UNIVERSITY OF CALIFORNIA
RIVERSIDE

Chemical Biology of DNA Guanine Quadruplex and its Binding Proteins

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Zi Gao

September 2023

Dissertation Committee:
Dr. Yinsheng Wang, Chairperson
Dr. Joseph C. Genereux
Dr. Wenwan Zhong

The Dissertation of Zi Gao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to the individuals who have been instrumental in the completion of my dissertation. First and foremost, I would like to extend my deepest appreciation to my graduate advisor, Dr. Yinsheng Wang, for his guidance, support, and invaluable mentorship throughout this research journey. His expertise, encouragement, and unwavering commitment to scientific excellence have been truly inspiring.

I would also like to thank my dissertation committee members, Dr. Joey Genereux and Dr. WenWan Zhong, for their insightful feedback, constructive criticism, and valuable suggestions, which have greatly enriched the quality of my work.

I am grateful to postdoctoral fellows in the lab, especially, Dr. Lin Li, Dr. Xiaomei He, Dr. Feng Tang, and Dr. Xiaochuan Liu for their expertise, technical assistance, and collaborative spirit. It has been invaluable in carrying out through most of experiments and data analysis. I would like to extend my appreciation to our dedicated lab manager, Shuli Zhai, for her outstanding efforts in organizing orders, ensuring timely delivery, and efficient lab management.

I am also thankful to my fellow graduate peers, Dr. Jiekai Yin, Jun Yuan, and Yinan Wang, for their constant support, stimulating discussions, and camaraderie throughout this challenging journey. I am also grateful to former graduate students, Dr. Preston Bryan

Williams, Dr. Ming Huang, Dr. Yuxiang Cui, Dr. Yenyu Yang, and Dr. Tianyu Qi, for their contributions and assistance during different stages of this project.

In addition to the individuals mentioned above, I would like to extend my heartfelt appreciation to my friends Wenxin Zhao, Dr. Xiaomei He, Ziting Gao, Siyi Ge, Dr. Zhisheng Hu, Dr. Yi Huang. From the gym runs to the memorable drinking parties, your companionship, encouragement, and support have been a constant source of motivation and positivity throughout this journey. Your willingness to lend an ear, offer advice, and provide much-needed moments of respite have made the challenges more manageable and the successes more meaningful. Your presence in my life has enriched both my personal and academic experiences, and I am truly grateful for your unwavering friendship. Thank you for being there for me every step of the way.

Last but not least, I am forever grateful to my mom and dad for their unconditional love, unwavering support, and belief in my abilities. Their encouragement, and constant presence have been the cornerstone of my academic endeavors. I am truly fortunate to have such loving and supportive parents, and I am eternally grateful for everything they have done for me.

To all those mentioned above and countless others who have contributed to my academic and personal growth, I extend my deepest appreciation. Your guidance, encouragement, and belief in me have been crucial in my accomplishments. This dissertation would not have been possible without your support, and I am honored to have had such amazing individuals in my life. Thank you from the bottom of my heart.

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures, in part or full, are a reprint of the material as they appear in the following publication:

Chapter 2: Gao, Z.; Williams, P.; Li, L.; Wang, Y. A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* **2021**, *20* (11), 4919-4924.

Chapter 4: Gao, Z.; Yuan, J.; He, X.; Wang, H.; Wang, Y. Phase Separation Modulates the Formation and Stabilities of DNA Guanine Quadruplex. *JACS Au* **2023**, 3, 1650−1657.

Chapter 5: Gao, Z.; Yang, Y. Y.; Huang, M.; Qi, T. F.; Wang, H.; Wang, Y. Targeted Proteomic Analysis of Small GTPases in Radioresistant Breast Cancer Cells. *Anal Chem* **2022**, *94* (43), 14925-14930.

DEDICATION

To my family!

ABSTRACT OF THE DISSERTATION

Chemical Biology of DNA Guanine Quadruplex and its Binding Proteins

by

Zi Gao

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2023
Dr. Yinsheng Wang, Chairperson

Guanine quadruplex (G4) structures are nonconical DNA conformations that play crucial roles in various cellular processes, including DNA replication and transcription regulation. G4 structures are frequently found in promoter regions of oncogenes, which render G4 structures promising therapeutic targets. However, the mechanisms through which G4 structures regulate biological functions remain underexplored. This dissertation focuses on the development of novel quantitative proteomic methods to identify putative G4-binding proteins, to characterize novel G4-binding proteins, to explore cellular modifiers of G4 stability, and to discover of novel biomarkers in breast cancer cells.

In chapter 2, we employed an affinity-based quantitative proteomics analysis to identify novel G4-binding proteins. By utilizing three different biotinylated G4-forming oligonucleotides as probes, I achieved a comprehensive analysis of proteins interacting with various G4 conformations, which led to the discovery of over 30 G4-binding proteins.

Building upon the pull-down experiments, I introduced, in Chapter 3, the use of photo-crosslinking G4 probes to capture G4-binding proteins. This approach enabled the capture of weak and transient interactions, where harsh washing conditions eliminate protein-protein interactions during the pull-down process. By Using this method, I identified 99 putative G4-binding proteins, and I also characterized one of these proteins, HELLS as a novel G4 helicase.

In Chapter 4, I demonstrated, for the first time, that G4 DNA structures can undergo phase separation. By utilizing immunofluorescence microscopy and ChIP-seq analysis, I observed that phase separation modulates the stabilities of G4 structures *in vitro* and in cells. This discovery provides new insights into factors that modulate the formation and stabilities of G4 structures in cells.

In Chapter 5, I identified novel small GTPases as biomarkers of radioresistance in breast cancer cells through a multiple-reaction monitoring-based targeted proteomics analysis. The study revealed ARFRP1 as a novel radioresistance biomarker in breast cancer cells, where its downregulation promotes radioresistance.

In conclusion, this dissertation presents novel proteomic approaches for the identification of G4-binding proteins, understanding G4 stability modulation, and the discovery of potential biomarkers in breast cancer cells. These findings enhance our understanding of G4 biology and offer new avenues for therapeutic interventions and biomarker-driven cancer research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 DNA Guanine Quadruplex

DNA encodes genetic information necessary for the survival and functions of organisms. The classic B-form DNA structure is well-known, and is characterized by a double helix stabilized by Watson-Crick base pairing (Figure 1.1A). Aside from the B-form DNA, it has been revealed that non-canonical DNA structures, such as guanine-quadruplexes (G4s), i-motifs, and Z-DNA also play important roles in cell biology as well as cancer and neurogenerative diseases[1, 2]. Among them, G4 is formed from stacking of two or more layers of guanine-tetrads, which consist of a plane of four guanines stabilized by Hoogsteen-base pairing and a monovalent metal ion, such as K^+ or Na^+ (Figure 1.1B). G4s can exhibit diverse topologies dictated by the arrangement of G-tetrads, the presence of bulges or loops, and the orientation of the DNA strands. The primary topological variations observed in G4 structures are parallel, anti-parallel, and mixed forms[3] (Figure 1.1C).

Putative G4-forming sequences have been identified in important genomic regions, including replication origins, promoters, and telomeres, indicating the crucial regulatory roles of G4s in DNA replication, gene expression, and telomere maintenance[4-6]. Furthermore, numerous studies have revealed that G4s are also involved in regulating DNA damage repair, chromatin remodeling, and DNA methylation (Figure 1.2). Additionally, G4 structures can exhibit high polymorphism, where they can adopt different

conformations depending on their environment[7]. These various topologies of G-quadruplexes contribute to their structural diversity and functional versatility. The specific topology adopted by a G4 structure can influence its stability, interaction with binding partners, and biological activities. Therefore, it is essential to employ comprehensive methods that can effectively characterize G4 and provide insights into their structural and functional properties as well as elucidating the relationship between G4s topology and function. Such studies are crucial for understanding their roles in cellular processes and exploring their potential applications in therapeutics.

1.1.1 Characterizations of DNA G-quadruplex

Since the first report of the formation of a stable guanine tetrad structure in 1962, the discovery and identification of G4 structures were not demonstrated until 1987[8, 9]. Subsequently, significant efforts have been devoted to studying the formation of G4 structures both *in vitro* and *in vivo*, spanning various organisms from plants and bacteria to eukaryotes[10]. Computational prediction was employed to detect putative G4 forming sequences with the motif of $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$[11]. Through these predictions, more than 376,000 putative G4-forming sequences have been identified from the human genome. Many of these sequences have been experimentally shown to have the ability to adopt G4 structures under physiological conditions *in vitro*[12-14].

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been utilized to gain insights into the structures of G4 DNA[15, 16]. While the crystallization of G4s poses challenges for X-ray crystallography, NMR spectroscopy employs magnetic field to examine specific atomic nuclei, such as ^{1}H, ^{13}C, or ^{15}N, with high resolution in a

solution under physiologically relevant conditions. In addition to X-ray crystallography and NMR spectroscopy, various other techniques have been employed to investigate G4 structures and their formation.

Circular dichroism (CD)[17], which measures the differential absorption of left- and right-circularly polarized light by G4 DNA, provides valuable information about its secondary structure. G4 structures with different topologies show different spectral patterns. For instance, parallel G4 shows a positive band at 264 nm and a negative band at 245 nm, and antiparallel G4 shows a positive band at approximate 295 nm and a negative band at 260 nm[18, 19] (Figure 1.3). The dimethyl sulfate (DMS) protection assay[20] is another technique used to identify G4 regions within DNA sequences by assessing the susceptibility of single-stranded regions to chemical modification. Furthermore, fluorescent small molecules like thioflavin T (ThT)[21] and *N*-methyl-mesoporphyrin IX (NMM)[22] have been utilized as probes to detect and visualize G4 structures due to their abilities to bind selectively to G4 DNA and emit fluorescence signals; while ThT recognizes all forms of G4 DNA, NMM only binds to parallel G4s. These complementary approaches offer valuable insights into the folding and stabilities of G4 structures, which are important for understanding their biological functions.

1.1.2 Functions of DNA G4s

G4s have been found to influence DNA replication and gene expression depending on their specific genomic locations[23, 24]. G4-forming sequences can be found at 80% of mapped replication origins in human and mouse cells[25]. In addition to their presence in promoters and 5'-untranslted regions (5′UTRs) of oncogenes, such as *KIT*, *MYC* and *KRAS*,

where they can act as transcriptional repressors or activators[26-28], G4s have been implicated in various other important biological functions, highlighting their diverse roles in cellular processes. One notable role of G4 structures is their involvement in chromatin modulation and epigenetic regulation[29]. The formation of G4 structures in regulatory regions can affect the accessibility of DNA to transcription factors and other regulatory proteins, leading to alterations in gene expression patterns and cellular phenotypes[30]. Understanding the intricate interplay between G4 structures and chromatin dynamics is important for unraveling the complexity of gene regulation.

G4 structures have also been shown to be associated with DNA repair processes. Studies have suggested that, other than induction of DNA damage[31], G4s may participate in DNA damage recognition and repair, potentially acting as signaling elements to recruit repair proteins to damage sites[32, 33]. The presence of G4 structures in close proximity to DNA lesions can influence the efficiency and accuracy of DNA repair, contributing to genome stability and integrity[34]. Elucidating the molecular mechanisms underlying the interplay between G4 structures and DNA repair pathways holds great promise for advancing our knowledge of genome maintenance and for developing targeted therapies.

Another intriguing aspect of G4 biology is their involvement in telomere maintenance. Telomeres, composed of tandem repeat sequences, play a critical role in protecting the ends of chromosomes. G4 structures formed by the telomeric repeat sequence of (TTAGGG)n help to stabilize and regulate telomere length[35]. These G4 structures contribute to the prevention of telomere erosion, end-to-end fusion, and genomic instability, thereby maintaining chromosomal integrity and cell viability[36-39].

Understanding the mechanisms by which G4 structures regulate telomere biology can provide valuable insights into aging, cancer development, and potential therapeutic interventions targeting telomerase activity.

Together, the functional repertoire of G4s range from replication and transcription to chromatin modulation, DNA repair, telomere maintenance, and G4s hold potential as therapeutic targets. Due to their unique structural characteristics and prevalence in disease-associated regions of the genome, G4s have emerged as attractive targets for drug development. Small molecules designed to selectively bind and stabilize or disrupt G4 structures have shown promise in modulating gene expression and inhibiting the growth of cancer cells. The development of G4-targeting therapeutics opens up new avenues for precision medicine and personalized treatment strategies, particularly in the field of cancer therapy.

1.1.3 DNA G4-targeting small molecules

Small-molecule G4 stabilizers have emerged as promising tools for studying and targeting DNA G4s. It has been observed that, when cells are treated with G4 stabilizers, there is a higher occurrence of double-strand breaks and aberrant DNA replication and gene expression. This suggests that G4 structures play a role as epigenetic regulators, influencing transcriptional control of genes[40]. The stabilization of G4 structures in non-telomeric regions, such as proto-oncogenes and promoter regions, has shown anti-proliferative and anti-tumor activities in various *in vitro* and *in vivo* models of human cancers.

Recent research has also shed light on the involvement of G4s in regulating autophagy in neurons [41]. This highlights the potential of G4 structures as targets for modulating cellular processes beyond cancer-related pathways. Thus, small molecules that specifically bind to G4 structures with high affinity have the potential to be anti-cancer drugs and contribute to maintaining and protecting genome stability.

Numerous small-molecule ligands have been designed to target G4 structures. Here, we present several notable examples (Figure 1.4). Telomestatin, for instance, binds selectively to telomeric G4 DNA, and was able to cause telomere dysfunction by inhibiting telomerase activity and disturbing the shelterin complex[42]. Another example is meso-Tetra(*N*-methyl-4-pyridyl) porphrine (TMPyP4), which binds to G4s by stacking on top of the G-tetrad, resulting in the downregulation of the *MYC* gene transcription[43]. Pyridostatin (PDS) and PhenDC3 could inhibit the transcription of G4-related genes through interacting with G4 structures via π- stacking with the top G tetrad, showing great enhancement on G4 thermal stability and selectivity towards G4 over other secondary structures[44-46]. In addition to small molecules aimed at stabilizing G4 structures, there have also been discoveries of small molecules designed to unfold G4 structures. One such example is the Phenylpyrrolocytosine (PhpC)-based G-clamp analog, a small molecule that has been investigated for its ability to disrupt G4 structures and alleviate helicase impairment[47]. Small molecules that specifically target DNA G4 structures provide valuable tools for studying the functions of G4s and exploring their potential as therapeutic targets. In fact, some of these molecules are currently in clinical trials as potential anticancer drugs. CX-5461, a G4 stabilizer, exhibits specific toxicity against BRCA-deficient cancer cells[48],

highlighting its potential as a novel therapeutic option. The development of G4-targeting small molecules holds promise for cancer treatment, genome stability maintenance, and understanding the intricate mechanisms of G4-mediated biological processes.

1.2. Identification of G4-binding proteins (G4BPs)

To gain a comprehensive understanding of the functional roles of G4 structures, one has to investigate their interactions with proteins. In this section, we will direct our attention to the methods and approaches utilized for the identification and characterization of these pivotal players in G4 biology. By employing various techniques, researchers have endeavored to unravel the intricate landscape of G4-protein interactions, shedding light on the molecular mechanisms underlying G4-mediated processes.

1.2.1 Biophysical methods

One commonly employed method for identifying novel G4BPs is through pull-down experiments using biotinylated oligonucleotides as bait. This approach takes advantage of the strong non-covalent interaction between biotin and streptavidin, allowing for sample incubation and washing under desired conditions. Control oligonucleotides with substitutions of guanines with other nucleobases are used to prevent G4 folding. Cells are subjected to lysis directly to obtain whole-cell lysate or separation into cytoplasmic and nuclear fractions, depending on the research objectives. The eluted proteins from the bait are then digested and the resultant peptide mixture analyzed using LC-MS/MS, which enables the identifications of G4BPs.

Several quantitative proteomics approaches can be employed to identify and characterize G4BPs. One widely used method is label-free quantitative proteomics, known for its cost-effectiveness and broad applicability. In label-free studies, G4BPs such as nucleolin and PARP have been identified by their exclusive detection in G4 probe pull-down samples, but not in the corresponding control samples [49, 50].

Stable isotope labeling by amino acids in cell culture (SILAC) is one of the most robust methods in quantitative proteomic analysis. Cells are cultured in media containing heavy (^{13}C and ^{15}N)- or light (^{12}C and ^{14}N)-labeled arginine and lysine, and these amino acids are incorporated into newly synthesized proteins. By applying the method, G4BPs, such as SLIRP, YY1 and VEZF1 were identified[51]. Isobaric tandem-mass-tag (TMT) labeling expands the number of labeling samples to 16 at the peptide level, as opposed to protein-level labeling in SILAC. This approach led to the identification of SMARCB1 as a cMYC-G4 binding protein[52].

While label-free quantitative proteomics, SILAC, and TMT labeling have improved peptide identification and quantification, these methods rely on affinity capture of proteins, which may result in the detection of indirect binding through protein-protein interactions and may not identify those proteins with transient and/or weak interactions with G4 DNA. To address this limitation, crosslinking probes have emerged as a potential solution. These probes are designed to form covalent bonds between oligonucleotides and their interacting proteins, thereby stabilizing the interaction and enabling their subsequent identification.

Recently, Zhang et al.[53] developed a co-binding-mediated protein profiling (CMPP) method to identify the G4 binders in live cells. Briefly, they designed a small-molecule

probe based on a G4-binding ligand, PDS, conjugated with a linker, a diazirine and an alkyne group. The probe can be photo-crosslinked to putative G4BPs mediated by proximity in cells. A biotin group can be attached to the probe for pulldown purpose after the ligand-protein crosslinking and nuclear protein extraction. The samples were subsequently subjected to a label-free proteomics analysis, which led to identification of over 200 putative G4BPs. Similarly, a new method named G4-LIMCAP, which utilizes a probe consisting of a derivative of pyridostatin (N,N′-bis(2-quinolinyl)pyridine-2,6-dicarboxamide), a diazirine group with photo-crosslinking ability and an alkyne group that is used for the Cu(I)-catalyzed azide−alkyne cycloaddition (CuAAC) reaction[54](Figure 1.5). With this method, not only a list of putative G4BPs were identified, but also confirmed the direct binding of newly identified G4-binding protein, SERBP1.

Human protein microarray was also used to identify G4BPs. By running G4 oligos through plates fixed with over 10,000 proteins, BRD3 was identified as a G4 binding protein.[55]

1.2.2 Genetic and bioinformatics methods

A genome-wide screening of human genes that are silenced by shRNA after the treatment with a small-molecule G4-stabilizer reveals the G4-involving pathways and G4-associated proteins[56] (Figure 1.6). In the study, cells were transfected with a genome-wide pool of shRNAs targeting the protein-coding genes followed by G4 ligand treatment to stabilize DNA and/or RNA G4 structures. In this respect, if a gene is not required in a G4-dependent process, there is no effect on cell viability upon its knockdown. On the other hand, gene silencing results in cell death either due to loss of a direct G4 interaction (e.g.,

binding/unwinding) or indirectly through gene loss in a G4-dependent pathway. In this case, in absence of ligand, cells are viable in presence of the shRNA. The genes were further identified by PCR and sequencing. The authors identified 758 genes whose losses sensitize cells toward PDS. Further analysis using Gene Ontology (GO) revealed that these genes are involved in biological processes including DNA replication, cell cycle, ubiquitin-mediated proteolysis, spliceosome and ribosome.

Computational analysis has emerged as a valuable approach for the identification of putative G4 binding proteins. As the discovery of novel G4BPs continues to expand, various G4-binding motifs have been uncovered. One such motif is the arginine/glycine-rich domain (RGG domain), which is commonly found in RNA-binding proteins and has been observed in numerous G4BPs. For instance, CIRBP was recently identified as a G4-binding protein through the investigation of its interaction with G4 nucleic acids, specifically involving the RGG domain[57]. Furthermore, a recent statistical study based on the analysis of 77 known human G4BPs showed that besides the known RG rich domain, a 20-amino acid G4-binding motif is conserved among G4BPs[58].

The development of G4-specific antibodies and genome-wide sequencing techniques have significantly advanced our ability to map the localization of G4 structures throughout the entire genome. G4 ChIP-seq, using the single-chain antibody BG4, enables the precise mapping of G4 loci in the genome[59]. Subsequently, other antibodies, such as D1, which specifically target parallel G4s, have been developed, expanding our repertoire of G4 mapping tools[60] (Figure 1.7). Additionally, the application of the CUT&Tag method has further enhanced our ability to detect G4 peaks with superior signal-to-noise ratio

compared to other G4 mapping methods, which results in higher reliability in identifying G4-rich regions in cells[61, 62] (Figure 1.8). By combining the existing G4-ChIP data with the ChIP-seq results of putative G4BPs, the overlap in peak regions between the G4 structures and the proteins of interest (POI) can be calculated. This approach allows for assessing the potential interactions between the putative G4BPs and the G4 structures. Notably, comparing G4 structure loci with transcription factor (TF) binding sites revealed significant enrichment of TFs at G4 sites, including known G4BPs such as TAF15, FUS, and SP1[63]. However, it is crucial to validate these predicted interactions using biochemical experiments by confirming the observed associations between the candidate G4BPs and G4 structures, ensuring the accuracy and reliability of computational predictions.

1.3 Functions of G4BPs

To gain a more comprehensive understanding of how G4s regulate biological functions within cells, it is crucial to study G4BPS. Exploring the interactions between G4 structures and specific proteins can provide insights into the complex regulatory networks and unravel the puzzle of G4-mediated cellular processes.

The discovery and characterizations of G4BPs have yielded significant insights into the intricate relationship between DNA G4 structures and cellular processes. In the following sections, we will discuss the functions of various known G4BPs, highlighting their involvement in specific biological processes.

1.3.1 G4BPs in telomere maintenance

G4BPs play crucial roles in regulating biological functions related to telomeres. Several proteins have been found to interact with G4s in the telomeric region. These proteins interact with G4 structures in the telomeric region and contribute to the maintenance of telomere integrity and function. One key protein involved in telomere regulation is TRF2, which is part of the telomere shelterin complex. TRF2 exhibits high affinity in binding to double-stranded telomeric DNA and protects chromosome ends [64]. Depletion of TRF2 in cells leads to telomere dysfunction, including the loss of telomere overhang, end-to-end chromosome fusions, and apoptosis induced by ATM and p53 [65, 66]. TRF2 also interacts with telomere DNA and telomeric repeat-containing RNA (TERRA) through G4 structures [67]. Another important G4BP in the telomeric region is POT1, which is also a component of the shelterin complex. POT1 binds strongly to the telomeric G4 with a dissociation constant (K_d) of 26.4 nM[36]. It also forms a complex with TPP1 and consequently induces dynamic folding and unfolding of G4 in the telomeric region [68]. The sliding motion of POT1-TPP1 complex plays an essential role in the dynamic regulation on the telomere structure [68].

Replication protein A (RPA), a single-stranded DNA binding protein, has also been shown to interact with G4 structures in the telomeric region. RPA1 and RPA2 bind to human telomeric G4 structures and unfold these structures [69]. RPA is involved in telomerase activation, and mutations in RPA can lead to telomere shortening [70]. Furthermore, *BRCA1*, a well-known tumor suppressor gene associated with breast cancer, has been implicated in telomere regulation. Overexpression of BRCA1 in cells resulted in

telomerase inhibition and telomere shortening[71]. Mutations in *BRCA1* increase notably the risk of female breast and ovarian cancers, which may be attributed to its role in DNA repair and transcription regulation[72]. A telomeric ChIP assay showed that BRCA1 is localized at the telomere, and knockdown of BRCA1 caused elevated HTERT expression and increased telomerase activity and telomere length [71, 73]. TRF1 and TRF2 interact with BRCA1 in the telomere region in a DNA-dependent manner, which is associated with the length regulation of the 3′ G-rich overhang [71, 73].

1.3.2 G4BPs in DNA replication

Loss-of-function studies of a helicase involved in recognizing and resolving G4 structures *in vitro* provided the initial evidence that replication fork function can be compromised by G4s. Studies of helicases (e.g., FANCJ and Pif1) have shed light on their roles in resolving G4 DNA structures and maintaining genomic stability. FANCJ, a member of the Fanconi anemia (FA) pathway, plays a critical role in the repair of DNA interstrand crosslinks and is involved in maintaining the stabilities of replication forks. Deletion of the *Fancj* gene in *Caenorhabditis elegans* resulted in the accumulation of small deletions upstream from G4 structures, indicating that FANCJ is essential for the proper replication of G4-rich regions[74]. Additionally, human cells lacking functional FANCJ exhibit an increased susceptibility to genomic instability and the accumulation of large deletions in the vicinity of G4 structures[75]. These observations suggest that FANCJ helicase is crucial for resolving G4 structures during DNA replication and preventing replication fork stalling or collapse.

On the other hand, the Pif1 helicase, found in various organisms including yeast and humans, is known for its ability to unwind G4 DNA structures *in vitro*. Deletion of the *Pif1* gene in *Saccharomyces cerevisiae* induces rearrangements within guanine-rich minisatellites, underscoring its role in maintaining the stabilities of G4-prone regions. Furthermore, studies using potent G4 ligands, such as Phen-DC3, which inhibit Pif1-dependent G4 unwinding *in vitro*, have phenocopied the genomic instability observed in the absence of Pif1. These findings highlight the functions of Pif1 helicase in preventing G4-induced genomic instability and suggest that its activity is crucial for the faithful replication of G4-containing DNA regions.

Together, the roles of FANCJ and Pif1 helicases in resolving G4 structures and maintaining genomic stability underscore the importance of G4BPs and helicases in preserving genome integrity. The loss or dysfunction of these helicases can lead to replication fork stalling, DNA damage, and genetic instability.

1.3.3 G4BPs in transcription regulation

G4BPs play crucial roles in regulating transcription of key oncogenes. *MYC* is a proto-oncogene that participates in normal cell growth and differentiation [76]. As a common feature of malignant human tumors, aberrant expression of MYC caused chromosomal translocation, gene amplification and abnormal transcription [76]. The DNA sequence derived from the promoter of *MYC* was shown to form G4 structures in cells[49]. Mao et al.[49] found that nucleolin binds to G4 structure in the *MYC* promoter, which promotes G4 formation *in vivo*, thereby regulating *MYC* expression. Cellular nucleic-acid-binding protein (CNBP) harbors zinc finger and RGG domains. A previous study revealed that

CNBP could bind specifically to the G4-forming sequence of the *MYC* promoter and promote the G4 formation [77]. Nucleophosmin (NPM1) plays important roles in a variety of fundamental biological processes and tumor malignancies [78], which could bind to G4 structure derived from *MYC* promoter in vitro [79]. Moreover, helicase DDX5 has been shown to bind and unwind G4 structure at *MYC* promoter to facilitate its expression [80]. Two stable G4s in equilibrium can be formed in the promoter of *KRAS*, which is the most mutated oncogene[81]. Cogoi et al.[27] reported that Myc-associated zinc finger (MAZ) could bind specifically to the duplex and quadruplex conformations of the G4-forming sequence, whereas poly (ADP-ribose) polymerase 1 (PARP1) binds specifically to the G4 conformation. Abrogating G4 formation by mutating G4-forming sequence in the *KRAS* promoter led to its reduced transcription [27]. Notably, it was also reported that PARP1 could bind to G4 derived from *VEGF* promoter, *cKIT* promoter and human telomere, MAZ could bind to G-quadruplex derived from *HRAS* promoter [82, 83].

Through an integrated analysis of ChIP-seq data set, Raiber et al.[84] reported that 87% of specificity protein 1 (SP1) binding sites overlap with the G4-forming sequences, indicating that SP1 may be a generic G4-binding protein. *In vitro* experiments revealed that SP1 binds to G4s derived from *cKIT* and *HRAS* promoters[83, 84]. Overexpression of SP1 activates *HRAS* transcription, whereas knockdown of SP1 represses it [83, 84].

1.3.4 G4BPs in epigenetic regulation

5-methylcytosine in DNA is an essential epigenetic mark in chromatin, which assumes roles in modulating gene transcription and disease development[85]. Bioinformatic analysis showed that G4 structures are associated with hypomethylation at CpG islands.

Mao et al.[86] demonstrated that DNMT1 binds directly with G4 at CpG islands. In vitro experiments showed that the interaction between DNMT1 and G4 structure impaired the methylation activity of DNMT1[87]. It was hypothesized that G4 formation at CpG islands is recognized by DNMT1, which inhibits the latter's methylation activity and protects the CpG islands from methylation [86].

1.3.5 G4BPs in 3D genome organization

Yin Yang-1 (YY1) is a ubiquitously expressed and multifunctional transcription factor[88]. Aside from its binding with a double-stranded consensus motif, YY1 binds strongly to G4 structures across the genome, with comparable binding affinity. Notably, YY1-G4 interaction also participates in the DNA looping formation and regulates downstream gene expression[89].

Similarly, CCCTC-binding factor (CTCF) is architectural protein that can mediate chromatin loops, enhancer–promoter interactions, transcriptional pausing and alternative mRNA splicing[90]. It is highly co-localized with G4 sites in cells[91]. Evidence suggests that G4 structures at CTCF-binding sites contribute to the establishment and maintenance of long-range chromatin interactions through facilitating the formation of chromatin loops, which are essential for proper gene regulation and genome organization.

1.3.6 G4BPs in DNA repair and genome stability

G4s act like a roadblock for replication fork progression, leading to fork stalling followed by DNA double strand break (DSB) induction. Treatment of cells with PDS, a G4 stabilizer, led to significantly increased formation of γH2AX, which are indicative of DNA damage, and activate the ATM and CHK1 signaling pathway involved in the DNA

damage response [46]. G4 helicases (e.g., BLM, WRN) can unwind G4s and prevent G4-induced replication stalling and recombination events[32, 92]. For example, mutations in BLM are associated with Bloom syndrome, a rare genetic disease. It was shown that BLM helicase is responsible for G4 structure unwinding, and the loss of its helicase activity results in a marked increase in recombination events, particularly at transcribed genomic loci[32, 93]. These observations underscored the crucial role of proper G4 structure regulation in maintaining genomic stability, as the stabilization of G4 structures can induce DSBs and promote recombination, ultimately leading to genome instability[93]. Together, these findings highlight the detrimental consequences of perturbation G4 dynamics and emphasize the significance of understanding and maintaining the delicate balance in G4-mediated epigenetic modulation for cellular homeostasis.

