to Biocatalytic Activity. *Chemistry—A European Journal*, 19(38), 12739-12747.
245. Müller, S., Kumari, S., Rodriguez, R., & Balasubramanian, S. (2010). Small-molecule-mediated G-quadruplex isolation from human cells. *Nature chemistry*, 2(12), 1095-1098.

246. Bugaut, A., & Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic acids research*, 40(11), 4727-4741.
247. Karsisiotis, A. I., Hessari, N. M. A., Novellino, E., Spada, G. P., Randazzo, A., & Webba da Silva, M. (2011). Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie*, 123(45), 10833-10836.
248. Selth, L. A., Gilbert, C., & Svejstrup, J. Q. (2009). RNA immunoprecipitation to determine RNA-protein associations in vivo. *Cold Spring Harbor protocols*, 2009(6), pdb-prot5234.
249. Wang, X., Spandidos, A., Wang, H., & Seed, B. (2012). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic acids research*, 40(D1), D1144-D1149.
250. Moore, M. J., Schultes, C. M., Cuesta, J., Cuenca, F., Gunaratnam, M., Tanious, F. A., ... & Neidle, S. (2006). Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3, 6-and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. *Journal of medicinal chemistry*, 49(2), 582-599.
251. Miglietta, G., Cogo, S., Marinello, J., Capranico, G., Tikhomirov, A. S., Shchekotikhin, A., & Xodo, L. E. (2017). RNA G-quadruplexes in Kirsten ras (KRAS) oncogene as targets for small molecules inhibiting translation. *Journal of Medicinal Chemistry*, 60(23), 9448-9461.
252. Hoffman, E. A., Frey, B. L., Smith, L. M., & Auble, D. T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. *Journal of Biological Chemistry*, 290(44), 26404-26411.

253. Verma, A., Yadav, V. K., Basundra, R., Kumar, A., & Chowdhury, S. (2009). Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic acids research*, 37(13), 4194-4204.
254. Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D. N., & Perou, C. M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC genomics*, 15(1), 1-11.
255. Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4), 357-360.
256. Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166-169.
257. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 550.
258. Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols*, 8(9), 1765.
259. Dhapola, P., & Chowdhury, S. (2016). QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. *Nucleic acids research*, 44(W1), W277-W283.
260. Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... & McDermott, M. G. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), W90-W97.
261. Neidle, S. (2009). The structures of quadruplex nucleic acids and their drug complexes. *Current opinion in structural biology*, 19(3), 239-250.

262. Ma, K. X., Wang, H. J., Li, X. R., Li, T., Su, G., Yang, P., & Wu, J. W. (2015). Long noncoding RNA MALAT1 associates with the malignant status and poor prognosis in glioma. *Tumor Biology*, 36(5), 3355-3359.
263. Gilbert, S. L., Pehrson, J. R., & Sharp, P. A. (2000). XIST RNA associates with specific regions of the inactive X chromatin. *Journal of Biological Chemistry*, 275(47), 36491-36494.
264. Cai, Y., Sun, Z., Jia, H., Luo, H., Ye, X., Wu, Q., ... & Wan, J. (2017). Rpph1 upregulates CDC42 expression and promotes hippocampal neuron dendritic spine formation by competing with miR-330-5p. *Frontiers in molecular neuroscience*, 10, 27.
265. Kwok, C. K., Marsico, G., & Balasubramanian, S. (2018). Detecting RNA G-quadruplexes (rG4s) in the transcriptome. *Cold Spring Harbor perspectives in biology*, 10(7), a032284.
266. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh Jr, R. M., & Mergny, J. L. (2016). G-quadruplexes and helicases. *Nucleic acids research*, 44(5), 1989-2006.
267. Brázda, V., Hároníková, L., Liao, J. C., & Fojta, M. (2014). DNA and RNA quadruplex-binding proteins. *International journal of molecular sciences*, 15(10), 17493-17517.
268. Neidle, S. (2017). Quadruplex nucleic acids as targets for anticancer therapeutics. *Nature Reviews Chemistry*, 1(5), 1-10.
269. Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology*, 20(3), 300-307.
270. Lane, A. N., Chaires, J. B., Gray, R. D., & Trent, J. O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic acids research*, 36(17), 5482-5515.

271. Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., ... & Cui, Q. (2012). LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic acids research*, 41(D1), D983-D986.
272. Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., ... & Lander, E. S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell*, 159(1), 188-199.
273. Sharma, E., Sterne-Weiler, T., O'Hanlon, D., & Blencowe, B. J. (2016). Global mapping of human RNA-RNA interactions. *Molecular cell*, 62(4), 618-626.
274. Yang, S. Y., Amor, S., Laguerre, A., Wong, J. M., & Monchaud, D. (2017). Real-time and quantitative fluorescent live-cell imaging with quadruplex-specific red-edge probe (G4-REP). *Biochimica Et Biophysica Acta (BBA)-General Subjects*, 1861(5), 1312-1320.
275. Laguerre, A., Levillain, M., Stefan, L., Haudecoeur, R., Katranji, F., Pirrotta, M., & Monchaud, D. (2015). Synthetic G-Quartets as versatile nanotools for the luminescent detection of G-Quadruplexes. *CHIMIA International Journal for Chemistry*, 69(9), 530-536.
276. Rousselin, Y., Sok, N., Boschetti, F., Guillard, R., & Denat, F. (2010). Efficient Synthesis of New C-Functionalized Macrocyclic Polyamines. *European Journal of Organic Chemistry*, 2010(9), 1688-1693.
277. De Cian, A., Guittat, L., Kaiser, M., Saccà, B., Amrane, S., Bourdoncle, A., ... & Mergny, J. L. (2007). Fluorescence-based melting assays for studying quadruplex ligands. *Methods*, 42(2), 183-195.
278. Rosu, F., Gabelica, V., Houssier, C., & De Pauw, E. (2002). Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-

- stranded oligodeoxynucleotides by electrospray ionization mass spectrometry. *Nucleic Acids Research*, 30(16), e82-e82.
279. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews Genetics*. 2010;11(5):319-330.
280. Sommer A, Royle NJ. ALT: A Multi-Faceted Phenomenon. *Genes*. 2020;11(2):133.
281. Kim JY, Brosnan-Cashman JA, An S, et al. Alternative Lengthening of Telomeres in Primary Pancreatic Neuroendocrine Tumors Is Associated with Aggressive Clinical Behavior and Poor Survival. *Clinical Cancer Research*. 2016;23(6):1598-1606.
282. Singhi AD, Liu T-C, Roncaioli JL, et al. Alternative Lengthening of Telomeres and Loss of DAXX/ATRX Expression Predicts Metastatic Disease and Poor Survival in Patients with Pancreatic Neuroendocrine Tumors. *Clinical Cancer Research*. 2016;23(2):600-609.
283. Lawlor RT, Veronese N, Pea A, et al. Alternative lengthening of telomeres (ALT) influences survival in soft tissue sarcomas: a systematic review with meta-analysis. *BMC Cancer*. 2019;19(1).
284. Episkopou H, Draskovic I, Van Beneden A, et al. Alternative Lengthening of Telomeres is characterized by reduced compaction of telomeric chromatin. *Nucleic Acids Research*. 2014;42(7):4391-4405.
285. AR;Yoo J. Increased alternative lengthening of telomere phenotypes of telomerase-negative immortal cells upon trichostatin--a treatment. *Anticancer research*. 2013;33(3).
286. Lovejoy CA, Li W, Reisenweber S, et al. Loss of ATRX, Genome Instability, and an Altered DNA Damage Response Are Hallmarks of the Alternative Lengthening of Telomeres Pathway. Scott HS, ed. *PLoS Genetics*. 2012;8(7):e1002772.

287. Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. ATRX and DAXX: Mechanisms and Mutations. *Cold Spring Harbor Perspectives in Medicine*. 2017;7(3):a026567.
288. Lee M, Hills M, Conomos D, et al. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Research*. 2013;42(3):1733-1746.
289. Conomos D, Stutz MD, Hills M, et al. Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *The Journal of Cell Biology*. 2012;199(6):893-906.
290. Bandaria JN, Qin P, Berk V, Chu S, Yildiz A. Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. *Cell*. 2016;164(4):735-746.
291. Cho NW, Dilley RL, Lampson MA, Greenberg RA. Interchromosomal Homology Searches Drive Directional ALT Telomere Movement and Synapsis. *Cell*. 2014;159(1):108-121.
292. Dilley RL, Verma P, Cho NW, Winters HD, Wondisford AR, Greenberg RA. Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature*. 2016;539(7627):54-58.
293. Min J, Wright WE, Shay JW. Alternative Lengthening of Telomeres Mediated by Mitotic DNA Synthesis Engages Break-Induced Replication Processes. *Molecular and Cellular Biology*. 2017;37(20).
294. Arora R, Lee Y, Wischniewski H, Brun CM, Schwarz T, Azzalin CM. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature Communications*. 2014;5(1).

295. Silva B, Pentz R, Figueira AM, et al. FANCM limits ALT activity by restricting telomeric replication stress induced by deregulated BLM and R-loops. *Nature Communications*. 2019;10(1).
296. Yang SY, Lejault P, Chevrier S, et al. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nature Communications*. 2018;9(1).
297. Heaphy CM, Subhawong AP, Hong S-M, et al. Prevalence of the Alternative Lengthening of Telomeres Telomere Maintenance Mechanism in Human Cancer Subtypes. *The American Journal of Pathology*. 2011;179(4):1608-1615.
298. Heaphy CM, de Wilde RF, Jiao Y, et al. Altered Telomeres in Tumors with ATRX and DAXX Mutations. *Science*. 2011;333(6041):425-425.
299. Schwartzenuber J, Korshunov A, Liu X-Y, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. 2012;482(7384):226-231.
300. Lu R, O'Rourke JJ, Sobinoff AP, et al. The FANCM-BLM-TOP3A-RMI complex suppresses alternative lengthening of telomeres (ALT). *Nature Communications*. 2019;10(1).
301. Henson JD, Lau LM, Koch S, Martin La Rotta N, Dagg RA, Reddel RR. The C-Circle Assay for alternative-lengthening-of-telomeres activity. *Methods*. 2017;114:74-84.
302. Zhao L, Au JL-S, Wientjes MG. Comparison of methods for evaluating drug-drug interaction. *Frontiers in bioscience (Elite edition)*. 2010;2:241-249.
303. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & development*, 18(13), 1618-1629. (312)

304. Wang Y, Yang J, Wild AT, et al. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nature Communications*. 2019;10(1).
305. Nguyen DT, Voon HPJ, Xella B, et al. The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats. *EMBO reports*. 2017;18(6):914-928.
306. Bettin N, Oss Pegorar C, Cusanelli E. The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. *Cells*. 2019;8(3):246.
307. Phan AT. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS Journal*. 2009;277(5):1107-1117.
308. Zeng S, Xiang T, Pandita TK, et al. Telomere recombination requires the MUS81 endonuclease. *Nature Cell Biology*. 2009;11(5):616-623.
309. Li S, Lu H, Wang Z, et al. ERCC1/XPF Is Important for Repair of DNA Double-Strand Breaks Containing Secondary Structures. *iScience*. 2019;16:63-78.
310. McQuin C, Goodman A, Chernyshev V, et al. CellProfiler 3.0: Next-generation image processing for biology. Misteli T, ed. *PLOS Biology*. 2018;16(7):e2005970.
311. Stauffer W, Sheng H, Lim HN. EzColocalization: An ImageJ plugin for visualizing and measuring colocalization in cells and organisms. *Scientific Reports*. 2018;8(1).
312. Tseng, T. Y., Chu, I. T., Lin, S. J., Li, J., & Chang, T. C. (2019). Binding of small molecules to G-quadruplex DNA in cells revealed by fluorescence lifetime imaging microscopy of o-BMVC foci. *Molecules*, 24(1), 35.
313. Yangyuru, P. M., Di Antonio, M., Ghimire, C., Biffi, G., Balasubramanian, S., & Mao, H. (2015). Dual Binding of an Antibody and a Small Molecule Increases the Stability of TERRA G-Quadruplex. *Angewandte Chemie International Edition*, 54(3), 910-913.

314. Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., ... & Ballester, B. (2019). Involvement of G-quadruplex regions in mammalian replication origin activity. *Nature communications*, 10(1), 1-16.

## Appendices

### Appendix A - Cellular detection of G-quadruplexes by optical imaging methods (*Current Protocols in Cell Biology*, 2018)

#### A.1 Summary

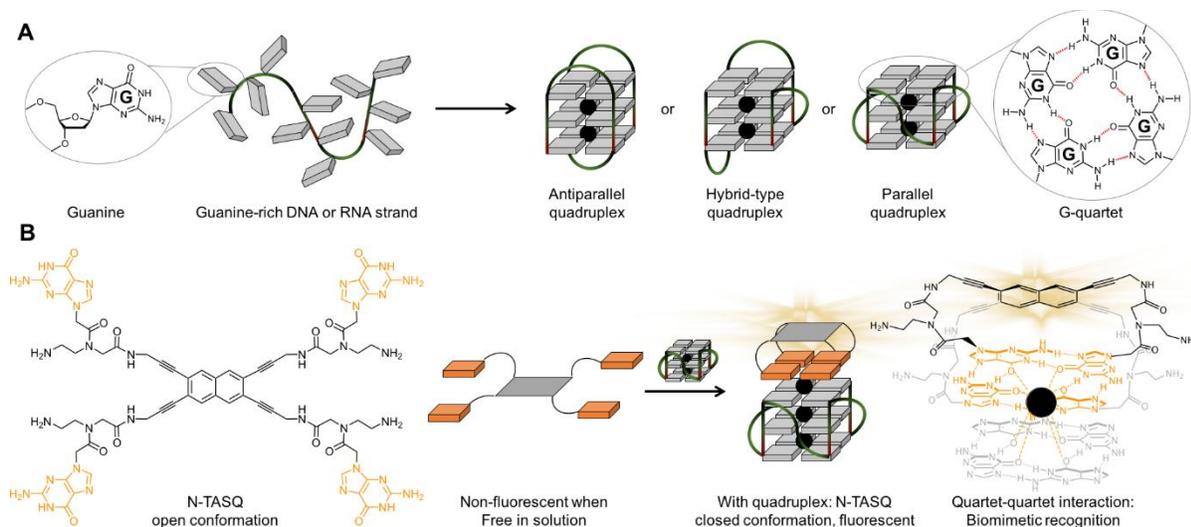
G-quadruplexes are higher-order nucleic acid structures that fold from guanine (G)-rich DNA and RNA strands. Their formation in biologically relevant G-rich sequences is suspected *in vitro* for more than 30 years now, but the first pieces of evidence of their existence *in cellulo* have been recently obtained through the development of new biochemical tools. Effective molecular devices have been designed, studied and applied, ranging from quadruplex-specific antibodies to smart fluorescent probes. This field of research is a major chemical biology area since it aims at uncovering the many key cellular mechanisms in which quadruplexes are involved; the wealth of knowledge acquired over the past three decades strongly advocates for pivotal roles in the regulation of gene expression at both transcriptional (DNA quadruplexes) and translational levels (RNA quadruplexes), but the most recent data argues for a myriad of additional actions, from chromosomal stability to the firing of replication origins, from telomere homeostasis to functional dysregulations underlying genetic diseases (including cancers and neurodegeneration). To get a comprehensive and accurate overview of the complex biology of quadruplexes, the development of cell-compatible quadruplex-selective fluorescent trackers is of great interest. We report here on the repertoire of protocols we have developed over the past years to visualize quadruplexes in cells. This achievement was possible thanks to the discovery of a novel family of versatile quadruplex-selective fluorophores, the twice-as-smart quadruplex ligands named TASQ (for

template-assembled synthetic G-quartet). TASQs interact with quadruplexes in a bioinspired manner, therefore being exquisitely selective, and label them with a high specificity. The versatility of this probe allows for multiple imaging techniques in both fixed and live cells including the use of the multiphoton microscopy, confocal microscopy and real-time fluorescent image collection.

## A.2 Introduction

**Behind the double helix of DNA.** The discovery of the structure of the DNA double-helix in 1953 by Watson, Crick, Wilkins and Franklin (1-3) created momentum across the field of genetics and biology since it provided the mechanism behind the repository of our genetic information with an identified and well-defined structure. Knowing the fine structural details of the double-helix (also called duplex-DNA, or B-DNA) allows for better understanding of how genetic information is inherited and decoded (*i.e.*, replication and transcription). It also opens new possibilities to manipulate the aforementioned genetic events with external agents. Deservingly so, this discovery attracted much attention across all scientific fields, from chemistry to genetics. Yet, this discovery also kept the limelight away from a series of results collected in the first few years following the duplex-DNA structure elucidation related to the topological diversity of nucleic acids. For instance, in 1957, Felsenfeld *et al.* reported on the self-association of three nucleic acid (RNA) strands in a triple-helix structure that was subsequently termed triplex. (4) In 1959, Hoogsteen described a base pairing arrangement that deviates from the canonical Watson-Crick association. (5) The now well-known Hoogsteen bonding utilizes other nucleotide faces, thereby offering possibilities of new nucleobase associations and forms the basis for subsequent discoveries, including base triplets (for triplexes) and quartets (for quadruplexes). This discovery was followed

by Gellert and co-workers' demonstration in 1962 by X-ray analyses that four guanines can self-associate *via* their Hoogsteen faces to form cyclic arrays of guanines consequently termed guanine quartet (or G-quartet, Fig. A.1A). (6) This is of particular interest since guanine quartet is the basic building block of a higher-order DNA/RNA-structure known as G-quadruplexes (*vide infra*, Fig. A.1). Now, more than 60 years after the discovery of the structure of the duplex-DNA, a dozen of alternative higher-order nucleic acid structures have been fully characterized *in vitro*, mainly differing from each other by the number of strands (2 for duplex *versus* 3 for triplex and three-way junction, 4 for quadruplex and four-way DNA junction). (7,8) However, while the presence of most of these alternate DNA structures is associated with genetic instability (with the notable exception of a special type of four-way DNA junction named the Holliday junction (9)), G-quadruplexes attracted most of the research interest due to their involvement in key cellular processes such as replication, transcription and translation. (10,11)



**Figure A.1.** (A) Structure of a guanine (left) and a G-quartet (right) and schematic representation of the quadruplex folding from a G-rich DNA/RNA sequence (along with the different quadruplex topologies). (B) Structure of N-TASQ and schematic representation of its open and closed conformations along with its nature-inspired interaction with a quadruplex that triggers its fluorescence.

**Quadruplexes: a new hope for cancer treatments?** For quite a long time after their structural characterization, G-quartets were merely considered a peculiarity of nucleobase association, devoid of biological relevance. However, Sen & Gilbert, in 1988, speculated that G-quartets could be involved in the organization of guanine-rich (G-rich) DNA strands, notably at telomeres. (12) They correctly anticipated that inter-strand G-quartet formation participates in the association of four G-rich strands to form higher-order architectures termed G4-DNA (subsequently named G-quadruplex-DNA or quadruplex-DNA, Fig. A.1A) whose stability is provided by the self-association (stacking) of G-quartets. The relevance of G4-DNA was eventually demonstrated in 1991 when Zahler *et al.* reported that the folding of telomeric quadruplexes prevents telomeres from being recognized by telomerase, (13) an enzyme involved in telomere repair and synthesis. (14) These results were especially interesting given that telomerase actions at the telomeres were found to be essential for oncogenesis in majority of the human cancers. (15,16) These observations have ushered in a novel way of thinking about DNA-interacting small-molecules as anticancer agents, whose therapeutic specificity relies on a structure-selective rather than a sequence-selective recognition of DNA.

**Quadruplexes: a novel level of genetic regulation.** From the beginning, quadruplexes have drawn attention from a wide scientific community including physicists, chemists, computational scientists, biologists and clinicians. New data indicate that quadruplex-forming sequences (QFS) are widespread in the human genome (more than 300,000 sequences identified *in silico*, (17-19) more than 700,000 sequences identified *via* high-throughput sequencing (20)) and in the transcriptome (more than 3,000 sequences identified in isolated polyA RNAs from HeLa cells *via* high-throughput sequencing (21)). Importantly, Hänsel-Hetsch *et al.* recently showed that the dense nuclear organization (chromatin) seriously impedes but does not abolish quadruplex

formation. (22) Interestingly, DNA QFS are particularly enriched in human genome regions with high structural significance (telomeres, replication origins, gene promoters) while being mostly absent from house-keeping gene regions. Similarly, RNA QFS are also found in strategic positions, being enriched in both 5' and 3' untranslated regions (UTR) of mRNA, and may be involved in translation regulation and mRNA processing. These results strongly suggest key structural roles for both DNA and RNA quadruplexes; however, these roles are quite difficult to comprehend, due to their antagonistic cellular consequences. (23) For instance, DNA quadruplexes might facilitate firing of replication origins (positive role) while triggering genetic instability when acting as roadblocks along DNA during DNA replication (negative role). Similarly, RNA quadruplexes might recruit initiation factors in IRES (internal ribosome entry site)-dependent translation (positive role) while hampering ribosome processivity in cap-dependent translation (negative role). While the biological relevance of quadruplexes appears to be quite paradoxical, they represent a novel level of genetic regulation, providing nucleic acids (DNA and RNA) with the possibility of spatiotemporally controlling their expression, based on mechanisms and recognition factors that remain to be fully understood.

**The quest of molecular tools to visualize quadruplexes *in vitro* and *in cellulo*.** In light of this complexity, molecular probes that could help understand the biological relevance of DNA and RNA quadruplexes are invaluable. Several quadruplex-specific fluorescent dyes have been reported over the past years, (24,25) including organic fluorophores (chiefly cyanine, porphyrin and carbazole derivatives) and metallo-organic luminophores (chiefly platinum, ruthenium and zinc complexes). However, only a few of them allow for detecting quadruplexes *i) in vitro* and *in cellulo*, *ii) in both fixed and live cells*, and *iii) in a way that is compatible with different optical imaging modalities including conventional fluorimeters and 96-well plate readers, confocal and*

two-photon microscopes as well as real-time live cell optical imagers. We recently reported on a series of biomimetic ligands named TASQ (template-assembled synthetic G-quartet, Fig. A.1B) that interact with quadruplexes according to a nature-inspired, “like-likes-like” recognition between native and synthetic G-quartets, the former belonging to the quadruplex target, the latter to the TASQ. This quadruplex/ligand interaction is thus driven by a G-quartet association reminiscent of what was initially surmised by Sen & Gilbert. The study of TASQ ligands, which is ongoing since 2011, allows us to sequentially develop a series of prototypes: the first biomimetic ligands (*i.e.*, synthetic G-quartets that interact with native G-quartets according to a bioinspired association), (26,27) the first smart quadruplex ligands (*i.e.*, synthetic G-quartets that actively assemble in the presence of their native G-quartet targets only), (28-30) and the first twice-as-smart quadruplex ligands (*i.e.*, a synthetic G-quartet that is both a smart quadruplex ligand and a smart fluorescent probe). (31,32) Of particular interest are the twice-as-smart quadruplex ligands, whose design was intended to meet the aforementioned challenges, *i.e.*, the direct detection of quadruplexes across multiple techniques, facilities and imaging conditions, with the use of a single ligand. One ligand from this family shown to be remarkably efficient was the naphthoTASQ or N-TASQ (naphthalene template-assembled synthetic G-quartet, Fig. A.1B) that was successfully used for labelling quadruplexes in quartz cuvettes and 96-well plates along with demonstrated applications in fixed and live cells using various optical imaging methods. (33-35) As discussed further below, each of these techniques/modalities/facilities was implemented to address a specific need, and the combination of these collected data is indispensable for gaining mechanistic insights into the relevance of DNA and RNA quadruplexes *in vitro* and *in cellulo*.

**Methodologies.** The implementation of different optical imaging techniques for studying quadruplexes provides answers to different, often interconnected or complementary, questions.

For example, fluorescence investigations carried out in quartz cuvettes allow for answering the question “*do this G-rich DNA (or RNA) sequence actually folds into a quadruplex structure?*” To date, only some technically demanding and relatively expensive methods, such as circular dichroism (CD), provide reliable and quick answers. Investigations performed in 96-well plates are specifically suited for cell-based studies, carried out with cellular extracts for instance, and may help determine whether quadruplex-forming oligonucleotides exist in mixtures of isolated nucleic acids. Cell imaging techniques can be used to demonstrate the existence of quadruplexes in either fixed or live cells. Quite surprisingly, this basic yet highly critical issue was not easy to address. It indeed required almost 25 years of research since the initial assumption about the biological relevance of quadruplexes to provide reliable demonstrations of their existence in human cells. Pioneering immunodetection studies in ciliates (Sty49, in 2001) (36) paved the way for subsequent investigations in human cells (BG4, in 2013). (37) These studies were however limited to fixed and permeabilized cells, which allow quadruplex-specific antibodies to access the cell nuclei. The detection of quadruplexes in live human cells was made possible by the development of small-molecule fluorophores, firstly imaged post-fixation with live-treated cells (PDS- $\alpha$ , in 2012) (38) and then imaged directly in live cells under treatment (N-TASQ, in 2015). (33) Results obtained with both quadruplex-specific antibodies and cell-permeable fluorophores were later substantiated by a growing number of reports and techniques. (*e.g.*, (39-43)) Finally, the studies performed with real-time optical imager were designed to take another leap in understanding the biological effects and kinetics of quadruplex formation as a function of time, over long incubation times (up to 6 days), as a way of better approximation of the therapeutic applications of quadruplexes (as targets) and quadruplex ligands (as agents).

### **A.3 Detection of Quadruplexes in cells: Confocal vs Two-photon Microscopes, Fixed vs Live-cell imaging**

To allow for a convenient and straightforward visualization of quadruplexes in cells, we developed a fluorescent probe named N-TASQ that can be used regardless of the cell viability state and across multiple optical imaging techniques. This probe belongs to the family of the twice-as-smart ligands, being both a smart quadruplex ligand and a smart fluorescent probe (Fig. A.1B, for a discussion about twice-as-smart ligand design and properties, see Background information). Straightforward protocols (hereafter described) have been developed for using N-TASQ in cells, which allow for labelling quadruplexes in fixed and/or live cells for optical imaging with confocal and/or two-photon microscopes in a convenient and reliable manner.

#### **Basic Protocol 1**

##### ***Materials***

- Cell lines of interest and appropriate media (according to the American Type Culture Collection, or any other cell repository/provider)
- Phosphate-buffered saline (PBS) pH 7.4
- Fixatives: cooled (-20°C) methanol (MeOH) or 4% paraformaldehyde (PFA) in PBS
- Permeabilizing solution: 0.1-0.5% (v/v) Triton X-100 in PBS
- Fluorophore N-TASQ (solubilised in sterile water)
- Mounting medium (Fluoromount-G, Southern Biotech)
- Sterilized plastic vessels routinely used in cell culture
- Removable chambered coverglass specially treated to ensure a consistent growth surface, staining and microscopic examination.

- Coverslips adapted for viewing at high magnification on the microscope
- Nail polish and absorbent paper (Kimwipes)

*Note: all solutions and equipment coming into contact with living cells must be sterile.*

*Note: volume indicated corresponds to the suppliers' instructions.*

### ***Cell culture protocol***

- 1) Cultivate cells in appropriate medium supplemented with the suited percentage (*i.e.* 5-10%) of foetal bovine serum (FBS) at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere according to standard procedures.
- 2) Transfer cells on removable chambered coverglass according to the manufacturer's instructions
- 3) Let cells recover for 24h at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere.

*Note: for optimal optical imaging, the cells should have grown to only ~50% confluency by the day of the experiment.*

### ***Cell preparation***

N-TASQ is unique in that it offers two different labelling possibilities, designed for two different purposes: N-TASQ is used either as live-cell staining agent, which requires cell incubation with a suited, non-toxic N-TASQ concentrations (2.5-5µM, cf. §I.1.c.1), or as post-fixation staining agent, which allows for using higher N-TASQ concentrations (100µM, cf. §I.1.c.2) since cells are first fixed and permeabilized. As further demonstrated below, this experimental choice has critical consequences since the former protocol leads to the straightforward visualization of RNA quadruplexes, while the latter offers a convenient way to detect both RNA and DNA quadruplexes.

#### ***N-TASQ as live-cell staining agent: visualization of RNA quadruplexes***

N-TASQ allows for the visualization of RNA quadruplexes when used as live-cell staining agent (for a discussion about the cellular targets of N-TASQ, see Background information).

- 1) Incubate cells with 2.5-10 $\mu$ M N-TASQ (depending on the selected experimental setup) overnight at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in appropriate medium supplemented with 10% FBS. Prepare controls (*i.e.*, cells without N-TASQ) concomitantly.
- 2) Remove medium and wash each chamber twice with PBS; remove the walls of the chambered coverglass according to the manufacturer's instructions (with a safety removal tool provided with the chamber).
- 3) Add a drop (~10 $\mu$ L) of PBS on each delimited chamber to make the slide ready for live cell imaging (§I.1.d).

*Note: PBS, which is not a culture medium, is used for the time required to collect images only. A fixation step is not required when N-TASQ is used as live-cell staining agent (Fig. A.2D); however, for conservation purposes, fixative and mounting agents can be employed following points 7-9 of §I.1.c.2 (Fig. A.2A,E).*

#### ***N-TASQ for post-fixation staining: visualization of DNA quadruplexes***

N-TASQ allows for the visualization of DNA quadruplexes when used as post-fixation staining agent (for a discussion about the cellular targets of N-TASQ, see Background information, for a discussion about the choice of the fixative agents, see Technical Considerations).

- 1) Fix and permeabilize cells with cold methanol (-20°C) for 15 min or with PFA 4% for 10 min at room temperature (250 $\mu$ L/well).
- 2) if PFA is used as fixative, permeabilize cells adding a solution of 0.5% Triton X-100 solution (250 $\mu$ L/well) in each chamber (15 min); wash cells carefully three times (5 min between washes) with PBS (500 $\mu$ L/well).
- 3) Remove the fixative agent and wash cells carefully three times (5min between washes) with PBS solution (500 $\mu$ L/well).

*Note: never let the cells dry out during or in between the different steps.*

- 4) Incubate cells with the required concentration of N-TASQ (5-100 $\mu$ M at 250 $\mu$ L/well, depending on the selected experimental setup) for 30min to 3h (according to the experimental set up), or with PBS only, for control samples.

*Note: along with the N-TASQ cell labelling step, a series of controls must be performed, the most important one being the evaluation of auto-fluorescence: to this end, untreated cells are fixed and permeabilized in a manner that is strictly identical to N-TASQ treated cells. Other controls might be required depending on the experimental procedures, such as the nuclear staining with DRAQ5 or DAPI dyes, for instance, with and without N-TASQ colabelling.*

- 5) Wash cells three times with PBS (500 $\mu$ L/well).
- 6) After PBS removal, remove the walls of the chambered coverglass according to the manufacturer's instructions (with a safety removal tool provided with the chamber).
- 7) Add a small drop (~10 $\mu$ L) of mounting medium (with anti-fade properties) in each delimited chamber.
- 8) Place a coverslip, adapted to the size of the coverglass slide, onto the drop; remove the excess liquid with absorbent paper.
- 9) Seal the edges of the coverslip with nail polish, to prevent desiccation and facilitate manipulation of the samples.

*Note: The fixed, mounted nail-sealed slides can be stored up to 1 year in the dark at 4°C.*

#### ***Double N-TASQ treatments: visualization of DNA/RNA quadruplexes***

N-TASQ allows for the visualization of both DNA and RNA quadruplexes when used as both live-cell and post-fixation staining agent (for a discussion about the cellular targets of N-TASQ, see Background information).

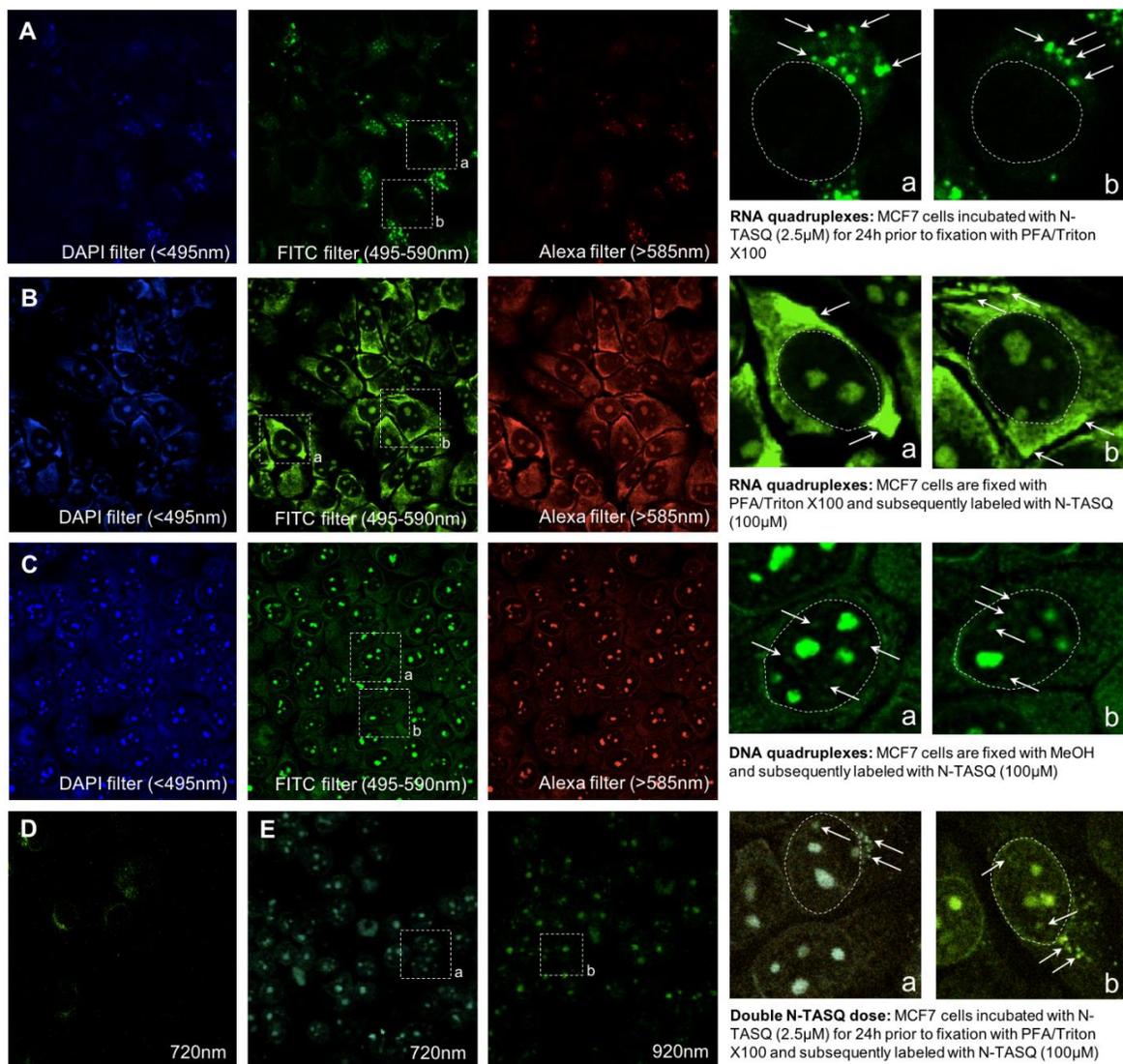
### ***Optical imaging:***

1) Analyse the prepared samples by confocal microscopy according to the manufacturer's instruction for the model of microscope.

*Note: if a Leica TCS SP8 confocal laser-scanning microscope is used, images are recorded through four detection channels: DAPI (excitation: 340-380nm, emission: <425nm), CFP (cyan fluorescent protein, excitation: 426-446nm, emission: 460-500nm), GFP (green fluorescent protein, excitation: 450-490nm, emission: 500-550nm) and CY3 (excitation: 517-562nm, emission: >580nm). If a Zeiss LSM 700 confocal laser scanning microscope is used (Fig. A.2A-C), images are collected through three detection channels, i.e., DAPI, Alexafluor 488 (AF488) and 594 (AF594) filters (lasers at 408, 488 and 555nm, and the tracks for each channel: 495 and below, 495-590 and 585nm and above).*

2) Alternatively, analyse the prepared samples by multiphoton confocal microscopy according to the manufacturer's instruction for the model of microscope.

*Note: if a Nikon AI-MP multiphoton confocal microscope is used (Fig. A.2D, E), two-photon images are recorded upon 720-920nm excitation (IR laser Chameleon, Coherent) and fluorescence emission collected through three detection channels, i.e., FF01-492/SP "DAPI channel", FF03-525/50 "FITC channel" (for fluorescein isothiocyanate) and, FF01-629/56 "Alexa channel" (Semrock).*



**Figure A.2.** Visualization of quadruplexes in cells: confocal (A-C) or two-photon images (D,E) collected with MCF7 cells live-incubated with sublethal concentration (2.5µM) of N-TASQ (A,D,E) or labelled by N-TASQ (100µM) after cell fixation (B,C), which can be done with either PFA/Triton X-100 (A,B,E) or methanol (C).

### *Quantitative analyses with automated foci count:*

Optical imaging allows for many informative measures, including the localization and the quantification of the fluorescence intensity of labeled cellular components. Several software programs can be used for quantification purposes, including ImageJ (FIJI) (44) or CellProfiler (45) (open-source software <https://fiji.sc/> and <http://cellprofiler.org/>). For the latter, algorithms for

image analysis are available as individual modules that can be placed in sequential order to form a pipeline, then used to identify and measure biological objects in optical images.

- 1) Export collected images collected as a TIFF file (with high resolution) for each separate channel and name them accordingly (*e.g.*, DAPI for excitation at 340-380nm and emission at 425nm, or GFP for excitation at 450-490nm and emission at 500-550nm).
- 2) Load a home-made pipeline for foci counting in CellProfiler for image analyses.
- 3) Import, filter and treat each image according to the pipeline's instructions.