1.4 Characterization of G4 DNA-protein Interactions

1.4.1 Circular dichroism (CD) spectroscopy

Since the first interpretation of G4 structures by circular dichroism by Spada GP et al. [94], CD has been a well-adopted technique for G4 structure detection. CD is also used to monitor the interaction between G4 nucleic acids and their binding proteins. Since proteins only give relatively small signal intensity emanating from some aromatic amino acids, conformational changes of G4 structures induced by interaction with G4BPs can be detected. The interactions of proteins, including hRPA and nucleolin, with DNA G4 were demonstrated by CD spectroscopy, in which proteins alone do not have strong signal; while the CD spectra are altered after the addition of proteins to G4 DNA probes.[17, 49] The assay

can be conducted under near-physiological conditions with preferred buffer. One of the limitations of CD spectroscopy on characterizing the G4-protein interaction is that it is not a direct observation of protein-G4 interaction.

1.4.2 Electrophoresis mobility shift assay (EMSA)

Electrophoresis mobility shift assay (EMSA), also known as 'gel shift assay', is a robust and widely used tool to study the interaction between proteins and nucleic acids. EMSA not only can be used for qualitative analysis, but also, because of the high sensitivity, can be employed for quantitative analysis of protein-DNA interactions to derive dissociation constants. In EMSA, purified candidate G4BPs are incubated with G4-forming nucleic acids and the samples are resolved on TAE (Tris-acetate-EDTA) or TBE (Tris-borate-EDTA) native polyacrylamide gels. The G4-forming nucleic acids can be labeled with fluorophores or radiolabeled ATP for detection purposes. By exploiting the differential mobility of free DNA and protein-DNA complexes, EMSA enables the separation of these species on the gel[95]. *Bona fide* G4BPs will form a complex with G4 DNA and run more slowly than unbound G4 DNA. EMSA was adopted in most of the identification of protein-G4 binding studies, such as TFAM, hnRNPA1 and hnRNPA2/B1[39, 96, 97]. EMSA provides a straightforward approach for accessing the interactions between proteins and the nucleic acids. However, it is important to note that EMSA does have a limitation wherein the running of samples on gels can potentially disrupt the stability of protein-G4 DNA complexes.

1.4.3 Fluorescence polarization/anisotropy

Fluorescence anisotropy is a powerful technique used to investigate changes in the rotational diffusion of fluorophores labeled on oligonucleotides[98]. By assessing the rotational mobility of the fluorophore, fluorescence anisotropy can provide valuable insights into protein-DNA interactions. When proteins bind to the fluorophore-labeled G4 probes, the rotational motion of the G4 probes becomes restricted, resulting in a decrease in the rotational diffusion rate. Compared to EMSA, fluorescence anisotropy offers several advantages[99]. Firstly, it can be performed under near physiological conditions, allowing for the study of protein-DNA interactions in more biologically relevant settings. Secondly, fluorescence anisotropy measurements can be conducted in real time, providing dynamic information about the kinetics of the binding process. This real-time monitoring enables the detection of rapid binding events and facilitates the investigation of binding kinetics and mechanisms. In addition, the dissociation constant can be derived from such measurements. Anisotropy was used in analysis of the interaction between G4 and proteins, including SLIRP, GRSF1, PC4 and Rif1.[51, 100-102]

1.4.4 Enzyme-linked immunosorbent assay (ELISA)

The binding affinities of proteins towards G4s can also be analyzed by ELISA. This method has been increasingly adopted in recent G4BP studies due to its versatility and high-throughput capabilities.[3, 53, 80] Briefly, biotinylated G4 probes are conjugated to streptavidin-coated plates and incubated with proteins of interest. The protein of interest is then recognized by its specific antibody. The measurement is achieved by assessing the activities of the conjugated enzyme (e.g., horseradish peroxidase) upon incubation with a

substrate (e.g., 3,3',5,5' tetramethylbenzidine (TMB)) to produce a measurable product. ELISA offers several advantages for characterizing protein-G4 interactions. Firstly, it is a plate-based assay, allowing for the analysis of multiple samples simultaneously, which enhances experimental throughput. This feature is particularly useful when screening large numbers of proteins for their binding affinities towards G4 structures. Additionally, ELISA offers high sensitivity and specificity due to the use of specific antibodies for protein detection. The signal generated by the enzyme-substrate reaction can be easily quantified using spectrophotometric or fluorometric measurements.

1.4.5 Förster resonance energy transfer (FRET)

FRET is a robust qualitative and quantitative tool for studying protein-ligand, protein-nucleic acids, and protein-protein interactions. In FRET, a donor fluorophore is excited by incident light, and if an acceptor is in close proximity, the excited state energy from the donor can be transferred to the acceptor. This leads to a reduction in the donor's fluorescence intensity and excited state lifetime, and an increase in the acceptor's emission intensity[103]. In this vein, G4 fluorescent probes were constructed to study the interactions between target protein and G4s. Both the interactions and the G4 unwinding activities can be illustrated. Mainly, two different FRET based assay were employed to study the unwinding kinetics of protein of interest on G4 structures, including single molecule Förster resonance energy transfer (smFRET) assays and bulk FRET assays[104] (Figure 1.9). The G4 unwinding activity of BLM, DDX5 and DHX36 were demonstrated by FRET. [80, 105, 106]

1.4.6 Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

Interrogation of DNA G4-protein interactions through ChIP-seq has provided valuable insights into the role of G4 structures in gene regulation and genome organization. By utilizing specific antibodies targeting G4BPs, such as SMARCA4[53] and YY1[89], ChIP-seq allows for mapping of interacting DNA sequence of the protein. The binding motif of the G4BPs can be found uncovered by such analysis. Furthermore, overlapping the sequencing results with G4 ChIP-seq, which utilizes G4 structure-specific antibody to map the G4 sites at the genome-wide scale will reveal whether the protein binding site is colocalized with G4 sites. In the meanwhile, perturbing protein expression and monitoring the G4 sites in cells or treating the cells with G4 targeted small molecules (e.g., PDS and TMPyP4) can also contributes to understanding the cellular functions of the G4 DNA-protein interactions[89]. This technique has shed light on the involvement of G4 structures in diverse biological processes, including transcriptional regulation, chromatin remodeling, and DNA replication. The integration of ChIP-seq with G4-specific antibodies has significantly advanced our understanding of the intricate interplay between G4 structures and proteins, providing new avenues for exploring the biological significance of G4s and their potential as therapeutic targets.

1.5 Roles of phase separation in cells

1.5.1 Introduction of phase separation

In eukaryotic cells, macromolecules are segregated into distinct compartments or organelles surrounded by membranes, such as nucleus, lysosome, endoplasmic reticulum,

etc. However, some cellular components adopt alternative non-membrane-bound structures known as membraneless organelles. These organelles, such as nucleoli, stress granules, and P bodies, are assembled by proteins, nucleic acids and other molecular components. It has been discovered that these dynamic, liquid-like organelles play essential roles in various biological processes[107] (Figure 1.10). For example, nucleoli, which are formed by the liquid-liquid phase separation of ribosomal RNA (rRNA) and its associated proteins, are responsible for ribosome biogenesis[108]. Similarly, stress granules, formed in response to cellular stress, function as storage sites for untranslated mRNAs, allowing cells to quickly resume translation once stress is alleviated[109]. The formation of the membraneless organelle is driven by liquid-liquid phase separation (LLPS)[110].

1.5.2 Phase-separated biomolecules

Phase separation in cells relies on weak, multivalent, and dynamic interactions between proteins and nucleic acids. It also occurs through the self-assembly of components driven by multivalent interactions and clustering. The assembly of these condensates and the partitioning of biomolecules can be attributed to the interactions including electrostatic interactions, π-π interactions, cation-π interactions, and hydrophobic interactions[107]. Proteins with intrinsically disordered regions (IDRs) or repetitive motifs, which are prone to oligomerization, have been observed to undergo condensate formation[111, 112]. Low-complexity domains (LCDs) are IDRs that contain limited types of amino acids. These regions, a.k.a. prion-like domains (PLDs), are frequently found in RNA-binding proteins (RBPs). An example of an RBP with an LCD is FUS (Fused in Sarcoma), which can undergo phase separation and further disrupt RNP granule function and impair new protein

synthesis in neuron terminals[113]. TDP-43 (TAR DNA-binding protein 43) is another IDR-containing RBP protein. It is demonstrated that TDP-43 mutations associated with amyotrophic lateral sclerosis (ALS) significantly enhanced phase separation[114].

Aside from proteins, DNA and RNA can also undergo phase separation. For instance, Shakya et al.[115] found that certain sequences of DNA, e.g., poly(GC), can form phase-separated condensates. Intermolecular RNA-RNA interactions leads to self-association and phase separation, resulting in the formation of distinct condensates enriched in nucleic acids and determining the composition of certain RNP granules [116]. It was also found that RNA with specific secondary structures, such as RNA G4s can assemble into phase-separated droplets under physiological conditions[117]. This highlights the dynamic and responsive nature of phase separation in orchestrating complex cellular activities, emphasizing its significance in cellular regulation and function. Overall, the phenomenon of phase separation in both proteins and nucleic acids provides a versatile framework for modulating cellular processes. It underscores the importance of LLPS as a fundamental mechanism in cell biology.

1.5.3 Phase separation in regulating biological functions

Recent studies revealed that, other than facilitating the formation of the membraneless granules, phase separation also plays a critical role in regulating a wide range of biological functions, including gene expression, cell signaling, cell division, and stress response[118-120]. For example, transcription factors can undergo phase separation to form transcriptional condensates called super-enhancers[121], which bring together necessary components, including transcription factors, coactivators, chromatin regulators and core

transcription apparatus for control of expression of important genes[122]. The formation of transcriptional condensates enhances the local concentration of transcription machinery[123], modulates chromatin architectures and promotes cooperative interactions among proteins[124].

Phase separation has also been implicated in the assembly of protein complexes involved in cell signaling. Signaling complexes, such as signalosomes and receptor clusters, can assemble through phase separation, bringing together key signaling molecules to enhance signal transduction and cellular responses[125]. It has been shown that LLPS drives the formation of signaling condensates and participates in many immune signaling pathways including T cell receptor (TCR)[126] and B cell receptor (BCR)[127], cGAS-STING[128] and insulin/IGF signaling[125]. Signaling regulation at the synapse, such as the postsynaptic density in neurons, is another example where LLPS may regulate multi-step biochemical processes. By organizing and concentrating signaling molecules, LLPS provides a mechanism for postsynaptic neurons to suppress signaling noise and avoid overexcitation[129]. It highlights the roles of LLPS in not only dense protein complexes but also synaptic signal transmission.

Phase separation also assumes a crucial role in regulating cell division. During cell division, phase separation contributes to the formation of mitotic spindle, heterochromatin and centrosome[130, 131]. Microtubules, such as γ-tubulin, undergo phase separation to promote efficient microtubule growth and establishment of bipolar spindle geometry, ensuring proper chromosome alignment and segregation during mitosis [132]. Moreover, phase separation regulates the spatial distribution of key cell cycle regulators, e.g., cyclins

and cyclin-dependent kinases (CDKs), which govern cell cycle progression[133]. This phase-separated compartment facilitates proper cell growth and division.

Importantly, phase separation is not only involved in cellular processes but also associated with many human diseases, including cancer, neurodegeneration, and infectious diseases. For instance, dysregulation of phase separation of transcription factors and coactivators can disrupt gene regulatory networks and promote uncontrolled cell growth. Mutations in proteins involved in phase separation, such as FUS, DAXX, and hnRNPA1, have been associated with certain types of cancer[37, 113, 134]. Cancer cells generate super-enhancers at oncogenes and other genes important in tumor pathogenesis. Moreover, disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types[135]. Dysregulation of phase separation can lead to the formation of pathological aggregates, such as amyloid fibrils, which are associated with neurodegenerative disorders like Alzheimer's and Parkinson's diseases. Proteins with IDRs or LCDs can undergo phase separation and form liquid-like droplets. However, under certain conditions, these droplets can transition into more stable and insoluble aggregates, leading to pathological accumulation of proteins, including tau and Aβ in AD, α-synuclein in Parkinson's disease, huntingtin protein in Huntington's disease, and FUS/TDP43 in ALS[136].

In summary, phase separation is a rapidly advancing field that has transformed our understanding of cellular organization and function. It is involved in cellular compartmentalization, gene regulation, cell signaling and many other important biological processes. Dysregulation of phase separation can contribute to the development of various

diseases. Further research will deepen our understanding of phase separation mechanisms and its implications in human diseases, potentially leading to new therapeutic approaches for phase separation-related disorders.

1.6 Mass spectrometry-based proteomics analysis

Proteomics is a rapidly advancing field that aims to understand the functions and interactions of proteins in biological systems[137]. MS is a powerful tool used in proteomics to identify and quantify proteins from complex biological samples. Here, we will discuss the different MS-based protein identification and quantification strategies, including discovery proteomics and targeted proteomics.

1.6.1 Discovery proteomics

Discovery proteomics is a powerful technique for the identification and characterization of proteins in biological samples. MS-based protein identification is a prevalent method used in discovery proteomics. The objective of discovery proteomics is to identify as many proteins as possible in a sample, which can offer valuable insights into the role and interactions of proteins in biological systems.

Discovery proteomics is also called shotgun proteomics or bottom-up proteomics[138]. In this technique, proteins are digested into peptides using proteases, then subjected to LC-MS/MS analysis. Peptide identification involves comparing experimental tandem mass spectra with theoretical spectra generated from *in silico* digestion of a protein database. Data-dependent acquisition (DDA) and data-independent acquisition (DIA) are two commonly used approaches for MS-based protein identification in discovery proteomics.

DDA entails the selection of the most abundant ions from the precursor ion scan for fragmentations, which generate tandem mass spectra (MS/MS) that are then used for peptide identification[139]. While this approach is sensitive and can identify numerous peptides and proteins in a single experiment, its reproducibility and selectivity are limited since the selection of precursor ions for fragmentation is dependent on their relative abundances.

A groundbreaking technique was developed by Gillet et al.[140] called SWATH-MS, a variation of data-independent acquisition (DIA) techniques, and the method yields over 89% peptide coverage vs. ~7% from DDA analysis of the same sample. SWATH-MS is capable of effectively integration of extensive proteome analysis with precise quantification, and reliable accuracy. With SWATH-MS, the mass spectrometer fragments the peptides within a defined *m/z* range using a user-defined retention time window. (Figure 1.11) This approach is highly reproducible due to its comprehensive and unbiased sampling. However, DIA is more complex and requires specialized software for data analysis due to its larger and more complex dataset. Peptide-centric scoring analysis was introduced, which requires prior knowledge about the chromatographic and mass spectrometric behaviors of all queried peptides in form of peptide query parameters (PQPs)[141]. It then tests every peptide queried and assigns a p value for each peptide for a confidence estimate of detection.

Softwares were developed for the analysis of DDA and DIA data, including MaxQuant[142], Proteome Discoverer[143], Skyline[144], MaxDIA[145] and DIA-NN[146]. These tools provide essential information on protein abundance, modifications, and interactions and

can be used for a wide range of applications, including biomarker discovery, drug target identification, and systems biology.

1.6.2 Targeted proteomics

Targeted proteomics is a powerful approach for the quantitative analysis of specific proteins or protein modifications in complex biological matrices. In contrast to shotgun proteomic methods, targeted proteomics focuses on achieving highly reproducible and sensitive measurements of specific peptides, which requires prior knowledge about the analytes of interest.

Multiple-reaction monitoring (MRM), also referred to as selected-reaction monitoring (SRM), is a targeted proteomics technique that has now been widely used for the quantification of specific peptides or proteins in complex biological samples. Before its application on proteomics, MRM has been used for small-molecule measurements since 1977[147]. The introduction of triple-quadrupole (QqQ) mass spectrometers with extended mass ranges allows for MRM-based proteomic analysis, enabling selective detection of specific precursor-product ion transitions of interest.

The design of MRM involves the selection of precursor and product ions based on their unique *m/z* ratios, which allows for the selective detection and quantification of specific peptides. The first quadrupole is used to selectively isolate the precursor ions of interest. The second quadrupole is operated in radiofrequency (RF)-only mode and serves as a collision cell for fragmenting the precursor ions. The third quadrupole is used to monitor the product ions that are generated by the fragmentation of the selected precursor ions (Figure 1.12).

One of the major advantages of MRM is its high sensitivity and reproducibility, which allows for the accurate and reproducible quantification of specific peptides or proteins in complex biological samples. However, one of the main challenges of MRM is the limited number of transitions that can be monitored in a single LC-MS/MS run, which can limit the number of peptides or proteins that can be analyzed simultaneously. To overcome this limitation, Escher et al.[148] introduced normalized retention time (iRT) in 2012. iRT facilitates the multiplexing of experiments by shortening the scheduled window, allowing for the analysis of a greater number of peptides without sacrificing cycle time.

Parallel-reaction monitoring (PRM) is a new targeted proteomics technique that has emerged as a promising alternative to traditional MRM-based methods. One major difference from MRM resides in that all the peptides within a predefined *m/z* range are simultaneously fragmented, generating a full-scan MS/MS that contains all the product ions[149], which can be used for the accurate and reproducible quantification of specific peptides or proteins.

One of the major advantages of PRM over MRM is its high selectivity, which allows for the detection and quantification of low-abundance peptides and proteins in complex biological samples, due to its utilization of a high-resolution mass spectrometer for MS/MS analysis. Unlike MRM, PRM does not require prior knowledge of the peptide transitions because it acquires full-scan MS/MS of the targeted peptides. PRM also offers improved reproducibility and accuracy compared to MRM, as all the transitions of interest are monitored simultaneously in a single acquisition, minimizing the effects of sample variability and drift in instrument conditions.

1.7 Small GTPases

Small guanosine triphosphatases (Small GTPases) are enzymes that catalyze the hydrolysis of guanosine triphosphate (GTP) to guanosine diphosphate (GDP)[150]. They act as molecular switches, cycling between an active, GTP-bound state and an inactive, GDP-bound state, thereby regulating downstream signaling pathways[151]. Small GTPases, comprising over 150 members in humans, are GTP-binding proteins with low molecular weight, with 5 subfamilies including Ras, Rho, Rab, Ran, and Arf[150]. These GTPases play crucial roles in various cellular processes and are involved in signal transduction, gene expression, cytoskeletal organization, receptor internalization, vesicular trafficking, and nucleocytoplasmic transport[152, 153]. Aberrant expression of small GTPases has been observed in different types of cancer, such as hepatocellular carcinoma, non-small cell lung carcinoma, pancreatic carcinoma, colorectal cancer, and prostate cancer[154]. Their dysregulation is associated with promoting or suppressing cancer cell migration, invasion, and metastasis.

Tumors were also found to harbor mutant *RAS* genes, including *KRAS*, *HRAS* and *NRAS*. There are specific associations between the various RAS oncogenes and particular types of human cancer[155]. Proteins in the Rho family, including Rac1, RhoA and Cdc42, were found to be involved in modulating the formation of cytoskeleton, cell polarity, cell cycle progression, membrane transport pathways and transcription factor activity[156].

Overall, small GTPases are key regulators of cellular processes and play vital roles in development, homeostasis, and disease. Their dysregulation can contribute to various human diseases, including cancer. Understanding the functions and mechanisms of small

GTPases is essential for deciphering the complexities of cancer biology and identifying potential therapeutic targets.

1.8 Scope of this dissertation

G4 structures, noncanonical DNA conformations, are intricately involved in fundamental cellular processes such as DNA replication, transcriptional regulation, epigenetic modulation, and chromatin remodeling. These structurally unique motifs are frequently present in the promoter regions of oncogenes, including *KRAS*, *MYC*, and *KIT*, as well as in telomeric regions. G4BPs, exemplified by WRN and BLM, are pivotal players in genetic diseases, underscoring the therapeutic potential of targeting G4 structures. However, despite their diverse and pivotal biological functions, the precise mechanisms through which G4 structures modulate cellular processes remain largely elusive. In order to gain insights into the functions of G4s, investigating the proteins that interact with them becomes imperative. Leveraging LC-MS analysis, we have devised a novel quantitative proteomic approach aimed at unraveling putative G4BPs. The identified candidate G4BPs were subjected to rigorous purification and thorough characterizations to validate their specific DNA-protein interactions. Moreover, we have explored the dynamic behavior of G4 structures in cellular environments, wherein protein-mediated phase separation emerges as a critical determinant influencing G4 stability both *in vitro* and in cellular contexts. Additionally, we have harnessed targeted proteomics strategies to unveil novel biomarkers with potential relevance to breast cancer cells. These comprehensive

investigations contribute to a deeper understanding of G4-mediated processes and their implications in disease contexts.

In chapter 2, we presented a comprehensive investigation involving the development and implementation of an affinity-based quantitative proteomics analysis. To facilitate the identification of proteins interacting with different G4 conformations, we employed three distinct biotinylated G4-forming oligonucleotides as probes. Robust and accurate protein quantification was achieved through the utilization of SILAC. To ensure high confidence in the identified protein candidates, a stringent cutoff criterion was imposed, resulting in a refined list of putative G4 binding proteins. Among them, three proteins exhibit binding affinities to all three DNA G4 structures, while 78 other proteins bind selectively to one or two of the three DNA G4 structures. Notably, additional validation experiments were conducted to confirm the DNA G4 binding abilities of selected proteins, including SLIRP, YY1, VEZF1, and GRSF1. These findings provide a better understanding of the repertoire of proteins involved in G4-DNA interactions and provide insights into their biological functional implications.

In Chapter 3, we present a novel approach using photo-crosslinking G4 probes to capture G4BPs through the formation of covalent bonds. Building upon the methodology employed in the affinity pull-down experiments described in Chapter 2, we implemented a rigorous washing condition to enhance the capture of weak and transient interactions, while effectively eliminating non-specific proteins that were pulled down through protein-protein interaction. Specifically, we introduced an o-nitrobenzyl alcohol (o-NBA) photoreactive group, conjugated to a thymine residue of a biotin-labeled G4 DNA probe. A T-linker was

incorporated to confer flexibility, facilitating the PANAC photoclick reaction with primary amines in close proximity to the G4 probe. Through the utilization of this methodology, we successfully identified 99 proteins as putative G4 binding proteins, employing the same stringent cutoff as described in Chapter 1. Notably, HELLS, a protein of interest, was subsequently purified and validated as a novel G4 helicase.

In Chapter 4, we present a groundbreaking observation regarding the phenomenon of phase separation in DNA G4 structures. For the first time, we demonstrate the ability of DNA G4 structures to undergo phase separation, a process characterized by the formation of distinct liquid-like droplets. To investigate the modulatory effect of phase separation on G4 stability, we employed 1,6-hexanediol to disrupt DNA G4 droplets. Through the utilization of immunofluorescence microscopy and ChIP-seq analysis, we obtained compelling evidence indicating that phase separation plays a significant role in modulating globally the stability of G4 structures within cells. Furthermore, by analyzing the ChIP-seq data obtained from both our study and publicly available data on transcription factors (TFs), we discovered that G4-mediated LLPS is involved in regulating the chromatin occupancy of TFs. This remarkable finding expands our understanding of G4 structure regulation, extending beyond the previously established protein-mediated mechanisms.

In Chapter 5, we employed a targeted proteomics approach to identify novel biomarkers associated with radioresistance in breast cancer cells. To accomplish this, we utilized a high-throughput scheduled MRM method coupled with the incorporation of synthetic stable isotope-labeled (SIL) peptides. Specifically, our study focused on investigating differentially expressed small GTPase proteins in two pairs of breast cancer

cell lines, i.e., MDA-MB-231 and MCF7 and their radio resistant clones. Small GTPases, known for their critical involvement in various cellular processes, were explored as potential biomarkers. Through our targeted proteomics analysis, we successfully identified seven small GTPases exhibiting consistent alterations in both pairs of parental and radioresistant cell lines. Notably, our study uncovered ARFRP1 as a novel biomarker associated with radioresistance in breast cancer cells, with its downregulation being linked to the promotion of radioresistance. These findings expand our knowledge on the molecular determinants of radioresistance and hold implications for the development of improved therapeutic strategies in breast cancer treatment.

In conclusion, this dissertation presents novel proteomic approaches for studying G4 structures, their binding proteins, and their implications in cancer biology. The developed methods led to the identification of a number of novel G4BPs, understanding G4 stability modulation, and the discovery of potential biomarkers in breast cancer cells. These findings enhance our understanding of G4 biology, and offer new avenues for therapeutic interventions of cancer and biomarkers for radiotherapy.

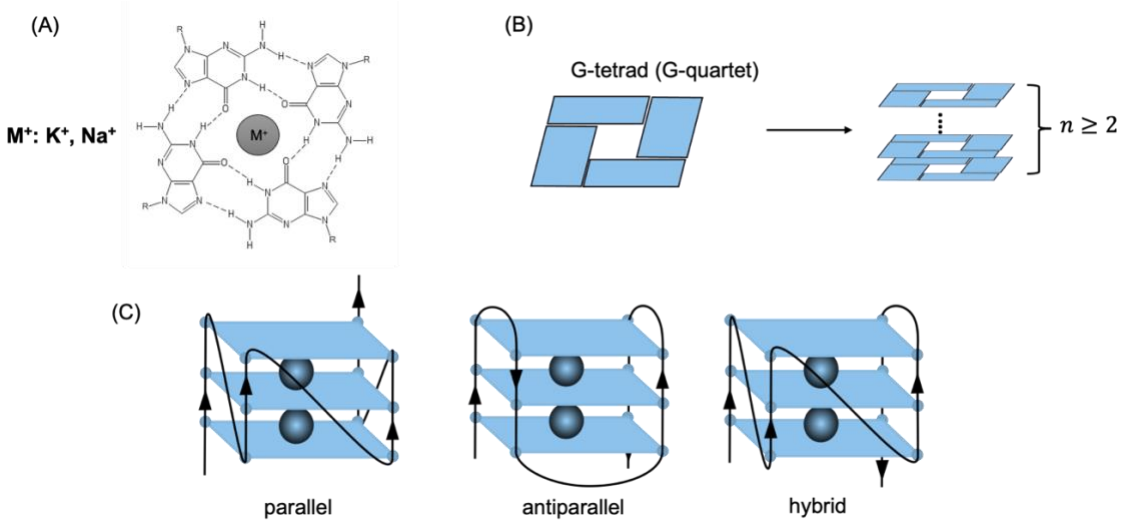

Figure 1.1: G-quadruplex structures. (A) Hoogsteen-base pairs formed between guanines and O^6G coordinate a metal ion. (B) four guanines form a G-tetrad and stack to form a G-quadruplex. (C) G-quadruplex can form into various topology, including parallel, antiparallel and hybrid.

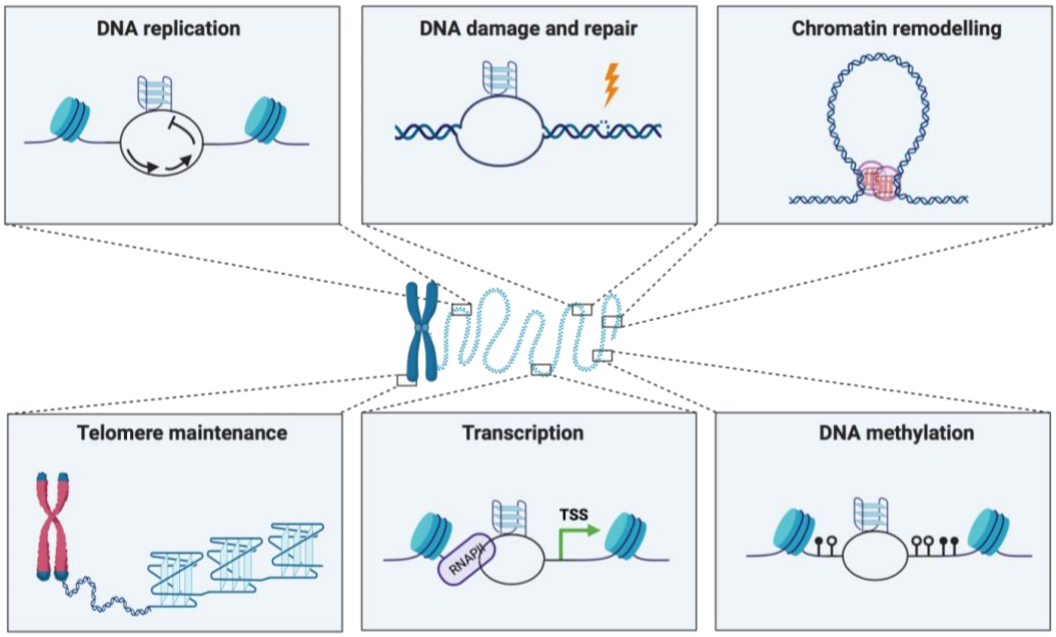

Figure 1.2. A schematic diagram illustrating the functions of G-quadruplex DNA.