*Note: for more information, see tutorials available at <http://cellprofiler.org/tutorials/>.*

## **Basic Protocol 2**

### ***Live-cell treatment***

- 1) Cultivate MCF-7 cells (provided by ATCC) in 75cm<sup>2</sup> tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in DMEM supplemented with 10% foetal bovine serum (Dominique Dutscher) according standard procedures.
- 2) After removal of the culture medium, harvest cells by treatment with 0.25% trypsin solution (Dominique Dutscher) and reseed them in appropriate density in chambered coverglass (8 chamber, Falcon), and let them recover for 24h at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere.
- 3) After removal of the culture medium, incubate cells with a N-TASQ solution (2.5µM in DMEM supplemented with 10% foetal bovine serum) for 24h at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere.
- 4) After removal of the culture medium, add a solution of PFA 4% (250µL) for 10 min at room temperature. After PFA solution removal (and a washing step), incubate cells with a solution

of 0.5% Triton X-100 in PBS (250 $\mu$ L) for 15min and wash them three times with PBS (500 $\mu$ L, 5min between washes).

- 5) After removal of the permeabilizing solution, remove the walls of the chamber according to the manufacturer's instructions.
- 6) Add a small drop (~10 $\mu$ L) of Fluoromount-G (Southern Biotech) onto each delimited chamber and place an adapted coverslip to cover the slide. After a small manual pressure, remove the excess of liquid with absorbent paper.
- 7) Seal the edges of the coverslip with nail polish (drying time ~10min). The slide is now ready for confocal analysis (cf. §I.1.d and Fig. A.2A).

### ***Post-fixation labelling***

- 1) Cultivate MCF-7 cells (provided by ATCC) in 75cm<sup>2</sup> tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in DMEM supplemented with 10% foetal bovine serum (Dominique Dutscher) according standard procedures.
- 2) After removal of the culture medium, harvest cells by treatment with 0.25% trypsin solution (Dominique Dutscher) and reseed them in appropriate density in chambered coverglass (8 chamber, Falcon); let them recover for 24h at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere.
- 3) After removal of the culture medium, add cold methanol (-20°C) for 15 min at room temperature to fix and permeabilize cells.
- 4) After removal of the methanol, wash the slide twice with PBS (500 $\mu$ L) and incubate cells with N-TASQ solution (5-100 $\mu$ M in 250  $\mu$ L PBS) for 1-3h at room temperature.
- 5) After removal of the N-TASQ solution, wash each well twice with PBS (500 $\mu$ L) and remove the walls of the chamber according to the manufacturer's instructions.

- 6) Add a small drop (~10 $\mu$ L) of Fluoromount-G onto each delimited chamber and place an adapted coverslip to cover the slide. After applying a small manual pressure, remove the excess of liquid with absorbent paper.
- 7) Seal the edges of the coverslip with nail polish (drying time ~10min). The slide is now ready for confocal analysis (cf. §I.1.d and Fig. A.2C).

### **Alternative Protocol**

An alternative to the removable 8-chamber coverglass is possible, using coverslips commonly used for optical imaging purposes.

### ***Additional Materials***

- Blocking solution: 4% bovine serum albumin (BSA) in PBS
- Nuclear staining: DAPI or DRAQ5 dyes
- Washing solution: PBS or PBS-0.1% Triton X-100 (PBS-T) (v/v)
- Sterilized 12-mm diameter round-glass coverslips adapted for images at high magnification
- 24 wells tissue culture plate
- Forceps and a needle
- Microscope slides
- 70% ethanol

### ***Cells preparation on routine coverslips***

- 1) Cultivate MCF-7 cells in 75 cm<sup>2</sup> tissue culture flasks (Corning) at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in DMEM supplemented with 5% foetal bovine serum (Gibco) and 100U Penicillin-Streptomycin mixture (1.0u.mL<sup>-1</sup> Pen/1.0 $\mu$ g.mL<sup>-1</sup> Strep, Gibco) according to standard protocols.

2) Harvest cells by treatment with 0.25% trypsin solution (Gibco) and reseed them in appropriate density on sterilized 12-mm diameter coverslips placed in a 24-well plate.

*Note: coverslips can be sterilized by soaking in 70% ethanol. The coverslips, handled with forceps, are placed in a sterile 24-well plate (1 coverslip per well) or in a Petri dish. Press down carefully on the coverslip to make a seal that will keep the coverslip adherent to the chamber, and dry off excess fluid. Rinse with PBS.*

**Cell preparation, i.e., cell fixation and permeabilization** (cf. also §I.1)

1) Remove the culture medium and wash cells with PBS. Add cold methanol (-20°C, 250µL) or PFA 4% (250µL) for 15 or 10 min, respectively, at room temperature to fix and permeabilize cells.

2) if PFA is used as fixative, permeabilize cells incubating them with a solution of 0.5% Triton X-100 in PBS (250µL) for 10min; wash three times carefully with PBS (500µL, 5min between washes).

*Note: fixed and permeabilized cells can be stored for further use at 4°C. Make sure that the coverslip is covered with enough PBS to avoid drying off and that the plate is sealed with paraffin.*

**Cell labelling with N-TASQ:**

*Note: particular care must be given to the cell-adherent surface faces: cells should be facing up in the 24-well plates during the staining process and down on the slides during the mounting step.*

1) Transfer the coverslips into a new 24-well plate and rinse once with PBS (500µL).

2) Incubate cells with the blocking solution (4% BSA in PBS, 500µL) for 1h at room temperature.

*Note: This step is optional for N-TASQ labelling but must be performed for costaining with antibodies (according to supplier's instructions)*

3) After removal of the blocking solution, incubate cells with 5-100 $\mu$ M N-TASQ solution (250 $\mu$ L) or with PBS only for controls samples.

*Note: the coverslip must be entirely covered by staining solution for complete staining.*

4) After removal of the N-TASQ solution, wash the coverslips three times with PBS (500 $\mu$ L, 5min between washes)

*Note: after this step, cells can be counterstained with a nuclear staining agent, such as DRAQ5, according to the supplier's instructions.*

5) The coverslip is then picked up from the 24 well-plate using forceps and a needle.

6) Add a small drop (~10 $\mu$ L) of FluoromountG on the coverglass, mount the coverslip face down on a drop of mounting medium. After a gentle manual pressure, remove the excess of liquid with absorbent paper.

7) Seal the edges of the coverslip with nail polish (drying time ~10min). The slide is now ready for optical imaging (*cf. §I.1.d*)

*Note: The fixed, mounted nail-sealed slides can be stored up to 1 year in the dark at 4°C.*

#### ***Co-staining with antibodies:***

N-TASQ can be used in combination with quadruplex-specific antibodies (such as BG4) or with antibodies against different cellular components, the probe being either used as live-cell or post-staining agent. In the following protocol, we describe the co-staining with BG4 implemented to demonstrate that N-TASQ targets quadruplexes in cells.

*Note: BG4 ScvF antibody was performed using an E. coli BL21(DE3) strain transfected with pSANG10-3F under ampicillin selection. 6xHis-tagged BG4 is purified through a single round of*

*Ni-NTA (Nickel nitrilotriacetic acid) resin binding and elution. Purified fractions were separated by SDS-PAGE and quantified with standard Bradford protein assay and validated by anti-FLAG western detection.*

- 1) Prepare cells according to the aforementioned protocol (cf. §I.1.c.1), incubating them with 2.5-5  $\mu\text{M}$  N-TASQ for 2-24h at 37 °C in a humidified, 5%  $\text{CO}_2$  atmosphere in appropriated medium supplemented with 10% FBS. Prepare controls (cells without N-TASQ) concomitantly. After culture medium removal, wash cells three times with PBS.
- 2) Fix and permeabilize cells with PFA 4% for 10 min at room temperature (250 $\mu\text{L}$ /well); wash them three times with PBS.
- 3) Incubate cells with the blocking solution (4% BSA) for 1h at room temperature.
- 4) Place the minimal volume of a solution of 10 $\mu\text{g.mL}^{-1}$  BG4 (ScFv antibody against quadruplex structures) to cover the coverslip and incubate for 3h at room temperature.
- 5) After BG4 solution removal and three PBS-T washings, incubate cells with a solution of 2 $\mu\text{g/mL}$  FLAG M2 monoclonal antibodies (Sigma) for 3h at room temperature or overnight at 4°C.
- 6) After M2 solution removal and three PBS-T washings, incubate cells with secondary anti-mouse IgG antibodies conjugated with AF594 (Molecular Probes) for 3h at room temperature.
- 7) After solution removal and three PBS-T washings, incubate cells with 100 $\mu\text{M}$  N-TASQ for 3h at room temperature and wash them 3 times in PBS.

*Note: after this step, cells can be counterstained with a standard nuclear staining agent.*

8) Mount the coverslips with Fluoromount-G on glass slides to collect confocal and/or two-photon images according to previous protocol (*cf. §I.1.d*).

*Note: when using N-TASQ, the appropriate controls are untreated cells for assessing auto-fluorescence. Other control samples are required to verify the specificity of labelling when co-localizations are performed with antibodies: unstained cells (as above) along with cells incubated with secondary antibodies only (for assessing non-specific cross-interactions).*

#### **A.4 Detection of Quadruplexes *in cellulo*: Real-time Monitoring in Live Cells**

Recent technical advances now permit real time recording of live-cell behavior (with automatic imaging and analysis over hours to weeks) within the controlled environment of a standard cell incubator *via* mostly non-invasive technology. This methodology was implemented with N-TASQ to assess whether N-TASQ could enter and/or label cells as a function of cellular transformation status (healthy versus cancerous cells, for a discussion about the cellular targets of N-TASQ, see Background information).

#### **Main Protocol**

##### ***Materials***

- Cell lines of interest and appropriate media, according to ATCC or Essen Bioscience for NucLight red lentivirus transduced cells.

*Note: we used breast cancer MDA-MB-231 cells transduced with NucLight Red lentivirus that directs the production of red fluorescent proteins (RFP) that label cell nuclei. The RFP can be visualized, for the quantification of cell growth purposes, through the red channel of the microscope ( $\lambda_{ex} = 585\text{nm}$ , emission filter: 625-705nm), leaving N-TASQ to be detected via the*

green channel ( $\lambda_{ex} = 460\text{nm}$ , emission filter: 504-544nm) where its fluorescence signal is highest. For comparison, non-cancerous foreskin fibroblast BJ cells immortalized by transducing in retrovirus-encoded hTERT (consequently termed BJ-hTERT) were used. Since these cells do not express fluorescent proteins, we used the phase-contrast channel to quantify cell confluency, as measure of growth rate.

- Growth support compatible with real-time facility (e.g., 96-well flat bottom plates)
- Counting device (e.g., cell counter Z2 Beckman Coulter)
- Sterilized plastic vessels routinely used in cell culture
- Fluorophore N-TASQ (solubilised in sterile water)
- Live Cell Monitoring and Analysis System (e.g., IncuCyte ZOOM)

*Note: among the real-time live cell optical imagers now commercially available, we used IncuCyte ZOOM system (Essen Bioscience), which comprises a microscope housed in a cell incubator and a networked external controller hard drive that gathers and processes image data (up to 2000 images per hour).*

### ***Preparation of the plates for real-time monitoring systems***

- 1) Cultivate cells at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere in appropriate medium supplemented with the suited percentage (*i.e.* 5-10%) of foetal bovine serum (FBS) according standard procedures.
- 2) Harvest cells with 0.25% Trypsin/EDTA solution (Gibco) and re-suspend them in growth media.
- 3) Count cells using the Beckman Coulter Z2 cell counter.

4) Seed cells onto a 96-well plate at the optimal density (1,000-10,000 cells/well for a 120h experiment).

*Note: preliminary experiments are necessary to determine the growth rate, to adjust the seeding density per well that would accommodate the optimal length of the experiment for each cell line.*

5) Place the 96-well plate inside the IncuCyte ZOOM, which itself is housed inside a cell incubator and let cells recover for 24h at 37°C in a humidified, 5% CO<sub>2</sub> atmosphere

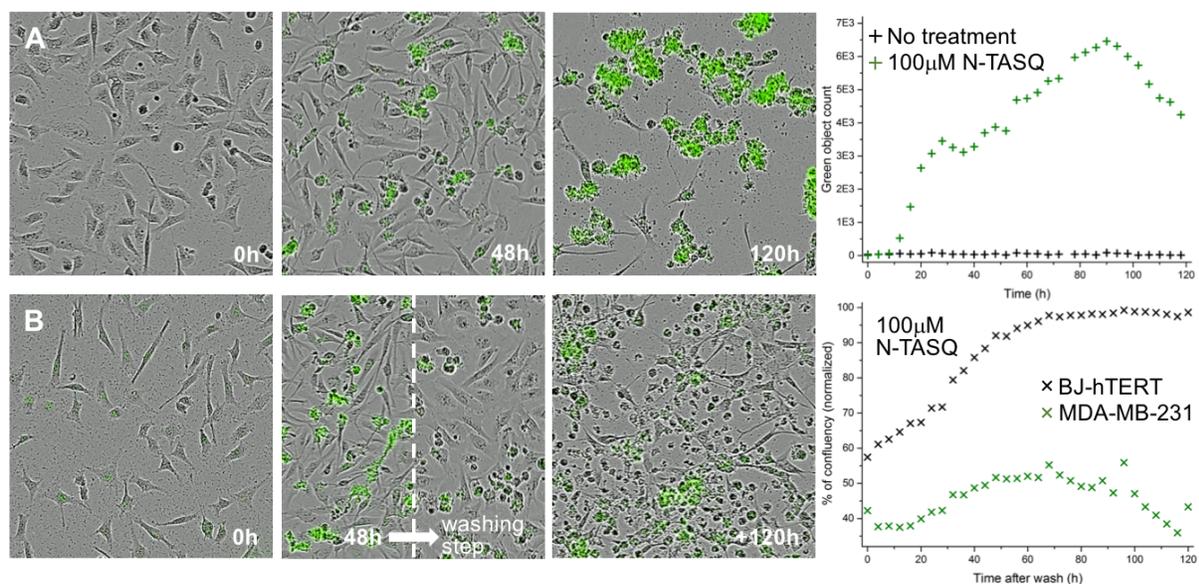
*Note: the culture plate is monitored with real-time imaging modality during this time to check for the consistency of seeding density and equal growth.*

6) Incubate cells with N-TASQ solutions (1-100µM, in triplicates) for up to 120h at 37°C in a humidified, 5% CO<sub>2</sub> atmosphere; collect images at 2 to 3-hour intervals on the red and/or phase-contrast channels for quantifying cell growth properties, and the green channel for visualizing N-TASQ fluorescence (Fig. A.3A). Prepare controls (*i.e.*, wells without N-TASQ) concomitantly.

### ***Analysis of Live-cell Imaging***

1) As indicated above, the channel for cell density quantification depends on whether cells express fluorescent proteins: for regular (parental, non-fluorescently labeled) cells, use the optical light/phase-contrast channel to calculate the cell growth confluency as a percentage of maximal growth obtained with vehicle control cells; for red fluorescent protein-expressing cells, use the red object count to determine the cell count (which can be converted to cell growth parameters during the analysis and calculations).

- 2) Normalize these output quantifications using values from control treatments. Designate the value calculated from vehicle-treated wells as the maximal reachable confluency (MaxRC), which represents the highest growth density value for the particular cell line and treatment. Designate the value calculated from wells treated with the highest N-TASQ concentration (100 $\mu$ M) as the minimal reachable confluency (MinRC), which represent the lowest growth density (background) value. Calculate the growth density values from each treatment concentrations (GD<sub>x</sub>) through the normalization of green object count (GOC) obtained for each concentration (x) to these minimum and maximum values:  $GD_x = [GOC_x - MinRC] / MaxRC$ , thus generating dose-response curves.
- 3) If multiple channels (*e.g.*, red *versus* green) are used, N-TASQ may affect the quantification of the output of the channels used to monitor cell growth. Remove the signal from N-TASQ from other channels using the spectral unmixing function. This is useful when N-TASQ is used in concert with another fluorescent marker.
- 4) Quantify the fluorescence level from bound N-TASQ using the Green Object Count function in the IncuCyte software. Export the values as a function of time and plot them using scientific graphing software (*e.g.*, Graphpad Prism, OriginPro.8).



**Figure A.3.** Real-time monitoring of cell labelling with N-TASQ over 120h-incubation, without (A) or with a washing step after 48h incubation (B), and quantification of the resulting influence on the cell growth *via* the N-TASQ fluorescence (green object count, IncuCyte ZOOM).

***Alternate protocol: including an additional washing step***

Images seen in Fig. A.3A highlight the N-TASQ propensity to adhere to and label cell surfaces. To further evaluate the anticancer activity of internalized N-TASQ, the aforementioned protocol can be modified introducing an additional washing step after 48h N-TASQ incubation. The image collection protocol remains the same. This step, shown here for both MDA-MB-231 and BJ-hTERT cells, is intended to remove externally bound N-TASQ. The internalized N-TASQ's anti-proliferative effects were demonstrated to trigger massive cell necrosis by 120h (Fig. A.3B). Importantly, N-TASQ markedly impedes MDA-MB-231 proliferation while accumulates in a subpopulation of BJ-hTERT cells without impeding cell growth, thus demonstrating treatment selectivity against transformed/cancerous cell types.

## A.5 Commentary

**Background Information.** The search for efficient quadruplex-specific dyes is currently the focus of an intense research effort, due to the fact that live cell-compatible quadruplex-probes uniquely provide information for the roles that DNA and RNA quadruplexes play in a functional genomic context. While quadruplex-specific antibodies such as BG4, (37,46,47) 1H6 (39,40,48) and more recently D1 (49) have given impetus across the field of quadruplexes thanks to the established, unambiguous target specificity of these antibodies, when applied to nuclear targets, their poor ability to cross membranes restricts their uses to applications where cellular structure and morphology are largely disrupted. Quadruplex-specific small molecule probes offer a reliable alternative, notably due to their improved cellular penetrance; however, their target specificity might be promiscuous, thereby casting doubts on their fair representation of quadruplex loci in global cellular fluorescence pattern. To tackle this issue, considerable efforts have been invested to design smart quadruplex-probes, that is, fluorophores that *actively* recognize their targets, in a manner that triggers changes in the structure of the probe itself. Some quadruplex-specific dyes such as BMVC, (50,51) DAOTA-M2 (41,52,53) and more recently c-exNDI (54) are invaluable probes to gain insights into the cellular distribution of quadruplexes but their *passive* quadruplex recognition might cast doubts on the reliability of the cellular imaging results. Alternative strategies such as hybridization (43,55) and bimolecular fluorescence complementation (56) approaches have been developed to allow for an *active* recognition of quadruplexes; however, their multipartner/multistep protocols might raise technical issues that, again, create uncertainties on the reliability of optical imaging results.

In this context, we have developed a series of nature-inspired quadruplex ligands that actively fold into their fluorescent conformation in presence of their quadruplex targets. The structure of these ligands, termed TASQ, relies on four guanines assembled around a template that form an intramolecular G-quartet upon interaction with the accessible G-quartet of a quadruplex, on the “like-likes-like” principle. A wise selection of the fluorogenic template (pyrene for PyroTASQ, naphthalene for N-TASQ) is important so that its fluorescence is released when the intramolecular G-quartet is formed only. This design makes TASQ both a smart ligand, which actively assembles when interacting with quadruplexes, and a smart probe, whose fluorescence is triggered by the ligand/quadruplex association. The preparation of the N-TASQ (used herein) has been described elsewhere; (33) its quadruplex-interacting properties have been thoroughly investigated *in vitro* along with its interesting spectroscopic properties which allows it to act both as a smart probe (*vide supra*) and as a red edge probe (also termed G4-REP, for quadruplex-selective red edge probe (34,35)). This latter property makes N-TASQ quite unique as the red edge effect (REE (57,58)) is a peculiar spectroscopic effect originating in the modification of the first hydration shell of the ligand upon intimate contacts with its nucleic acid target. This gives rise to various intermediate fluorogenic systems that could be excited by photons of lower energy, that is, shifted toward the long wavelength edge of the its absorption signature. The practical application of the REE is that it alleviates the optical restriction from designated excitation and emission wavelengths, thereby offering the possibility of using G4-REP in virtually any experimental setup. Consequently, the cell-compatible N-TASQ appears to meet all required specifications as an ideal molecular tool to track and visualize quadruplexes in live cells.

N-TASQ can label RNA and/or DNA-quadruplexes, depending on the implemented protocol. Live-incubation of cells with N-TASQ diverts RNA from normal cellular processes and triggers

the accumulation of RNA-quadruplexes in discrete cytoplasmic foci, presumably in higher-order ribonucleoprotein assemblies and/or granular cytoplasmic P-bodies. The resulting higher-order TASQ/quadruplex-RNA assemblies are easily detectable by confocal (Fig. A.2A-B) or two-photon microscopy (Fig. A.2D-E). Alternatively, when N-TASQ is used as post-fixation staining agent, it allows for DNA-quadruplex visualization. Of note, methanol fixation has been shown to be the method of choice since it allows for the detection of discrete fluorescent foci, making the analysis of the collected images more accurate and reliable (Fig. A.2C); the PFA/Triton X-100 protocol leads to rather diffuse and ill-defined cytoplasmic labelling that makes precise interpretation difficult (Fig. A.2B). Finally, N-TASQ can also be used as both live-cell and post-fixation staining agent; this double N-TASQ treatment allows for the visualization of both DNA and RNA quadruplexes, albeit with a better contrast for the latter owing to the accumulation of TASQ/quadruplex-RNA assemblies over the 24-h incubation in contrast to the snapshot of the low population density of clustered quadruplex-DNA found in cells. As above, non-toxic N-TASQ concentrations (*e.g.*, 2.5 $\mu$ M) are used for live-cell incubation, while a large excess of N-TASQ (*e.g.*, 100 $\mu$ M) can be used after cell fixation (Fig. A.2E).

N-TASQ is also suited to live-cell imaging facilities. Indeed, detection of quadruplexes through the use of antibodies or molecular probes allows for the visualization of spatial distribution of quadruplexes in cells but these techniques can only provide a snapshot or endpoint evaluation of any biological events. This poses a major limitation in our ability to resolve the temporal distribution of quadruplexes in living cells. Recent technical advances now permit real time recording of live-cell behavior (with automatic imaging and analysis over hours to weeks) within the controlled environment of a standard cell incubator *via* mostly non-invasive technology. To assess whether N-TASQ could enter and/or label cells as a function of cellular transformation

status (healthy versus cancerous cells), we performed experiments with N-TASQ incubation in a live-cell imaging facility. Through this series of experiments, we showed that while N-TASQ enters both the cancerous (MDA-MB-231) and non-transformed human cell lines (BJ-hTERT) equally well at low concentrations and for short incubation times (10 $\mu$ M, 24h), N-TASQ demonstrated propensity to adhere to the surface of cancer cells at higher doses and during extended incubation period (100 $\mu$ M, 120h). This difference in N-TASQ distribution results in massive cell death in MDA-MB-231 cells. Our data suggested the existence of tumor-specific membrane receptors, which may also promote and/or stabilize the closed conformation of N-TASQ to make it fluorescent. Further, we showed that cell surface nucleolin could be the privileged N-TASQ target, as they are being over-expressed at the surface of cancerous cells (MDA-MB-231 strongly express nucleolin while BJ cells do not), (59) and displaying high-affinity for G-quartet-containing binding agents. (60) We hypothesized that N-TASQ may interact with cell surface nucleolin through its closed, fluorescent conformation, triggering cell death in a way that is similar to the guanine-rich aptamer AS1411 developed by P. Bates and colleagues. (60) The nucleolin/TASQ interaction was investigated through co-staining experiments with N-TASQ and TAMRA-labeled AS1411. (35)

**Technical considerations.** Good fixative agents accurately preserve the three-dimensional geometry of cells. Fixation methods fall generally into two classes. On one hand, organic solvents, such as methanol, which remove lipids, precipitate proteins and dehydrate the cells; on the other hand, cross-linking agents, such as paraformaldehyde (PFA), form intermolecular bridges, normally through free amino groups, thus creating a network of linked antigens. The appropriate fixation method should be chosen according to the relevant application. While cross-linkers

preserve cell structure better than organic solvents, they might reduce the antigenicity of some cell components, and require an additional permeabilization step: this step removes more cellular membrane lipids to allow large molecules, such as antibodies, to access their cellular target. Triton X-100 is commonly used at 0.1–0.5% (v/v, in a suitable buffer, generally PBS) to permeabilize nuclear membranes, making them accessible to a variety of probes. With organic solvents, the fixation and permeabilization steps are combined but some epitopes might be very sensitive to methanol, resulting in disruption of epitope structure and pattern. Of note, mounting medium is also of high importance and Fluoromount-G was found to be the medium of choice for N-TASQ studies.

**Critical Parameters.** TASQ dyes, including PyroTASQ and N-TASQ, are not commercially available at present, thus precluding their wider, large-scale use for cell imaging purposes. Beyond our own studies detailed here, many laboratories worldwide are currently implementing TASQ as quadruplex-specific probes for their respective investigations; we hope that these ongoing studies and their resulting outcomes will further support the generality of N-TASQ as cell-compatible quadruplex trackers, while we are working on making TASQ accessible to all. Of note, TASQ are stable compounds when solid (pale yellow powder) if stored under inert atmosphere (preferentially at 4°C); the corresponding water solution (usually 1-10mM) must be kept at 4°C (preferentially -30°C) and used within 2-6 months after preparation (avoiding freezing-thawing cycle).

## A.6 References

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids. *Nature*, **171**, 737-738.
2. Wilkins, M.H.F., Stokes, A.R. and Wilson, H.R. (1953) Molecular structure of deoxyribose nucleic acids. *Nature*, **171**, 738-740.

3. Franklin, R.E. and Gosling, R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740-741.
4. Felsenfeld, G., Davies, D.R. and Rich, A. (1957) Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.*, **79**, 2023-2024.
5. Hoogsteen, K. (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.*, **12**, 822-823.
6. Gellert, M., Lipsett, M.N. and Davis, R.D. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013-2018.
7. Wang, G. and Vasquez, K.M. (2006) Non-B DNA structure-induced genetic instability. *Mutation Res.*, **598**, 103-119.
8. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature Rev. Genet.*, **13**, 770-780.
9. Liu, Y. and West, S.C. (2004) Happy Hollidays: 40th anniversary of the Holliday junction. *Nat. Rev. Mol. Cell Biol.*, **5**, 937-944.
10. Maizels, N. (2015) G4-associated human diseases. *EMBO Rep.*, e201540607.
11. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627-8637.
12. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364-366.
13. Zahler, A.M. and Williamson, J.R. (1991) Inhibition of telomerase by G-quartet DNA structure. *Nature*, **350**, 718.
14. Blackburn, E.H., Greider, C.W. and Szostak, J.W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nature Med.*, **12**, 1133-1138.
15. Counter, C.M., Hirte, H.W., Bacchetti, S. and Harley, C.B. (1994) Telomerase activity in human ovarian carcinoma. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 2900-2904.
16. Kim, N.W., Piatyszek, M.A., Prowse, K.R. and Harley, C.B. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011.
17. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901-2907.
18. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908-2916.
19. Bedrat, A., Lacroix, L. and Mergny, J.-L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.*, **44**, 1746-1759.
20. 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. *Nature Biotechnol.*, **33**, 877-881.

21. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S. and Balasubramanian, S. (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature Meth.*, **13**, 841-844.
22. Hänsel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H. and Narita, M. (2016) G-quadruplex structures mark human regulatory chromatin. *Nature Genet.*, **48**, 1267-1272.
23. Tarsounas, M. and Tijsterman, M. (2013) Genomes and G-Quadruplexes: For Better or for Worse. *J. Mol. Biol.*, **425**, 4782-4789.
24. Largy, E., Granzhan, A., Hamon, F., Verga, D. and Teulade-Fichou, M.-P. (2013) Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes. *Quadruplex Nucleic Acids*, **330**, 111-177.
25. Vummidi, B.R., Alzeer, J. and Luedtke, N.W. (2013) Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem*, **14**, 540-558.
26. Stefan, L., Guedin, A., Amrane, S., Smith, N., Denat, F., Mergny, J.-L. and Monchaud, D. (2011) DOTASQ as a prototype of nature-inspired G-quadruplex ligand. *Chem. Commun.*, **47**, 4992-4994.
27. Xu, H.-J., Stefan, L., Haudecoeur, R., Sophie, V., Richard, P., Denat, F., Barbe, J.-M., Gros, C.P. and Monchaud, D. (2012) Porphyrin-templated synthetic G-quartet (PorphySQ): a second prototype of G-quartet-based G-quadruplex ligand. *Org. Biomol. Chem.*, **10**, 5212-5218.
28. Haudecoeur, R., Stefan, L., Denat, F. and Monchaud, D. (2013) A Model of Smart G-Quadruplex Ligand. *J. Am. Chem. Soc.*, **135**, 550-553.
29. Laguerre, A., Desbois, N., Stefan, L., Richard, P., Gros, C.P. and Monchaud, D. (2014) Porphyrin-Based Design of Bioinspired Multitarget Quadruplex Ligands. *Chemmedchem*, **9**, 2035-2039.
30. Haudecoeur, R., Stefan, L. and Monchaud, D. (2013) Multitasking Water-Soluble Synthetic G-Quartets: From Preferential RNA-Quadruplex Interaction to Biocatalytic Activity. *Chem. Eur. J.*, **19**, 12739-12747.
31. Laguerre, A., Stefan, L., Larrouy, M., Genest, D., Novotna, J., Pirrotta, M. and Monchaud, D. (2014) A Twice-As-Smart Synthetic G-Quartet: PyroTASQ Is Both a Smart Quadruplex Ligand and a Smart Fluorescent Probe. *J. Am. Chem. Soc.*, **136**, 12406-12414.
32. Zhou, J., Roembke, B.T., Paragi, G., Laguerre, A., Sintim, H.O., Guerra, C.F. and Monchaud, D. (2016) Computational understanding and experimental characterization of twice-as-smart quadruplex ligands as chemical sensors of bacterial nucleotide second messengers. *Sci. Rep.*, **6**, 33888.
33. Laguerre, A., Hukezalie, K., Winckler, P., Katranji, F., Chanteloup, G., Pirrotta, M., Perrier-Cornet, J.-M., Wong, J.M. and Monchaud, D. (2015) Visualization of RNA-quadruplexes in live cells. *J. Am. Chem. Soc.*, **137**, 8521-8525.

34. Laguerre, A., Wong, J.M.Y. and Monchaud, D. (2016) Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic mechanism. *Sci. Rep.*, **6**, 32141.
35. Yang, S.Y., Amor, S., Laguerre, A., Wong, J.M. and Monchaud, D. (2016) Real-time and quantitative fluorescent live-cell imaging with quadruplex-specific red-edge probe (G4-REP). *Biochim. Biophys. Acta*, DOI: 10.1016/j.bbagen.2016.11.046
36. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8572-8577.
37. 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.
38. 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.
39. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860-869.
40. Hoffmann, R.F., Moshkin, Y.M., Mouton, S., Grzeschik, N.A., Kalicharan, R.D., Kuipers, J., Wolters, A.H.G., Nishida, K., Romashchenko, A.V., Postberg, J. *et al.* (2015) Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.*, **44**, 152-163.
41. Shivalingam, A., Izquierdo, M.A., Marois, A.L., Vysniauskas, A., Suhling, K., Kuimova, M.K. and Vilar, R. (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.*, **6**, 8178.
42. Huang, W.-C., Tseng, T.-Y., Chen, Y.-T., Chang, C.-C., Wang, Z.-F., Wang, C.-L., Hsu, T.-N., Li, P.-T., Chen, C.-T., Lin, J.-J. *et al.* (2015) Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.*, **43**, 10102-10113.
43. Chen, S.-B., Hu, M.-H., Liu, G.-C., Wang, J., Ou, T.-M., Gu, L.-Q., Huang, Z.-S. and Tan, J.-H. (2016) Visualization of NRAS RNA G-quadruplex structures in cells with an engineered fluorogenic hybridization probe. *J. Am. Chem. Soc.*, **138**, 10382-10385.
44. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Meth.*, **9**, 676-682.
45. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J. and Golland, P. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.*, **7**, R100.
46. Biffi, G., Di Antonio, M., Tannahill, D. and Balasubramanian, S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.*, **6**, 75-80.

47. 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.
48. Artusi, S., Perrone, R., Lago, S., Raffa, P., Di Iorio, E., Palù, G. and Richter, S.N. (2016) Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. *Nucleic Acids Res.*, **44**, 10343-10353.
49. Liu, H.-Y., Zhao, Q., Zhang, T.-P., Wu, Y., Xiong, Y.-X., Wang, S.-K., Ge, Y.-L., He, J.-H., Lv, P. and Ou, T.-M. (2016) Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chemical Biology*, **23**, 1261-1270.
50. Huang, W.-C., Tseng, T.-Y., Chen, Y.-T., Chang, C.-C., Wang, Z.-F., Wang, C.-L., Hsu, T.-N., Li, P.-T., Chen, C.-T. and Lin, J.-J. (2015) Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.*, **43**, 10102-10113.
51. Tseng, T.-Y., Chien, C.-H., Chu, J.-F., Huang, W.-C., Lin, M.-Y., Chang, C.-C. and Chang, T.-C. (2013) Fluorescent probe for visualizing guanine-quadruplex DNA by fluorescence lifetime imaging microscopy. *J. Biomed. Opt.*, **18**, 101309.
52. Shivalingam, A., Vyšniauskas, A., Albrecht, T., White, A.J., Kuimova, M.K. and Vilar, R. (2016) Trianguleniums as Optical Probes for G-Quadruplexes: A Photophysical, Electrochemical and Computational Study. *Chem. Eur. J.*, **22**, 4129-4139.
53. Kotar, A., Wang, B., Shivalingam, A., Gonzalez-Garcia, J., Vilar, R. and Plavec, J. (2016) NMR Structure of a Triangulenium-Based Long-Lived Fluorescence Probe Bound to a G-Quadruplex. *Angew. Chem. Int. Ed.*, **55**, 12508-12511.
54. Doria, F., Nadai, M., Zuffo, M., Perrone, R., Freccero, M. and Richter, S.N. (2017) A red-NIR fluorescent dye detecting nuclear DNA G-quadruplexes: in vitro analysis and cell imaging. *Chem. Commun.*, **53**, 2268-2271.
55. Meguellati, K., Koripelly, G. and Ladame, S. (2010) DNA-Templated Synthesis of Trimethine Cyanine Dyes: A Versatile Fluorogenic Reaction for Sensing G-Quadruplex Formation. *Angew. Chem., Int. Ed.*, **49**, 2738-2742.
56. Liu, H.-h., Zheng, K.-w., Chen, Q., Hao, Y.-h. and Tan, Z. (2016) RNA G-quadruplex formation in defined sequence in living cells detected by bimolecular fluorescence complementation. *Chem. Sci.*, **7**, 4573-4581.
57. Demchenko, A.P. (2002) The red-edge effects: 30 years of exploration. *Lumin.*, **17**, 19-42.
58. Haldar, S., Chaudhuri, A. and Chattopadhyay, A. (2011) Organization and dynamics of membrane probes and proteins utilizing the red edge excitation shift. *J. Phys. Chem. B*, **115**, 5693-5706.
59. Hovanesian, A.G., Soundaramourty, C., El Khoury, D., Nondier, I., Svab, J., Krust, B. (2010) Surface expressed nucleolin is constantly induced in tumor cells to mediate calcium-dependent ligand internalization, *PLoS One*, **5**, e15787.

60. Bates, P.J., Laber, D.A., Miller, D.M., Thomas, S.D., Trent, J.O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer, *Exp. Mol. Pathol.*, **86**, 151-164.

## **Appendix B - Transcriptomic Profiling of G-quadruplexes using G4RP: a detailed protocol and optimization guide**

### **B.1 Summary**

Guanine-rich RNAs can fold into four-stranded structures, termed G-quadruplexes (G4-RNAs), and participate in a wide range of biological processes. Here, we describe in detail the G4-RNA-specific precipitation (G4RP) protocol, which allows the transcriptomic profiling of G4-RNAs. The G4RP protocol consists of a chemical crosslinking step, followed by affinity capture with the G4-specific probe, BioTASQ. G4RP can be coupled with sequencing to capture a comprehensive global snapshot of folded G4-RNAs. This method can also be used to profile induced changes (i.e. through G4 ligand treatments) within the G4-RNA transcriptome. The entire protocol can be completed in 1-2 weeks and can be scaled up or down depending on the specific experimental goals. In addition to the protocol details, we also provide here a guide for optimization in different lab setups.

### **B.2 Introduction**

G-quadruplexes (G4s) are alternate nucleic acid structures formed by Hoogsteen bonding of four guanines to form planar guanine quartet (G-quartet) units, which  $\pi$ -stack on each other to form the overall columnar structure. G4s forming in DNAs (G4-DNAs) and RNAs (G4-RNAs) can affect the corresponding biological processes. G4-forming sequences are within the genome and transcriptome of several prokaryotes, eukaryotes and even viruses (1). Importantly, G4-forming sequences are highly saturated in sites of high disease-relevance such as oncogenes, telomere, and repeat expansion regions. Biochemical and genomic techniques have shown that G4 formation is widespread in both the human genome (2-4) and the transcriptome (5-6). However,

accurately quantifying and mapping G4 sites formed within living cells continue to be an ongoing challenge. It is important to note that G4s formation is often outcompeted by Watson-crick duplex formation, protein-binding, and other G4-destabilizing events; thus, the G4 folding process requires windows of opportunities when the polynucleotides are single-stranded and in more openly accessible conformations. Formation of G4s likely permeates throughout most, if not all, important biological events in the flow of genetic information. G4-DNAs may play roles in various biological functions, including telomere maintenance, transcription regulation, epigenetic regulation, replication and recombination (7). Despite the increasing volume of research in the G4 field, much of the previous focus has been on G4-DNA rather than G4-RNA. Given the readily single-stranded nature of RNAs and superior thermostability of G4-RNA (8), it is anticipated that G4-RNAs constitute a significant proportion of all folded G4s within living cells. G4-RNA has also been implicated in a wide range of biological processes, including translation, RNA processing, splicing, and ribonucleoprotein (RNP) formation (8). Here, we discuss some of the current techniques used to study G4s and highlight our recently developed affinity-capture technique, called the G4-RNA-specific precipitation (G4RP) with sequencing (G4RP-seq), for studying G4-RNAs on a transcriptome-wide level.