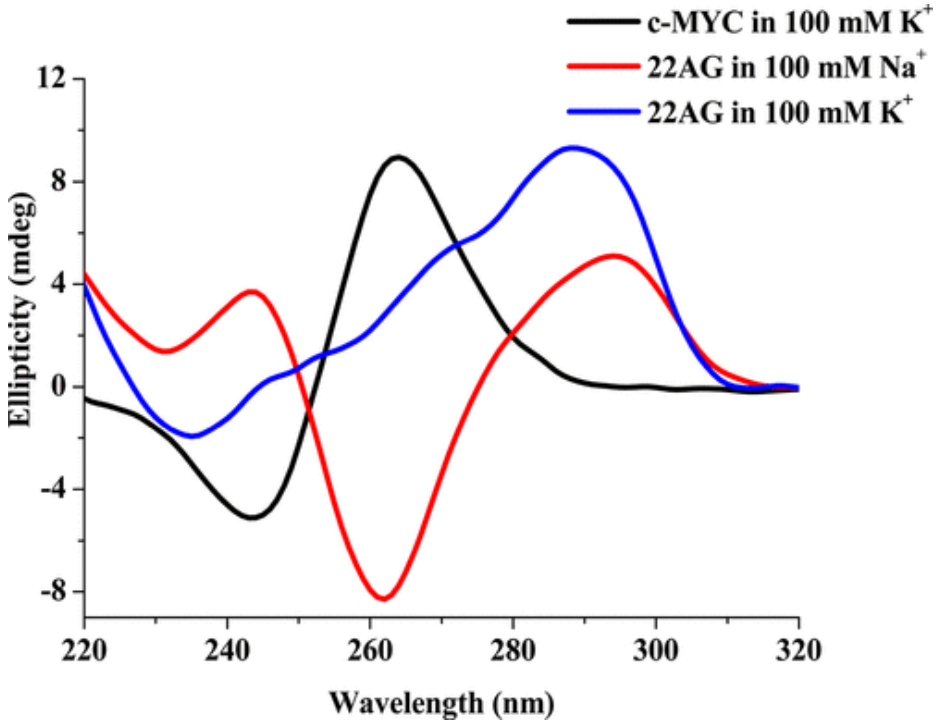

Figure 1.3. Representative CD spectra of G4 structures c-MYC (parallel) and 22AG in 100 mM NaCl (antiparallel) and 100 mM KCl (hybrid). (Adopted from Ref.[19])

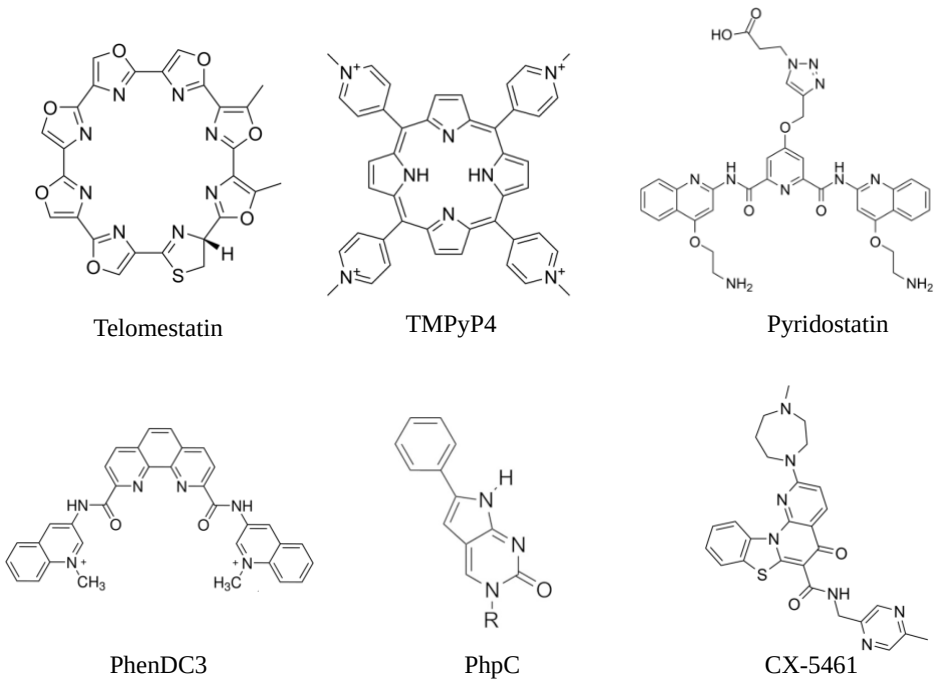

Figure 1.4: G4-binding small molecules. Telomestation, TMPyP4, Pyridostatin, PenDC3 and CX-5416 are G4 stabilizers. PhPC is a G4 disrupter.

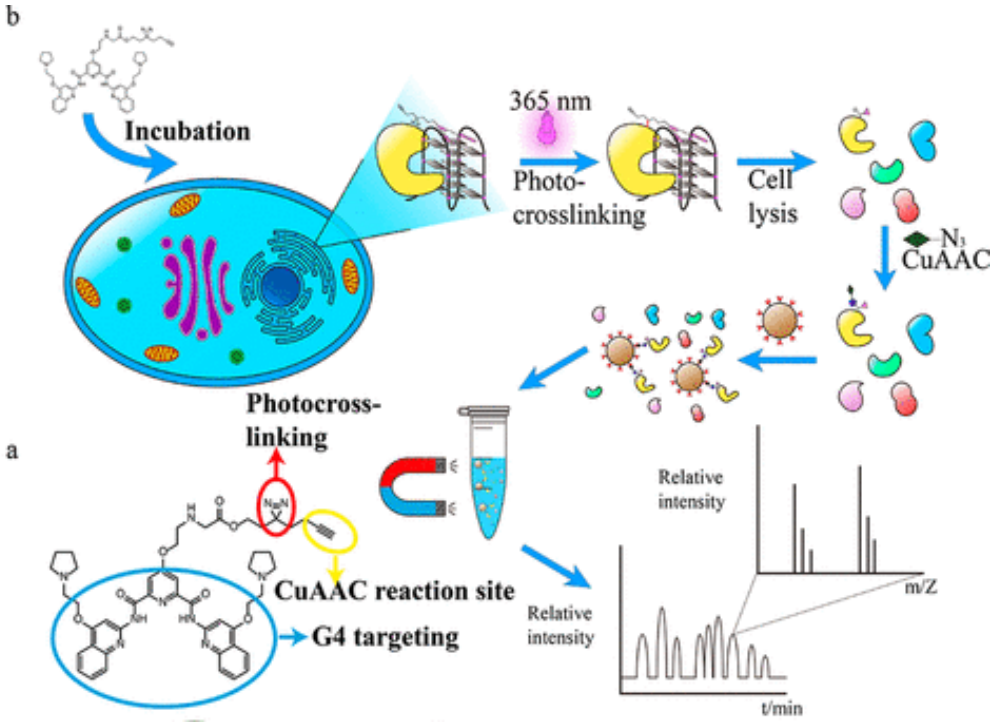

Figure 1.5. Schematic of G-quadruplex ligand-mediated cross-linking and pull-down (G4-LIMCAP) (adopted from Ref.[54])

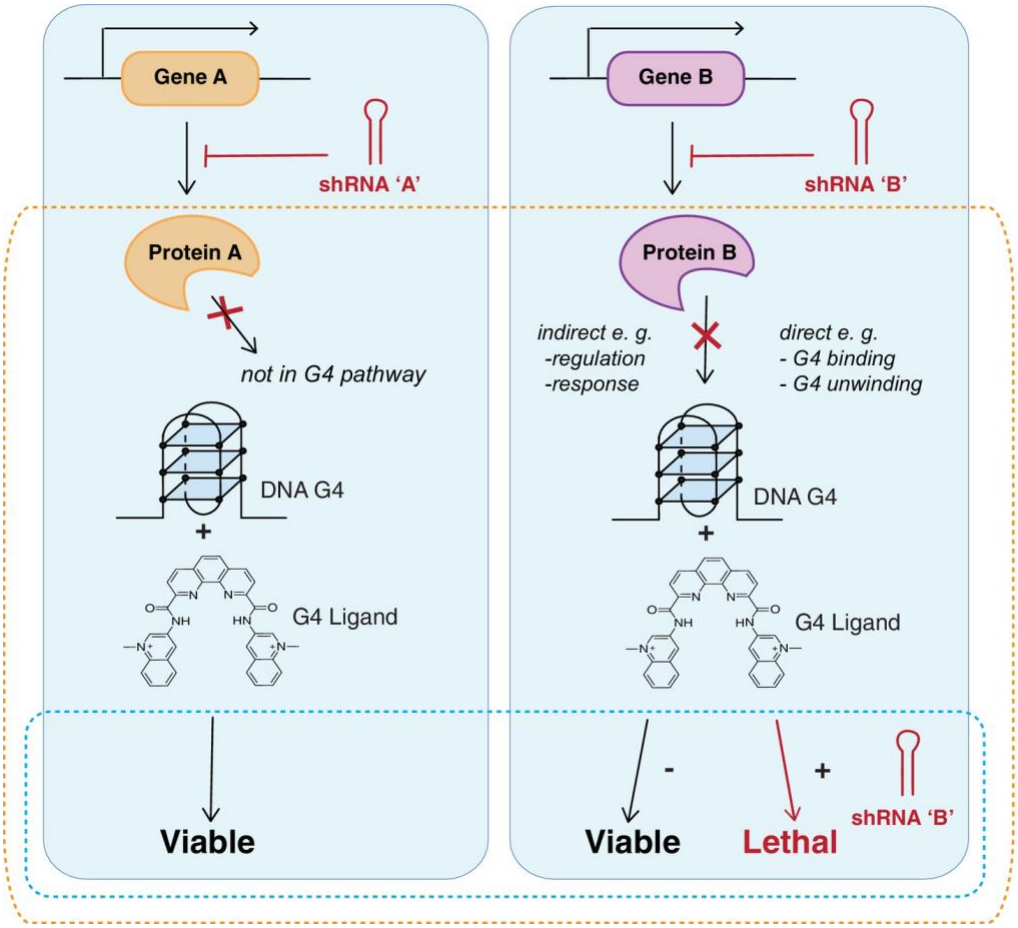

Figure 1.6. a shceme of design of the genome-wide screening shRNA silencing combined with small molecular G4-stablization, identifies genes that when depleted compromise cell viability. (Adopted from Ref[56])

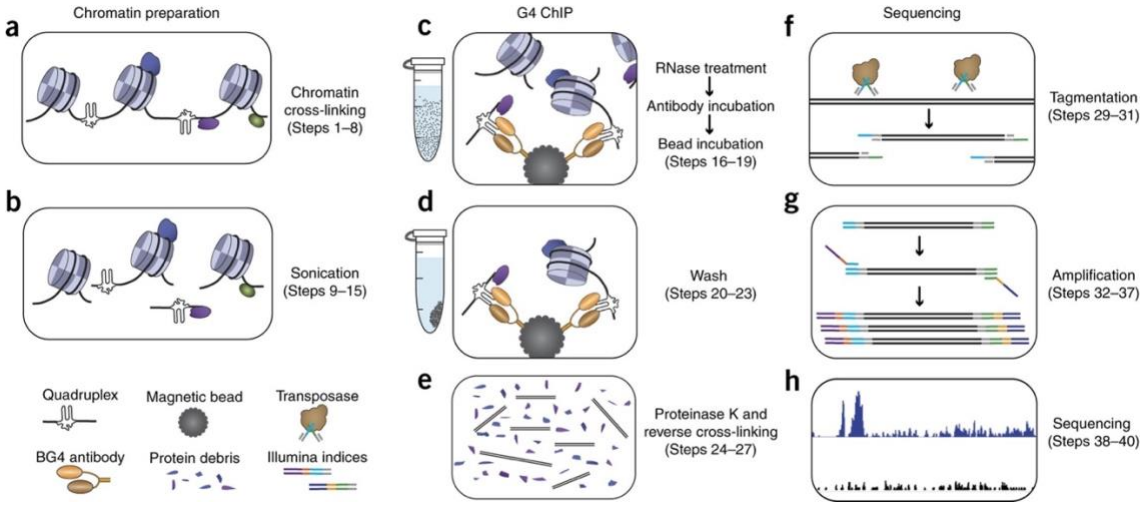

Figure 1.7. G4 ChIP-seq workflow. The procedure of ChIP-seq is demonstrated sequentially from a-h. (Adopted from Ref[59])

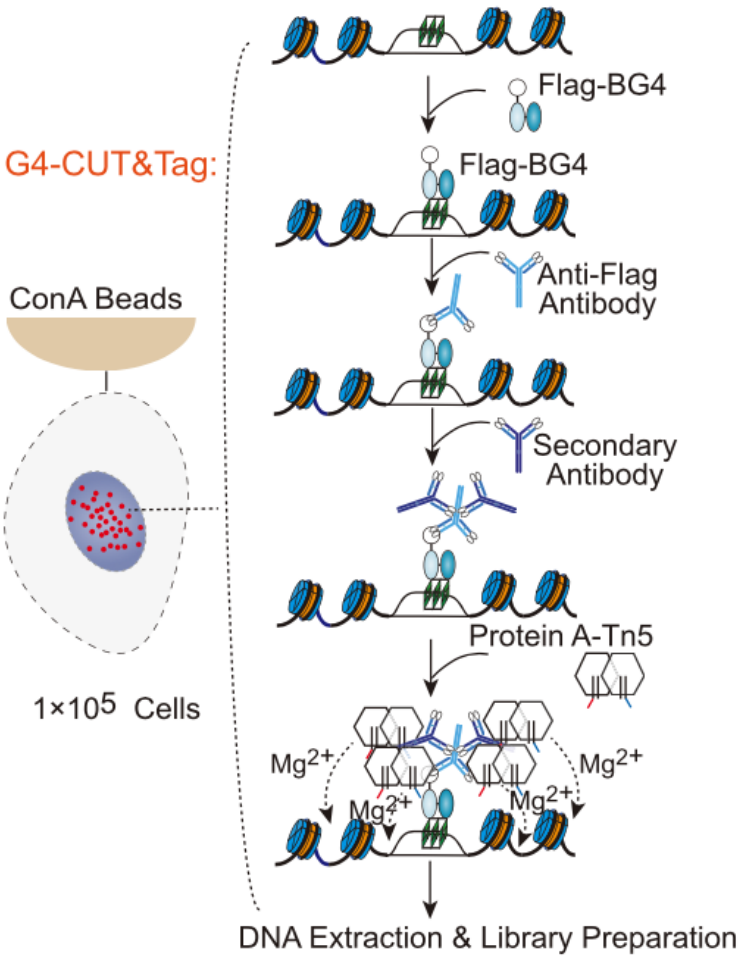

Figure 1.8. G4 CUT &Tag workflow. (Adopted from Ref[62])

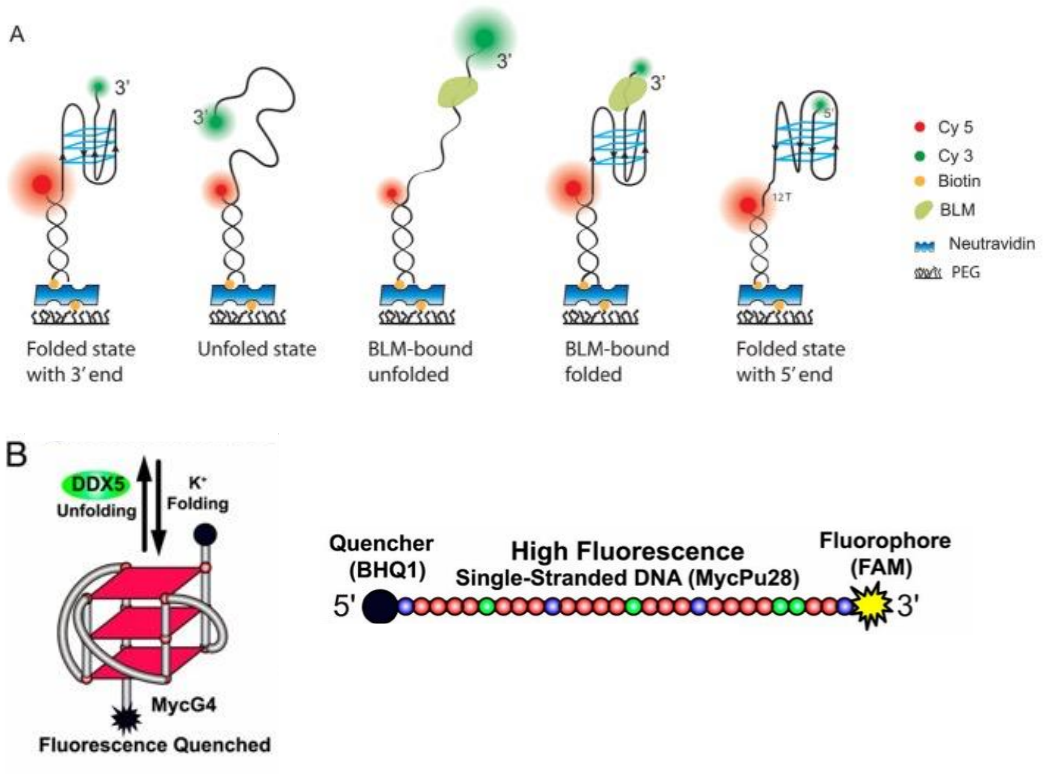

Figure 1.9 (A) Cartoons depicting smFRET assay of different DNA and BLM-DNA complex conformations. (Adopted from Ref[106]) (B) A black hole quencher containing FRET probe was designed for DDX5 helicase unfolding assays. (Adopted from Ref[80])

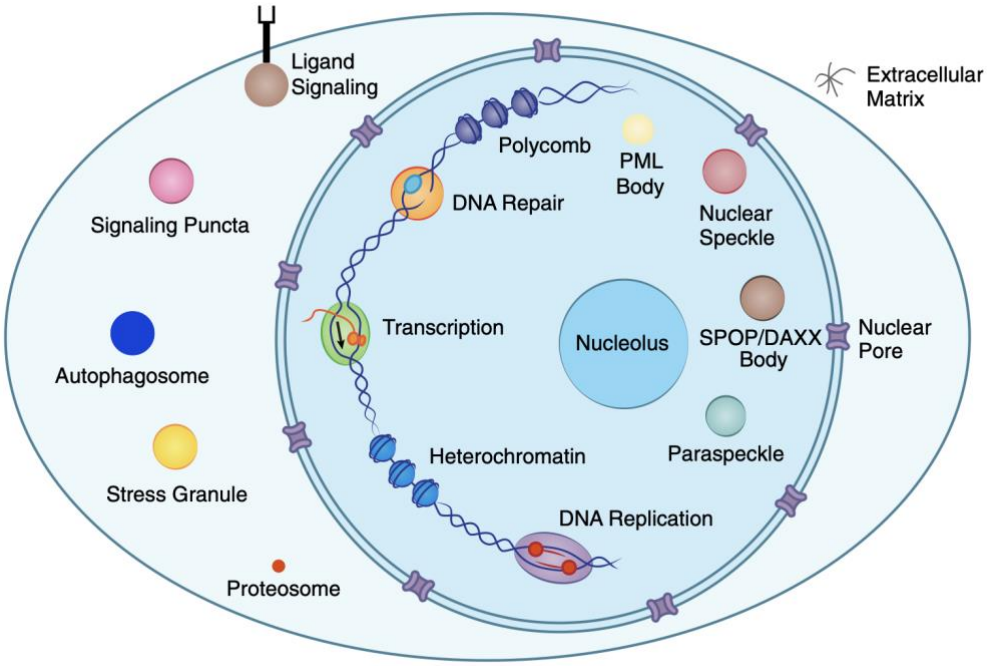

Figure 1.10. Biomolecular condensates in cells. (Adopted from Ref[107])

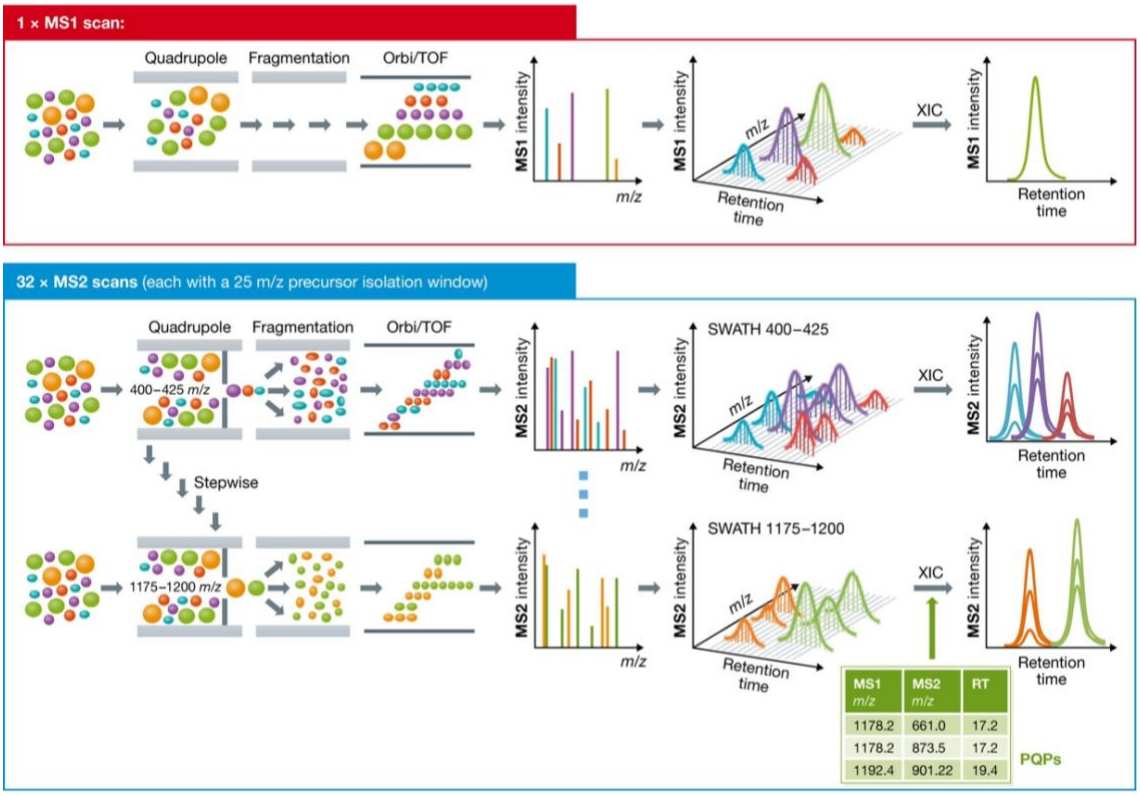

Figure 1.11. A schematic diagram of a SWATH-MS maeasuremnt, where a single precursor MS scan was recorded followed by a serious of MS2 spetrum with a defined precursor isolation window. Peptide query parameters (PQPs) are assigned for data analysis. (Adopted from Ref[157])

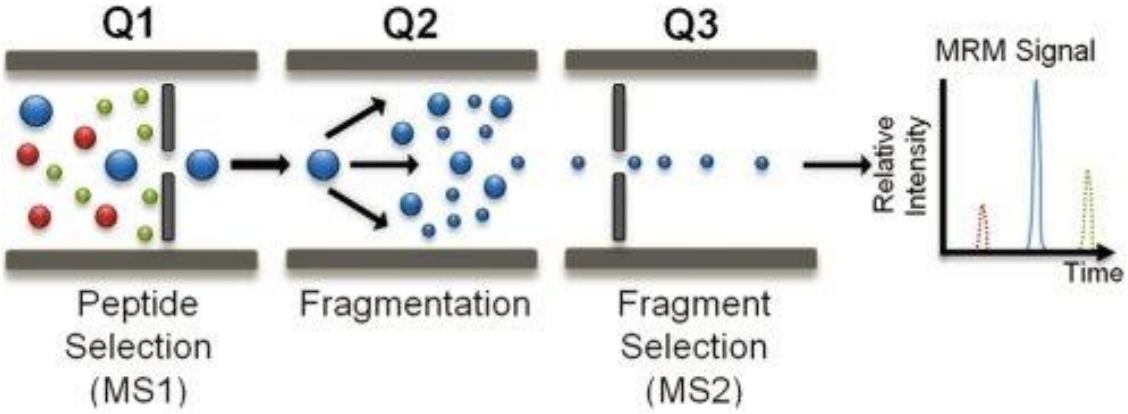

Figure 1.12. A schematic diagram of a QqQ-MS commonly used in MRM analysis. (Adopted from Ref [158])

References

(1) Makova, K. D.; Weissensteiner, M. H. Noncanonical DNA structures are drivers of genome evolution. *Trends Genet* **2023**, *39* (2), 109-124.

(2) Tateishi-Karimata, H.; Sugimoto, N. Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. *Nucleic Acids Res* **2021**, *49* (14), 7839-7855.

(3) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **2013**, *5* (3), 182-186.

(4) Lago, S.; Nadai, M.; Cernilogar, F. M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S. N. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat.Commun.* **2021**, *12* (1), 3885.

(5) Valton, A. L.; Prioleau, M. N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet* **2016**, *32* (11), 697-706.

(6) Paeschke, K.; Juranek, S.; Simonsson, T.; Hempel, A.; Rhodes, D.; Lipps, H. J. Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol* **2008**, *15* (6), 598-604.

(7) Rigo, R.; Groaz, E.; Sissi, C. Polymorphic and Higher-Order G-Quadruplexes as Possible Transcription Regulators: Novel Perspectives for Future Anticancer Therapeutic Applications. *Pharmaceuticals (Basel)* **2022**, *15* (3).

(8) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **1987**, *51* (6), 899-908.

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

(10) Vannutelli, A.; Schell, L. L. N.; Perreault, J. P.; Ouangraoua, A. GAIA: G-quadruplexes in alive creature database. *Nucleic Acids Res* **2023**, *51* (D1), D135-D140.

(11) Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **2005**, *33* (9), 2908-2916.

(12) Parkinson, G. N.; Lee, M. P.; Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002**, *417* (6891), 876-880.

(13) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **2005**, *44* (6), 2048-2058.

(14) Wei, D.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res* **2012**, *40* (10), 4691-4700.

(15) Lin, C.; Dickerhoff, J.; Yang, D. NMR Studies of G-Quadruplex Structures and G-Quadruplex-Interactive Compounds. *Methods Mol Biol* **2019**, *2035*, 157-176.

(16) Campbell, N. H.; Parkinson, G. N. Crystallographic studies of quadruplex nucleic acids. *Methods* **2007**, *43* (4), 252-263.

(17) Fan, J. H.; Bochkareva, E.; Bochkarev, A.; Gray, D. M. Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures. *Biochemistry* **2009**, *48* (5), 1099-1111.

(18) Del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2018**, *57* (24), 7171-7175.

(19) Carvalho, J.; Queiroz, J. A.; Cruz, C. Circular Dichroism of G-Quadruplex: a Laboratory Experiment for the Study of Topology and Ligand Binding. *Journal of Chemical Education* **2017**, *94* (10), 1547-1551.

(20) Onel, B.; Wu, G.; Sun, D.; Lin, C.; Yang, D. Electrophoretic Mobility Shift Assay and Dimethyl Sulfate Footprinting for Characterization of G-Quadruplexes and G-Quadruplex-Protein Complexes. *Methods Mol Biol* **2019**, *2035*, 201-222.

(21) Mohanty, J.; Barooah, N.; Dhamodharan, V.; Harikrishna, S.; Pradeepkumar, P. I.; Bhasikuttan, A. C. Thioflavin T as an efficient inducer and selective fluorescent sensor for the human telomeric G-quadruplex DNA. *J. Am. Chem. Soc.* **2013**, *135* (1), 367-376.

(22) Nicoludis, J. M.; Barrett, S. P.; Mergny, J. L.; Yatsunyk, L. A. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res* **2012**, *40* (12), 5432-5447.

(23) Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J. M.; Lemaitre, J. M. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* **2012**, *19* (8), 837-844.

(24) Dahan, D.; Tsirkas, I.; Dovrat, D.; Sparks, M. A.; Singh, S. P.; Galletto, R.; Aharoni, A. Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res* **2018**, *46* (22), 11847-11857.

(25) Picard, F.; Cadoret, J. C.; Audit, B.; Arneodo, A.; Alberti, A.; Battail, C.; Duret, L.; Prioleau, M. N. The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet* **2014**, *10* (5), e1004282.

(26) D'Aria, F.; Pagano, B.; Petraccone, L.; Giancola, C. KRAS Promoter G-Quadruplexes from Sequences of Different Length: A Physicochemical Study. *Int J Mol Sci* **2021**, *22* (1).

(27) Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K. K.; Xodo, L. E. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J Biol Chem* **2010**, *285* (29), 22003-22016.

(28) Ducani, C.; Bernardinelli, G.; Hogberg, B.; Keppler, B. K.; Terenzi, A. Interplay of Three G-Quadruplex Units in the KIT Promoter. *J Am Chem Soc* **2019**, *141* (26), 10205-10213.

(29) Hansel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48* (10), 1267-1272.

(30) Schiavone, D.; Guilbaud, G.; Murat, P.; Papadopoulou, C.; Sarkies, P.; Prioleau, M. N.; Balasubramanian, S.; Sale, J. E. Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* **2014**, *33* (21), 2507-2520.

(31) Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M. P.; Foiani, M.; Nicolas, A. G-quadruplex-induced instability during leading-strand replication. *EMBO J* **2011**, *30* (19), 4033-4046.

(32) Wu, W. Q.; Hou, X. M.; Li, M.; Dou, S. X.; Xi, X. G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* **2015**, *43* (9), 4614-4626.

(33) De Magis, A.; Gotz, S.; Hajikazemi, M.; Fekete-Szucs, E.; Caterino, M.; Juranek, S.; Paeschke, K. Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nat Commun* **2020**, *11* (1), 3907.

(34) Brosh, R. M., Jr.; Wu, Y. An emerging picture of FANCJ's role in G4 resolution to facilitate DNA replication. *NAR Cancer* **2021**, *3* (3), zcab034.

(35) Yang, S. Y.; Chang, E. Y. C.; Lim, J.; Kwan, H. H.; Monchaud, D.; Yip, S.; Stirling, P. C.; Wong, J. M. Y. G-quadruplexes mark alternative lengthening of telomeres. *NAR Cancer* **2021**, *3* (3), zcab031.

(36) Chaires, J. B.; Gray, R. D.; Dean, W. L.; Monsen, R.; DeLeeuw, L. W.; Stribinskis, V.; Trent, J. O. Human POT1 unfolds G-quadruplexes by conformational selection. *Nucleic Acids Res* **2020**, *48* (9), 4976-4991.

(37) Ghosh, M.; Singh, M. RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res* **2018**, *46* (19), 10246-10261.

(38) Zhang, Q. S.; Manche, L.; Xu, R. M.; Krainer, A. R. hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **2006**, *12* (6), 1116-1128.

(39) Wang, F.; Tang, M. L.; Zeng, Z. X.; Wu, R. Y.; Xue, Y.; Hao, Y. H.; Pang, D. W.; Zhao, Y.; Tan, Z. Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc Natl Acad Sci U S A* **2012**, *109* (50), 20413-20418.

(40) Zimmer, J.; Tacconi, E. M. C.; Folio, C.; Badie, S.; Porru, M.; Klare, K.; Tumiati, M.; Markkanen, E.; Halder, S.; Ryan, A.; et al. Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Mol Cell* **2016**, *61* (3), 449-460.

(41) Moruno-Manchon, J. F.; Lejault, P.; Wang, Y.; McCauley, B.; Honarpisheh, P.; Morales Scheihing, D. A.; Singh, S.; Dang, W.; Kim, N.; Urayama, A.; et al. Small-molecule G-quadruplex stabilizers reveal a novel pathway of autophagy regulation in neurons. *Elife* **2020**, *9*.

(42) Kim, M. Y.; Vankayalapati, H.; Shin-Ya, K.; Wierzba, K.; Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J Am Chem Soc* **2002**, *124* (10), 2098-2099.

(43) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat Chem Biol* **2005**, *1* (3), 167-173.

(44) Halder, R.; Riou, J. F.; Teulade-Fichou, M. P.; Frickey, T.; Hartig, J. S. Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res Notes* **2012**, *5*, 138.

(45) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M. P.; Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (4), 999-1002.