#### *Current methods for studying genomic and transcriptomic G4s in cells*

While interest in G4 has been multiplying, there is still a limited number of techniques for studying G4s on a genome- and transcriptome-wide scale. For genome-wide studies, G4-seq combines polymerase-stop assay with next-generation sequencing under cation conditions that favour (using K<sup>+</sup>) or disfavour G4 formation (using Li<sup>+</sup>) using purified, protein-free DNA (3). As folded G4s can cause polymerase-stalling, quantification of these polymerization stop sites can infer G4 folding state and generate a genomic G4 map. While the resolution is high, a major

limitation of G4-seq is the *in vitro* experimental setting, which is not entirely representative of the native chromatin state. Genome-wide G4 mapping by performing chromatin-immunoprecipitation sequencing (G4-ChIP-seq) using BG4, a single-chain variable fragment antibody specific for G4, allows approximation of a more biologically relevant (i.e. “*in vivo*”) setting (4). In comparison, G4-seq observed over 750,000 potential G4 sites (3), whereas G4-ChIP-seq reported only 10,000 enriched sites (4). This large discrepancy between the two techniques emphasizes the importance of the local chromatin environment in the formation of G4-DNAs. A recently developed sequencing technique, rG4-seq, which is conceptually similar to G4-seq, has been implemented for studying transcriptome-wide G4-RNAs (5). This technique exploits the effect of G4-RNAs on the processivity of reverse-transcriptase (RT) in complementary DNA synthesis. By mapping and comparing RT polymerization stop sites under G4-favourable and unfavourable conditions *in vitro*, thousands of potential G4-RNA sites were discovered (5, 9). As the query of G4 formation using purified RNA will not recapitulate the contributions of other biological factors (e.g. RNA-binding proteins, other nucleic acid structures.), rG4-seq possesses much of the same limitations as G4-seq. To that end, the RT stop-profiling method has been adapted to study G4-RNAs *in vivo* with an additional step of treating cells with dimethyl sulfate (DMS) before purification and RT. DMS-modified guanine bases are unable to fold into G4s *in vitro* and thus can be used to infer folding state *in vivo* indirectly (9). Surprisingly, DMS/RT-stop profiling showed that at equilibrium, G4-RNA are nearly all unfolded *in vivo*, a significant departure from live-cell imaging and *in vitro* quantification methods. This apparent discrepancy is likely due to the nature of the DMS/RT-stop profiling assay, which may be more suitable for establishing the shift of G4 equilibrium rather than capturing a snapshot of the folded G4 transcriptome. Drawing inspirations from G4-ChIP-seq,

we developed the G4RP-seq protocol, which allows a broad-level snapshot view of the folded G4 transcriptome (6). Using G4RP-seq, we showed that G4-RNAs fold in a sequence-dependent manner and that treatment with G4-stabilizing ligands enhances these G4 folded states. Thus, the current model depicts G4-RNAs as short-lived structures which are continuously being folded and unfolded by various regulatory mechanisms inside living human cells. These findings present us with a new perspective in which G4-RNAs are portrayed more as events rather than simple physical structures.

*Development of the Biotinylated Template-Assembled Synthetic G-quartet (BioTASQ) and G4RP Protocol*

A common recurring challenge in designing G4 ligands has been the lack of specificity due to unintended off-target binding. A class of biomimetic probes, called the Template-Assembled Synthetic G-quartet (TASQ), was developed to mitigate these off-target binding events. The chemical design of TASQ, inspired by the natural formation of G-quartets in G4s, consists of a central core with four arms, each attached to guanine. These guanines self-assemble into a synthetic G-quartet upon association with a G4 target through end-quartet stacking. Due to the use of the synthetic G-quartet itself as the recognition molecule, it provides high specificity and defined binding properties. A caveat of TASQ is its preference for targeting G4s with a parallel conformation due to the accessible end-quartets for TASQ binding without steric hindrance from the loops. This binding preference is especially useful for targeting G4-RNAs due to their parallel-conformational constraints (8). The extra utility of TASQ molecules is demonstrated by swapping the central core with different application-specific moieties (10-13). The newest generation of TASQ molecules is a biotinylated version of TASQ (BioTASQ) (3,14), allowing for affinity capture of G4 targets using streptavidin-biotin interactions. G4RP is derived from a

standard RNA immunoprecipitation (RIP) protocol (15); however, instead of antibody-protein recognition in RIP, G4RP uses a ligand-nucleic acid structure recognition with BioTASQ. Like RIP and other similar methods, the G4RP protocol can be coupled with high-throughput sequencing to profile for transcriptomic G4s.

#### Advantages and Limitations of G4RP

Using G4 ligands as baits for G4 targets has previously been explored (16); however, specificity for the intended G4 targets has been an ongoing challenge for most, if not all, G4 ligands.

BioTASQ was designed with these limitations in mind by providing higher specificity due to the strictly defined binding mechanism (10). The use of a synthetic G-quartet as the bait provides the best-predicted interactions given that G-quartet stacking interaction is obligatory and exclusive for all G4 formations and stabilization.

Some of the key advantages and limitations of G4-RNA quantification assays stem from whether the approach is direct or indirect. Table B.1 outlines some of the key differences between the two methods. We consider G4RP a direct approach to detect G4s due to the probe's direct association with the G4 target. However, direct approaches do have the potential pitfall of probe-assisted stabilization of the G4 targets. Since BioTASQ has properties of a G4 ligand, it conceivably can stabilize G4 targets, resulting in quantification biases in capturing G4-RNAs that are not representative of their true biological states. Therefore, the *in vivo* crosslinking step, prior to cell lysis and BioTASQ-assisted pulldown in G4RP, is crucial in minimizing this bias by preventing G4-folding *in vitro*. In contrast, indirect approaches, including rG4-seq (5) and DMS/RT-stop profiling (9), involves nucleic acid modifications prior to subsequent screening using reverse transcriptase. Indirect approaches heavily rely on the following two assumptions. First, modifying chemicals are assumed to access all cellular RNA and modify unprotected sites

uniformly within the transcriptome of living cells, which is unlikely. Second, *in vitro* folded G4s are presumed to be the sole cause of all RT stop sites, which may lead to overestimation of the actual G4 sites due to the inclusion of potential polymerase-stalling effects by other RNA structures. As indirect approaches are likely biased against highly dynamic RNA structures, the extent to which they can be used to interrogate stably structured RNAs is also unclear (17).

Depending on the experimental conditions, when the labelling chemicals are usually not the limiting reagent, chemical modification in live cells can often lead to overwhelmingly modified RNAs. Thus, G4RP is considered a more appropriate method for capturing transiently formed G4 structures, uncoupled from the influence from the equilibrium of the folded versus unfolded state. Finally, the currently reported indirect methods do not consider the formation of intermolecular G4-RNA since the principle of RT-stop profiling is based on the effect of the cis-acting intramolecular G4s. In this sense, G4RP provides measurements that are more representative of the biologically relevant interactions due to its ability to reporting on both intra- and inter-molecular folded G4-RNAs.

A major limitation of G4RP is the lack of mapping resolution due to the sonication step. The main challenge lies in striking a balance between over-sonication and under-sonication. Under-sonication prevents the proper release of cellular content, reduction in resolution, and lower pulldown efficiency. In contrast, over-sonication increases mapping resolution but lead to inconsistency and disruption of the epitope necessary for BioTASQ binding. Optimization to reach the specific goals of the experiment is highly recommended. While the sequence analysis of BioTASQ-enriched RNA fraction in G4RP shows robust enrichment of G-rich sequences, it should be noted that a minor portion of indirect interactions could be crosslinked and captured.

While G4RP does offer several advantages over the indirect polymerase-dependent methods, the choice of G4-RNA quantification method depends on the biological questions of interest. It is important to emphasize that no individual method is without limitations and caveats, and that necessary future steps should be invested in combining complementary methods in order to mitigate some of the individual limitations, thus providing a more unbiased view of the G4-RNA transcriptome.

<b>Table B.1. Indirect vs direct methods in studying transcriptomic G4-RNAs</b>		
	<b>Indirect Method (rG4-seq, DMS/RT-profiling)</b>	<b>Direct method (G4RP-seq)</b>
<b>Readout method</b>	Reverse-transcription stalling + sequencing	Probe-binding/extraction + sequencing
<b>Methods to capture biological state</b>	Modification of nucleic acids in live cells	Chemical crosslinking
<b>Resolution</b>	High	Low
<b>Targets</b>	-Intramolecular G4s	-Intramolecular G4s -Intermolecular G4s
<b>Enrichment method</b>	-Poly (A) enrichment	-Ribo-depletion
<b>Biases/ Limitations</b>	- <i>in vitro</i> biases -DMS modification could lead to underestimation of dynamic structures Overestimate by counting stops induced by other structures Not able to read 5' G4 if RT stop by 3' G4 signal is strong	-Probe-assisted stabilization could lead to overestimation -a small portion of crosslinked indirect interactions could be captured

### Future applications

Theoretically, G4RP could be combined with chemical modification and *in vitro* RT-stop profiling to generate a high-resolution map of transcriptome-wide G4-RNA sites; however, sample processing compatibilities and assay optimization between these methods have not yet been tested. The use of different crosslinking agents (eg. Psoralen) with different effective distance could also be informative to elucidating a complete picture of the G4-RNA interactions. Integration of G4RP-seq with other methods for identifying RNA-RNA interactions, such as LIGR-seq (18), MARIO (19), SPLASH (20), and PARIS (21), may aid in identifying long-range interactions through G4s. G4RP may also be potentially useful in identifying protein partners that may interact with G4-RNAs.

## **B.3 Materials**

### **Biological materials**

Cell lines of interest. We have used G4RP successfully in variety of human cell lines of different types including embryonic kidney (HEK293), human breast cancer (MCF7, MDA MB-231), human cervical cancer (HELA), and colon cancer (HT29). With proper optimization of later steps, we expect G4RP to be compatible with most organisms. Here, our protocol focuses on cultured cell lines. The use of tissue may also be compatible but will require additional optimization of tissue homogenization, cross-linking and sonication steps.

### **Reagents**

Most standard molecular and cell biology grade reagents can be obtained from various vendors and do not need to be from the ones specified here.

### Cell culturing

- Dulbecco's Modified Eagle Medium (DMEM) (Gibco, cat. no. LS11965092 or vendors of choice)
- Fetal bovine serum (FBS) (GE Lifescience, cat. no. SH3039602 or vendors of choice)
- Penicillin-streptomycin (P/S) (Gibco, cat. no. 151140148 or vendors of choice)

### Standard buffer components in the G4RP Protocol

- Nuclease-free water (from vendors or DEPC-treated inhouse)
- Formaldehyde (37% wt/vol) (Sigma) (keep away from light and air; aliquot into vials)
- Glycine (Fisher, cat. no. G45-212 or vendors of choice)
- 10X PBS (Fisher, cat no. BP39920 or vendors of choice)
- HEPES (Sigma, cat. no. H3375 or vendors of choice)
- NaCl (Fisher, cat. no. S271-10 or vendors of choice)
- EDTA (Sigma, cat. no. E5134 or vendors of choice)
- EGTA (Sigma, cat. no. E3889 or vendors of choice)
- KCl (Fisher, cat. no. P330 or vendors of choice)
- DTT (Fisher, cat. no. R0861 or vendors of choice)
- IPEGAL (NP40) (Sigma, cat. no. I3021 or vendors of choice)
- Tris (Fisher, cat. no. BP152 or vendors of choice)
- Chloroform (Sigma, cat. no. C2432 or vendors of choice)
- Isopropyl alcohol (Sigma, cat. no. I9516 or vendors of choice)
- Ethyl alcohol
- RiboLock RNA inhibitor (Thermos, cat. no. EO0381, or vendors of choice)

- Protease-inhibitor cocktail (PIC) (optional)
- TRIZOL (Invitrogen, cat. no. 15596026)

#### G4RP-specific reagents

- Biotin (Sigma, cat. no. 2031)
- BioTASQ
- Promega Magnosphere Paramagnetic (Streptavidin beads) (Promega, cat. no. Z5481)

#### Reverse Transcription and PCR

- Superscript III (Invitrogen, cat. no. 18080093)
- Random hexamer pd(N)<sub>6</sub> (Invitrogen, cat no. N8080127, or vendors of choice)
- dNTP mix (NEB, cat no. N0447L, or vendors of choice)
- SYBR green PCR-mix (BioRad, cat. no. 1725120, or vendors of choice)

#### Other consumables

- Cell and tissue culture dishes (vendor of choice)
- Cell culture treated flasks (vendor of choice)
- Cell scrapers (vendor of choice)
- Wide-mouth tips (vendor of choice)

#### Sequencing kits

- RNEasy Plus RNA purification kit (Qiagen, cat. no. #74134)

Note: Illumina Ribo-Zero has been discontinued and incorporated into the TruSeq Stranded Total RNA. Use EITHER the old Illumina Ribo-Zero + TruSeq RNA library kit OR the new Illumina Stranded Total RNA library prep kit.

- Old version: Illumina Ribo-Zero (discontinued and incorporated into the TruSeq Stranded Total RNA library prep) + TruSeq RNA library prep kit (Illumina, cat. no. RS-122-2001)

- New version: TruSeq Stranded Total RNA library prep (Illumina, cat. no. 20020596)

## **Equipment**

### Lab equipment

- Standard cell culture incubator (eg. HERACell)
- Standard Benchtop centrifuge
- Temperature controlled centrifuge (or a standard unit in a refrigerated environment)
- Standard laboratory rocker
- Standard benchtop aspirator
- Cell sonicator of choice (eg. Covaris m220 ultrasonicator)
- Tube rotator
- Nanodrop spectrophotometer (Thermos)
- Thermal cycler (Bio-Rad)
- Real-time PCR detection system (Bio-Rad)
- Illumina HiSeq, NextSeq or NovaSeq sequencer

### Bioinformatics Tools

- Trimmomatics (22) - <https://github.com/timflutre/trimmomatic>
- HISAT2 (23) - <http://daehwankimlab.github.io/hisat2/>
- Integrative Genomics Viewer (24) - <http://software.broadinstitute.org/software/igv/>
- HTSeq-count (25) - <https://htseq.readthedocs.io/en/master/>
- DESeq2 (26) - <https://doi.org/doi:10.18129/B9.bioc.DESeq2>

## Reagent Setup

CAUTION: All buffers should be sterile and carefully prepared using RNase-free water.

- DEPC H<sub>2</sub>O (0.1% DEPC-treated water can be made by adding 1mL DEPC to 1000mL of ddH<sub>2</sub>O. Mix and incubate at room temp for 1h before autoclaving.)
- DEPC-PBS (Dilute 10X PBS stock in DEPC H<sub>2</sub>O or nuclease-free water from a vendor; sterilized)
- Fixing buffer: 50 mM HEPES KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0 in DEPC H<sub>2</sub>O (A 5X stock solution can be made.)
- 1M Glycine (in DEPC H<sub>2</sub>O and sterile filtered, stored at 4°C for a few weeks)
- G4RP Buffer: 150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40 (made with DEPC H<sub>2</sub>O or nuclease-free water) (Transfer amount required for the experiment from the bottle to a conical tube and add RNAase inhibitor fresh for every new experiment. Use the manufacturer-recommended amount of RNAase ~100 U/mL. 1X PIC may be added if the downstream application involves protein).
- Biotin: make 1 mM stock using pure ddH<sub>2</sub>O
- BioTASQ: make 1 mM stock using pure ddH<sub>2</sub>O (can be stored at 4°C for a few months)
- Streptavidin MagneSphere Paramagnetic particles (Promega): Always mix well before transferring beads with a pipette using wide-mouth pipette tips. The magnetic beads are stored in a stabilizing buffer (with BSA). You may gently mix well with tapping and aliquot to small volumes for later use. For a new experiment, wash in DEPC-PBS 3x with gentle tapping, use a magnetic rack, aspirate the buffer and resuspend in fresh DEPC-PBS to make working stock (for one supplied 0.6 mL tube of beads, resuspend in 100 µL

of DEPC-PBS to make 6 mg/mL). For optimal performance, new working stocks should be prepared for each new experiment.

#### **B.4 Procedures**

The G4RP protocol itself can be completed in two days. Prior to G4RP, cell culture and live-cell chemical treatment steps can take a few days to a week to set up. It is recommended that the weekly schedule is planned to avoid any interruption in the experimental pipeline. In case there are time constraints or other factors that cause anticipated interruptions, we have included a recommended pause point in which the experiments can be temporarily halted and resumed later. Please refer to Fig. B.1 for a pictorial overview of the protocol.

##### **Step 1: Seeding Cell & Treatment (TIMING: 2 days to 1 week)**

- 1) Seed and treat cells as necessary in DMEM culture medium supplemented with FBS and P/S.
- 2) Generally, for most cell lines, 250,000-1,000,000 cells are seeded per 10cm dish for a 3-4 day experiment. Seeding density can be scaled for a shorter or longer experimental period. The exact timeframes need to be optimized for different cells and treatments. Typically, the cells are harvested while in the log-phase growth with confluency at around 60-90%.

CRITICAL STEP: If G4 ligand treatment is used, make sure a suitable dose is chosen. Cell lines can respond differently to different G4 ligands. A dose-finding experiment should be performed, and a low-toxic (LD15 to LD25) dose should be chosen for subsequent experiments.

##### **Step 2: Crosslinking (TIMING: 1-2 h)**

CAUTION: Formaldehyde is highly toxic. Always handle a high concentration of formaldehyde inside the chemical fume hood. After the crosslinking step, all formaldehyde waste should be

placed in separate waste containers to be disposed appropriately according to Institutional procedures.

- 1) Make fresh 1% formaldehyde in 1X Fixing Buffer before a new experiment.
- 2) Aspirate media and wash once with DEPC-PBS
- 3) Aspirate PBS and add 10 mL of 1% formaldehyde/1X Fixing Buffer to a 10 cm dish for 5 min on a rocker at room temperature.
- 4) Quench crosslinking by adding glycine to a final concentration of 125 mM and incubate 5 min on a rocker at room temp.

**TIMING:** The duration of the formaldehyde crosslinking step is important. Try to keep it as consistent as possible between samples. Stagger the addition of formaldehyde and quenching if processing several samples at the same time. You may want to optimize the duration of crosslinking. It should generally be between 5 to 10 minutes. Typically, crosslinking over 10 minutes is not recommended as over-crosslinking may limit the accessibility of epitopes.

**CRITICAL STEP:** After the crosslinking step, keep the sample at 4°C as much as possible. The centrifugation should be done at 4°C as well. All buffer from this point on needs to be RNase-free.

- 5) Remove the fixing solution and add 0.8 mL of DEPC-PBS (pre-chilled) and scrape cells into an Eppendorf tube. Keep on ice.
- 6) Spin tubes at 1.5k rpm for 3 min at 4°C.
- 7) Aspirate supernatant (avoid disturbing the cell pellet) and wash again with 1 mL of DEPC-PBS.

PAUSE POINT: You may freeze down the cell pellets at this stage and store them at  $-20^{\circ}\text{C}$  to be processed later. However, for optimal results, we recommend that you do not pause at this point as processing fresh sample generally have better and more consistent yield. The thawed pellet may also become clumpy and difficult to resuspend, which can affect sonication.

**Step 3: Sonication (TIMING: 1-2.5 h)**

- 8) Aspirate the supernatant and resuspend the pellet in 400  $\mu\text{L}$  G4RP buffer (supplemented with 100U/mL RNase inhibitor; it is optional to add 1XPIC).
- 9) Sonicate on a Covaris m220 focused ultrasonicator using 10 % duty factor “Chromatin Shearing” preset for 2 min at  $7^{\circ}\text{C}$  temperature setpoint. The Covaris microtube-130 is used for the sonication. Mix well before adding the resuspended cells to the tube. Make sure there are no bubbles when adding the sample to the AFA microtubes prior to sonication. Each sonication tube should contain exactly 130  $\mu\text{L}$  corresponding to approximately 2-3 million cells. Repeat until all the resuspended cells are sonicated.

CRITICAL STEP: Optimization with specific sonicators (Covaris, Diagenode, and other brands) is required. The sonication protocol should be similar to that recommended for ChIP according to sonicator-specific published protocols. We recommend that different durations of sonication be tested. Generally, the sonication duration should not exceed that recommended for ChIP, and a good starting point for optimizing the duration is to use half of which is recommended for ChIP. Make sure the temperature is kept around  $2-7^{\circ}\text{C}$ .

- 10) Spin sonicate at 13200 rpm for 10 min at  $4^{\circ}\text{C}$ . Move the supernatant to new tubes. The supernatant (sonicate), containing the cellular materials that are successfully released from sheared cells, will be used for the subsequent steps.

PAUSE POINT: It is possible to freeze down the sonicate. However, for optimal results, it is recommended to continue directly to the next step.

**Step 4: Binding and Washes (TIMING: O/N + 4 h)**

- 1) Each sonicate of ~390  $\mu\text{L}$  can be aliquoted into two tubes (each with 180  $\mu\text{L}$ ): one for the biotin control and the other for BioTASQ. Save 8.75  $\mu\text{L}$  of the remaining sonicate as 5 % input control. The input control can be stored temporarily at 4°C then process later with the other samples.
- 2) Add 20  $\mu\text{L}$  of 1 mM stock biotin (20 nmol) to a final concentration of 100  $\mu\text{M}$  to the control tube. Add 20  $\mu\text{L}$  of 1mM stock BioTASQ (20 nmol) to a final concentration of 100  $\mu\text{M}$  to the G4RP tube.
- 3) To prepare a 5% input, you can take out 5% of your starting material (5% of 180  $\mu\text{L}$  = 9  $\mu\text{L}$ ) and leave it on the side at 4°.
- 4) Incubate overnight at 4°C on rotator.

CRITICAL STEP: The amount of magnetic bead added can be optimized. The recommended amount is between 10-100  $\mu\text{g}$ , depending on the brand of the magnetic beads. An optimization experiment using a different amount of beads is higher recommended.

Generally, a higher amount increases yield but also increases non-specific background binding. Always mix well before transferring the beads to other tubes.

- 5) Next day, use wide-mouth tips to add 10  $\mu\text{L}$  of MagneSphere magnetics beads (60  $\mu\text{g}$ ) to each tube.
- 6) Incubate for 2 h at 4°C on a rotator.

- 7) Use the magnet to pull beads to the side by leaving them on the magnet for 2 min. Shift the tubes slightly to pull beads closer together. Aspirate carefully without removing the beads.
- 8) Wash beads 3 times using G4RP buffer. Put on rotator at room temp for 5 min for each wash.
- 9) Wash 2 times in DEPC-PBS for 2 min each.
- 10) Resuspend beads in 100  $\mu$ L of DEPC-PBS (supplemented with RNase inhibitor). Take out the 5% input that stored away in the earlier steps and add 100  $\mu$ L DEPC-PBS.
- 11) Incubate at 70°C for 1 h to reverse crosslink.

**TROUBLESHOOTING:** Sufficient reversal of crosslink is essential here as crosslinked material can affect proper TRIZOL extraction. You may need to increase the duration of heat to 2 h. Make sure there is no RNase contamination as it can lead to RNA degradation during the heating step.

#### **Step 5: RNA Purification (TIMING: 2 h)**

**CAUTION:** TRIZOL and chloroform should be handled in the chemical fume hood. Waste should be disposed of appropriately according to Institutional procedures.

- 1) Add 1 mL of TRIZOL to each tube directly to the beads and gently mixed.
- 2) Incubate the sample in TRIZOL for 5-10 min at room temp to dissociate any ribonucleoproteins (RNPs) and other structures.
- 3) Add 0.1 mL of chloroform. Make sure the cap is sealed and vigorously shake for 15 sec. Leave at room temp for 3 min.
- 4) Centrifuge at 12000 x g for 15 min at 4°C. The mixture will separate into the top aqueous phase, the middle interphase and the lower pink-coloured organic phase. The aqueous phase contains the RNA.

- 5) Transfer the aqueous phase to a new tube by tilting the tube at 45° and slow pipetting. Avoid disrupting the interphase/organic phase.
- 6) Add 0.5 mL of fresh isopropanol to the tube with the aqueous phase. Incubate at room temp for 10 min.
- 7) Centrifuge for 10 min at 12000 x g at 4°C.

CRITICAL STEP: The RNA pellet will be very small and unlikely to be seen by naked eye. Do not touch the bottom of the tube with the pipette. Pipette out slowly by drawing against the side of the tube.

- 8) Carefully and slowly discard the supernatant but leave 10-15 µL of liquid at the bottom.
- 9) Add 1 mL of fresh 75% ethanol.

PAUSE POINT: You may store pellets in ethanol at 4°C for up to 3 days or -20°C for months.

- 10) Vortex briefly and centrifuge at 7500 x g for 5 min at 4°C.

CRITICAL STEP: It is unlikely that you will see a pellet here, as the amount of RNA is very low. Avoid touching the pipette tip to the bottom or creating any motion that could disturb the pellet. The addition of linear acrylamide may also be considered to add recovery.

- 11) Slowly remove the ethanol leaving approximately 10 µL of liquid at the bottom of the tube.
- 12) Keep the lids of the tube open and air dry for 10 min at room temp.
- 13) Add 30 µL of RNase-free water and pipette up and down.
- 14) Incubate at 55°C for 15 min.
- 15) Store the purified RNA samples at -70°C.
- 16) Check RNA concentration and quality using Nanodrop.

PAUSE POINT: RNA sample can be stored long term after this step. Thawing should be minimized from here on.

**Step 6: Reverse Transcription & qPCR Quantification (TIMING: 4-5 h)**

CRITICAL STEP: After adding different components, always mix by gently flicking the tube, then a quick pulse with a centrifuge. Do not vortex.

- 1) Thaw purified RNA from G4RP on ice.
- 2) A master mix can be made with pdn6 and dNTP. Add to each tube and mix:
  - a. 13  $\mu$ L of G4RP RNA sample
  - b. 1  $\mu$ L of 50  $\mu$ M random hexamer (pdn6)
  - c. 1  $\mu$ L of 10 mM dNTP mix
- 3) Incubate at 65°C for 5 min.
- 4) Place on ice for 5 min.
- 5) Make a master mix of the following component and gently mix. Add to each tube:
  - a. 4  $\mu$ L of 5X first strand buffer
  - b. 1  $\mu$ L of 0.1 M DTT
  - c. 0.2  $\mu$ L of RNAse inhibitor
  - d. 0.2  $\mu$ L of Superscript III
- 6) Gently mix the RT reaction tube and pulse spin by brief centrifugation.
- 7) Incubate at room temp for 5 min.
- 8) Incubate at 55°C for 45 min. These steps can be performed on a thermocycler.
- 9) Inactivate reaction by incubating at 70°C for 15 min and cool down to room temp.
- 10) The RT sample can be stored at -20°C long term and can be quantified using qPCR.

- 11) Proceed to real-time qPCR with an appropriate primer set (Table B.2) and SYBR mix of your choice.
- 12) Make a master mix with the appropriate primer set. An example of a simple qPCR tube for real-time is as followed:
- a. 1  $\mu$ L of RT sample
  - b. 0.5  $\mu$ L of forward primer (10  $\mu$ M)
  - c. 0.5  $\mu$ L of reverse primer (10  $\mu$ M)
  - d. 5  $\mu$ L of 2X SYBR mix
- 13) Run the real-time qPCR using the default amplification protocol. It should typically take around 20-30 cycles to reach exponential phase (~2-3 hours).

Gene	Forward Primer Sequence	Reverse Primer Sequence
<i>NRAS</i>	ATGACTGAGTACAAACTGGTGGT	CATGTATTGGTCTCTCATGGCAC
<i>VEGFA</i>	CCTTGCCTTGCTGCTCTACC	AGATGTCCACCAGGGTCTCG
<i>MALAT1</i>	AAAGCAAGGTCTCCCCACAAG	GGTCTGTGCTAGATCAAAAGGCA

**Step 7: Ribo-depletion and Library Preparation (TIMING: 6-8 h)**

- 1) Starting from the TRIZOL-purified RNA from G4RP, thaw frozen RNA on ice.
- 2) It is recommended to have an additional purification step before RNA-seq to remove any potential residual contamination (e.g. phenol, guanidium salt, etc.), as well as DNA contamination using DNase treatment coupled with RNA column purification. Follow the manufacturer's recommended protocol. Quantify RNA for concentration and purity.

**CRITICAL STEP:** Use a higher amount of input for the library preparation to minimize the number of PCR cycles required (preferably <7 cycles). G4 sequences may be sensitive to

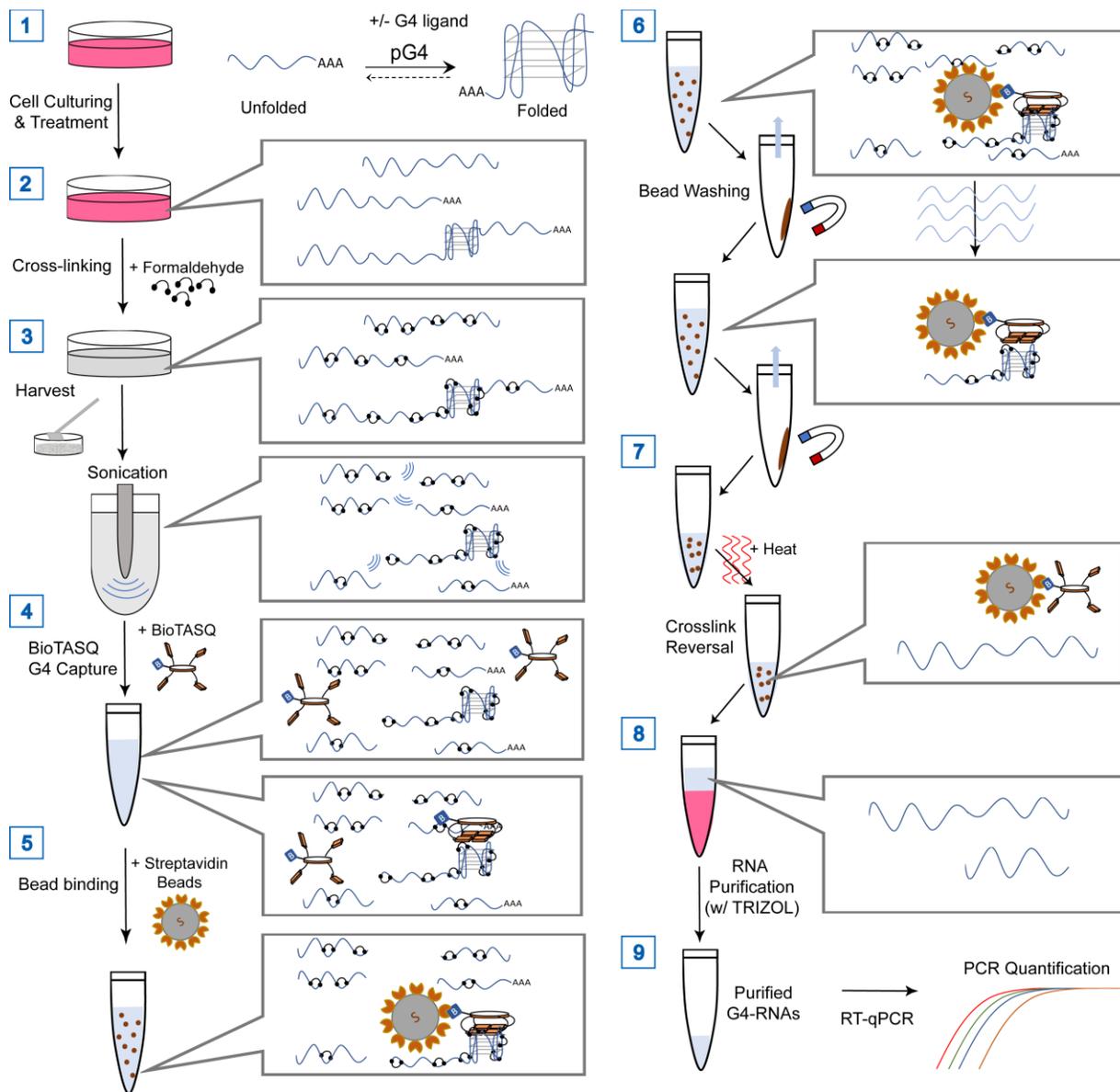
PCR and could be underrepresented after a higher number of cycles. If the input is too low, it is recommended to pool multiple samples together.

**CRITICAL STEP:** The fragmentation step can be modified depending on the fragment size. Due to the sonication step earlier, the average length of RNA fragments will be reduced. To avoid over-fragmentation, the duration of the thermal fragmentation should be reduced. We recommend 2-4 min of thermal fragmentation.

- 3) For the library preparation, use the Ribo-Zero Prep along with Illumina TruSeq RNA library prep kit. Ribo-zero has been discontinued and incorporated into the new Illumina TruSeq Stranded Total RNA library prep kit; the new library prep should be equally compatible.

Briefly, the steps are as followed:

- a. Deplete rRNA using Ribo-Zero oligos and beads.
  - b. Thermal fragment briefly at 4 min.
  - c. Synthesize the first-strand of cDNA, followed by the second cDNA strand.
  - d. Adenylate 3' ends and ligate adapters.
  - e. Enrich DNA fragments by PCR amplification. Use minimal cycles of PCRs (5-7 cycles).
  - f. Quality-check library
- 4) Sequence on a HiSeq, NextSeq or NovaSeq platform at 2x125bp and depth of 15-20 million reads.



**Figure B.1. Overview of the G4RP workflow.** 1) Cells were cultured in 10 cm culturing dishes. If required, cells may be treated with G4 ligands at this stage. 2) Cells were crosslinked using formaldehyde. 3) Cells were harvested by scraper and sonicated using an ultrasonicator. 4) BioTASQ was added to sonicate. 5) Streptavidin beads were added. 6) Beads were washed multiple times. 7) Beads were heated to reverse the crosslink. 8) TRIZOL was added to the sample and the RNA fraction is purified. 9) Reverse-transcription was performed on the RNA sample before quantification using real-time quantitative PCR.

## Step 8: Sequencing Bioinformatics Analyses (TIMING: varies)

The analyses are done using a series of bioinformatics tools. An overview of the G4RP-seq analysis steps is depicted in Fig. B.3A. For investigators who are less experienced with bioinformatics or lack a powerful local workstation computer, performing data analysis through a remote Galaxy server (27) (usegalaxy.org), which contains all the required bioinformatics packages, is highly recommended.

- 1) Quality-check raw FASTQ file and trimmed using Trimmomatic (22).
- 2) Align to the target genome (e.g. Human genome hg19 or hg38 assembly) using HISAT2 (23) using default settings. Integrative Genome Viewer (IGV) (24) can be used to visualize the alignments.
- 3) Prepare and download gene annotation file (in GTF format) from UCSC matching that of the target genome.
- 4) Feature count and annotate using HTSeq-count (25) with “Union” mode and values (feature type=exon, ID attribute=gene\_id). Depending on the library prep kit used, you can set either “stranded” or “non-stranded.”
- 5) Normalize and calculate differential features counts using DESeq2 (26).
- 6) DESeq2 should be run for BioTASQ sample vs Input control for each treatment condition to generate a relative BioTASQ enrichment score (ES), the log fold change readout from DESeq2. The gene transcript can be ranked according to enrichment score. It is important to note that all subsequent comparisons are relative measures.
- 7) The enrichment score between conditions can be compared to generate ligand-induced changes. The enrichment score change ( $\Delta$ ES) is calculated as followed: ES (treated) - ES (untreated).

8) Using Excel, you can filter by counts (as a measure of RNA abundance) and fold change.

The filter threshold can differ depending on the quality and depth of sequencing.

## **TIMING**

The duration per step is estimated here and can be highly variable depending on the pace at which the investigator is running the experiment. There are also several pause points within each step in which the investigator can temporarily halt the experiment and resume later.

Step 1 – 2 days to 1 week (depending on cell culturing experimental setup)

Step 2 – 1-2 hours (depending on the number of samples)

Step 3 – 1-2.5 hours (depending on the number of samples)

Step 4 – overnight + 4 hours

Step 5 – 2 hours

Step 6 – 4-5 hours

Step 7 – 6-8 hours

Step 8 – varies (depending on computer power)

## B.5 Troubleshooting

<b>Table B.3. Troubleshooting &amp; Solutions</b>		
<b>Issues</b>	<b>Potential Reasons</b>	<b>Solution</b>
Low levels of all extracted samples including input control	Cells too clumpy; cells not sonicated properly; cellular components not released into the supernatant.	Do not freeze down pellet and directly proceed to the sonication step. Reoptimize the sonication step. Increase sonication duration. Add 0.1% SDS to resuspended cells before sonication may also help.
High signal in biotin control	Streptavidin beads concentration too high	Lower amount of beads added; reoptimize using different amounts of beads
	Input too high; too many cells used; saturated signal	Dilute sample before G4RP; distribute to multiple tubes
Low yield with BioTASQ	Incorrect concentration of BioTASQ and beads are used	Confirm that the stock concentrations are correct. Remake stock concentration if unsure.
	Insufficient amount of beads	Increase amount of beads; reoptimize using different amounts of beads
	Over cross-linking; leading to loss of accessibility to the G4 epitopes	Cross-linking duration should not exceed 10 min. Some formaldehyde contains additives (eg. Methanol) that may increase permeation and cross-linking efficiency. Use fresh formaldehyde without additives. Make fresh glycine to quench effectively. Slightly increase glycine concentration.
	Insufficient reversal of cross-linking; lead to loss of RNA in RNPs during TRIZOL step	Increase duration of the heat cross-link reversal step (up to 2h).
	RNase contamination	Remake new buffers and ensure no RNase contamination is present. Add in fresh RNase inhibitors in every new experiment.
Low ligand-induced change detected	Under cross-linking; leading to bias towards BioTASQ stabilization	Increase cross-linking duration between 5 min to 10 min. Keep samples on ice or 4°C after cross-linking.
	Treatment with an insufficient dose of G4 ligand	Cell lines can exhibit vary degree of sensitivity to G4 ligands. Perform a dose-response curve to profile the cytotoxicity of a specific G4 ligand in a specific cell line. Chose a
Low A260/280 ratio for RNA	Contamination from TRIZOL components (eg. Phenol), protein or DNA.	Wash twice, instead of once, with ethanol during the TRIZOL extraction. Run an additional column purification at later steps before sequencing.