(46) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol* **2012**, *8* (3), 301-310.

(47) Mitteaux, J.; Lejault, P.; Wojciechowski, F.; Joubert, A.; Boudon, J.; Desbois, N.; Gros, C. P.; Hudson, R. H. E.; Boule, J. B.; Granzhan, A.; et al. Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J Am Chem Soc* **2021**, *143* (32), 12567-12577.

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

(49) Gonzalez, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J Biol Chem* **2009**, *284* (35), 23622-23635.

(50) Pagano, B.; Margarucci, L.; Zizza, P.; Amato, J.; Iaccarino, N.; Cassiano, C.; Salvati, E.; Novellino, E.; Biroccio, A.; Casapullo, A.; et al. Identification of novel interactors of human telomeric G-quadruplex DNA. *Chem Commun (Camb)* **2015**, *51* (14), 2964-2967.

(51) Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J Am Chem Soc* **2017**, *139* (36), 12426-12429.

(52) Makowski, M. M.; Grawe, C.; Foster, B. M.; Nguyen, N. V.; Bartke, T.; Vermeulen, M. Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. *Nat Commun* **2018**, *9* (1), 1653.

(53) Zhang, X.; Spiegel, J.; Martinez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **2021**, *13* (7), 626-633.

(54) Su, H.; Xu, J.; Chen, Y.; Wang, Q.; Lu, Z.; Chen, Y.; Chen, K.; Han, S.; Fang, Z.; Wang, P.; et al. Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J Am Chem Soc* **2021**, *143* (4), 1917-1923.

(55) Vlasenok, M.; Levchenko, O.; Basmanov, D.; Klinov, D.; Varizhuk, A.; Pozmogova, G. Data set on G4 DNA interactions with human proteins. *Data Brief* **2018**, *18*, 348-359.

(56) Zyner, K. G.; Mulhearn, D. S.; Adhikari, S.; Martinez Cuesta, S.; Di Antonio, M.; Erard, N.; Hannon, G. J.; Tannahill, D.; Balasubramanian, S. Genetic interactions of G-quadruplexes in humans. *Elife* **2019**, *8*.

(57) Huang, Z. L.; Dai, J.; Luo, W. H.; Wang, X. G.; Tan, J. H.; Chen, S. B.; Huang, Z. S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J Am Chem Soc* **2018**, *140* (51), 17945-17955.

(58) Brazda, V.; Cerven, J.; Bartas, M.; Mikyskova, N.; Coufal, J.; Pecinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **2018**, *23* (9).

(59) Hansel-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. *Nat Protoc* **2018**, *13* (3), 551-564.

(60) Liu, H. Y.; Zhao, Q.; Zhang, T. P.; Wu, Y.; Xiong, Y. X.; Wang, S. K.; Ge, Y. L.; He, J. H.; Lv, P.; Ou, T. M.; et al. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **2016**, *23* (10), 1261-1270.

(61) Lyu, J.; Shao, R.; Kwong Yung, P. Y.; Elsasser, S. J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res* **2022**, *50* (3), e13.

(62) Li, C.; Wang, H.; Yin, Z.; Fang, P.; Xiao, R.; Xiang, Y.; Wang, W.; Li, Q.; Huang, B.; Huang, J.; et al. Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res.* **2021**, *31* (9), 1546-1560.

(63) Spiegel, J.; Cuesta, S. M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **2021**, *22* (1), 117.

(64) Sfeir, A.; de Lange, T. Removal of shelterin reveals the telomere end-protection problem. *Science* **2012**, *336* (6081), 593-597.

(65) van Steensel, B.; Smogorzewska, A.; de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **1998**, *92* (3), 401-413.

(66) Karlseder, J.; Broccoli, D.; Dai, Y.; Hardy, S.; de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **1999**, *283* (5406), 1321-1325.

(67) Biffi, G.; Tannahill, D.; Balasubramanian, S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J Am Chem Soc* **2012**, *134* (29), 11974-11976.

(68) Hwang, H.; Buncher, N.; Opresko, P. L.; Myong, S. POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure* **2012**, *20* (11), 1872-1880.

(69) Safa, L.; Delagoutte, E.; Petruseva, I.; Alberti, P.; Lavrik, O.; Riou, J. F.; Saintome, C. Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie* **2014**, *103*, 80-88.

(70) Luciano, P.; Coulon, S.; Faure, V.; Corda, Y.; Bos, J.; Brill, S. J.; Gilson, E.; Simon, M. N.; Geli, V. RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *EMBO J* **2012**, *31* (8), 2034-2046.

(71) Xiong, J.; Fan, S.; Meng, Q.; Schramm, L.; Wang, C.; Bouzahza, B.; Zhou, J.; Zafonte, B.; Goldberg, I. D.; Haddad, B. R.; et al. BRCA1 inhibition of telomerase activity in cultured cells. *Mol. Cell. Biol.* **2003**, *23* (23), 8668-8690.

(72) Ballal, R. D.; Saha, T.; Fan, S.; Haddad, B. R.; Rosen, E. M. BRCA1 localization to the telomere and its loss from the telomere in response to DNA damage. *J. Biol. Chem.* **2009**, *284* (52), 36083-36098.

(73) Ballal, R. D.; Saha, T.; Fan, S.; Haddad, B. R.; Rosen, E. M. BRCA1 localization to the telomere and its loss from the telomere in response to DNA damage. *J Biol Chem* **2009**, *284* (52), 36083-36098.

(74) Cheung, I.; Schertzer, M.; Rose, A.; Lansdorp, P. M. Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. *Nat Genet* **2002**, *31* (4), 405-409.

(75) Wu, C. G.; Spies, M. G-quadruplex recognition and remodeling by the FANCJ helicase. *Nucleic Acids Res* **2016**, *44* (18), 8742-8753.

(76) Dang, C. V. MYC on the path to cancer. *Cell* **2012**, *149* (1), 22-35.

(77) Chen, S.; Su, L.; Qiu, J.; Xiao, N.; Lin, J.; Tan, J. H.; Ou, T. M.; Gu, L. Q.; Huang, Z. S.; Li, D. Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. *Biochim Biophys Acta* **2013**, *1830* (10), 4769-4777.

(78) Grisendi, S.; Mecucci, C.; Falini, B.; Pandolfi, P. P. Nucleophosmin and cancer. *Nat Rev Cancer* **2006**, *6* (7), 493-505.

(79) Scognamiglio, P. L.; Di Natale, C.; Leone, M.; Poletto, M.; Vitagliano, L.; Tell, G.; Marasco, D. G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. *Biochim Biophys Acta* **2014**, *1840* (6), 2050-2059.

(80) Wu, G.; Xing, Z.; Tran, E. J.; Yang, D. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc Natl Acad Sci U S A* **2019**, *116* (41), 20453-20461.

(81) Marquevielle, J.; Robert, C.; Lagrabette, O.; Wahid, M.; Bourdoncle, A.; Xodo, L. E.; Mergny, J. L.; Salgado, G. F. Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Res* **2020**, *48* (16), 9336-9345.

(82) Soldatenkov, V. A.; Vetcher, A. A.; Duka, T.; Ladame, S. First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem Biol* **2008**, *3* (4), 214-219.

(83) Membrino, A.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One* **2011**, *6* (9), e24421.

(84) Raiber, E. A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res* **2012**, *40* (4), 1499-1508.

(85) Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **2002**, *16* (1), 6-21.

(86) Mao, S. Q.; Ghanbarian, A. T.; Spiegel, J.; Martinez Cuesta, S.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **2018**, *25* (10), 951-957.

(87) Cree, S. L.; Fredericks, R.; Miller, A.; Pearce, F. G.; Filichev, V.; Fee, C.; Kennedy, M. A. DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett* **2016**, *590* (17), 2870-2883.

(88) Gordon, S.; Akopyan, G.; Garban, H.; Bonavida, B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* **2006**, *25* (8), 1125-1142.

(89) Li, L.; Williams, P.; Ren, W.; Wang, M. Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **2021**, *17* (2), 161-168.

(90) Ong, C. T.; Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **2014**, *15* (4), 234-246.

(91) Tikhonova, P.; Pavlova, I.; Isaakova, E.; Tsvetkov, V.; Bogomazova, A.; Vedekhina, T.; Luzhin, A. V.; Sultanov, R.; Severov, V.; Klimina, K.; et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. *Int. J. Mol. Sci.* **2021**, *22* (13).

(92) Johnson, J. E.; Cao, K.; Ryvkin, P.; Wang, L. S.; Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res* **2010**, *38* (4), 1114-1122.

(93) van Wietmarschen, N.; Merzouk, S.; Halsema, N.; Spierings, D. C. J.; Guryev, V.; Lansdorp, P. M. BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat Commun* **2018**, *9* (1), 271.

(94) Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular dichroism of quadruplex structures. *Top Curr Chem* **2013**, *330*, 67-86.

(95) Carey, M. F.; Peterson, C. L.; Smale, S. T. Experimental strategies for the identification of DNA-binding proteins. *Cold Spring Harb Protoc* **2012**, *2012* (1), 18-33.

(96) Xodo, L.; Paramasivam, M.; Membrino, A.; Cogoi, S. Protein hnRNPA1 binds to a critical G-rich element of KRAS and unwinds G-quadruplex structures: implications in transcription. *Nucleic Acids Symp Ser (Oxf)* **2008**, (52), 159-160.

(97) Lyonnais, S.; Tarres-Sole, A.; Rubio-Cosials, A.; Cuppari, A.; Brito, R.; Jaumot, J.; Gargallo, R.; Vilaseca, M.; Silva, C.; Granzhan, A.; et al. The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci Rep* **2017**, *7*, 43992.

(98) Anderson, B. J.; Larkin, C.; Guja, K.; Schildbach, J. F. Using fluorophore-labeled oligonucleotides to measure affinities of protein-DNA interactions. *Methods Enzymol* **2008**, *450*, 253-272.

(99) Shih, P.-C.; Yang, Y.; Parkinson, G. N.; Wilderspin, A.; Wells, G. A high-throughput fluorescence polarization assay for discovering inhibitors targeting the DNA-binding domain of signal transducer and activator of transcription 3 (STAT3). *Oncotarget* **2018**, *9* (66), 32690-32701.

(100) Griffin, W. C.; Gao, J.; Byrd, A. K.; Chib, S.; Raney, K. D. A biochemical and biophysical model of G-quadruplex DNA recognition by positive coactivator of transcription 4. *J Biol Chem* **2017**, *292* (23), 9567-9582.

(101) Mattarocci, S.; Reinert, J. K.; Bunker, R. D.; Fontana, G. A.; Shi, T.; Klein, D.; Cavadini, S.; Faty, M.; Shyian, M.; Hafner, L.; et al. Rif1 maintains telomeres and mediates DNA repair by encasing DNA ends. *Nat Struct Mol Biol* **2017**, *24* (7), 588-595.

(102) Gao, Z.; Williams, P.; Li, L.; Wang, Y. A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* **2021**, *20* (11), 4919-4924.

(103) Selvin, P. R. The renaissance of fluorescence resonance energy transfer. *Nat Struct Biol* **2000**, *7* (9), 730-734.

(104) Mendoza, O.; Bourdoncle, A.; Boulé, J. B.; Brosh, R. M.; Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res* **2016**, *44* (5), 1989-2006.

(105) Chen, M. C.; Tippana, R.; Demeshkina, N. A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D'Amare, A. R. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **2018**, *558* (7710), 465-469.

(106) Budhathoki, J. B.; Ray, S.; Urban, V.; Janscak, P.; Yodh, J. G.; Balci, H. RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res* **2014**, *42* (18), 11528-11545.

(107) Boija, A.; Klein, I. A.; Young, R. A. Biomolecular Condensates and Cancer. *Cancer Cell* **2021**, *39* (2), 174-192.

(108) Brangwynne, C. P.; Mitchison, T. J.; Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc Natl Acad Sci U S A* **2011**, *108* (11), 4334-4339.

(109) Piotrowska, J.; Hansen, S. J.; Park, N.; Jamka, K.; Sarnow, P.; Gustin, K. E. Stable formation of compositionally unique stress granules in virus-infected cells. *J Virol* **2010**, *84* (7), 3654-3665.

(110) Jin, X.; Lee, J. E.; Schaefer, C.; Luo, X.; Wollman, A. J. M.; Payne-Dwyer, A. L.; Tian, T.; Zhang, X.; Chen, X.; Li, Y.; et al. Membraneless organelles formed by liquid-liquid phase separation increase bacterial fitness. *Sci Adv* **2021**, *7* (43), eabh2929.

(111) Banjade, S.; Rosen, M. K. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *Elife* **2014**, *3*.

(112) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166* (3), 651-663.

(113) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Strohl, F.; et al. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-pi Interactions. *Cell* **2018**, *173* (3), 720-734 e715.

(114) Conicella, A. E.; Dignon, G. L.; Zerze, G. H.; Schmidt, H. B.; D'Ordine, A. M.; Kim, Y. C.; Rohatgi, R.; Ayala, Y. M.; Mittal, J.; Fawzi, N. L. TDP-43 alpha-helical structure tunes liquid-liquid phase separation and function. *Proc Natl Acad Sci U S A* **2020**, *117* (11), 5883-5894.

(115) Shakya, A.; King, J. T. DNA Local-Flexibility-Dependent Assembly of Phase-Separated Liquid Droplets. *Biophys. J.* **2018**, *115* (10), 1840-1847.

(116) Van Treeck, B.; Parker, R. Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell* **2018**, *174* (4), 791-802.

(117) Zhang, Y.; Yang, M.; Duncan, S.; Yang, X.; Abdelhamid, M. A. S.; Huang, L.; Zhang, H.; Benfey, P. N.; Waller, Z. A. E.; Ding, Y. G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.* **2019**, *47* (22), 11746-11754.

(118) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **2018**, *28* (6), 420-435.

(119) Ong, J. Y.; Torres, J. Z. Phase Separation in Cell Division. *Mol Cell* **2020**, *80* (1), 9-20.

(120) Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* **2022**.

(121) Sabari, B. R.; Dall'Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **2018**, *361* (6400).

(122) Hnisz, D.; Abraham, B. J.; Lee, T. I.; Lau, A.; Saint-Andre, V.; Sigova, A. A.; Hoke, H. A.; Young, R. A. Super-enhancers in the control of cell identity and disease. *Cell* **2013**, *155* (4), 934-947.

(123) Cho, W. K.; Spille, J. H.; Hecht, M.; Lee, C.; Li, C.; Grube, V.; Cisse, II. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **2018**, *361* (6400), 412-415.

(124) Wang, W.; Qiao, S.; Li, G.; Cheng, J.; Yang, C.; Zhong, C.; Stovall, D. B.; Shi, J.; Teng, C.; Li, D.; et al. A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Res.* **2022**, *50* (9), 4917-4937.

(125) Gao, X. K.; Rao, X. S.; Cong, X. X.; Sheng, Z. K.; Sun, Y. T.; Xu, S. B.; Wang, J. F.; Liang, Y. H.; Lu, L. R.; Ouyang, H.; et al. Phase separation of insulin receptor substrate 1 drives the formation of insulin/IGF-1 signalosomes. *Cell Discov* **2022**, *8* (1), 60.

(126) Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **2016**, *352* (6285), 595-599.

(127) Wong, L. E.; Bhatt, A.; Erdmann, P. S.; Hou, Z.; Maier, J.; Pirkuliyeva, S.; Engelke, M.; Becker, S.; Plitzko, J.; Wienands, J.; et al. Tripartite phase separation of two signal effectors with vesicles priming B cell responsiveness. *Nat Commun* **2020**, *11* (1), 848.

(128) Su, Z.; Dhusia, K.; Wu, Y. Coarse-grained simulations of phase separation driven by DNA and its sensor protein cGAS. *Arch Biochem Biophys* **2021**, *710*, 109001.

(129) Zeng, M.; Shang, Y.; Araki, Y.; Guo, T.; Huganir, R. L.; Zhang, M. Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell* **2016**, *166* (5), 1163-1175 e1112.

(130) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase separation drives heterochromatin domain formation. *Nature* **2017**, *547* (7662), 241-245.

(131) Vega, F. M.; Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett* **2008**, *582* (14), 2093-2101.

(132) Schneider, M. W. G.; Gibson, B. A.; Otsuka, S.; Spicer, M. F. D.; Petrovic, M.; Blaukopf, C.; Langer, C. C. H.; Batty, P.; Nagaraju, T.; Doolittle, L. K.; et al. A mitotic chromatin phase transition prevents perforation by microtubules. *Nature* **2022**, *609* (7925), 183-190.

(133) Krasinska, L.; Fisher, D. A Mechanistic Model for Cell Cycle Control in Which CDKs Act as Switches of Disordered Protein Phase Separation. *Cells* **2022**, *11* (14).

(134) Lin, D. Y.; Huang, Y. S.; Jeng, J. C.; Kuo, H. Y.; Chang, C. C.; Chao, T. T.; Ho, C. C.; Chen, Y. C.; Lin, T. P.; Fang, H. I.; et al. Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* **2006**, *24* (3), 341-354.

(135) Jia, Q.; Chen, S.; Tan, Y.; Li, Y.; Tang, F. Oncogenic super-enhancer formation in tumorigenesis and its molecular mechanisms. *Exp Mol Med* **2020**, *52* (5), 713-723.

(136) Knowles, T. P.; Vendruscolo, M.; Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **2014**, *15* (6), 384-396.

(137) Al-Amrani, S.; Al-Jabri, Z.; Al-Zaabi, A.; Alshekaili, J.; Al-Khabori, M. Proteomics: Concepts and applications in human medicine. *World J Biol Chem* **2021**, *12* (5), 57-69.

(138) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113* (4), 2343-2394.

(139) Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, *11* (4), 535-553.

(140) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **2012**, *11* (6), O111 016717.

(141) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Kall, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol Cell Proteomics* **2015**, *14* (9), 2301-2307.

(142) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016**, *11* (12), 2301-2319.

(143) Orsburn, B. C. Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* **2021**, *9* (1).

(144) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966-968.

(145) Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat Biotechnol* **2021**, *39* (12), 1563-1573.

(146) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **2020**, *17* (1), 41-44.

(147) Baty, J. D.; Robinson, P. R. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. *Biomed Mass Spectrom* **1977**, *4* (1), 36-41.

(148) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12* (8), 1111-1121.

(149) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **2012**, *11* (11), 1475-1488.

(150) Johnson, D. S.; Chen, Y. H. Ras family of small GTPases in immunity and inflammation. *Curr Opin Pharmacol* **2012**, *12* (4), 458-463.

(151) Cherfils, J.; Zeghouf, M. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* **2013**, *93* (1), 269-309.

(152) Kidd, A. R., 3rd; Snider, J. L.; Martin, T. D.; Graboski, S. F.; Der, C. J.; Cox, A. D. Ras-related small GTPases RalA and RalB regulate cellular survival after ionizing radiation. *Int J Radiat Oncol Biol Phys* **2010**, *78* (1), 205-212.

(153) Connelly, T. M.; Berg, A. S.; Harris, L. R., 3rd; Hegarty, J. P.; Ruggiero, F. M.; Deiling, S. M.; Brinton, D. L.; Koltun, W. A. T-cell activation Rho GTPase-activating protein expression varies with inflammation location and severity in Crohn's disease. *J Surg Res* **2014**, *190* (2), 457-464.

(154) Jung, H.; Yoon, S. R.; Lim, J.; Cho, H. J.; Lee, H. G. Dysregulation of Rho GTPases in Human Cancers. *Cancers (Basel)* **2020**, *12* (5).

(155) Karnoub, A. E.; Weinberg, R. A. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol* **2008**, *9* (7), 517-531.

(156) Etienne-Manneville, S.; Hall, A. Rho GTPases in cell biology. *Nature* **2002**, *420* (6916), 629-635.

(157) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol* **2018**, *14* (8), e8126.

(158) Boja, E. S.; Rodriguez, H. The path to clinical proteomics research: integration of proteomics, genomics, clinical laboratory and regulatory science. *Korean J Lab Med* **2011**, *31* (2), 61-71.

Chapter 2: A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins

2.1 Introduction

Regions of genomic DNA with contiguous runs of guanines exhibit the ability to fold into non-B form secondary structures known as guanine quadruplexes (G4).[1] The G4 structures are assembled from multiple G-tetrads stacked upon one another, where a monovalent cation, primarily K^+ or Na^+, further stabilizes the G tetrad structure.[2]

Bioinformatic and experimental studies have revealed the widespread occurrence of G4 structures in the human genome. In this vein, computational analyses uncovered more than 300,000 putative G4-forming motifs in the human genome.[3-5] With the use of a G4 structure-specific antibody (BG4) and fluorescence microscopy analysis, Biffi et al.[6] revealed the presence of G4 structures in chromosomal DNA of human cells. Moreover, chromatin immunoprecipitation using BG4 followed by next-generation sequencing (ChIP-Seq) analyses led to the discovery of approximately 10,000 G4 structure sites in chromatin of cultured human cells.[6-8] These G4 structure sites are enriched at loci of important biological relevance and regulatory functions, including more than 2000 gene promoters and telomeric regions.[6, 7]

DNA G4 structures have been shown to assume important roles in many biological processes, including DNA replication, transcription, alternative polyadenylation, and maintenance of genomic stability.[9-14] In this vein, promoter sequences with the ability to

fold into G4 structures are of particular importance owing to the potential roles of these G4s in gene regulation. For instance, the nuclease hypersensitivity element III_1, which is found within the promoter of c-*MYC* oncogene and regulates 85-90% of its transcriptional activity, harbors a G4 motif.[15] Likewise, the c-*KIT* proto-oncogene harbors two different G4 sequence motifs upstream to its core promoter, and these G4 structures are involved in regulating the expression of the c-*KIT* gene.[16, 17] Moreover, a recent study showed that G4 structure can remotely modulate gene expression by enabling DNA looping.[13] Apart from gene promoters, the human telomere is known to fold readily into G4 structure,[18-20] which modulates telomere integrity.[21]

Many proteins, including nucleolin, Pif1, PARP1, SLIRP, SUB1, Rif1, VEZF1, WRN and YY1, were found to interact with G4 structures.[13, 14, 22-30] We reason that a better understanding about how DNA G4 structures function in gene regulation and human diseases entails a systematic investigation about how these structures are recognized by cellular proteins. Additionally, since the turn loop sizes and primary DNA sequence for each G4 are unique, we reason that cells may also be equipped with proteins that interact selectively with only certain G4-folding pattern(s).

In this study, we conducted an exhaustive quantitative proteomics-based interaction screening using three pairs of DNA probes that are capable or incapable of folding into G4 structures (Figure 2.1 and Table 2.1). Through these experiments, we identified more than 80 candidate G4BPs (Table 2.2 and Figure 2.2). Interestingly, some of these proteins display preferential binding to all three G4 structures than their corresponding mutated sequences, whereas others interact uniquely with certain G4 structures.

2.2 Materials and Methods

Oligodeoxyribonucleotides (ODNs)

The biotinylated G4-forming sequences derived from the human telomere and the promoters of c-*MYC* and c-*KIT* genes and the corresponding mutated sequences unable to fold into G4 structures were purchased from Integrated DNA Technologies (IDT) and purified by HPLC (Table S1). The 5′-carboxytetramethylrhodamine (TAMRA)-labeled DNA sequences used for the fluorescence anisotropy measurements were also purchased from IDT and purified by HPLC.

G-Quadruplex Formation

The biotinylated G4 DNA probes were dissolved in buffer A containing 10 mM Tris-HCl (pH 7.5), 100 mM KCl and 0.1 mM EDTA, and annealed by heating the solution to 95°C for 5 min, followed by the cooling to room temperature slowly over 3 hr. The formation of G4 structures was confirmed by circular dichroism (CD) spectropolarimetric measurements, as previously reported [30].

Cell Culture

HeLa cells were cultured in SILAC DMEM medium (Thermo) supplemented with 10% dialyzed fetal bovine serum (FBS, Invitrogen) and 1% penicillin and streptomycin (Invitrogen). The SILAC media were prepared by supplementing arginine- and lysine-depleted DMEM medium with unlabeled L-arginine (Sigma) and L-lysine (Sigma), or [$^{13}C_6$]-L-arginine and [$^{13}C_6$,$^{15}N_2$]-L-lysine (Cambridge Isotope Laboratories), which are designated as light and heavy media, respectively [31]. The cells were cultured in complete

heavy SILAC media for 10 cell doublings to ensure complete labeling. All cells were maintained at 37°C with 5% CO_2.

Preparation of Nuclear Proteome

HeLa cells, when reached 80% confluency, were harvested using trypsin-EDTA (Invitrogen) and pelleted by centrifugation. The cell pellet was then washed twice with 1× phosphate-buffered saline (PBS). The nuclear proteome was prepared from heavy- and light-labeled cells using the Thermo Pierce NE-PER nuclear and cytoplasmic extraction reagents following the manufacturer's guidelines. The protein concentrations were measured using Bradford assay (Bio-Rad), and the nuclear lysate was stored at -80°C until use.

Affinity Purification of G4BPs

The annealed biotin-conjugated G4 DNA probes and the corresponding mutant probes, at a concentration of 0.5 μM, were incubated separately with high-capacity streptavidin agarose beads (Thermo) with rocking for 60 min following the manufacturer's guidelines. The beads were then washed and equilibrated for three times with a 1-mL aliquot of the aforementioned buffer A. After each washing, the beads were centrifuged at 700*g* for 1 min and the supernatant discarded.

The DNA-bound streptavidin beads were then incubated with 500 μg of nuclear proteome in buffer B, which contained 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 0.5 mM EDTA and 10% glycerol, at 4°C with rocking for 2 hr. In the forward SILAC experiment, the light and heavy nuclear proteomes were incubated with the G4-containing DNA probe and the corresponding mutated probe incapable of folding into G4, respectively. To remove

any experimental bias, we also performed the reverse SILAC experiment where the heavy and light nuclear protein lysates were incubated with the G4-containing probe and the mutant control probe, respectively. After the incubation, the DNA-protein mixture was washed for three times with 1-mL buffer C, which contained 20 mM Tris-HCl (pH 7.5), 50 mM KCl, 0.5 mM EDTA, 10% glycerol, along with elevating concentrations of NaCl (50, 100, and 200 mM). After the washing, the beads were combined and the bound proteins were eluted with the addition of 30 μL of 2× SDS-PAGE loading buffer (Bio-Rad) with 5 min of boiling. The resulting mixture was centrifuged and the supernatant loaded onto a 12% SDS-PAGE gel. After a very short separation, gel band containing the proteins was excised and cut into small pieces. The proteins were then digested in-gel with trypsin, as described previously [32]. Briefly, excess SDS in the gel was removed with overnight shaking in an equal-volume mixture of 25 mM NH_4HCO_3 and acetonitrile. The supernatant was removed and the gel pieces were dehydrated with acetonitrile. Proteins were then reduced with 10 mM dithiothreitol (DTT) at 37°C for 1 hr and subsequently alkylated by incubating with 55 mM iodoacetamide (IAA) (Sigma) in the dark for 1 hr. Gel pieces were washed for three times with 25 mM NH_4HCO_3 (1 mL) with 5 min of shaking. Proteins were then digested in-gel with trypsin at 37°C overnight, and the peptides were subsequently eluted from the gel by incubating, with vigorous shaking for 15 min, first in 25 mM NH_4HCO_3 containing 5% acetic acid for two times, then in 25 mM NH_4HCO_3 containing 5% acetic acid and 50% acetonitrile, and finally in 25 mM NH_4HCO_3 containing in 5% acetic acid and 95% acetonitrile. The eluted peptide fractions were subsequently

pooled, evaporated to dryness, and desalted using OMIX C_{18} Tips (Agilent), following the manufacturer's recommended procedures.

Mass Spectrometry

On-line LC-MS/MS analysis of the peptide samples was performed on an LTQ-Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source and coupled with an EASY-nLC II HPLC system (Thermo, San Jose, CA, USA). The HPLC separation was performed using a trapping column followed by a separation column, both packed in-house with ReproSil-Pur C18-AQ resin (3 μm, Dr. Maisch HPLC GmbH, Germany). The peptides were separated using a 170-min linear gradient of 2-40% acetonitrile in 0.1% formic acid at a flow rate of 230 nL/min and electrosprayed (spray voltage 1.8 kV) into the mass spectrometer operated in the positive-ion mode. Full-scan MS (*m/z* 300-1500) were acquired at a resolution of 60,000 (at *m/z* 400), followed by data-dependent acquisition of MS/MS for the 20 most abundant ions found in the full-scan MS exceeding a threshold of 1000 counts. The normalized collision energy for MS/MS was 35.0.

Data Analysis

All raw data were analyzed in parallel with MaxQuant Version 1.6.15.0 for protein identification and quantification [33]. MaxQuant multiplicity was set to 2, and Lys8 and Arg6 were selected as heavy amino acids. Protein N-terminal acetylation and methionine oxidation were set as variable modifications, and cysteine carboamidomethylation was set as a fixed modification. The maximum number of missed cleavages for trypsin was set to two per peptide. The tolerances in mass accuracy for MS and MS/MS were 20 ppm and

0.5 Da, respectively. Raw MS data were searched against the UniProt human proteome database (Proteome ID: UP000005640_9606) to which contaminations and reverse sequences were added. The match between runs option was enabled with the alignment window being 5 min. Raw output results were analyzed and known contaminant proteins were removed from analysis. Proteins exhibiting a G4/ssDNA SILAC ratio of at least 1.5 were categorized as putative G4BPs.

Generation of Recombinant GRSF1 Protein

The coding sequence (CDS) of human *GRSF1* gene was amplified by PCR and inserted into the pGEX plasmid. The plasmid was introduced into Rosetta (DE3) pLysS *Escherichia coli* cells and cultured in LB medium at 37°C. Induction was conducted with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG, Sigma) at 20°C for about 12 hr. The cells were then harvested and sonicated for 8 min. After centrifugation, the supernatant was collected and GST-tagged GRSF1 protein was purified with glutathione superflow agarose (Pierce) following the manufacturer's recommended procedures. The concentration of the GRSF1 proteins was quantified using Bradford protein assay kit (Bio-Rad), and its purities verified by SDS-PAGE analysis.

Fluorescence Anisotropy

Fluorescence anisotropy measurements were conducted on a Horiba QuantaMaster-400 spectrofluorometer (Photon Technology International). Fluorescently labeled DNA (50 nM) was diluted into a 50-mM HEPES (pH 7.5) buffer containing 150 mM potassium acetate and different concentrations of recombiN.D.t GRSF1 protein. The excitation wavelength was 550 nm, and the fluorescence anisotropy was recorded at 580 nm. The

instrument G factor was determined prior to anisotropy measurements. The entrance and exit slits were set at 6 nm for excitation, and 7.8 nm for emission. The data were fitted to derive the dissociation constant (K_d) following previously published procedures [34]

2.3 Results

To discover systematically novel G4-interacting proteins and to assess their binding specificities, we employed three G4 DNA probes derived from the G-rich sequences of the human telomere and the promoters of c-*KIT* and c-*MYC* genes, and these sequences were previously characterized by solution-phase NMR studies to adopt well-defined G4 foldings *in vitro*.[35-37] We also obtained the corresponding mutated probes incompetent in G4 folding.[30] The proper folding of the G4-containing probes and the inabilities of the mutant probes in G4 folding were confirmed by circular dichroism (CD) measurements, as described elsewhere.[30] In this vein, the sequences derived from the c-*KIT* and c-*MYC* promoters form parallel G4 folding topology, whereas that from the human telomere exhibits a G4 folding pattern with both parallel and antiparallel strands.[30]

Equal amounts of the heavy- and light-labeled nuclear proteomes, which were obtained from stable isotope labeling by amino acids in cell culture (SILAC),[38] were passed through streptavidin columns immobilized with biotin-conjugated G4 DNA and the corresponding mutated sequence (M4), respectively, which we designate as the reverse experiment (Figure 2.1). In this vein, by culturing cells in a medium in which arginine and lysine are replaced with their stable isotope-labeled counterparts, i.e., [$^{13}C_6$]-L-arginine and [$^{13}C_6$,$^{15}N_2$]-L-lysine, SILAC facilitates the labeling of the proteome with these heavy-

labeled amino acids during protein synthesis. To remove potential experimental bias arising from incomplete SILAC labeling, we also conducted the forward experiment, where the light- and heavy-labeled nuclear proteomes were passed through streptavidin columns immobilized with biotin-conjugated G4 and M4 DNA probes, respectively.[39] In total, we conducted a minimum of 4 independent (2 forward and 2 reverse) SILAC-based interaction screening experiments for each pair of G4/M4 probes (Table 2.1).

After incubation with the nuclear protein lysate, the DNA-bound avidin beads were washed, and the proteins captured on the beads were eluted, combined, trypsin-digested, and subjected to LC-MS/MS analysis, as detailed in the online Supporting Information. By performing this experiment on multiple G4 folding patterns, we could achieve a quantitative comparison about the binding selectivities of candidate proteins toward the three G4 structures.

We were able to identify many proteins exhibiting preferential binding towards G4 probes over the mutated single-stranded DNA probes (Figure 2.2). We employed a stringent criterion for considering a protein to be a G4-binding protein, where the protein needs to be enriched on the G4 over the corresponding M4 probes in both forward and reverse SILAC experiments with an average G4/M4 ratio being greater than 1.5. With this criterion, we identified 41, 19 and 33 proteins that can bind preferentially to the G4 sequences derived from the promoters of c-*KIT* and c-*MYC* genes and the human telomere, respectively, over their mutant counterparts (Figure 2.2). In this context, it is worth noting that the SILAC-based proteomic approach provides a quantitative measure about relative, but not absolute binding affinities of proteins toward G4 over the corresponding M4

probes. The method, therefore, does not offer insights into the relative binding affinities of different G4BPs toward any specific G4 probe employed in this study.

Among these proteins, 11 were previously described to interact directly with G4 DNA structures, including DDX5, MAZ, NPM1, etc., where known DNA G4BPs are highlighted in red in Figure 2.[40-50] Interestingly, we also observed several proteins (e.g., FUS and SRSF1) that were previously characterized as RNA G4BPs.[51, 52] Aside from proteins that bind specifically to all three G4 structures, i.e., SLIRP, YY1, and YY2 (Figure 2 and Table S2), we identified a number of proteins that bind exclusively to one or two of the G4 structures, e.g., NSUN2 (to c-*MYC* and c-*KIT* G4 structures) and GRSF1 (to c-*MYC* G4 structure) (Figures 2.2-2.3). In this vein, it is worth noting that a more comprehensive proteomic datasets are employed in the present study; hence, the SILAC ratios for YY1 and SLIRP proteins differ slightly from our previously published results.[13, 30]

The results from our quantitative proteomics-based interaction screening showed that GRSF1 binds selectively to the G4 structure derived from the c-MYC promoter, but not that from the human telomere or c-KIT promoter (Figures 2.2-2.3). Thus, we next asked whether this protein can bind directly and selectively to the G4 structure derived from the promoter of the c-MYC gene by using fluorescence anisotropy measurements. It turned out that GRSF1 indeed displays strong and selective binding towards the c-MYC G4 probe over the corresponding mutated probe, as manifested by the Kd values of 59 nM and 1.28 μM for the G4 and M4 probes, respectively (Figure 2.4).

Our proteomic data also led to the discovery of a number of proteins that bind more strongly to M4 over the corresponding G4 DNA probes (Figure 2.5). Interestingly, several

of these proteins (BLM, nucleolin, HNRNPA2, and SUB1) were previously characterized as DNA G4BPs, and TARDBP was characterized as a RNA G4-binding protein (Figure 2.5).[23, 24, 53-56] Along this line, BLM was shown to unfold G4 DNA, which may explain their preferential binding toward M4 over G4 DNA probes.[54] While these proteins were shown to bind to G4 DNA, it will be important to examine their relative affinities in binding to G4 DNA vs. mutated single-stranded DNA that cannot fold into G4 structure. It will also be important to assess the functions of these proteins in modulating the biology of G4 DNA.

Our proteomic data revealed that CNBP binds preferentially with M4 over G4 probe derived from the c-*MYC* promoter, whereas it binds more strongly to G4 over M4 probe derived from the c-*KIT* promoter. The exact reason for the different selectivities of this protein in binding with G4 vs. M4 probes derived from these two promoter sequences is not clear, though we reason that the nucleobases not involved in G tetrad formation and G4 folding (i.e., those residing in the loop regions of the G4 structure) may also modulate the differential interactions between CNBP and G4/M4 probes. In this respect, it is worth noting that purified CNBP protein was found to be capable of unfolding G4 structures; where the protein binds more strongly to unfolded than folded G4 DNA derived from the c-*MYC* promoter, but exhibits similar affinities to unfolded and folded G4 DNA derived from the c-*KIT* promoter.[49]

Our proteomic data showed that XRCC5 and XRCC6, a.k.a. Ku70 and Ku86, respectively, display stronger binding to M4 over G4 probes derived from all three sequences. These two proteins form a heterodimer and function in the repair of DNA

double-strand breaks through the non-homologous end-joining pathway.[57] At first glance, this seems to be incongruent with the heterodimer' ability in binding with and sliding along broken ends of DNA at the DSB sites.[57] However, Yuan et al. [58] showed that the Ku complex can also bind directly with single-stranded DNA, albeit at a lower affinity than that with the corresponding double-stranded DNA. In addition, Shao et al.[59] revealed that Ku heterodimer can bind to RNA and assume an important function in 18S ribosomal RNA processing. Thus, the Ku complex's ability in binding with single-stranded nucleic acids may contribute to its preferential interaction with single-stranded M4 DNA over the corresponding folded G4 DNA.

2.4 Conclusion

Although *in vitro* formation of G4 structure has been known for decades, only recently has it come into light the widespread presence of these DNA structures in human cells.[6, 7] Previous studies also suggested the functions of G4 structures in many different biological processes. In this vein, G4 sequence motifs are highly enriched in genomic regions of biological importance, e.g., promoters of genes.[7] Hence, it is important to have a more complete understanding about how G4s are sensed by cellular proteins and how these proteins modulate the biological functions of G4 DNA. Many studies have attempted to address this question using diverse approaches and have led to the discovery of a number of proteins that bind directly and strongly with G4 DNA. High-resolution mass spectrometry-based techniques are particularly well suited to explore the interaction proteomes of G4 DNAs and they were previously employed for the identification of a

diverse set of G4 DNA-binding proteins.[23, 55] Different folding patterns have been described for G4 DNA; nevertheless, many of these prior interactome studies only employed one G4 folding pattern. Given the high structural diversity of G4 folding patterns, we aimed to expand this knowledge by directly comparing the interaction proteomes of three unique G4 folding patterns. Interestingly, we discovered three proteins that can bind specifically and recognize G4 structures derived from all three G-rich sequences (Figure 2.2).

Among the identified G4BPs, we recently validated that SLIRP and YY1 can bind directly to all three G4 structures with low-nM binding affinity *in vitro*, and ChIP-Seq results showed that SLIRP and YY1 can also bind with G4 structures in chromatin.[13, 30] It will be important to assess whether YY2, a closely related protein of YY1, can bind directly with G4 DNA structures and to explore the biological functions of such interactions.

Aside from the generic G4BPs, our method allowed for the discovery of proteins that specifically recognize selected G4 structure(s). We also revealed that purified GRSF1 protein can bind strongly and selectively to G4 structures derived from the G4 sequence contained in the c-*MYC* promoter, but not to that from the human telomere or c-*KIT* promoter (Figure 2.4).

In summary, we identified, from exhaustive SILAC-based quantitative proteomic experiments, many novel putative G4BPs that recognize all G4 folding patterns and or select G4 folding patterns. We further demonstrated that GRSF1 can interact directly and selectively with G4 DNA derived from c-*MYC* promoter region with high affinity *in vitro*. Hence, our study revealed that G4BPs hold the ability to differentiate and selectively bind

only certain G4-folding patterns, which may bear a significant impact in understanding the biological functions of G4 DNA. Moreover, the large number (~ 80) of candidate G4 DNA-binding proteins identified in the present study constitute an important resource for the research community to assess the functions of these proteins in the biology of G4 DNA. Our work also revealed a number of putative 'anti-reader' proteins for G4 DNA, which calls for the assessment of the functions of these proteins in the future.

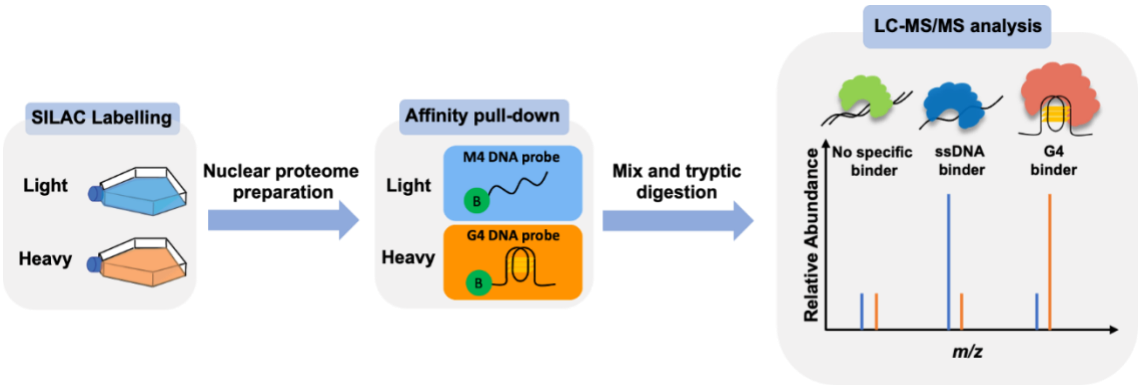

Figure 2.1. The experimental workflow for the identification of novel DNA G4BPs. Shown in the scheme is a reverse SILAC-labeling experiment, where the heavy- and light-labeled nuclear protein lysates are incubated with 5′-biotinylated G4 DNA probe and the corresponding single-stranded DNA probe (M4), respectively. The ‘B’ in green circle denotes 5′-biotin labeling.

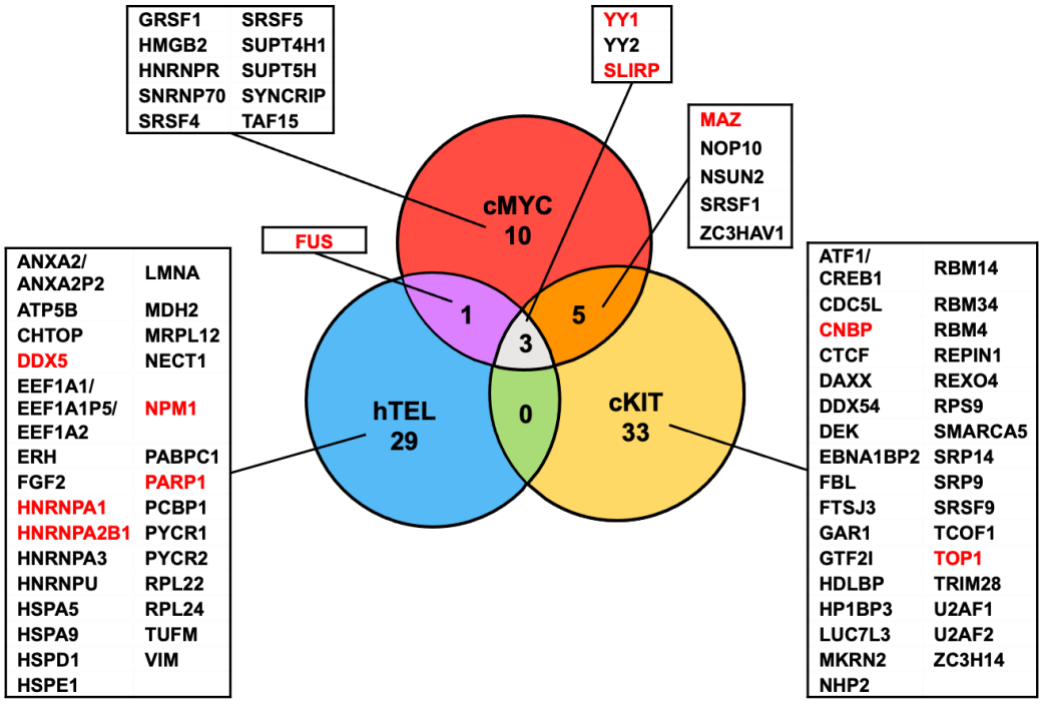

Figure 2.2. A Venn diagram displaying the overlap in interacting proteins among the three G4 folding patterns studied. Candidate G4BPs identified from SILAC-based affinity screening are listed. Among the identified candidate G4BPs, unique peptides were detected for YY1, the peptides detected for YY2 are shared with YY1. Common peptides were detected for ATF1 and CREB1; for ANXA2 and ANXA2P2; and for EEF1A1, EEF1A1P5, and EEF1A2. Proteins highlighted in red are known to bind to DNA G4 structures.

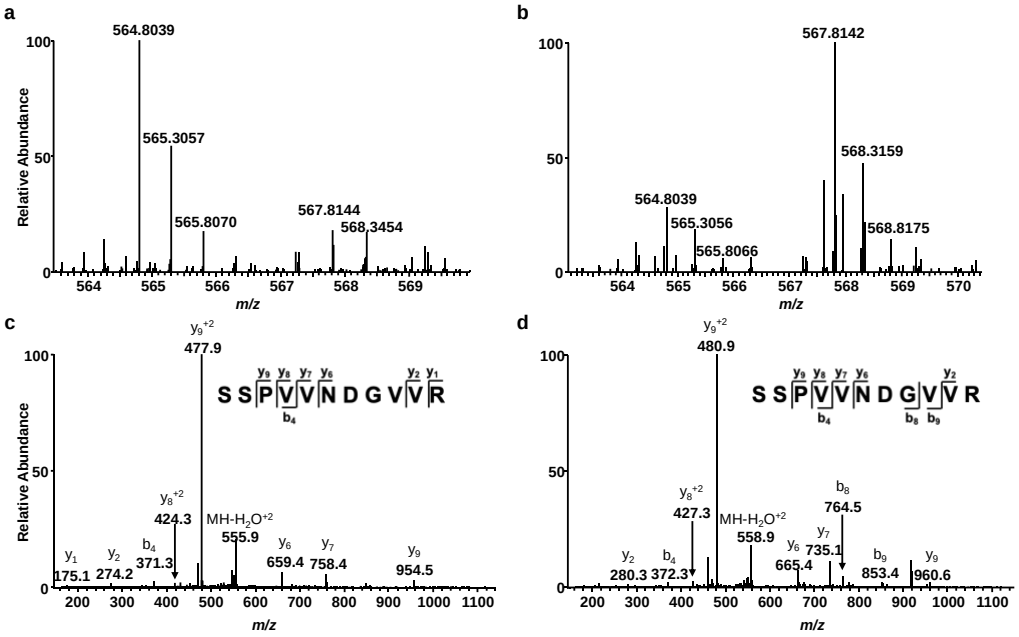

Figure 2.3. GRSF1 binds preferentially to G4 structures derived from the promoter of the c-MYC gene. (a-b) ESI-MS showing the [M + 2H]2+ ions of light and heavy arginine-containing peptide SSPVVNDGVVR with monoisotopic m/z values of ~ 564.8 and 567.8, respectively, obtained from forward (a) and reverse (b) SILAC-based interaction screening experiments. (c-d) MS/MS for the [M+2H]2+ ions of the light (c) and heavy (d) arginine-containing peptide, SSPVVNDGVVR, derived from GRSF1.

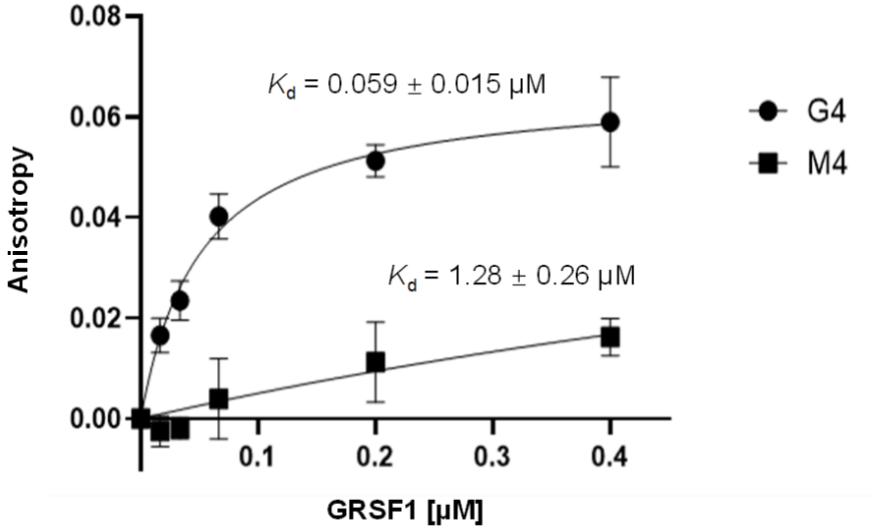

Figure 2.4. Fluorescence anisotropy for measuring the K_d values for the binding of the GRSF1 protein toward G4 structures derived from the promoter of c-*MYC* gene.

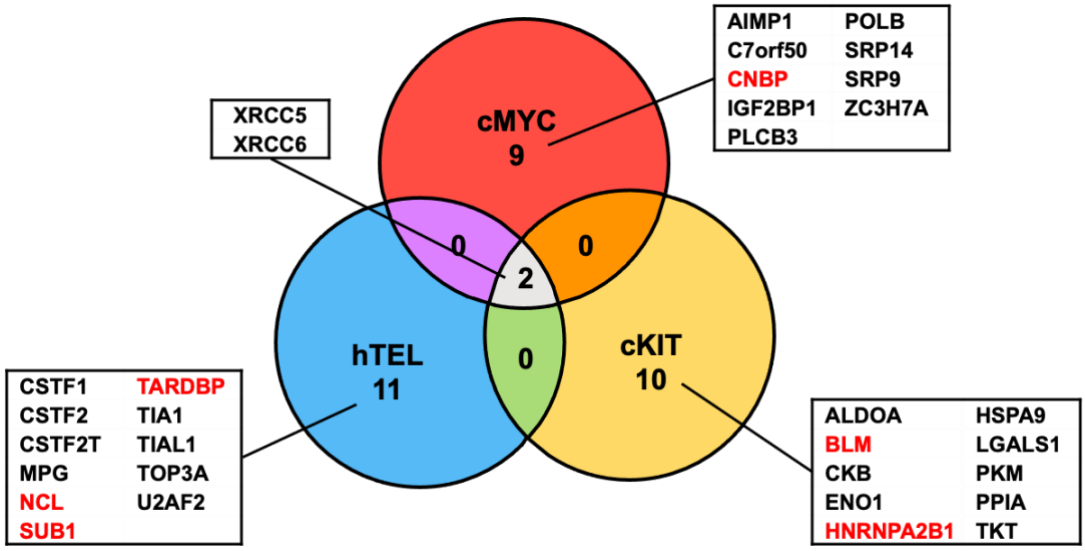

Figure 2.5. A Venn diagram displaying the overlap in proteins that bind more strongly to mutated single-stranded DNA probes (M4) over the corresponding G4 DNA probes derived from human telomere (hTEL) and the promoters of c-MYC and c-KIT genes. Putative anti-reader proteins for G4 DNA are listed. Proteins highlighted in red are known to bind to DNA G4 structures.

Sequence Name	DNA Sequence
c-*KIT* G4	5′-Biotin-T6-AGG GAG GGC GCT GGG AGG AGG G-3′
c-*KIT* ssDNA	5′-Biotin-T6-AGG GAG GGC TCT GTG AGG AGG G-3′
c-*MYC* G4	5′-Biotin-T6-TGA GGG TGG GGA GGG TGG GGA AGG-3′
c-*MYC* ssDNA	5′-Biotin-T6-TGA GGG TGA GGA GTG TGG GGA AGG-3′
HumTel26 G4	5′-Biotin-T6-AAA GGG TTA GGG TTA GGG TTA GGG AA-3′
HumTel26 ssDNA	5′-Biotin-T6-AAA GGG TTA GTG TTA GTG TTA GGG AA-3′

Table 2.1. The DNA sequences employed for the affinity purification pull-down of cellular proteins that can bind to G4 DNA. The differences in sequences between the G4 and the corresponding single stranded DNA are underlined.

Gene names	c-*KIT* ratio [G4/M4]		c-*MYC* ratio [G4/M4]		hTEL ratio [G4/M4]	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
SLIRP	1.60	0.85	2.67	0.81	3.47	0.98
YY1;YY2	1.95	0.19	2.37	1.26	2.28	0.31
MAZ	2.13	1.65	2.03	0.29	N.D.	N.D.
NOP10	1.71	0.72	2.13	0.43	N.D.	N.D.
NSUN2	2.51	1.31	2.01	0.55	N.D.	N.D.
SRSF1	2.30	1.21	1.63	0.81	N.D.	N.D.
ZC3HAV1	5.49	3.36	2.48	0.59	N.D.	N.D.
HNRNPU	N.D.	N.D.	1.16	0.27	2.94	1.42
FUS	1.26	1.00	2.10	0.82	1.60	0.81
ATF1;CREB1	1.78	0.34	0.79	0.42	N.D.	N.D.
CDC5L	1.60	0.46	N.D.	N.D.	N.D.	N.D.
CNBP	1.82	0.98	0.45	0.05	1.15	0.47
CTCF	1.96	1.08	1.25	0.69	N.D.	N.D.
DAXX	5.88	5.20	0.97	0.11	N.D.	N.D.
DDX54	2.13	0.93	N.D.	N.D.	N.D.	N.D.
DEK	1.74	0.86	N.D.	N.D.	N.D.	N.D.
EBNA1BP2	1.70	1.02	0.99	0.14	N.D.	N.D.
FBL	1.69	0.93	1.19	0.51	N.D.	N.D.
FTSJ3	2.01	0.10	0.93	0.40	N.D.	N.D.
GAR1	1.63	0.83	N.D.	N.D.	N.D.	N.D.
GTF2I	3.76	1.89	N.D.	N.D.	N.D.	N.D.
HDLBP	2.99	1.29	N.D.	N.D.	N.D.	N.D.
HP1BP3	2.61	1.36	N.D.	N.D.	N.D.	N.D.
LUC7L3	1.65	0.78	N.D.	N.D.	N.D.	N.D.
MKRN2	5.77	2.69	N.D.	N.D.	N.D.	N.D.
NHP2	1.64	0.90	1.41	0.72	N.D.	N.D.
RBM14	6.83	4.12	1.11	0.44	1.37	0.55
RBM34	2.45	1.06	N.D.	N.D.	N.D.	N.D.
RBM4	4.87	2.75	1.18	0.47	N.D.	N.D.
REPIN1	1.68	0.81	N.D.	N.D.	N.D.	N.D.
REXO4	1.88	0.79	1.22	0.63	N.D.	N.D.
RPS9	1.74	0.87	N.D.	N.D.	N.D.	N.D.
SMARCA5	2.47	0.93	N.D.	N.D.	N.D.	N.D.

SRP14	1.74	0.74	0.52	0.06	N.D.	N.D.
SRP9	1.81	0.93	0.47	0.04	N.D.	N.D.
SRSF9	1.90	0.95	N.D.	N.D.	N.D.	N.D.
TCOF1	2.09	1.04	N.D.	N.D.	N.D.	N.D.
TOP1	4.43	2.27	1.04	0.52	1.26	0.50
TRIM28	6.33	6.71	N.D.	N.D.	N.D.	N.D.
U2AF1	2.11	1.19	1.20	0.11	N.D.	N.D.
U2AF2	2.21	1.70	1.04	0.07	0.41	0.27
ZC3H14	3.62	1.85	N.D.	N.D.	N.D.	N.D.
ANXA2;ANXA2P2	0.78	0.60	1.05	0.41	2.73	1.82
ATP5B	N.D.	N.D.	1.18	0.61	2.52	1.46
CHTOP	1.24	0.92	N.D.	N.D.	3.19	1.68
DDX5	N.D.	N.D.	N.D.	N.D.	3.25	1.76
EEF1A1;EEF1A1P5;EEF 1A2	1.22	1.24	0.93	0.24	3.22	1.86
ERH	0.77	0.39	N.D.	N.D.	2.41	0.60
FGF2	1.03	0.34	N.D.	N.D.	2.76	0.46
HNRNPA1	0.63	0.28	0.85	0.15	4.17	3.63
HNRNPA2B1	0.53	0.27	0.96	0.10	4.87	2.69
HNRNPA3	0.80	0.36	0.84	0.19	4.78	4.48
HSPA5	0.77	0.40	0.95	0.15	1.75	1.03
HSPA9	0.62	0.39	1.11	0.57	1.85	0.91
HSPD1	0.80	0.54	1.28	0.88	2.41	1.58
HSPE1	0.76	0.48	1.16	0.56	3.70	2.43
LMNA	0.89	0.45	1.05	0.48	2.28	1.42
MDH2	N.D.	N.D.	N.D.	N.D.	2.53	1.49
MRPL12	N.D.	N.D.	N.D.	N.D.	2.70	0.38
NECT1	N.D.	N.D.	N.D.	N.D.	1.79	0.88
NPM1	1.03	0.68	1.16	0.19	3.06	1.80
PABPC1	1.31	0.65	1.01	0.37	1.59	0.78
PARP1	1.33	0.47	1.13	0.17	4.14	2.49
PCBP1	0.81	0.39	N.D.	N.D.	2.18	1.10
PYCR1	N.D.	N.D.	N.D.	N.D.	12.92	11.67
PYCR2	N.D.	N.D.	N.D.	N.D.	15.04	11.41
RPL22	1.25	0.81	1.24	0.48	1.93	0.95
RPL24	0.96	0.55	0.93	0.10	2.18	1.30

TUFM	0.68	0.37	0.95	0.51	1.59	0.43
VIM	1.27	1.20	1.47	1.15	3.00	1.69
GRSF1	1.42	0.92	3.24	0.34	N.D.	N.D.
HMGB2	0.83	0.34	1.88	0.20	N.D.	N.D.
HNRNPR	1.44	1.19	3.02	1.11	N.D.	N.D.
SNRNP70	N.D.	N.D.	1.58	0.53	N.D.	N.D.
SRSF4	1.00	0.48	2.49	1.16	N.D.	N.D.
SRSF5	1.23	0.84	1.81	0.90	N.D.	N.D.
SUPT4H1	1.17	0.72	2.45	0.27	N.D.	N.D.
SUPT5H	1.26	0.66	2.21	0.92	N.D.	N.D.
SYNCRIP	N.D.	N.D.	3.90	1.95	N.D.	N.D.
TAF15	1.19	1.13	2.33	1.18	N.D.	N.D.

Table 2.2. A complete list of G4BPs identified from the SILAC-based interaction screening and their G4/M4 ratios. ‘N.D.’, not detected.

References

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

(2) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **2006**, *34* (19), 5402-5415.

(3) Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **2005**, *33* (9), 2908-2916.

(4) Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **2005**, *33* (9), 2901-2907.

(5) Bedrat, A.; Lacroix, L.; Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* **2016**, *44* (4), 1746-1759.

(6) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5* (3), 182-186.

(7) Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet* **2016**, *48* (10), 1267-1272.

(8) Spiegel, J.; Cuesta, S. M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **2021**, *22* (1), 117.

(9) Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol* **2014**, *10* (4), 313-318.

(10) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **2012**, *13* (11), 770-780.

(11) Ribeyre, C.; Lopes, J.; Boulé, J. B.; Piazza, A.; Guédin, A.; Zakian, V. A.; Mergny, J. L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet* **2009**, *5* (5), e1000475.

(12) Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* **2006**, *34* (9), 2536-2549.

(13) Li, L.; Williams, P.; Ren, W.; Wang, M. Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol.* **2021**, *17* (2), 161-168.

(14) Li, L.; Williams, P.; Gao, Z.; Wang, Y. VEZF1-guanine quadruplex DNA interaction regulates alternative polyadenylation and detyrosinase activity of VASH1. *Nucleic Acids Res.* **2020**, *48* (21), 11994-12003.

(15) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A* **2002**, *99* (18), 11593-11598.

(16) Bejugam, M.; Sewitz, S.; Shirude, P. S.; Rodriguez, R.; Shahid, R.; Balasubramanian, S. Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. *J Am Chem Soc* **2007**, *129* (43), 12926-12927.

(17) Bejugam, M.; Gunaratnam, M.; Müller, S.; Sanders, D. A.; Sewitz, S.; Fletcher, J. A.; Neidle, S.; Balasubramanian, S. Targeting the c-Kit Promoter G-quadruplexes with 6-Substituted Indenoisoquinolines. *ACS Med Chem Lett* **2010**, *1* (7), 306-310.

(18) Sundquist, W. I.; Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989**, *342* (6251), 825-829.

(19) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res* **2007**, *35* (7), 2440-2450.

(20) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A* **2001**, *98* (15), 8572-8577.

(21) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol* **2005**, *12* (10), 847-854.