## **B.6 Anticipated Results**

### **Sonication Check**

Due to the different sonication techniques used in different labs, it is advised that a sonication check be performed using different sonication conditions. For example, for the Covaris ultrasonicator, we divided a crosslinked sample and performed a parallel comparison between different sonication conditions (sonication duration: 0 min, 2 min and 4 min) under the same settings. Sonicated samples can be heated for 1h to reverse crosslink, and the RNA can be extracted using TRIZOL or RNA column purifications. The purified RNA can then be observed on an RNA bleach gel stained with SYBR (28). As depicted in Fig. B.2A, no sonication would result in an inefficient release of RNA material from crosslinked material. Note that crosslink reversal is inefficient in unsonicated materials and thus would result in a loss of extracted RNA, as shown by the lack of RNA band on the gel. Generally, a sonication duration of 2-4 min is sufficient, yielding a range of bands between 200-800 nt. Longer sonication duration also leads to a shorter average length of RNA fragments, as seen by the shifting of median band length between the 2 and 4 min sonication conditions. It is important to note that, while longer sonication may help increase resolution, it may also disrupt some interactions. Thus, a balance is required to prevent under- or over-sonication.

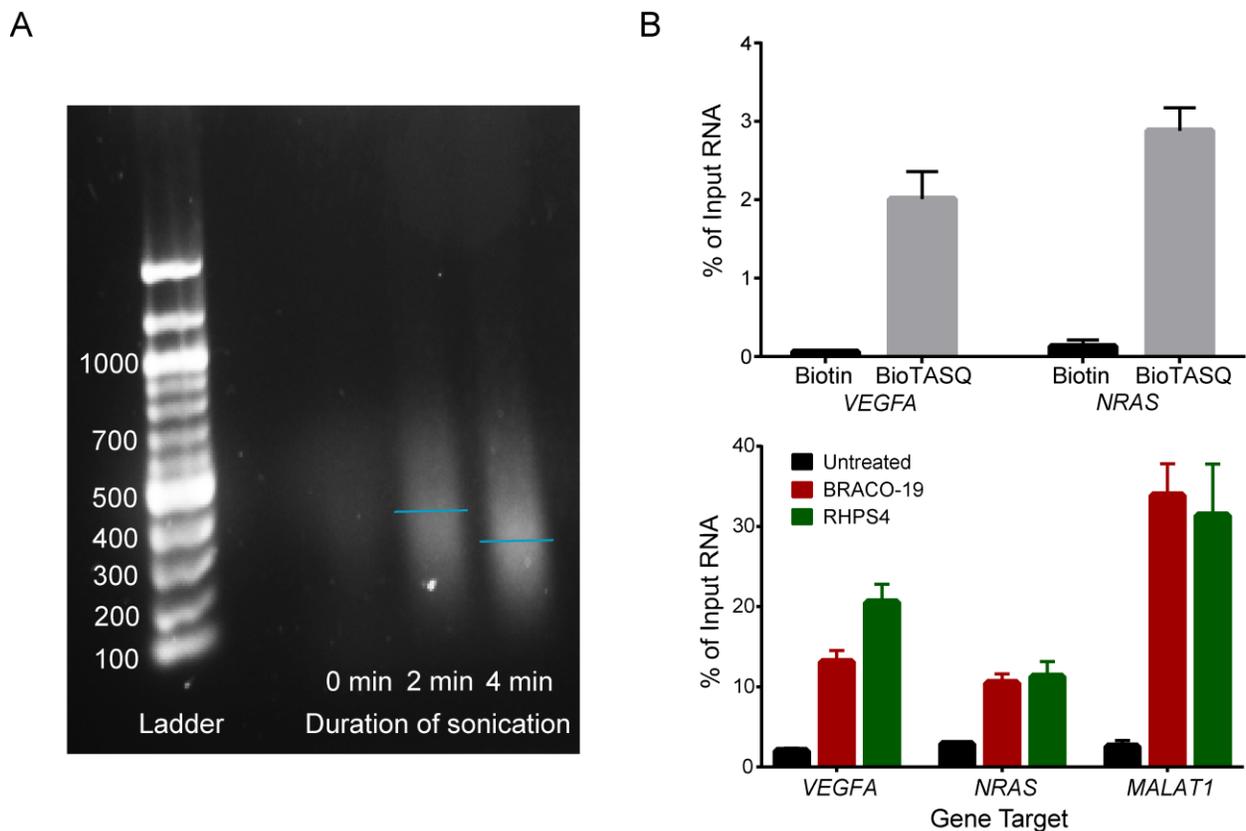
### **RNA yield from G4RP**

After the step with TRIZOL-purification of the RNA fraction, the purified RNA can be quantified using Nanodrop or other comparable methods. A lower A260/280 ratio is indicative of some contamination from the TRIZOL procedures. In that case, column repurification is recommended before further processing for RNA-seq. Typically, the 5% input concentration is generally around 100 ng/ $\mu$ L (with a 30  $\mu$ L yield = 3  $\mu$ g). The BioTASQ-enriched sample

concentration varies around 100-200 ng/ $\mu$ L (30  $\mu$ L yield of 3-6  $\mu$ g). Biotin control concentration is much lower at  $\sim$ 50 ng/ $\mu$ L (30  $\mu$ L yield of 1.5  $\mu$ g). The biotin control consists mostly of non-specific rRNA binding.

### G4RP-qPCR quantification

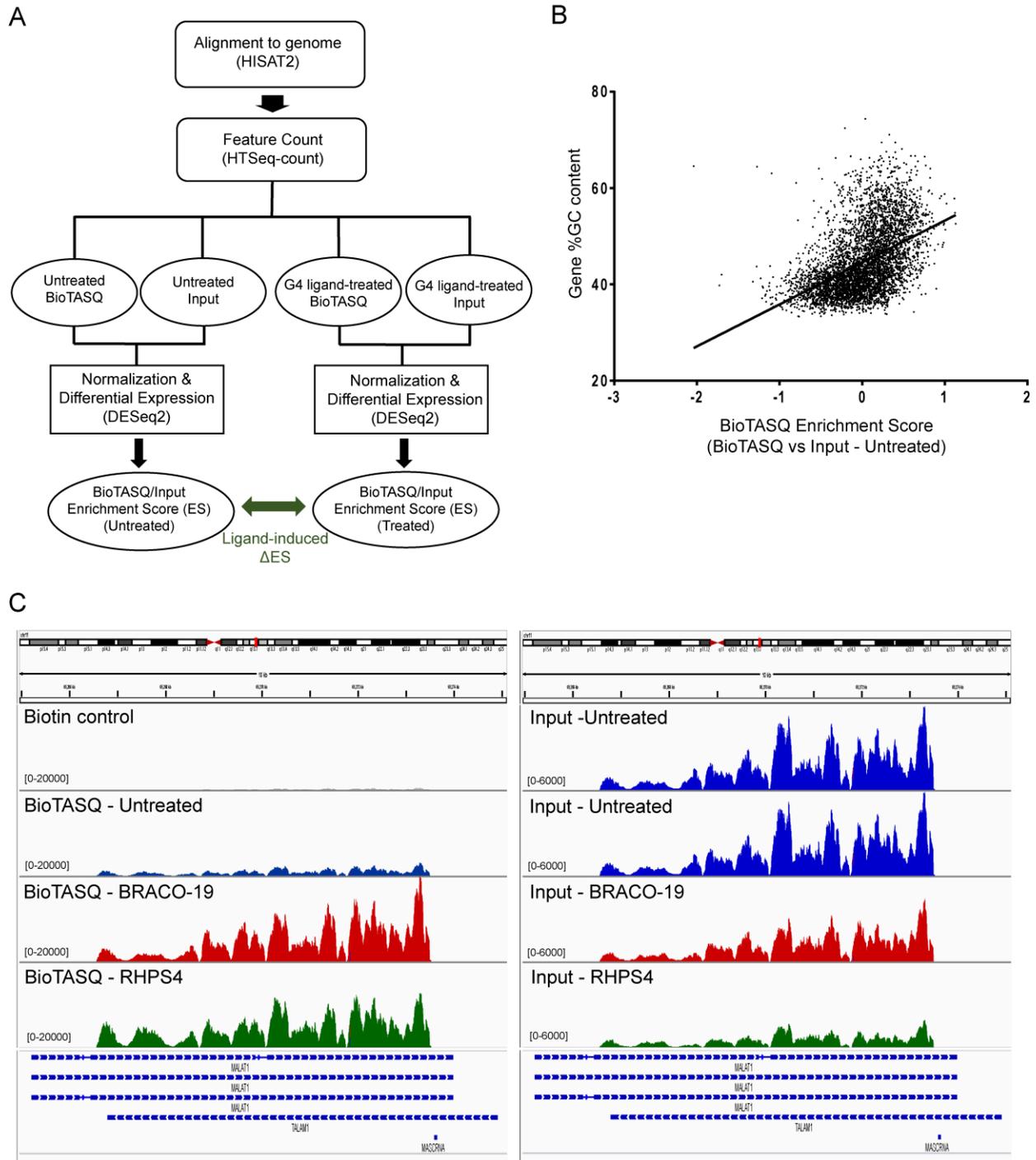
RT-qPCR and real-time quantification of the G4RP samples typically yield high signal-to-background ratios ( $>20$  fold) when comparing BioTASQ versus the biotin control sample (Fig. B.2B). Ligand-induced change ranges from 2-fold to as high as 12-fold, depending on the RNA target and G4 ligand used.



**Figure B.2. Expected results from G4RP.** A) RNA gel of samples that have undergone different duration of sonication (0, 2, 4 min). B) (Top) G4RP-qPCR signal of BioTASQ vs biotin control for RNA targets, VEGFA and NRAS. (Bottom) G4 ligand (BRACO-19 and RHPS4)-induced changes in the G4RP-qPCR signals for targets, VEGFA, NRAS and MALAT1.)

## **G4RP-seq analysis**

After processing the files through a series of bioinformatics tools (Fig. B.3A), the differential expression values (output from DESeq2) are used as the BioTASQ enrichment score. A quick indicator of a successful G4RP-seq is to rank the transcripts by enrichment score and check for an increase in GC content. At baseline levels (derived from untreated samples), the enrichment score is expected to be positively correlated with the GC% of the gene transcript (Fig. B.3B). Signals of aligned reads can be visualized using IGV (Fig. 3B.C). The biotin control is expected to show negligible levels with reads primarily aligned to rRNA regions. By comparing the BioTASQ with that of input control, the relative level of BioTASQ interaction with the target can be elucidated (Fig. B.3C). An example of a high G4-ligand-inducible target is the long non-coding RNA, MALAT1. The raw signals show much higher alignments in the G4-ligand-treated samples compared to that of the untreated.



**Figure B.3. Expected results from G4RP-seq.** A) Schematics of the G4RP-seq bioinformatics analyses. B) Scatterplot comparing the BioTASQ Enrichment Score with GC content of respective gene transcripts at baseline (derived from untreated sample). Each dot represents a gene transcript. C) Signal of aligned sequencing reads in the MALAT1 region for (Left) biotin control, BioTASQ in untreated sample, BioTASQ in BRACO-19-treated sample and BioTASQ in RHPS4-treated sample and (Right) the corresponding input controls.

## B.7 References

1. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Research* **47**, 3862–3874 (2019).
2. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry* **5**, 182–186 (2013).
3. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology* **33**, 877–881 (2015).
4. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature Protocols* **13**, 551–564 (2018).
5. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature Methods* **13**, 841–844 (2016).
6. Yang, S. Y. *et al.* Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nature Communications* **9**, (2018).
7. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology* (2020). doi:10.1038/s41580-020-0236-x
8. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of Molecular Biology* **429**, 2127–2147 (2017).
9. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371–aaf5371 (2016).

10. Stefan, L. *et al.* DOTASQ as a prototype of nature-inspired G-quadruplex ligand. *Chemical Communications* **47**, 4992 (2011).
11. Haudecoeur, R., Stefan, L., Denat, F. & Monchaud, D. A Model of Smart G-Quadruplex Ligand. *Journal of the American Chemical Society* **135**, 550–553 (2013).
12. A Twice-As-Smart Synthetic G-Quartet: PyroTASQ Is Both a Smart Quadruplex Ligand and a Smart Fluorescent Probe. *Acs.org* (2014). Available at: <https://pubs.acs.org/doi/abs/10.1021/ja506331x>. (Accessed: 8th June 2020)
13. Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *Journal of the American Chemical Society* **137**, 8521–8525 (2015).
14. Renard, I. *et al.* Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. *Nucleic Acids Research* **47**, 5502–5510 (2019).
15. Selth, L. A., Gilbert, C. & Svejstrup, J. Q. RNA Immunoprecipitation to Determine RNA-Protein Associations In Vivo. *Cold Spring Harbor Protocols* **2009**, pdb.prot5234-pdb.prot5234 (2009).
16. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. *Nature Chemistry* **2**, 1095–1098 (2010).
17. Kubota, M., Tran, C. & Spitale, R. C. Progress and challenges for chemical probing of RNA structure inside living cells. *Nature Chemical Biology* **11**, 933–941 (2015).
18. Sharma, E., Sterne-Weiler, T., O’Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Molecular Cell* **62**, 618–626 (2016).
19. Nguyen, T. C. *et al.* Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nature Communications* **7**, (2016).

20. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell* **62**, 603–617 (2016).
21. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).
22. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
23. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357–360 (2015).
24. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2012).
25. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).
26. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, (2014).
27. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy Team, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**, R86 (2010).
28. Aranda, P. S., LaJoie, D. M., & Jorcyk, C. L. (2012). Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis*, **33**(2), 366-369.

## **Appendix C - Genome-wide detection of G-quadruplexes and the effects of G4 ligands on transcription**

### **C.1 Brief Introduction and Methods**

Genome-wide G4 mapping was performed here using ChIP-seq and the G4-specific antibody, BG4. Here, I present a collection of data generated from preliminary ChIP-seq experiments. The optimization of this method was done prior to the publication of the G4-ChIP methodologies by the developer of the BG4 antibody, the Balasubramanian Lab (1-2). Notably, there are several deviations between our protocol and the one presented by Balasubramanian Lab as listed in Table C.1. Our method is more similar to that of the traditional ChIP procedure, which are often used for studying transcription factor binding. Thus, in comparison to the published Balasubramanian Lab's version of G4-ChIP-seq, our protocol takes additional days to complete. Also, our protocol was tested in the presence of various G4 ligands whereas it has been not been done with Balasubramanian's protocol.

MCF7 cells were first cultured and was either mock-treated or treated with G4 ligands, BRACO-19 (2.5  $\mu\text{g}/\text{mL}$ ) or RHPS4 (1.25  $\mu\text{M}$ ), for 72 h. ChIP was performed as described in Section 2.4 and the library construction was performed using NebNext DNA library prep kit. The G4-ChIP samples were sequenced at a depth of 30 million reads with 2x150bp paired-end mode. Raw reads were quality-check and cleaned. Sequence alignment to hg19 human genome reference was performed using the Bowtie2 algorithm (3). ChIP-seq peaks were called using the MACS2 program (4) by processing ChIP samples against input control.

**Table C.1.** G4-ChIP methodology comparisons with G4-ChIP (Balasubramanian Lab)

<b>Key Differences in ChIP-seq approaches</b>	<b>Our G4-ChIP Protocol</b>	<b>G4-ChIP Protocol from Balasubramanian Lab (104-105)</b>
<b>Cell lines used</b>	MCF7	HaCaT, NHEK, K562, IMR90, U2OS
<b>Ligand treatment</b>	G4 ligands: RHPS4, BRACO-19, TmPyP4	Histone deacetylase inhibitor, entinostat
<b>Crosslinking</b>	1% formaldehyde for 5 min	1% formaldehyde for 10 min
<b>Sonication tool</b>	Covaris m220 ultrasonicator	Bioruptor Plus
<b>Antibody incubation</b>	Overnight at 4°C	1 h at 16°C
<b>Beads</b>	Anti-FLAG M2 agarose beads	Anti-FLAG M2 magnetic beads
<b>Washing</b>	3X washes each in ChIP Wash Buffers A, B, C and TE at room temperature	3X washes in Wash buffer at 4°C followed by one wash at 37°C
<b>Crosslinking reversal</b>	Overnight in 1% SDS at 65°C	2h in TE buffer supplemented by proteinase K
<b>DNA purification</b>	Qiagen PB binding buffer + BioBasic Kit	Qiagen MinElute Kit
<b>Library prep kit</b>	NebNext DNA library prep kit	Nextera Library prep kit (Tagmentation)
<b>DNA Input used by library construction</b>	50-100 ng	1-10 ng
<b>Read Length and read mode</b>	2 x 150 bp paired end	75-bp single end
<b>Sequence Alignment tools</b>	Bowtie 2	BWM-MEM

## C.2 Preliminary ChIP-seq Results

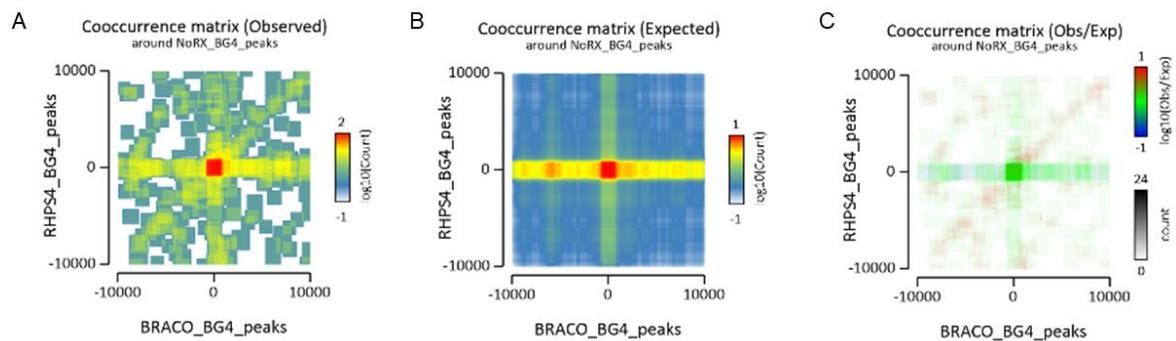
Sequence alignment and peak-calling against input control resulted in 100000-200000 initial peaks. Around 4000-10000 peaks remained after filtering by >2.5 fold-enrichment (Table C.2). The peaks were then annotated to the closest gene. Distance to the nearest transcription start site (TSS) was calculated for each ChIP-seq peak to generate a distribution profile for each sample (Fig. C.2A). For all three samples, majority of the peaks were distributed close to TSSs as shown by a higher number of peaks associated close to the 0 position. Genomic feature annotations showed that approximately half of the peak sites were located in intragenic regions with the other half distributed in intergenic regions (Fig. C.2B). Compared to the non-treated sample, G4-ligand treated samples showed an increase in intragenic peaks (39% vs 45% or 43%). A snapshot at a region within chromosome 16 showed wide distribution of ChIP-seq peaks (Fig. C.2B). Using a *de novo* motif discovery tool, MEME-ChIP (5), a highly ranked G-rich motif was generated (Fig. C.2C), suggesting BG4 binding to G4 sites.

**Table C.2.** Number of raw and filtered peak-called regions.

<b>Sample</b>	<b>Number of raw peak-called regions</b>	<b>Number of filtered peak-called regions</b>
No Rx – BG4	129693	4793
BRACO-19 – BG4	202783	5907
RHPS4 – BG4	110315	9853



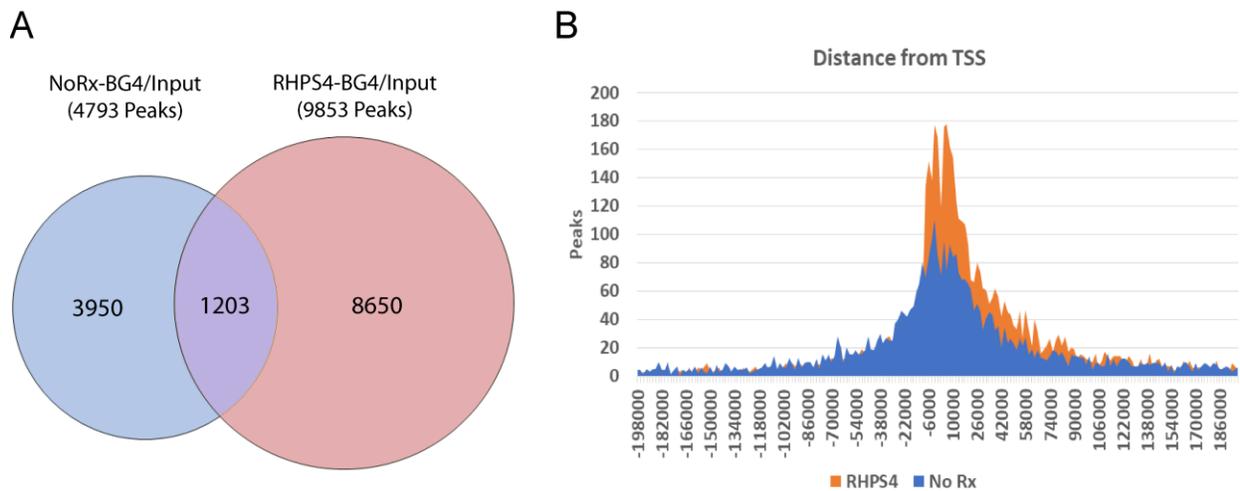
Peak co-occurrence analysis (6) between BRACO-19-treated and RHPS4-treated samples centered on peaks of mock-treated samples showed strong concordance as seen by high enrichment at the 0 position (Fig. C.2). This indicated that peaks from BRACO-19-treated and RHPS4-treated sample are more likely to associate at the same locations as peaks from the mock-treated sample (Fig. C.3). Interestingly, the observed profile differed from the expected profile seen by the diagonal enrichment, suggesting associations of sites between BRACO-19 and RHPS4 sample independent of those found in mock-treated samples. These represented potential G4 ligand-induced *de novo* sites.



**Figure C.2.** Cooccurrence of ChIP-seq peaks between BRACO-19 and RHPS4-treated samples centered around non-treated samples.

While there were observable overlaps, BRACO-19- and RHPS4-treated samples showed a relatively different ChIP-seq profile. RHPS4 treatment led to an overall increase in number of peaks, whereas BRACO-19 treatment showed only a modest increase. Upon visualization of peaks, BRACO-19 treatments appeared to have decrease peak signals at the some of the same sites as mock-treated sample suggesting potential interference between BRACO-19 and BG4. On the other hand, RHPS4 had stronger peaks in general but also showed a small portion of decreased peaks when compared to mock-treated samples. This suggests that the interference

between G4 ligand and BG4 is dependent on the type of ligand itself and the specific peak region (or G4-forming site). Whether there are biological reasons other than interference is unclear. It may also be possible that ligand treatment reduces accessibility of the chromatin loci, which may result in fewer G4 formations. When the list of peaks from non-treated and RHPS4-treated samples were overlapped, more than 1200 sites were shared. Overlaying the TSS profiles of the two conditions revealed more *de novo* peaks near the TSS in the RHPS4-treated sample.

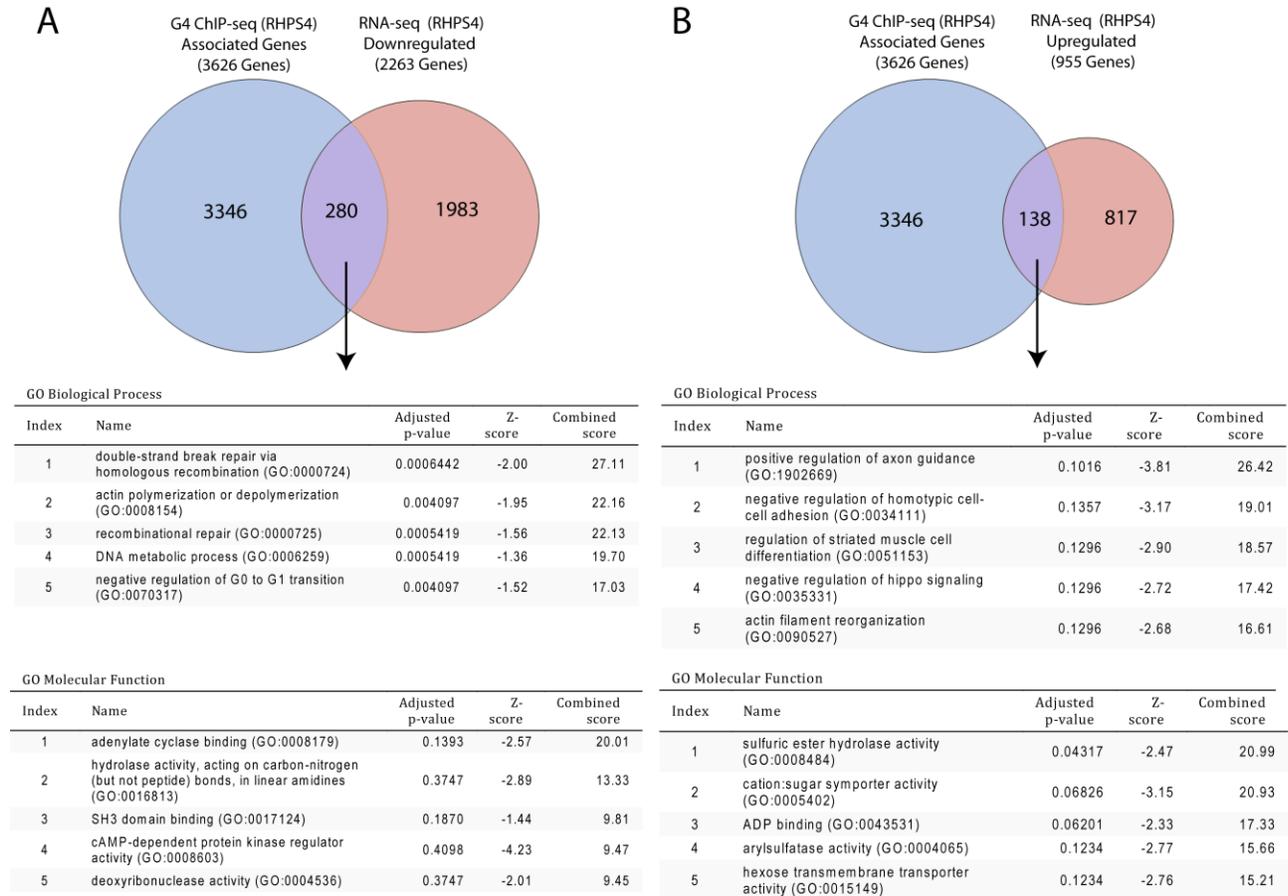


**Figure C.3.** Peak overlap analysis between non-treated and RHPS4-treated samples.

### C.3 ChIP-seq and RNA-seq intersection

Since the RNA-seq data from samples treated with RHPS4 (in the study for Chapter 4) could also be used as a reference for RHPS4-induced gene expression changes, we checked if the ligand-induced increases in ChIP-seq peaks corresponds with gene expression changes of associated genes. RHPS4 ChIP-regions corresponded with 3346 genes. Overlap of these ChIP-associated genes with the downregulated or upregulated gene list from RNA-seq resulted in 280 and 128 genes respectively (Figure C.4). Gene ontology analyses were performed on these

overlapped gene lists. Interestingly, several of these downregulated genes were found to be related to DNA repair. The full list of overlapped genes can be found in Table C.3.



**Figure C.4.** Overlap analysis of RHPS4-induced changes in ChIP-seq and RNA-seq. Associated genes in ChIP-seq were intersected with either A) downregulated genes or B) upregulated genes from RNA-seq. Gene ontology analyses were performed on the overlapped gene list.

**Table C.3.** Full list of overlapped gene for downregulated genes (top) and upregulated genes (bottom)

**Full Overlapped Gene list (Downregulated)**

ABI1	CACYBP	DEPTOR	G2E3	KIFAP3	MYBL1	PLEKH2	SDCCAG3	TM2D3	ZMYM1
ACTR3B	CALM1	DHFR	GALNT11	L2HGDH	MYBL2	PLEKHG5	SEN8	TM6SF1	ZNF273
ACTR3C	CAP2	DHRS7	GEMIN6	LBR	NASP	PLOD2	SF1	TMEM135	ZNF277
ADAM22	CBX1	DLGAP5	GINS1	LMAN2L	NBPF10	PRCP	SGTB	TMEM168	ZNF383
ADAT2	CCDC124	DNAJC24	GK	LOC202181	NBPF15	PRKAR2A	SH2B3	TMEM38B	ZNF43
ADCK5	CCDC43	DNAJC25	GLCE	LOC284379	NCAPG2	PRPSAP1	SIL1	TMEM60	ZNF449
AGK	CCDC77	DNAL1	GNA12	LYRM4	NCK1	PSMG1	SLC16A1	TNK1	ZNF675
AGMAT	CDC14B	DNASE1	GNPNAT1	MAN1A2	NDUFA5	PTPN3	SLC24A1	TRMT5	ZNF850
AKAP5	CDKAL1	DPY19L1P1	GOLIM4	MANEAL	NDUFAB1	PUS7	SLC25A15	TROAP	ZNF90
AKR1A1	CECR2	DPY19L2P1	GPR137C	MAP2K1	NDUFAB1	RAB30	SLC25A19	TSKU	
ANKRD20A5P	CEP57L1	DSCR8	HDHC2	MASTL	NDUFB3	RAB3IP	SLC25A29	TTF2	
ANKS1B	CEP78	DYNC2H1	HELB	MAX	NEB	RAB6B	SLC35A5	TUBE1	
AR	CFDP1	ELL	HIST1H2AL	MCM3AP-AS1	NOD1	RABGEF1	SMC2	TXNDC16	
ARHGAP18	CHPT1	EMP2	HIST2H2AB	MCU	NOL7	RACGAP1	SMYD3	ULK2	
ARHGAP26	CINP	EPHA3	HMGB2	MEGF8	NOMO3	RAD51AP1	SNORA38B	UMPS	
ARHGEF26	CMC2	ESD	HOMER2	MIS18BP1	NRF1	RAD51B	SNX10	USP13	
ARL5A	CNTRL	ESR1	HP1BP3	MMP16	NSMCE4A	RAD54B	SPC24	USP25	
ARPC5L	COIL	EVL	HSDL2	MMS22L	NUCKS1	RAVER2	SREBP1	WASF3	
ARSB	COQ6	EXD2	ID2	MNAT1	NUP205	RBM8A	ST7L	WDHD1	
ATF1	CREB3L2	EZH2	IFT81	MOCS2	NXN	RELT	STAM	WDPCP	
BBS9	CRIPAK	FAH	INTS4	MPP6	ORC5	RNF2	SUV39H2	WDR41	
BMS1P4	CRISPLD2	FAM172A	IPO7	MR1	PA2G4P4	RNF32	SYK	WDR70	
BRCA1	CRTC1	FAM207A	JAK2	MRPL3	PCGF6	RNF41	SYTL4	WDR89	
BRI3BP	CSE1L	FAM98B	JPH1	MRPL48	PCHO	RNH1	TATDN3	WDYHV1	
BRMS1L	CUEDC1	FANCD2	KATNAL1	MRPS18A	PCTP	RPA1	TBC1D2	WRB	
C15orf40	DAB2IP	FANCM	KCTD3	MRPS21	PDSS1	RPA3	TBCA	XBP1	
C17orf53	DCTPP1	FBXO5	KIAA0391	MRPS36	PGAP1	RSRC1	TCF7L2	XRRA1	
C3orf33	DDAH1	FDX1	KIF11	MSRB2	PIMI1	RUFY2	TEAD4	ZCCHC10	
C7orf50	DDX39A	FILIP1L	KIF14	MTA3	PLCH1	RUSC2	TIMELESS	ZDHHHC24	
C8orf37	DEPDC1B	FSD1L	KIF27	MTBP	PLCXD3	SCAND1	TIMM44	ZFP64	

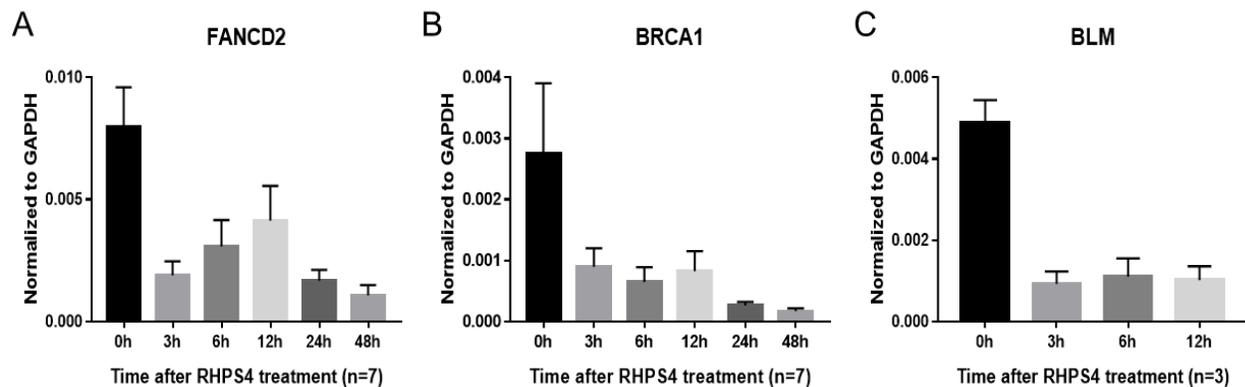
**Full Overlapped Gene list (Upregulated)**

43167	AUH	CROT	FNBP1	LACC1	NRIP1	RNASEL	SND1-IT1	TMEM219	ZNF629
43352	BCAR1	CSRNP2	FRA10AC1	LHFPL2	NRP1	RPS6KL1	SPECC1	TMTC2	ZNF701
AGAP1	BLZF1	CTNND2	FRMD6	LMF1	NTN4	RRM2B	SPECC1L	TP53INP1	ZSWIM6
AHCYL2	C12orf66	DOLK	FZD1	LURAP1L	PARP14	RUNX2	SQSTM1	TPRN	
AJUBA	CABLES1	DSCAM	GNS	MAK	PHF21A	SAMD12	ST3GAL1	TRAF6	
AKAP13	CASP8	EGLN3	GPHN	MAP2K5	PIEZO1	SEC61A2	STK3	TRIB1	
AP4S1	CCDC144B	EIF2B2	GRAMD1A	MAPK14	PLCB4	SEMA4D	STOM	TRIM14	
APH1B	CCDC92	ELMO2	GUSBP11	ME1	PMEP1	SERTAD4	SULF2	TUBB1	
APOPT1	CCT6P3	EPHA7	HDAC5	MYO18A	PPP1R12B	SGPP1	SUPV3L1	VPS13C	
ARID5B	CDK6	ERN1	HMCN1	NAV1	PRKG1	SIPA1L1	SUSD1	VPS53	
ARSG	CDYL2	ESRP2	IFRD1	NBEA	PRSS23	SLC16A6	TARSL2	WDR37	
ATG2B	CENPBD1	ESYT2	KCNK2	NEDD9	RASA2	SLC2A13	TAX1BP1	WWC3	
ATP13A2	CHMP4C	EXD3	KIAA1217	NEK1	RDX	SLC2A8	TFAM	ZHX2	
ATXN1	CMYA5	FBLIM1	KIF13B	NINL	RIOK3	SLC35C1	THNSL1	ZMAT3	
ATXN7L1	CPEB4	FGD6	KLF10	NRCAM	RIPK4	SLC39A10	TIMP3	ZNF425	

**C.4 Effects of G4 stabilization on transcriptome – a time course study**

The ChIP-seq/RNA-seq overlapped gene list for downregulated genes showed several DNA repair-related genes. Interestingly, a study has suggested that human DNA repair genes harbour a higher density of pG4 sequences than other groups of genes (7). Thus, we surmised that one of the cytotoxic mechanisms of G4 ligands may be through the transcriptional inhibition of DNA

repair genes. Since increases in transcription could lead to an increase opportunity for G4 formation, we wondered whether these DNA repair genes are first transcriptionally induced by G4 ligand-induced DNA damage then later transcriptionally inhibited by the very same G4 ligand due to increased accessibility. To test this hypothesis, a time-course experiment was performed to evaluate gene expression changes over time in selected DNA repair genes (from the list of overlapped gene list) after treatments with RHPS4. The hypothesized changes in gene expression pattern, however, was not detected within the chosen timepoints (0h, 3h, 6h, 12h, 24h and 48h). Transcription inhibition of FANCD2, BRCA1 and BLM occurred rapidly within 3h and remained suppressed for all later timepoints (Fig. C.5). FANCD2 appeared to display some increases in later timepoint suggesting potential induction of gene expression that may be outcompeted by G4 ligand binding. Due to high variabilities even with 7 independent biological repeats, it is difficult to make reliable conclusions with current results. It is possible that the peak induction may have occur outside these tested timepoints, possibly earlier than even 3h.



**Figure C.5.** Gene expression changes in mRNAs of A) *FANCD2*; B) *BRCA1*; C) *BLM* after treatment of RHPS4.

## C.5 Future considerations for G4-ChIP

As seen in these preliminary experiments, ligand-induced changes may be difficult to interpret using G4-ChIP as a readout. One possible way to circumvent potential interference is to reduce the dose of G4 ligands and to increase antibody concentration to promote favourable stoichiometry. In future G4-ChIP analyses for ligand-treated samples, negative peaks, which are detected as a loss of peak in the ligand-treated sample compared to non-treated sample, may also need to be considered. The addition of both positive and negative peaks could contribute to the discovery of more G4 ligand-interaction sites. G4-ChIP may require another complementary method for confirmation, perhaps using probes, for more confident identification of G4s. Finally, a direct comparison of different G4-ChIP-seq method using the same cell lines and culturing conditions would be essential to determine whether there are method-specific biases.

## C.6 References

1. Hänsel-Hertsch, R., Beraldi, D., Lensing, S. V., Marsico, G., Zyner, K., Parry, A., ... & Tannahill, D. (2016). G-quadruplex structures mark human regulatory chromatin. *Nature genetics*, 48(10), 1267-1272.
2. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., & Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nature protocols*, 13(3), 551.
3. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357.

4. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), 1-9.
5. Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27(12), 1696-1697.
6. Lerdrup, M., Johansen, J. V., Agrawal-Singh, S., & Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nature structural & molecular biology*, 23(4), 349-357.
7. Fleming, A. M., Zhu, J., Ding, Y., Visser, J. A., Zhu, J., & Burrows, C. J. (2018). Human DNA repair genes possess potential G-quadruplex sequences in their promoters and 5'-untranslated regions. *Biochemistry*, 57(6), 991-1002.