(22) Mendoza, O.; Bourdoncle, A.; Boulé, J. B.; Brosh, R. M.; Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res* **2016**, *44* (5), 1989-2006.

(23) Gonzalez, V.; Guo, K.; Hurley, L.; Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **2009**, *284* (35), 23622-23635.

(24) Gao, J.; Zybailov, B. L.; Byrd, A. K.; Griffin, W. C.; Chib, S.; Mackintosh, S. G.; Tackett, A. J.; Raney, K. D. Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA. *Chem. Commun.* **2015**, *51* (33), 7242-7244.

(25) Kanoh, Y.; Matsumoto, S.; Fukatsu, R.; Kakusho, N.; Kono, N.; Renard-Guillet, C.; Masuda, K.; Iida, K.; Nagasawa, K.; Shirahige, K.; et al. Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat Struct Mol Biol* **2015**, *22* (11), 889-897.

(26) Soldatenkov, V. A.; Vetcher, A. A.; Duka, T.; Ladame, S. First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem Biol* **2008**, *3* (4), 214-219.

(27) Pagano, B.; Margarucci, L.; Zizza, P.; Amato, J.; Iaccarino, N.; Cassiano, C.; Salvati, E.; Novellino, E.; Biroccio, A.; Casapullo, A.; et al. Identification of novel interactors of human telomeric G-quadruplex DNA. *Chem Commun (Camb)* **2015**, *51* (14), 2964-2967.

(28) Zhang, T.; Zhang, H.; Wang, Y.; McGown, L. B. Capture and identification of proteins that bind to a GGA-rich sequence from the ERBB2 gene promoter region. *Anal Bioanal Chem* **2012**, *404* (6-7), 1867-1876.

(29) Sanders, C. M. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem J* **2010**, *430* (1), 119-128.

(30) Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G quadruplex-binding protein. *J. Am. Chem. Soc.* **2017**, *139* (36), 12426-12429.

(31) Guo, L.; Xiao, Y.; Wang, Y. Monomethylarsonous acid inhibited endogenous cholesterol biosynthesis in human skin fibroblasts. *Toxicol Appl Pharmacol* **2014**, *277* (1), 21-29.

(32) Bing, T.; Shangguan, D.; Wang, Y. Facile Discovery of Cell-Surface Protein Targets of Cancer Cell Aptamers. *Mol Cell Proteomics* **2015**, *14* (10), 2692-2700.

(33) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26* (12), 1367-1372.

(34) Rossi, A. M.; Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **2011**, *6* (3), 365-387.

(35) Wang, Y.; Patel, D. J. Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure* **1993**, *1* (4), 263-282.

(36) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res* **2006**, *34* (9), 2723-2735.

(37) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **2007**, *129* (14), 4386-4392.

(38) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376-386.

(39) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002**, *1* (5), 376-386.

(40) Arimondo, P. B.; Riou, J. F.; Mergny, J. L.; Tazi, J.; Sun, J. S.; Garestier, T.; Hélène, C. Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res* **2000**, *28* (24), 4832-4838.

(41) Marchand, C.; Pourquier, P.; Laco, G. S.; Jing, N.; Pommier, Y. Interaction of human nuclear topoisomerase I with guanosine quartet-forming and guanosine-rich single-stranded DNA and RNA oligonucleotides. *J Biol Chem* **2002**, *277* (11), 8906-8911.

(42) Takahama, K.; Takada, A.; Tada, S.; Shimizu, M.; Sayama, K.; Kurokawa, R.; Oyoshi, T. Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem Biol* **2013**, *20* (3), 341-350.

(43) Wu, G.; Xing, Z.; Tran, E. J.; Yang, D. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc Natl Acad Sci U S A* **2019**, *116* (41), 20453-20461.

(44) Xodo, L.; Paramasivam, M.; Membrino, A.; Cogoi, S. Protein hnRNPA1 binds to a critical G-rich element of KRAS and unwinds G-quadruplex structures: implications in transcription. *Nucleic Acids Symp Ser (Oxf)* **2008**, (52), 159-160.

(45) Wang, F.; Tang, M. L.; Zeng, Z. X.; Wu, R. Y.; Xue, Y.; Hao, Y. H.; Pang, D. W.; Zhao, Y.; Tan, Z. Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc Natl Acad Sci U S A* **2012**, *109* (50), 20413-20418.

(46) Federici, L.; Arcovito, A.; Scaglione, G. L.; Scaloni, F.; Lo Sterzo, C.; Di Matteo, A.; Falini, B.; Giardina, B.; Brunori, M. Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *J Biol Chem* **2010**, *285* (48), 37138-37149.

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

(48) Cogoi, S.; Shchekotikhin, A. E.; Xodo, L. E. HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic Acids Res* **2014**, *42* (13), 8379-8388.

(49) David, A. P.; Pipier, A.; Pascutti, F.; Binolfi, A.; Weiner, A. M. J.; Challier, E.; Heckel, S.; Calsou, P.; Gomez, D.; Calcaterra, N. B.; et al. CNBP controls transcription by unfolding DNA G-quadruplex structures. *Nucleic Acids Res.* **2019**, *47* (15), 7901-7913.

(50) Fan, J. H.; Bochkareva, E.; Bochkarev, A.; Gray, D. M. Circular dichroism spectra and electrophoretic mobility shift assays show that human replication protein A binds and melts intramolecular G-quadruplex structures. *Biochemistry* **2009**, *48* (5), 1099-1111.

(51) Yagi, R.; Miyazaki, T.; Oyoshi, T. G-quadruplex binding ability of TLS/FUS depends on the beta-spiral structure of the RGG domain. *Nucleic Acids Res* **2018**, *46* (12), 5894-5901.

(52) von Hacht, A.; Seifert, O.; Menger, M.; Schutze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res.* **2014**, *42* (10), 6630-6644.

(53) Lopez, C. R.; Singh, S.; Hambarde, S.; Griffin, W. C.; Gao, J.; Chib, S.; Yu, Y.; Ira, G.; Raney, K. D.; Kim, N. Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA. *Nucleic Acids Res.* **2017**, *45* (10), 5850-5862.

(54) Mohaghegh, P.; Karow, J. K.; Brosh, R. M., Jr.; Bohr, V. A.; Hickson, I. D. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res* **2001**, *29* (13), 2843-2849.

(55) Wang, F.; Tang, M. L.; Zeng, Z. X.; Wu, R. Y.; Xue, Y.; Hao, Y. H.; Pang, D. W.; Zhao, Y.; Tan, Z. Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (50), 20413-20418.

(56) Ishiguro, A.; Kimura, N.; Watanabe, Y.; Watanabe, S.; Ishihama, A. TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells* **2016**, *21* (5), 466-481.

(57) Grundy, G. J.; Moulding, H. A.; Caldecott, K. W.; Rulten, S. L. One ring to bring them all--the role of Ku in mammalian non-homologous end joining. *DNA Repair* **2014**, *17*, 30-38.

(58) Yuan, Y.; Britton, S.; Delteil, C.; Coates, J.; Jackson, S. P.; Barboule, N.; Frit, P.; Calsou, P. Single-stranded DNA oligomers stimulate error-prone alternative repair of DNA double-strand breaks through hijacking Ku protein. *Nucleic Acids Res.* **2015**, *43* (21), 10264-10276.

(59) Shao, Z.; Flynn, R. A.; Crowe, J. L.; Zhu, Y.; Liang, J.; Jiang, W.; Aryan, F.; Aoude, P.; Bertozzi, C. R.; Estes, V. M.; et al. DNA-PKcs has KU-dependent function in rRNA processing and haematopoiesis. *Nature* **2020**, *579* (7798), 291-296.

Chapter 3: Identification of HELLS as a Novel G-quadruplex Helicase

3.1 Introduction

Guanine quadruplexes (G4s) are non-canonical DNA structures characterized by the arrangement of four strands of nucleic acids, where guanine bases are interconnected through Hoogsteen base pairing and stabilized by monovalent metal ions[1]. These unique secondary structures have garnered considerable attention due to their unique features and potential functional roles in various biological processes. G4 structures exhibit a diverse range of regulatory mechanisms, including their involvement in DNA replication, gene transcription, telomere maintenance, and DNA repair. The interaction of G4s with specific proteins, such as BLM[2], WRN[3] and FANCJ[4], has been implicated in genetic diseases. Moreover, abnormal expression of several G4 binding-proteins (G4BPs), including FUS[5] and hnRNPA1[6], has been observed in neurodegenerative diseases, underscoring the importance of understanding the functional properties of G4s and their interactions with G4BPs. Such insights hold promise for elucidating the underlying cellular processes and exploring novel avenues for therapeutic interventions.

In previous study on identifying G4BPs, we employed affinity pull-down assays with biotin-conjugated DNA probes capable or incapable of folding into G4 structures[7] coupled with LC-MS/MS analysis. To quantify the differential binding preferences between these

probes, we utilized robust stable isotope labeling by amino acids in cell culture (SILAC). With this method, we identified successfully a total of 78 putative binding proteins, including well-known G4BPs, e.g., hnRNPA1, PARP1, Pif1, and WRN[6, 8-10]. Subsequently, we conducted further investigations to validate the abilities of some of the newly discovered proteins in binding directly with G4 DNA, including SLIRP, YY1, GRSF1, and VEZF1[7, 11-13]. This method provides a valuable tool for uncovering G4BPs and results in a comprehensive list of candidate G4BPs. The approach, nevertheless, has some inherent limitations. In particular, those proteins exhibiting weak and transient interactions with G4 structures may not be effectively captured for subsequent LC-MS/MS detection. Additionally, this method may also result in the discovery of those proteins that interact indirectly with G4 DNA through protein-protein interactions, which could result in false-positive identification of G4BPs.

Photoaffinity labeling-based methods have been applied to uncover the regulatory mechanism of gene transcription, DNA replication, and DNA repair by identification and characterizations of the protein–protein, protein–nucleic acid, and protein–small molecule interactions that control the underlying processes[14-19]. By leveraging photoirradiation, these techniques employ photo-crosslinking functional groups, such as benzophenone (BP)[20], aryl azide (AA)[21], and diazirine (DA)[22], to generate highly reactive species capable of forming covalent bonds with neighboring molecules upon activation. These covalent modifications provide direct insights into the molecular interactions and dynamics underlying the studied processes. The investigation of G4BPs using photo-crosslinking probes has been reported, employing G4-targeted small molecules. Notably, the G4-

LIMCAP method has been developed[23], utilizing a probe consisting of a derivative of pyridostatin, i.e., (*N*,*N'*-bis(2-quinolinyl)pyridine-2,6-dicarboxamide). This probe incorporates a diazirine group capable of photo-crosslinking and an alkyne group for subsequent copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reactions. Similarly, Zang et al. [24] reported the co-binding-mediated protein profiling (CMPP) method, which utilizes pyridostatin (PDS) conjugated with a linker, diazirine, and alkyne group. Both probes enable photo-crosslinking to putative G4BPs in cells through proximity-based interactions. Following crosslinking and nuclear protein extraction, a biotin group can be attached to the probe using click chemistry via the alkyne function for subsequent pulldown. However, it is important to note that both methods, employing PDS-based probes, do not discriminate between DNA G4BPs and RNA G4BPs, as PDS exhibits comparable affinity for both DNA and RNA G4 structures[25]. In addition, binding of PDS with G4 DNA may prohibit the binding of some G-binding proteins, as documented previously for YY1[12] and VEZF1[13]. Moreover, upon photoactivation, the diazirine undergoes conversion to a reactive intermediate, i.e., carbene, which can either form covalent bonds with nearby biomolecules or be quenched by the surrounding aqueous environment[26]. This introduces challenges of non-specific interactions with proteins and reduced crosslinking efficiency.

Herein, inspired by recently published work of Guo et al.[27, 28] in PANAC photoclick reaction to capture proteins, we developed a novel approach utilizing click-chemistry-based crosslinking for quantitative proteomics-based interaction screening of G4BPs. This new approach offers several advantages over traditional affinity pull-down method,

including increased sensitivity and specificity, as well as the ability to capture low affinity and transient DNA-protein interactions. With this new method, we were able to identify a number of novel G4BPs including HELLS. We also found that HELLS can unfold DNA guanine quadruplex structures, thereby regulating the genome stability and cellular homeostasis.

3.2 Materials and Methods

Cell culture

HeLa and U2OS cells were maintained in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher) and 1% penicillin-streptomycin solution (PS, GE Healthcare). The cells were incubated at 37°C in a 5% CO_2 environment.

Syntheses of o-nitrobenzyl alcohol (o-NBA) G4 DNA probes

The G4-forming sequences derived from the human telomere, with a biotin label on the 5′ termini and an aliphatic amine-derived thymidine and the corresponding mutated sequences incapable of folding into G4 structures were purchased from Integrated DNA Technologies (IDT). The conjugation of oNBA (4-hydroxymethyl-3-nitrobenzoic acid) to the oligonucleotides was carried out using an hexafluorophosphate azabenzotriazole tetramethyl uronium (HAUT)-based coupling, following the procedures outlined by Guo et al.[28]. In brief, the oligonucleotide sample was mixed with an equal volume of 500 mM sodium borate buffer (pH 9.4). A mixture containing equal amounts of HAUT, 4-hydroxymethyl-3-nitrobenzoic acid, and *N*,*N*-Diisopropylethylamine (DIPEA) (200 mM

in dimethylacetamide) was prepared and added to the oligonucleotide sample. The reaction mixture was incubated on a thermomixer at room temperature overnight. To the mixture were subsequently added a solvent mixture of 5.0-M aqueous solution of NaCl and absolute ethanol (1:2.5, v/v), and the mixture was incubated at -20°C for 1 hr to precipitate the oligodeoxynucleotide. Lastly, the sample was buffer exchanged into ddH_2O using ultrafiltration (3 kDa MWCO centrifugal filter). The product was further subjected to LC-MS and MS/MS analysis on an LTQ linear ion-trap mass spectrometer (Thermo fisher).

The oNBA-G4 and TAMRA-labeled G4 probes were annealed in a buffer containing 10 mM Tris-HCl (pH 7.5), 10 mM KCl, and 0.1 mM EDTA by heating to 95°C for 5 min, followed by cooling down to room temperature slowly over 6 h.

Photoaffinity labeling

The nuclear proteome was isolated from HeLa cells labeled with heavy and light isotopes using the NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific), following the manufacturer's instructions. Protein concentrations were determined using the Bradford Quick Start Protein Assay kit (Bio-Rad), and the nuclear lysate was stored at -80°C until further use.

For the pull-down experiments, 0.5 μM o-NBA-conjugated probes and 500 μg of nuclear lysate in a binding buffer (20 mM HEPEs pH 7.5, 100 mM KCl, and 1 mM $MgCl_2$) was incubated at 4°C with rotation for 30 min. The mixture was then transferred to a 24-well plate and exposed to 15 W 366 nm UV light for 10 min at a distance of 5 cm on ice. Subsequently, the samples were collected and incubated with pre-washed streptavidin beads in the binding buffer at 4°C with rotation for 2 hr. After the incubation, the unbound

proteins were removed by washing three times with a washing buffer consisting of 20 mM HEPES (pH 7.5), 500 mM NaCl, and 0.1% SDS. The beads were subsequently combined, and the bound proteins were eluted by using 2×SDS-PAGE loading buffer (Bio-Rad) and boiling the mixture for 5 min. The resulting mixture was then subjected to in-gel digestion, following a previously published protocol, and LC-MS/MS analysis.

Mass Spectrometry

On-line LC-MS/MS analysis of the peptide samples was performed on an Orbitrap Fusion Lumos tribrid mass spectrometer (Thermo Fisher Scientific) equipped with a Flex nanoelectrospray ion source (Thermo) and coupled with an EASY-nLC 1000 system (Thermo, San Jose, CA, USA). A high-field asymmetric-waveform ion mobility spectrometry (FAIMS) was incorporated into the mass spectrometer configuration. The compensation voltages (CV) were precisely adjusted to -40, -60, and -80 V. The carrier gas flow rate was set at 4.2 L/min. Cycle time of each compensation voltage is set to 1 second. The HPLC separation was performed using a 5 µm trapping column followed by a 3 µm analytical column, both packed in-house with ReproSil-Pur C18-AQ resin (Dr. Maisch HPLC GmbH, Germany). The peptides were separated using a 160-min linear gradient of 8-34% acetonitrile in 0.1% formic acid at a flow rate of 300 nL/min and electrosprayed (spray voltage 2 kV) into the mass spectrometer operated in the positive-ion mode. Full-scan MS (*m/z* 300-1200) were acquired at a resolution of 60,000, followed by MS/MS acquisition in linear ion trap, where the scan rate set as rapid. Fragmentation was conducted with higher-energy collisional dissociation (HCD) at a fixed collisional energy of 30%.

Data Analysis

Protein identification and quantification was conducted by running raw data MaxQuant (Version 2.1.2.0) [29, 30]. Raw MS data were searched against the UniProt human proteome database (Proteome ID: UP000005640_9606). For quantification at MS level, the multiplicity was set to 2, and Lys8 and Arg6 were selected as heavy amino acids. Cysteine carboamidomethylation was set as a fixed modification, along with N-terminal protein acetylation and methionine oxidation were set as variable modifications. The maximum number of missed cleavages for trypsin was set to two per peptide, and peptide length range was 7-25aa. The tolerances in mass accuracy for MS and MS/MS were 20 ppm and 0.5 Da, respectively. The match between runs option was enabled with the alignment window being 3 min.

Plasmid construction and protein purification

The coding sequence of human HELLS was amplified from the pBac-HELLS plasmid, which was generously provided by Dr. Kathrin Muegge at the National Cancer Institute (NCI). Subsequently, the coding sequence was inserted into the pRK7 vector, incorporating three consecutive Flag epitope tags at the carboxyl terminus. HEK293T cells were transfected with these plasmids using TransIT-2020 transfection reagent (Mirus Bio, Madison, WI) and harvested 36 hr later for protein extraction. CelLytic M Cell Lysis Reagent (C2978, Sigma) supplemented with 1× protease inhibitor cocktail (P8340, Sigma) was used for protein extraction. Protein purification was conducted according to the manufacturer's instructions. Briefly, the cell lysate was incubated with Anti-Flag M2 magnetic beads (M8823, Sigma) at 4°C for one hour, followed by washing with TBS (50

mM Tris HCl (pH 7.4), 150 mM NaCl), and elution with 3× Flag peptide (NC0792928, Thermo Fisher Scientific).

Fluorescence anisotropy

Fluorescence anisotropy measurements were conducted on a microplate reader (Synergy H1, Agilent). TAMRA-labeled DNA (2 nM final concentration) was incubated with different concentrations of recombinant HELLS in a buffer containing 20 mM HEPES (pH 7.5), 50 mM KCl, 2.5mM $MgCl_2$ and 0.5 mM DTT. Anisotropy was recorded using the red filter set with the excitation and emission wavelength at 530 nm and 590 nm respectively.

Western blot

For detection of HELLS knockdown efficiency in U2OS cells, cells were lysed with CelLytic M cell lysis reagent (Sigma) supplemented with 1% protease inhibitor cocktail. Protein concentration was measured by Quick Start Bradford Protein Assay (Bio-Rad) and denatured at 95 °C for 5-min in SDS-PAGE loading buffer (Bio-Rad). The lysates were separated using SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was blocked with 5% non-fat milk in PBS-T for 1 hr and then incubated with the corresponding primary antibodies, including HELLS Polyclonal antibody (Proteintech, 11955-1-AP, 1:500 dilution), anti-tubulin (Santa Cruz, SC-32293, 1:5000). The secondary antibodies were donkey anti-rabbit secondary antibody (Sigma, A0545, 1:5000) or anti-mouse secondary antibody (Santa Cruz, m-IgGκ BP-HRP, 1:5000). The Western blot signal was detected using ECL Western blotting detection reagent (Amersham, Little Chalfont, UK) and imaged in a LI-COR Odyssey system.

Immunofluorescence

The immunofluorescence microscopy experiments were conducted according to previously established protocols[31]. In brief, control and knockdown cells were fixed in methanol and acetic acid mixture (3:1, v/v) at room temperature for 10 minutes. Subsequently, the cells were permeabilized performed using 0.1% triton-X100 in 1x PBS for 15 minutes on ice, followed by treatment with RNase A (Thermo Fisher Scientific) at 37ºC for 30 min. After blocking with 2% BSA for 1 hour at room temperature, the immunofluorescence microscopy experiments were carried out using standard methods. Antibodies against BG4 (MABE917, Sigma-Aldrich), anti-FLAG (14793S, Cell Signaling Technology), and anti-rabbit Alexa 594-conjugated secondary antibody (A11037, Invitrogen) were used. Nuclei were stained using DAPI (D9542, Sigma-Aldrich). Finally, the coverslips were mounted with ProLong Diamond Antifade Mountant (Invitrogen). Images were captured using an LSM880 confocal laser scanning microscope (Carl Zeiss) equipped with a 100× objective, and subsequent image analysis was performed using ZEN software. The number of foci per nucleus was quantified using the Find Maxima feature in ImageJ software.

G4-ChIP qPCR

G4-ChIP was conducted by following previously described procedures with slight modifications, using the custom-purified BG4 antibody[31]. Briefly, the chromatin samples were sonicated and diluted in a blocking buffer containing 25 mM HEPES (pH 7.5), 10.5 mM NaCl, 110 mM KCl, 1 mM $MgCl_2$, and 1% BSA, and treated with RNase A. Subsequently, the chromatin samples were incubated with 500 ng of the BG4 antibody

under rotation at 1400 rpm for 1.5 hr at 16 °C. Pre-blocked Anti-Flag M2 magnetic beads (Sigma, M8823) were added to the mixture, which was then incubated under the same conditions for 1 hr. After washing the beads with ice-cold washing buffer (10 mM Tris, pH 7.4, 100 mM KCl, 0.1% Tween 20) for 6 cycles, the DNA was eluted with TE buffer containing proteinase K while rotating at 1400 rpm for 6 hr at 65 °C. The eluted DNA was purified using the DNA Clean and Concentrator-5 kit (Zymo). For quantitative PCR, the immunoprecipitated samples and the input control were used to determine the enrichment of G4 structures. qPCR was performed using Luna Universal qPCR Master Mix (NEB) on a CFX96 touch real-time PCR detection system (Bio-Rad). The primer sequences used are listed in the Supporting Information, Table S1.

3.3 Results

Identification of G4BPs by using photo-crosslinking

The experimental procedures employed herein are similar to our previously reported method[7, 11], albeit with some modifications on probes design and washing conditions (Figure 3.1A). To facilitate the crosslinking, a o-NBA (ortho-nitrobenzylamine) group is conjugated to a thymine of a biotin-labeled G4 DNA probes derived from the human telomere G4 through amine coupling (Figure 3.1B). A T-linker is placed between the o-NBA group and provide flexibility for primary amines and o-nitrobenzyl alcohols cyclization (PANAC) reaction through reaction with primary amines in proximal proteins[27] (Figure 3.1C). We also synthesized the corresponding mutated sequences (M4) incapable of folding into G4 structures, thereby mitigating the effects of non-specific binding arising

from photocrosslinking reactions. Recognizing that G4BPs can interact with G4 structures in different conformations, we designed two pairs of probes: 5'T G4 probes with the o-NBA group placed on the 5' end thymidine of the G4 structure and LoopT G4 probes with the o-NBA group being located on the thymidine in the loop region of the G4 structure (Figure 3.1D). We synthesized the probes, purified them using HPLC, and characterized them by LC-MS/MS (Figure 3.2).

Afterwards, we incubated light and heavy SILAC-labeled nuclear lysates separately with G4 and M4 probes, and exposed the resulting mixtures to 365-nm UVA light to induce crosslinking to the mixtures were subsequently added streptavidin beads, which, after incubation, were washed rigorously with a buffer containing 500 mM NaCl and SDS to eliminate unbound components. The beads containing the heavy- and light-labeled proteins were subsequently combined, and the proteins were eluted from the beads, digested with trypsin, and the ensuing peptides subjected to LC-MS/MS analysis. To define putative G4-binding proteins in this study, we imposed a G4/M4 cutoff ratio of 1.5. This comprehensive approach allowed for a thorough analysis of G4BPs, considering their potential interactions with different regions of the G4 structure.

Such analysis led to the identification of 296 and 252 candidate G4BPs with the 5'T and loopT probe pairs, respectively (Figure 3.3). Proteins detected in both sets of pulldown experiments exhibited a strong correlation, indicating the robustness and reliability of the method in capturing G4BPs (Figure 3.4). Amongst the identified proteins, 99 were found to be common to both probe pairs, which included previously reported G4BPs such as, e.g., hnRNPA1[6], hnRNPA2[32], and SERBP1[23](Figure 3.3). Gene ontology analysis was

conducted on the putative G4BPs, revealing a predominant localization of these proteins within the nuclear lumen, chromosome and nuclear speck. Furthermore, their functional roles were found to encompass ribosome biogenesis, ncRNA metabolic processes, chromatin organization, as well as catalytic activities involving nucleic acids (Figure 3.3). We further identified HELLS as a candidate G4-binding protein, with a SILAC protein ratio (G4/M4) of 3.30 in 5'T pulldown and 2.00 in LoopT pulldown. Representative electrospray ionization-mass spectrometry (ESI-MS) and MS/MS data for a tryptic peptide of HELLS are illustrated in Figure 3.4.

HELLS binds directly with DNA G4s

HELLS, also known as LSH (helicase, lymphoid specific), belongs to the SNF2 family of ATP-dependent chromatin remodeling proteins. A mutation in the HELLS will result in Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF), an inherited disease with early mortality[33]. Dysregulation of HELLS expression or function has been associated with various diseases, including cancer and immunodeficiency diseaes[34, 35]. Therefore, HELLS has emerged as an important target for therapeutic intervention in cancer treatment.

To gain insights into the interaction between HELLS and DNA G4 structures, we purified recombinant Flag-tagged HELLS protein from HEK293T cells (Figure 3.5A) and employed fluorescence anisotropy to measure the binding affinities between HELLS and DNA G4s. The results revealed a strong and preferential binding of HELLS towards DNA G4 structures in comparison to single-stranded DNA, with the dissociation constant (K_d) being 2.1±0.6 nM for its finding with G4 DNA probe (Figure 3.5B). The robust and

selective binding of HELLS toward DNA G4s underscores its potential significance in G4-related biological processes.

We also conducted a competitive binding assay to investigate the effect of TMPyP4, a widely used small-molecule G4 ligand[36], on the interaction of HELLS with DNA G4s. TMPyP4 is known for its ability to stabilize G4 structures and has been extensively utilized in both *in vitro* and *in vivo* studies. The competitive binding assay involved the preincubation of HELLS with G4 DNA, followed by the addition of 15 nM TMPyP4. The results, as depicted in Figure 3.5C, demonstrated a reduction in the binding affinity of HELLS towards G4s upon the introduction of TMPyP4. This reduction in binding affinity was evidenced by an increase in apparent K_d to 46 ± 16 nM. This observation further confirms the direct binding of HELLS with G4 DNA, as the weakened interaction in the presence of TMPyP4 validates the specific binding of HELLS to G4 structures.

On the grounds that HELLS has been reported to modulate chromatin and facilitate the binding of transcription factor to chromatin, we next investigated the effect of TMPyP4 treatment on the displacement of HELLS from chromatin in cells. To this end, we treated U2OS cells with TMPyP4, and examined the levels of HELLS in chromatin fraction at 0.5, 6, and 12 hr following the treatment. As illustrated in Figure 3.5D, we observed a notable displacement of HELLS from chromatin following a 0.5-hr treatment with TMPyP4. This finding suggests that the binding of HELLS with chromatin occurs through its recognition of G4 structures in chromatin, suggesting that HELLS may facilitate the binding of transcription factors to chromatin through G4s.

Downregulation of HELLS results in elevated levels of G4s in cells

To investigate the roles of HELLS in modulating G4 structures in cells, we established U2OS cells with stable knockdown of HELLS using short-hairpin RNA (shRNA), where the successful knockdown was confirmed by Western blot analysis (Figure 3.6A). Next, we monitored the levels of G4 structures in cells using immunofluorescence microscopy with BG4 antibody[37] (Figure 3.6B). Our results revealed a > 2-fold increase in the global G4 levels upon HELLS knockdown compared to shControl cells (Figure 3.6C). Moreover, there was a consistent correlation between the knockdown efficiency of HELLS and the number of G4 foci observed, where a higher knockdown efficiency confers a greater number of G4 foci in cells. The observation suggests the ability of HELLS as a helicase in unwinding G4 structures in cells.

HELLS is enriched at G4 regions in cells

The functional significance of HELLS extends beyond its participation in modulation of chromatin architecture and extends to the transcriptional regulatory networks[38, 39]. Notably, HELLS has demonstrated a pivotal role in facilitating the binding of transcription factors to chromatin, with a particular emphasis on promoter regions[40].

To gain a comprehensive understanding of the binding of HELLS to G4 structures within chromatin across the entire genome, we performed bioinformatic analysis using chromatin immunoprecipitation sequencing (ChIP-seq) data obtained from HELLS ChIP-seq in mouse fibroblast cells (MEFs)[40] and G4 ChIP-seq in mouse embryonic stem cells (ESCs)[41], acquired from previously published datasets. Such analysis unveiled a significant co-occupancy of HELLS at G4 structure loci. In particular, approximately 83.6% of the

peaks identified in the HELLS ChIP-seq overlapped with the BG4 ChIP-Seq peaks. Importantly, the majority of the overlapping peaks were found to be located within promoter regions, consistent with the prior reports on the characteristic distribution of both G4 structures and HELLS binding (Figure 3.7).

To further validate these findings in human cells, we conducted G4 ChIP-qPCR experiments to monitor G4s at promoter region of several specific genes (*KRAS*, *ZNF76*, and *HSPD1*) in U2OS cells. The enrichment of G4 in shCtrl cells indicates the consistency of G4 loci in mouse and human cells at those targeted genes. Moreover, the G4 ChIP-qPCR analysis performed on HELLS knockdown cells exhibited an augmented enrichment of G4 compared to control cells, indicating that the downregulation of HELLS resulted in an elevation of G4 structures at these particular gene loci (Figure 3.8). Collectively, these findings support the regulatory role of HELLS in modulating G4 dynamics at specific genomic loci.

3.4 Conclusion

We present here a novel approach for the identification of novel G4BPs by incorporating the photoactivatable o-NBA group into G4 DNA probes derived from the human telomere G4 repeat sequence. Through the use of two pairs of probes targeting distinct regions of the G4 structure, we successfully captured and identified a repertoire of G4BPs from the nuclear lysate. The application of UVA irradiation for rapid crosslinking, followed by stringent washing, ensured efficient capture of G4-associated proteins, especially for those interacting transiently and/or weakly with G4 structures.

Utilizing LC-MS/MS analysis, we identified 99 proteins exhibiting greater enrichment with G4 over M4 in both pairs of probes. Among these proteins, several well-known G4BPs and G4 helicases, such as DDX5[42], DHX36[43], DDX24[24], were identified, underscoring the effectiveness of our approach. Importantly, our study unveiled HELLS, also known as LSH (helicase, lymphoid specific), belongs to the SNF2 family of ATP-dependent chromatin remodeling proteins, as a novel G4 helicase. A mutation in the HELLS will result in Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF), an inherited disease with early mortality[33]. Dysregulation of HELLS expression or function has been associated with various diseases, including cancer and immunodeficiency diseaes[34, 35]. Therefore, HELLS has emerged as an important target for therapeutic intervention in cancer treatment.

On top of that, HELLS has been known to facilitates transcription factor binding to chromatin, particularly at promoters. It interacts with regulatory proteins such as p53, DNMT1, and HP1, regulating DNA methylation, histone modification, and chromatin structure[35, 44, 45]. Notably, our study unveils a significant convergence between HELLS and G4 structures within the chromatin context, particularly at promoter regions. This convergence suggests a collaborative relationship between HELLS and G4 structures, implying their potential collective involvement in modulating chromatin architecture and gene expression, in addition to the protein-protein interactions described earlier. Furthermore, HELLS actively participates in DNA replication by being recruited to the replication fork, where it prevents replication fork stalling and promotes efficient replication processes[38]. HELLS is also involved in homologous recombination-mediated

DNA damage repair[45]. The identification of HELLS as a DNA G4 helicase provides valuable insights into the possibility of G4 structures playing a regulatory role in genome stability and cellular homeostasis.

In conclusion, our photo-crosslinking based pulldown approach offers a complementary and valuable addition to traditional affinity pulldown-based methods, enabling the discovery of novel G4-binding protein that may have been previously overlooked due to transient and weak interaction. The comprehensive list of putative binding proteins identified through our study serves as a valuable resource for future investigations aimed at unraveling the functional roles of G4BPs and expanding our understanding of the biological significance of G4 structures in various cellular processes. Moreover, our findings provide significant insights into the intricate interplay between HELLS and G4 structures, highlighting their potential synergistic mechanisms in modulating chromatin organization, governing gene expression regulation and maintain genome stability. By shedding light on these cooperative processes, our study contributes to a more comprehensive understanding of the dynamic and complex interrelationship between HELLS, G4 structures, and cellular function.

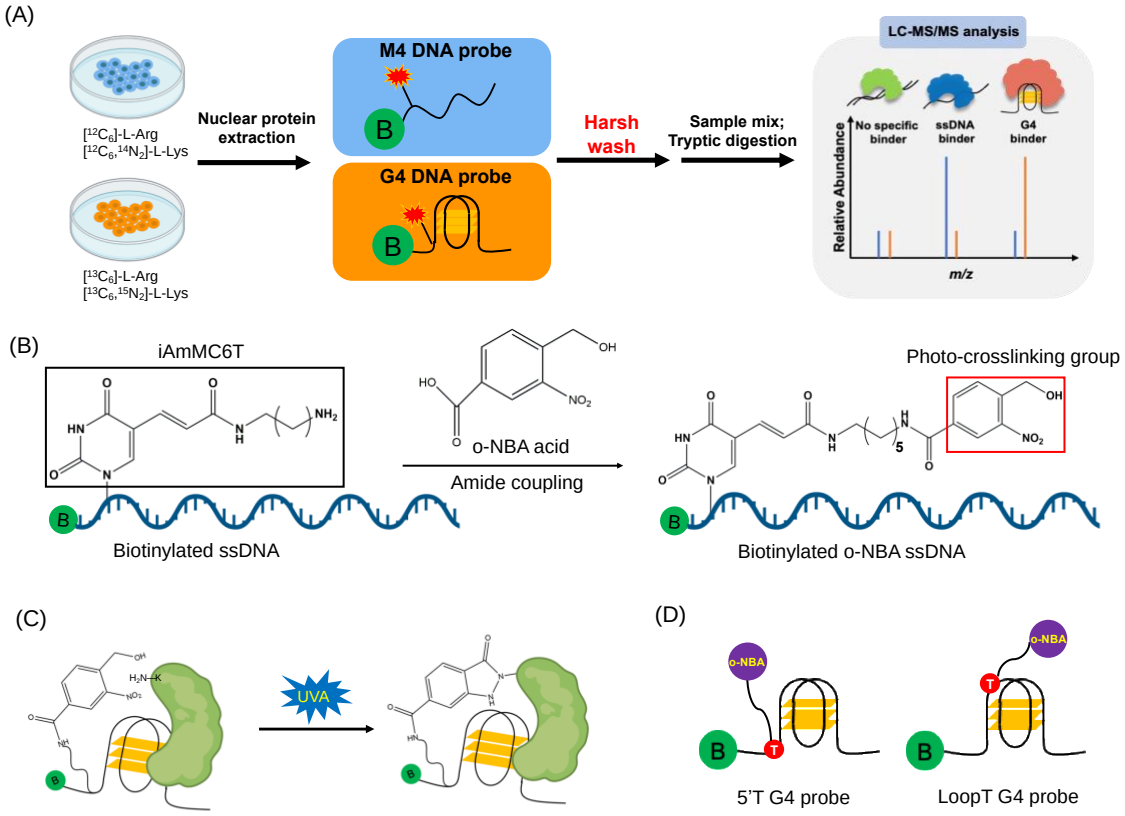

Figure 3.1. Identification of putative G4BPs with photo-crosslinking o-NBA group conjugated DNA probes. (A) a schematic workflow of the pulldown experiment for G4BPs identification. (B) Structures of biotinylated DNA G4 forming oligonucleotides with an internal amino modifier C6 dT (iAmMC6T), consist of a C6 linker on a thymidine and the structure of the o-NBA (ortho-nitrobenzylamine) photo-activating group conjugation through amide coupling. (C) The primary amine in G4BPs in a proximal will be captured by o-NBA group after photoactivation. (D) o-NBA group was placed at two different positions. 5'T G4 probes with the o-NBA group placed on the 5'end thymidine of the G4 structure and LoopT G4 probes with the o-NBA group located on the thymidine in the loop region of the G4 structure.

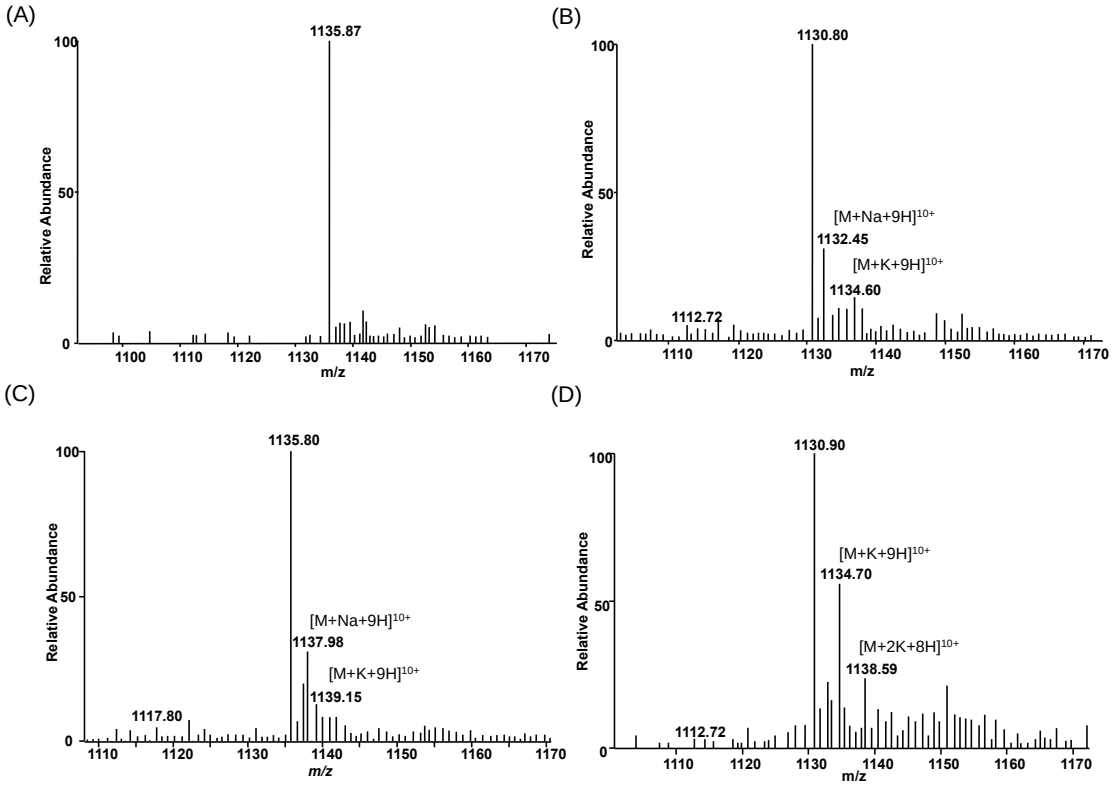

Figure 3.2: ESI-MS showing the $[M + 10H]^{10+}$ ions of successful synthesis of 5'T o-NBA G4/M4 as well as LoopT o-NBA G4/M4.

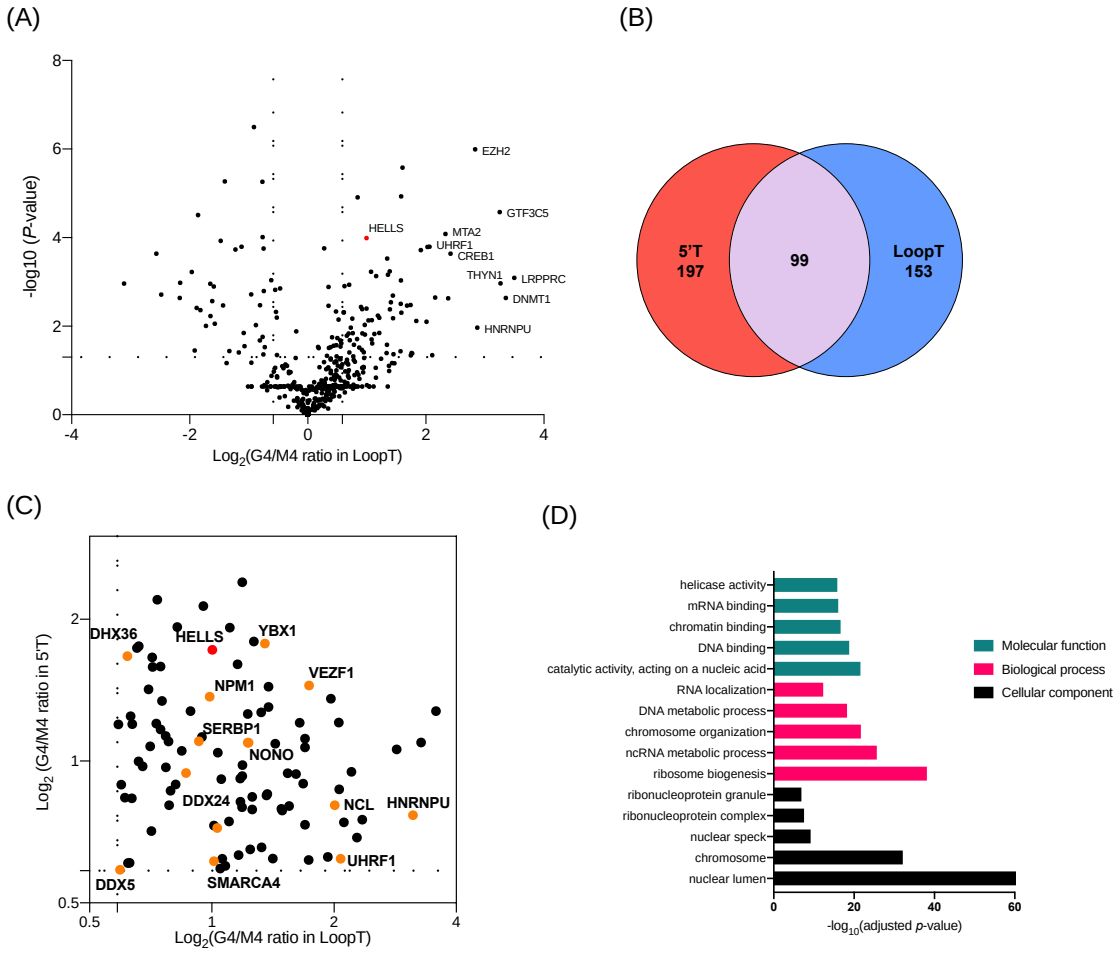

Figure 3.3: Putative G4-binding proteins identified in two sets of photo-crosslinking pulldown experiments. (A) volcano plot of identified G4-binding proteins in pulldown experiments using probes with o-NBA group placed at the thymidine in the loop region of the G4 structure. (B) Venn diagram of putative G4-binding protein identified in two sets of pull-down experiments with G4/M5 ration larger than 1.5. (C) A scatter plot of the common putative G4BPs identified in both 5'T and LoopT pull-down. Previously reported G4BPs are highlighted in orange. (D) Gene ontology (GO) enrichment analysis was performed using g: Profiler, providing GO terms including molecular functions, biological process and cellular component.

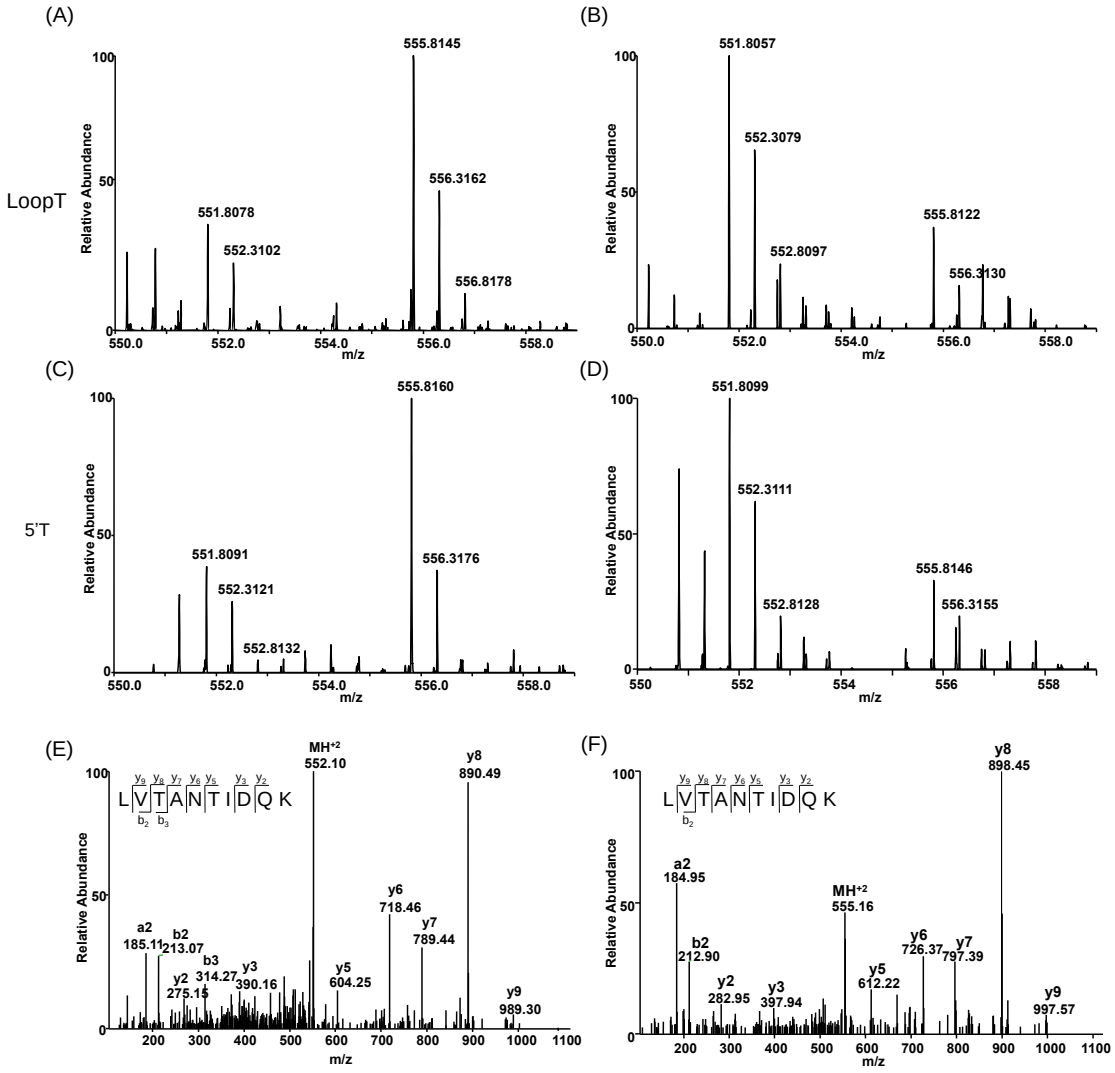

Figure 3.4. HELLS binds preferentially to G4 structures verses the mutant M4 sequences. ESI-MS showing the [M + 2H]2+ ions of light and heavy arginine-containing peptide LVTANTIDQK with monoisotopic m/z values being ~ 551.81 and 555.82, respectively, obtained from forward (A, C) and reverse (B, D) SILAC-based photo-crosslinking experiments with A and B from LoopT pulldown, C and D from 5'T pulldown. MS/MS for the [M+2H]2+ ions of the light (E) and heavy (F) lysine-containing peptide, LVTANTIDQK.

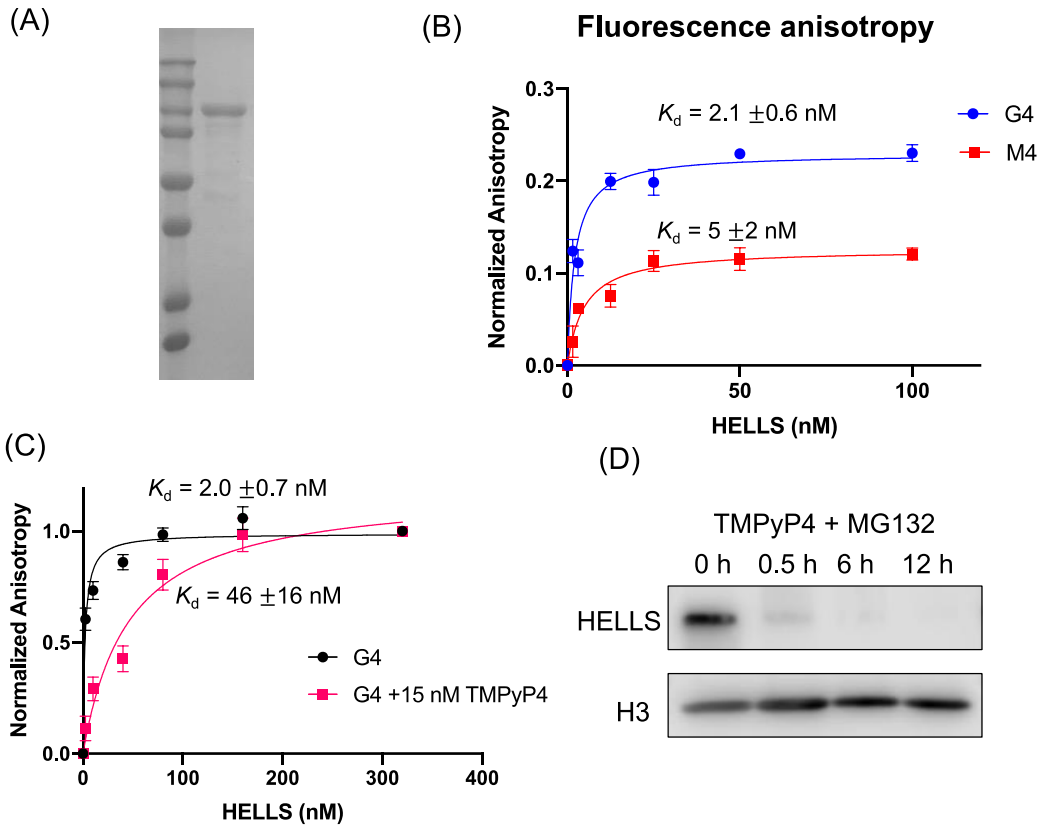

Figure 3.5: HELLS binds directly to G-quadruplex. (A) SDS-PAGE gel of purified Flag-tagged HELLS. (B) Fluorescent anisotropy for monitoring the binding of HELLS with G4 and M4 DNA probes. (C) Competitive binding assay with addition of 15 nM TMPyP4 in the HELLS-G4 binding samples. (D) Western blot analysis of HELLS in chromatin fraction of U2OS cells after treatment with 5 μM TMPyP4.

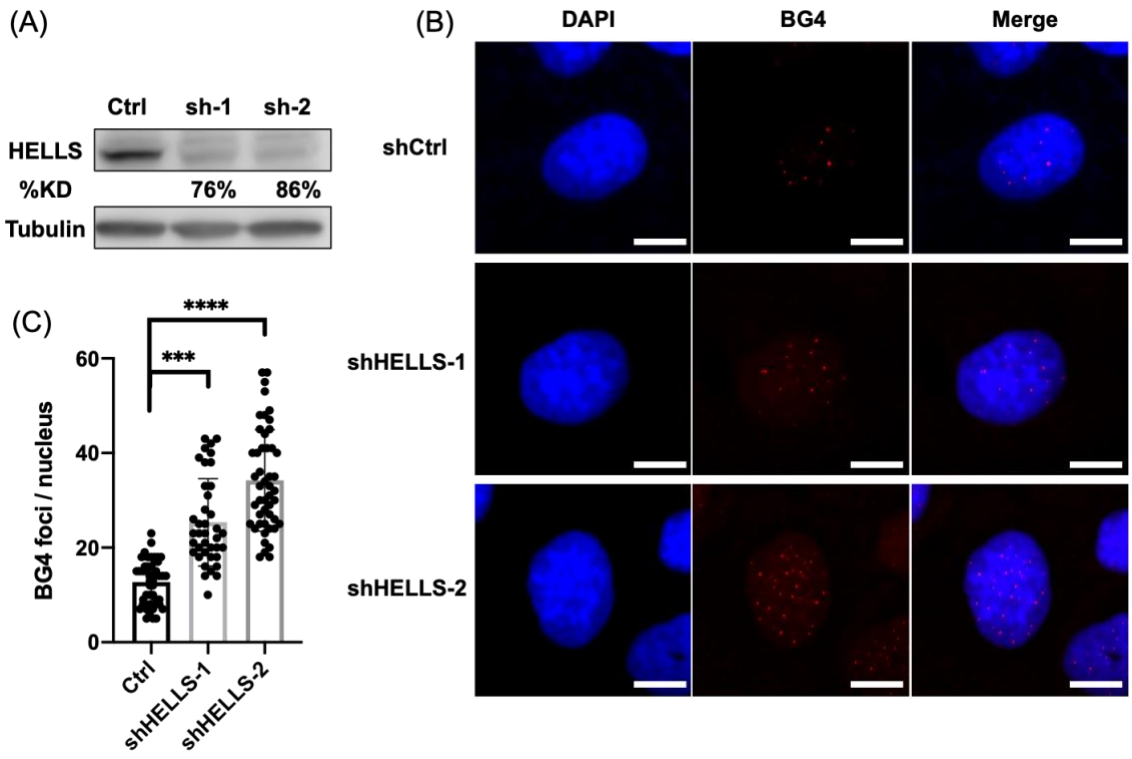

Figure 3.6: Knockdown of HELLS led to elevated numbers of G4 foci in U2OS cells. (A) Western blot analysis for monitoring the knockdown efficiencies of HELLS in U2OS cells with two different sequences of shRNA. (B) Immunofluorescence images of G4s in control and HELLS knockdown cells. (C) Quantification data of the numbers of G4 foci per nucleus in cells treated with shCtrl and shHELLS. (Scale bar:10 μm)

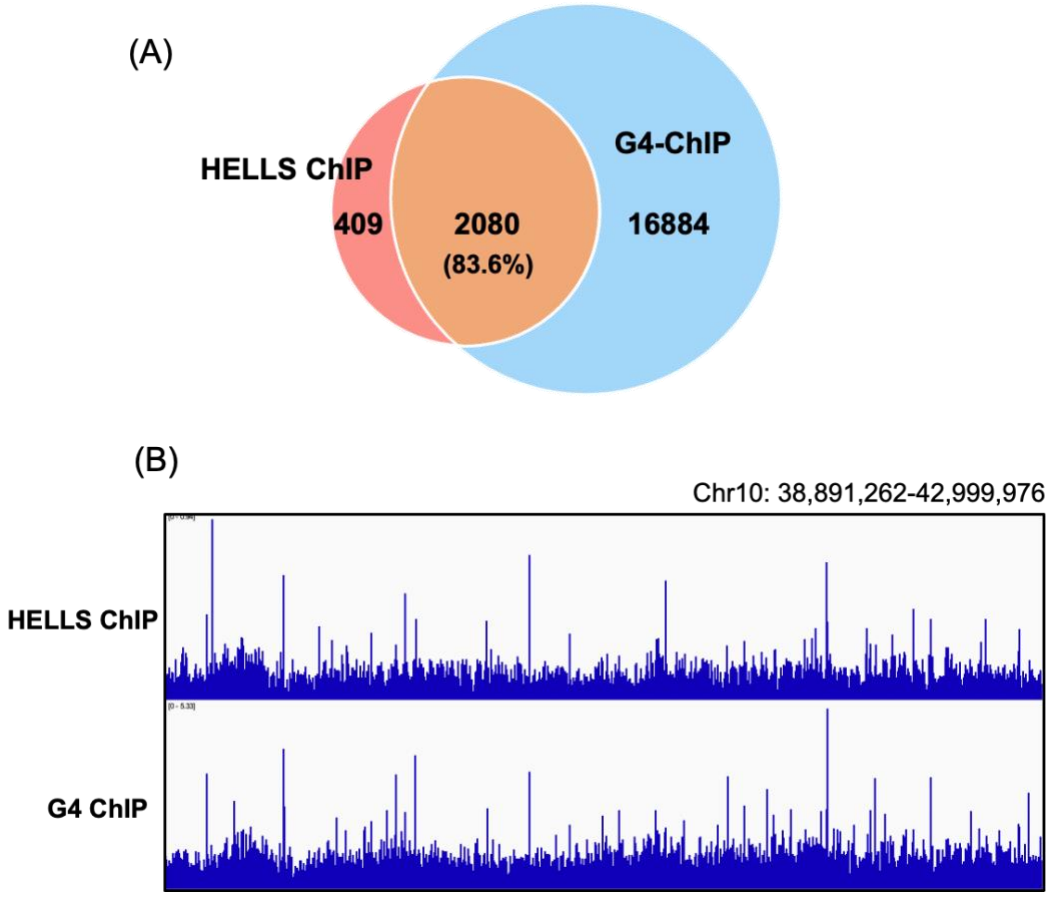

Figure 3.7: HELLS co-localized with G4 structure in mouse cells. (A) A venn diagram showing overlap of ChIP peaks in HELLS ChIP-seq and G4 ChIP-seq. (B) representative IGV plots of ChIP peaks in HELLS ChIP- seq and G4 ChIP-seq results.

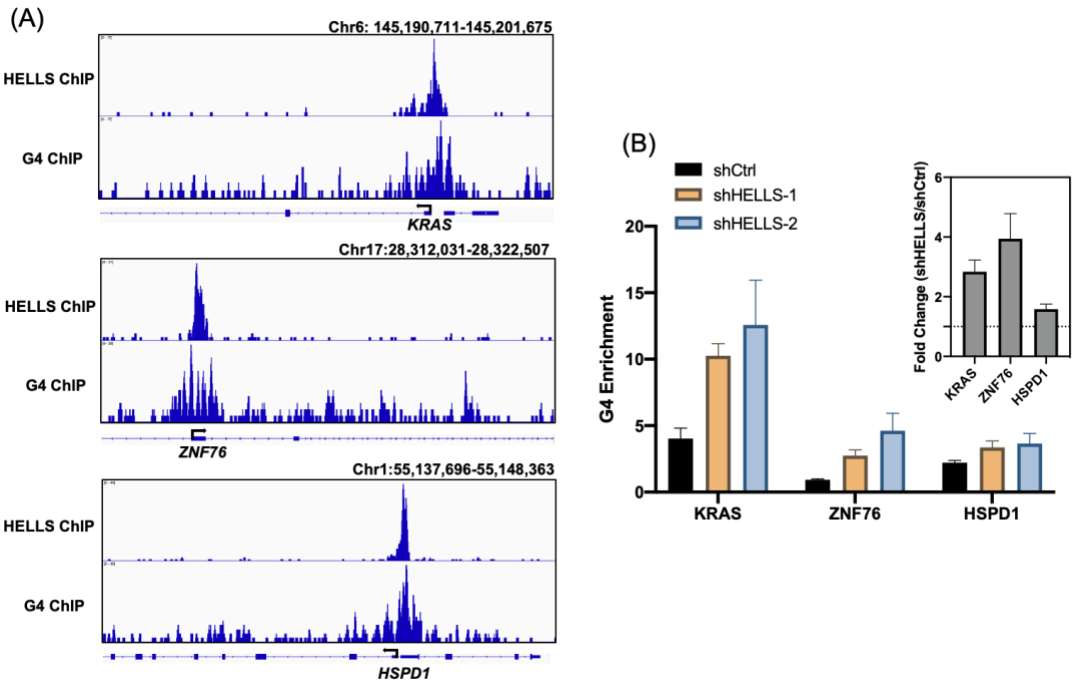

Figure 3.8: Knockdown of HELLS results in elevated enrichment of G4 structures at promoter regions in U2OS cells. (A) IGV plots showing the enrichments of HELLS binding sites and G4 structures near the transcription start site (TSS) of the genes (*KRAS, ZNF76, HSPD1*). (B) G4 ChIP-qPCR results showing enrichment of G4 structures in the promoter regions of *KRAS, ZNF76,* and *HSPD1* genes; and knock-down of HELLS results in augmented enrichment of G4 at these specific G4-containing regions.

Gene Name	5'T ratio (G4/M4)		LoopT ratio (G4/M4)	
	Avg.	Stdev.	Avg.	Stdev.
MYO1C	5.27	3.73	2.28	NaN
DDX54	4.60	3.68	1.66	NaN
MORC2	4.39	NaN	1.94	NaN
MKI67	3.80	3.10	1.77	0.40
XPC	3.79	0.02	2.15	0.54
YBX1	3.47	2.71	2.41	1.32
FBL	3.42	2.31	2.55	0.48
GNL2	3.37	3.20	1.58	0.88
ZNF638	3.34	NaN	1.57	NaN
HELLS	3.30	1.20	2.00	0.01
DHX36	3.18	1.58	1.54	0.89
RPL4	3.16	2.24	1.64	0.67
RPL17	3.04	NaN	2.23	0.31
MYO1B	3.00	1.39	1.68	NaN
SFPQ	3.00	2.27	1.64	0.29
VEZF1	2.73	NaN	3.33	NaN
ZC3H14	2.71	NaN	2.60	NaN
GNL3	2.68	1.72	1.62	0.35
NPM1	2.58	1.82	1.98	0.41

PSIP1	2.56	1.97	3.91	1.57
KIF2A	2.53	NaN	1.69	NaN
THAP11	2.47	NaN	2.60	NaN
TOP2A	2.42	1.56	1.85	0.42
LRPPRC	2.42	0.71	11.84	4.55
DDX49	2.41	NaN	2.50	0.21
SF3B1	2.39	1.23	2.34	0.53
GTPBP4	2.37	1.44	1.55	NaN
FBLL1	2.31	1.00	4.16	NaN
KIF22	2.31	1.91	3.13	NaN
TMPO	2.30	1.46	1.66	0.47
CALD1	2.29	1.45	1.55	0.25
MYBBP1A	2.29	1.33	1.50	0.26
RPS3A	2.25	0.83	1.68	0.68
RRP1B	2.19	1.32	1.70	0.88
POLRMT	2.18	0.36	1.93	0.30
DDX18	2.17	1.15	3.24	0.51
SERBP1	2.15	1.93	1.90	0.60
ATAD3A	2.14	1.19	1.72	0.29
GTF3C5	2.14	0.03	9.70	1.48
NONO	2.14	1.10	2.34	0.80

SF3B2	2.13	0.67	2.35	1.94
SUGP2	2.13	0.03	2.70	0.65
PRPF8	2.11	1.02	1.63	0.72
NCAPH2	2.10	0.44	3.24	1.28
XRCC5	2.08	0.17	7.22	2.35
TMEM201	2.07	NaN	1.79	NaN
TOP2B	2.06	1.18	2.05	0.23
RFC1	2.00	1.18	1.58	0.09
RECQL	1.97	0.78	2.28	0.05
SPTBN1	1.97	0.79	1.60	0.34
CSDE1	1.96	0.87	1.71	1.14
MTA2	1.93	1.06	4.62	1.37
DDX24	1.92	0.98	1.82	NaN
TRIM28	1.92	0.79	2.90	0.29
POLR1E	1.92	0.82	3.06	NaN
EXOSC10	1.90	0.63	2.28	0.61
LBR	1.89	0.57	2.26	1.83
IGF2BP3	1.89	0.94	2.08	0.32
ZC3HAV1	1.86	1.48	3.20	1.90
RRP7A	1.86	1.25	1.76	NaN
NOP56	1.85	0.81	1.51	0.34

PHF6	1.83	0.81	4.17	0.93
SUPT5H	1.82	0.56	1.73	0.45
PRRC2A	1.80	1.11	2.59	1.40
SCAF11	1.80	0.35	2.57	1.02
SMC4	1.79	0.63	2.39	NaN
CEBPZ	1.79	0.78	1.53	0.06
UTP20	1.78	0.82	1.55	0.25
SART1	1.76	0.89	2.26	NaN
NCL	1.75	0.88	4.02	0.68
SEC61A1; SEC61A2	1.75	0.99	1.72	0.69
BCLAF1	1.74	0.69	2.93	0.29
TEX10	1.74	0.66	2.28	0.55
ABCF1	1.73	0.81	2.80	1.19
RIF1	1.73	0.72	2.39	0.58
MDC1	1.72	0.90	2.81	0.75
HNRNPU	1.70	0.85	8.74	5.01
GTF3C3	1.68	0.05	5.08	1.20
THOC2	1.67	0.56	2.15	0.67
GRWD1	1.67	0.65	4.33	0.89
LARP1	1.66	0.78	3.24	1.20
ILF3	1.66	0.41	2.02	0.49

LAS1L	1.65	0.54	2.04	0.26
KIF4A	1.64	0.84	1.64	0.22
ZNF629	1.61	0.54	4.85	NaN
NCAPD3	1.58	0.52	2.51	0.72
BAZ1B	1.57	0.62	2.37	0.54
SMC2	1.55	0.47	2.24	0.33
GTF3C1	1.54	0.17	3.81	0.79
POLR2A	1.54	NaN	2.66	NaN
CPSF1	1.54	0.12	2.09	0.85
UHRF1	1.54	0.36	4.22	0.64
POLR1B	1.53	NaN	3.32	NaN
SMARCA4	1.53	0.43	2.02	0.50
RCC2	1.52	0.67	1.54	0.38
PELP1	1.52	0.74	1.54	0.58
PRKDC	1.51	0.46	2.11	0.82
PDCD11	1.51	0.54	2.07	1.50
DDX5	1.50	0.75	1.51	0.18

Table 3.1 Common proteins show G4/M4 ratio over 1.5 in both 5’T and LoopT pulldown experiments.

DNA sequence in pulldown experiments		
5’ T G4	5’-Biotin-T6 /iAmMC6T/AAAGGGTTAGGGTTAGGGTTAGGGAA-3’	
5’ T M4	5’-Biotin-T6 /iAmMC6T/AAAGGGTTAGTGTTAGTGTTAGGGAA-3’	
LoopT G4	5’-Biotin-T6 TAAAGGGTTAGGGT/iAmMC6T/AGGGTTAGGGAA-3’	
LoopT M4	5’-Biotin-T6 TAAAGGGTTAGTGT/iAmMC6T/AGTGTTAGGGAA-3’	
Fluorophore-labeled DNA		
G4 DNA	5'- TAMRA-TGAGGGTGGGGAGGGTGGGGAAGG-3'	
M4 DNA	5'- TAMRA-TGAGGGTGAGGAGTGTGGGGAAGG-3'	
ChIP-qPCR primers		
Gene	Forward	Reverse
KRAS	5'-TCTGGGCGAGAGGTCGG-3'	5'-CTGAAGAAGAATCGAGCGCGG-3'
HSPD1	5'-GGGGTAGTTCTTTCACCTCGG-3'	5'-ACGCTGACGCGAAGACTC-3'
ZNF76	5'-AAGAGCTCGGTTCAGATCCC-3'	5'-CCCCGGAACCCGTATAGAAA-3'
ESR	5'-GAAACAGCCCCA AATCTCAA-3'	5'-TTGTAGCCAGCAAGCAAATG-3'

Table 3.2 A list of G4 DNA sequences and ChIP-qPCR primers used in this study.

References

(1) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry* **2020**, *2* (2), 123-136.

(2) Nguyen, G. H.; Tang, W.; Robles, A. I.; Beyer, R. P.; Gray, L. T.; Welsh, J. A.; Schetter, A. J.; Kumamoto, K.; Wang, X. W.; Hickson, I. D.; et al. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc Natl Acad Sci U S A* **2014**, *111* (27), 9905-9910.

(3) Kamath-Loeb, A. S.; Loeb, L. A.; Johansson, E.; Burgers, P. M.; Fry, M. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J Biol Chem* **2001**, *276* (19), 16439-16446.

(4) Wu, C. G.; Spies, M. G-quadruplex recognition and remodeling by the FANCJ helicase. *Nucleic Acids Res* **2016**, *44* (18), 8742-8753.

(5) Yagi, R.; Miyazaki, T.; Oyoshi, T. G-quadruplex binding ability of TLS/FUS depends on the beta-spiral structure of the RGG domain. *Nucleic Acids Res* **2018**, *46* (12), 5894-5901.

(6) Ghosh, M.; Singh, M. RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res* **2018**, *46* (19), 10246-10261.

(7) Gao, Z.; Williams, P.; Li, L.; Wang, Y. A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* **2021**, *20* (11), 4919-4924.

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

(9) Dahan, D.; Tsirkas, I.; Dovrat, D.; Sparks, M. A.; Singh, S. P.; Galletto, R.; Aharoni, A. Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res* **2018**, *46* (22), 11847-11857.

(10) Johnson, J. E.; Cao, K.; Ryvkin, P.; Wang, L. S.; Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res* **2010**, *38* (4), 1114-1122.

(11) Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J Am Chem Soc* **2017**, *139* (36), 12426-12429.

(12) Li, L.; Williams, P.; Ren, W.; Wang, M. Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **2021**, *17* (2), 161-168.

(13) Li, L.; Williams, P.; Gao, Z.; Wang, Y. VEZF1-guanine quadruplex DNA interaction regulates alternative polyadenylation and detyrosinase activity of VASH1. *Nucleic Acids Res* **2020**, *48* (21), 11994-12003.

(14) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **2002**, *99* (17), 11020-11024.

(15) Tanaka, Y.; Bond, M. R.; Kohler, J. J. Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* **2008**, *4* (6), 473-480.

(16) Murale, D. P.; Hong, S. C.; Haque, M. M.; Lee, J. S. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome Sci* **2016**, *15*, 14.

(17) Saghatelian, A.; Jessani, N.; Joseph, A.; Humphrey, M.; Cravatt, B. F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc Natl Acad Sci U S A* **2004**, *101* (27), 10000-10005.

(18) Lavrik, O. I.; Kolpashchikov, D. M.; Prasad, R.; Sobol, R. W.; Wilson, S. H. Binary system for selective photoaffinity labeling of base excision repair DNA polymerases. *Nucleic Acids Res* **2002**, *30* (14), e73.

(19) Lavrik, O. I.; Kolpashchikov, D. M.; Weisshart, K.; Nasheuer, H. P.; Khodyreva, S. N.; Favre, A. RPA subunit arrangement near the 3'-end of the primer is modulated by the length of the template strand and cooperative protein interactions. *Nucleic Acids Res* **1999**, *27* (21), 4235-4240.

(20) Dorman, G.; Nakamura, H.; Pulsipher, A.; Prestwich, G. D. The Life of Pi Star: Exploring the Exciting and Forbidden Worlds of the Benzophenone Photophore. *Chem. Rev.* **2016**, *116* (24), 15284-15398.

(21) Xia, Y.; Peng, L. Photoactivatable lipid probes for studying biomembranes by photoaffinity labeling. *Chem Rev* **2013**, *113* (10), 7880-7929.

(22) Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg Med Chem* **2012**, *20* (2), 554-570.

(23) Su, H.; Xu, J.; Chen, Y.; Wang, Q.; Lu, Z.; Chen, Y.; Chen, K.; Han, S.; Fang, Z.; Wang, P.; et al. Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J Am Chem Soc* **2021**, *143* (4), 1917-1923.

(24) Zhang, X.; Spiegel, J.; Martinez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **2021**, *13* (7), 626-633.

(25) Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E. A.; Rodriguez, R.; Balasubramanian, S. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew Chem Int Ed Engl* **2012**, *51* (44), 11073-11078.

(26) West, A. V.; Muncipinto, G.; Wu, H. Y.; Huang, A. C.; Labenski, M. T.; Jones, L. H.; Woo, C. M. Labeling Preferences of Diazirines with Protein Biomolecules. *J Am Chem Soc* **2021**, *143* (17), 6691-6700.

(27) Guo, A. D.; Wei, D.; Nie, H. J.; Hu, H.; Peng, C.; Li, S. T.; Yan, K. N.; Zhou, B. S.; Feng, L.; Fang, C.; et al. Light-induced primary amines and o-nitrobenzyl alcohols cyclization as a versatile photoclick reaction for modular conjugation. *Nat Commun* **2020**, *11* (1), 5472.

(28) Guo, A. D.; Yan, K. N.; Hu, H.; Zhai, L.; Hu, T. F.; Su, H.; Chi, Y.; Zha, J.; Xu, Y.; Zhao, D.; et al. Spatiotemporal and global profiling of DNA-protein interactions enables discovery of low-affinity transcription factors. *Nat Chem* **2023**, *15* (6), 803-814.

(29) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26* (12), 1367-1372.

(30) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016**, *11* (12), 2301-2319.

(31) Gao, Z.; Yuan, J.; He, X.; Wang, H.; Wang, Y. Phase Separation Modulates the Formation and Stabilities of DNA Guanine Quadruplex. *JACS Au* **2023**.

(32) Wang, F.; Tang, M. L.; Zeng, Z. X.; Wu, R. Y.; Xue, Y.; Hao, Y. H.; Pang, D. W.; Zhao, Y.; Tan, Z. Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc Natl Acad Sci U S A* **2012**, *109* (50), 20413-20418.

(33) Thijssen, P. E.; Ito, Y.; Grillo, G.; Wang, J.; Velasco, G.; Nitta, H.; Unoki, M.; Yoshihara, M.; Suyama, M.; Sun, Y.; et al. Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat Commun* **2015**, *6*, 7870.

(34) He, Y.; Ren, J.; Xu, X.; Ni, K.; Schwader, A.; Finney, R.; Wang, C.; Sun, L.; Klarmann, K.; Keller, J.; et al. Lsh/HELLS is required for B lymphocyte development and immunoglobulin class switch recombination. *Proc Natl Acad Sci U S A* **2020**, *117* (33), 20100-20108.

(35) Zocchi, L.; Mehta, A.; Wu, S. C.; Wu, J.; Gu, Y.; Wang, J.; Suh, S.; Spitale, R. C.; Benavente, C. A. Chromatin remodeling protein HELLS is critical for retinoblastoma tumor initiation and progression. *Oncogenesis* **2020**, *9* (2), 25.

(36) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat Chem Biol* **2005**, *1* (3), 167-173.

(37) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **2013**, *5* (3), 182-186.

(38) Kollarovic, G.; Topping, C. E.; Shaw, E. P.; Chambers, A. L. The human HELLS chromatin remodelling protein promotes end resection to facilitate homologous recombination and contributes to DSB repair within heterochromatin. *Nucleic Acids Res* **2020**, *48* (4), 1872-1885.

(39) Zhang, G.; Dong, Z.; Prager, B. C.; Kim, L. J.; Wu, Q.; Gimple, R. C.; Wang, X.; Bao, S.; Hamerlik, P.; Rich, J. N. Chromatin remodeler HELLS maintains glioma stem cells through E2F3 and MYC. *JCI Insight* **2019**, *4* (7).

(40) von Eyss, B.; Maaskola, J.; Memczak, S.; Mollmann, K.; Schuetz, A.; Loddenkemper, C.; Tanh, M. D.; Otto, A.; Muegge, K.; Heinemann, U.; et al. The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *EMBO J* **2012**, *31* (4), 972-985.

(41) Lyu, J.; Shao, R.; Kwong Yung, P. Y.; Elsasser, S. J. Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res* **2022**, *50* (3), e13.

(42) Wu, G.; Xing, Z.; Tran, E. J.; Yang, D. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc Natl Acad Sci U S A* **2019**, *116* (41), 20453-20461.

(43) Chen, M. C.; Tippana, R.; Demeshkina, N. A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D'Amare, A. R. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **2018**, *558* (7710), 465-469.

(44) Baumann, C.; Ma, W.; Wang, X.; Kandasamy, M. K.; Viveiros, M. M.; De La Fuente, R. Helicase LSH/Hells regulates kinetochore function, histone H3/Thr3 phosphorylation and centromere transcription during oocyte meiosis. *Nat Commun* **2020**, *11* (1), 4486.

(45) Xu, X.; Ni, K.; He, Y.; Ren, J.; Sun, C.; Liu, Y.; Aladjem, M. I.; Burkett, S.; Finney, R.; Ding, X.; et al. The epigenetic regulator LSH maintains fork protection and genomic stability via MacroH2A deposition and RAD51 filament formation. *Nat Commun* **2021**, *12* (1), 3520.

Chapter 4: Phase Separation Modulates the Formation and Stabilities of DNA Guanine Quadruplex

4.1 Introduction

Guanine quadruplexes (G4s) are non-canonical DNA structures comprised of two or more layers of G-tetrads, each of which contains four guanines stabilized by Hoogsteen base pairing and a monovalent metal ion, e.g., K^+ and Na^+. The formation of G4 structures in human cells was first reported in 1987,[1] and many subsequent studies were conducted to assess G4 formation *in vitro* and in cells. In particular, it was revealed bioinformatically and experimentally that DNA G4 is ubiquitously present in the human genome.[2-4] These studies also unveiled the enrichment of G4 structures at replication origins, oncogene promoters and telomeres, suggesting the important functions of G4 structures in regulating DNA replication, transcription, telomere maintenance, and other biological processes.[5]

G4 structures are highly dynamic in cells,[6, 7] and effective regulations of G4s are crucial for maintaining genomic and epigenetic stability, as manifested by the functions of known G4BPs in chromatin remodeling (e.g., ATRX and REV1),[8, 9] long-range DNA interactions (e.g., YY1 and CTCF),[10, 11] and genetic diseases (e.g., WRN and BLM).[12, 13] Moreover, G4 structures are known to be dynamic during cell cycle progression, where higher levels of G4 structures were observed in S-phase cells.[14] Therefore, it is important to understand how G4 structures are regulated in cells.

Liquid-liquid phase separation (LLPS) promotes the formation of membraneless compartments in cells, such as nucleolus, nuclear speckles, Cajal bodies, processing bodies (P bodies), and stress granules, etc.[15] LLPS has been shown to participate in a wide range of cellular processes, including transcription, translation, DNA damage repair, cell signaling and spatial genome organization.[15-17] Although many studies are focused on LLPS of proteins or protein compartments, it was recently revealed that, in the absence of proteins, DNA can also undergo phase separation. For instance, Shakya et al.[18] found that certain sequences of DNA, e.g., poly(GC), can form phase-separated condensates. In addition, recent studies on G4BPs unveiled important functions of G4s in many different biological processes.[19-21] Interestingly, many of these proteins, including FUS,[22] hnRNPA1,[23] hnRNPA2,[24] and YY1,[25] can undergo phase separation. Furthermore, G4 structures can promote the condensation of G4BPs, including histone H1 and SERBP1.[26, 27]

In light of these previous studies, we hypothesized that phase separation may modulate the stabilities of G4 structures in cells. In this study, we observed LLPS of DNA G4s *in vitro* and demonstrated that phase-separated G4 DNA droplets can be disintegrated by 1,6-hexanediol (1,6-HD), an agent known to disrupt LLPS.[28, 29] We also found that disruption of phase separation diminished the enrichment of G4 structures in chromatin of cultured human cells.

4.2 Materials and Methods

Cell culture

U2OS human osteosarcoma cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher) and 1% penicillin-streptomycin solution (PS, GE Healthcare) at 37°C with 5% CO_2.

Oligodeoxyribonucleotides and G4 formation

Unlabeled and 5′-TAMRA-labeled oligodeoxyribonucleotides (ODNs) were purchased from Integrated DNA Technologies (IDT), and their sequences are listed in Supplementary Table 1. These ODNs were annealed in a buffer containing 10 mM Tris-HCl (pH 7.5), 10 mM KCl and 0.1 mM EDTA by heating to 95°C for 5 min, followed by cooling down to room temperature slowly over 6 hr.

Circular dichroism (CD) spectroscopy

The CD spectra for the ODNs (10 μM) in a binding buffer containing 10 mM HEPES (pH 7.4) and 0.1 mM EDTA with or without 0.1 mg/ml poly-L-lysine (P2636, Sigma) were recorded at room temperature on a Jasco-815 spectrometer (Easton, MD) at a scan rate of 1 nm/sec. The CD spectra were averaged from signal of two repetitive scans collected in the wavelength range of 220-320 nm. The final spectra were obtained by subtraction of signal acquired for the buffer solution and signal smoothing, and plotted using GraphPad Prism.

Phase separation assay

Pre-annealed 5′-TAMRA labeled DNA and unlabeled DNA were mixed at a molar ratio of 1:50 in the above-mentioned buffer. For imaging of DNA in the presence of NaCl, NaCl was added to the buffer until its desired concentration was reached. The samples were incubated with or without 0.1 mg/ml poly-L-lysine at 25ºC for 30 min. The samples were subsequently dropped onto a glass microscope slide and covered with a 12 mm coverslip (CG15NH1, Thorlabs) immediately before imaging. Fluorescence and differential interference contrast (DIC) imaging was conducted on an LSM 880 upright confocal microscope (Carl Zeiss) with a 60× objective. The images were analyzed by ZEN software.

Turbidity measurements

Unlabeled *cMYC* DNA (20 µM) was prepared under the same conditions as in phase separation assay. After incubation, the absorbance at 400 nm was measured using a Nanodrop One Spectrophotometer (Thermo Scientific). Triplicate measurements were performed. The average of the three readouts for each replicate was recorded.

1,6-hexanediol treatment

In all *in vitro* assays, including phase separation and turbidity measurements, stock solutions of 20%, 40% and 60% of 1,6-hexanediol (240117, Sigma) in a binding buffer were prepared before treatment. After addition of an equal volume of 1,6-HD-containing buffer, sample solution was mixed thoroughly by pipetting and incubated at 25ºC for another 30 min before measurement.

G4 immunofluorescence

U2OS cells were treated with 6% 1,6-hexanediol (240117, Sigma) prepared in DMEM for 1 min; JQ1 (HY-13030, Medchemexpress) was added to U2OS cells until its final concentration reached 1.0 μM, and the cells were incubated for 12 hr before immunofluorescence microscopy imaging. The immunofluorescence microscopy experiments were performed following previously published procedures.[14] Briefly, cells on coverslip were fixed in methanol: acetic acid (3:1, v/v) for 10 min, permeabilized with 0.1% triton-X100/PBS for 15 min, and treated with 50 μg/ml RNase A (EN0531, Thermo Fisher). After blocking with 2% BSA at room temperature for 1 hr, immunofluorescence microscopy experiments were conducted using standard methods with BG4 (MABE917, Sigma-Aldrich), anti-FLAG (14793S, Cell Signaling Technology) and anti-rabbit Alexa 594-conjugated (A11037, Invitrogen) antibodies. Nuclei were stained with DAPI (D9542, Sigma). Finally, the coverslips were mounted with ProLong™ Diamond Antifade Mountant (Invitrogen). Images were recorded using a LSM880 Confocal Laser Scanning Microscope (Carl Zeiss) with a 100x objective and analyzed with ZEN. The foci number per nucleus was counted using Find maxima in ImageJ. The graphs were plotted using GraphPad Prism8.

G4-ChIP sequencing and qPCR

G4-ChIP was performed using custom-purified BG4 antibody and conducted as previously described with minor modifications.[30] DNA was fragmented following the protocol described previously.[31] Briefly, chromatin samples were diluted in a blocking buffer containing 25 mM HEPES (pH 7.5), 10.5 mM NaCl, 110 mM KCl, 1 mM $MgCl_2$,

1% BSA, and treated with RNase A. The chromatin sample was subsequently incubated with 500 ng BG4 antibody with rotation at 1400 r.p.m for 1.5 hr at 16ºC. To the mixture were then added 5 µl pre-blocked Anti-Flag M2 magnetic beads (Sigma, M8823), and the suspension was incubated under the same conditions for 1 hr. After washing with ice-cold wash buffer (10 mM Tris, pH 7.4, 100 mM KCl, 0.1% Tween 20) for 7 times, the captured DNA was eluted with TE buffer containing Proteinase K while rotating at 1400 r.p.m. (6 hr, 65ºC). The eluted DNA was then purified by DNA Clean and Concentrator-5 (Zymo).

The DNA-sequencing library was prepared using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) following the manufacturer's instructions. The quantity and quality of the purified DNA libraries were assessed by Qubit and Agilent 2100 Bioanalyzer. The samples were multiplexed for next-generation sequencing on a MGISEQ-2000 platform (BGI).

For quantitative PCR, immunoprecipitated samples and the input control were used to quantify the enrichment of G4 structures. qPCR was carried out using Luna® Universal qPCR Master Mix (NEB) on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). The primers are listed in Supplementary Table 1.

Cell Counting Kit-8 (CCK8) assay

The cell cytotoxicity after 1,6-HD and JQ1 treatment was analyzed by using Cell Counting Kit-8 (CK04, Dojindo) following the manufacturer's protocols. Cells were seeded at a density of 5×10^3/well in 100 µL of medium into 96-well microplates. Cells were incubated overnight before treatment. After treatment, 10 µL of CCK-8 reagent was added to each well and then incubated for 3 hr. The absorbance was then measured at 450 nm

using a BioTek Synergy H1 microplate reader. All experiments were performed in triplicate.

Bioinformatics analysis

FastQC (version 0.11.9) was employed for the quality assessment of ChIP-seq data, and the reads were subsequently mapped to human hg38 reference genome using Bowtie2 (Version 2.5.0)[32]. The two replicates were merged using SAMtools for analysis[33]. The unaligned and repeated reads were filtered out, and the remaining reads were subjected to peak calling using MACS2 (v2.2.7.1)[34] with a false discovery rate of 0.05. Bigwig files were generated by using the bamCompare tool.[35] The results were visualized using Integrative Genomics Viewer (IGV)[36].

ChIP-seq datasets of transcription factors were retrieved from ChIP-atlas portal (https://chip-atlas.org/) using "hg38", Experiment type "ChIP: TFs and others", Cell type Class "Bone", Threshold for significance "50", Cell type "U2OS". Experiments with treatment or genetic manipulation were removed from analysis. BG4 peaks that can be detected in both control and 1,6-hexanediol-treated cells were separated into "Augmented G4" and "Diminished G4" based on enrichment score. Overlapping percentage of G4 with TFs binding sites were calculated using bedtools.[37] Fold change was defined as the ratio between TFs binding sites' overlapping percentage with "Augmented G4" and "Diminished G4" in 1,6-hexanediol- vs. mock-treated cells. Profiles and heatmaps were generated by using computeMatrix tool.[35] The ChIP-seq data reported in this paper are available at the NCBI GEO repository under accession number GSE225772 (Reviewer token: wdgpsgkqrxuzlwp).

4.3 Results

To investigate whether G4s undergo phase separation and assemble into droplets in the absence of proteins, we tested two different G4-forming sequences derived from the promoter regions of c*MYC* and c*KIT* proto-oncogenes. In this vein, the *cKIT* promoter harbors three G4-forming sequences, and the first sequence, KIT1, was used in the present study.[38] Oligodeoxynucleotides were pre-annealed in a buffer containing 10 mM KCl and confirmed by circular dichroism spectroscopy (Figure 4.1) to ensure the formation of G4 structures before the LLPS assays. As depicted in Figure 1A, both G4 sequences can undergo phase separation and assemble into droplets in the presence of poly-L-lysine (PLL), a cationic polymer used for studying phase separation,[39] where droplets start to form at 10 μM G4 DNA. We also examined G4 droplet formation by using a previously reported turbidity assay.[26, 40] Upon incubation with PLL, the turbidity of the DNA sample increased significantly, as manifested by the formation of droplets in PLL-containing DNA solution (Figure 4.2).

We next assessed the influence of salt concentration on phase separation of G4 DNA. Our results showed that, after annealing G4 DNA in a K^+-containing buffer, the droplets started to form at DNA concentrations as low as 10 μM (Figure 4.2). At a low concentration (5 μM), G4 DNA does not assemble into droplets until NaCl concentration is at least 100 mM (Figure 4.3). We also observed that, at higher DNA concentrations (20-50 μM), G4 DNA assembles into droplets at both low and high NaCl concentrations, but form aggregates at intermediate NaCl concentrations (Figure 4.2-4.3). We reason that the addition of salt increases the ionic strength of the solution and attenuated the repulsive

electrostatic interactions between negatively charged G4 DNA, thereby resulting in droplet formation at low DNA concentration and aggregation at high DNA concentration.[41] Further increases in salt concentration augmented the solubility of DNA in the PLL-DNA mixture;[42] as a result, we observed precipitate dissolution and droplet reformation at high salt concentrations. Together, we showed that G4 DNA can undergo phase separation at a concentration as low as 10 μM. In this respect, it is worth noting that the intracellular concentration of Na^+ is about 10-15 mM, residing in the low concentration range of NaCl used in this study.[43] In addition, the average local concentration of DNA in metaphase chromosomes is approximately 0.17 g/ml,[44] which is much higher than the 10 μM G4 DNA probe used in the current study (~ $7*10^{-5}$ g/ml). Thus, DNA in chromatin experiences an environment with much more molecular crowding, which favors LLPS,[45] than the *in vitro* conditions employed in the current study.

1,6-HD has been frequently employed to disrupt phase separation through inhibition of hydrophobic interactions.[28, 29] We found that addition of 1,6-HD led to disintegration of DNA G4 droplets (Figure 4.4), where we observed a marked attenuation in droplet size when the solution contained 20% 1,6-HD. We also assessed droplet disruption by using a turbidity assay. As displayed in Figure S3B, treatment with increasing concentrations of 1,6-HD led to decreased turbidity of the DNA solution, underscoring the disintegration and/or dissolution of G4 DNA droplets. As 1,6-HD exhibits a more pronounced impact on turbidity at low than high NaCl concentration, phase separation of G4 DNA at low NaCl concentration depends more on hydrophobic interaction. This result demonstrated that, aside from electrostatic interaction, hydrophobic interaction also contributes to LLPS of

G4 DNA. Together, these results revealed that 1,6-HD treatment promotes the disintegration of G4 droplets *in vitro*.

We next investigated how phase separation influences the stability of G4 structures in cells. To this end, we treated U2OS cells with 6% 1,6-HD for 1 min, and assessed G4 formation in cells by using immunofluorescence microscopy with the use of BG4 antibody.[14] In line with the *in vitro* observations, we detected pronounced diminutions in G4 structure foci in cells upon 1,6-HD treatment (Figure 4.5). In this context, 1,6-HD is known to be capable of disrupting protein-protein interactions;[46] hence, the diminished G4 levels in cells may also arise partly from attenuated protein-protein interactions involving G4BPs. It is of note that 1,6-HD exposure did not appreciably influence the survival of U2OS cells (Figure 4.4).

Several studies demonstrated the formation of G4s in promoters of transcriptionally active genes.[47-49] In addition, transcription co-activators form liquid-like condensates, which recruit other proteins to enhancer regions to activate transcription.[50, 51] Therefore, we hypothesized that the assembly of these phase-separated condensates associated with promoter regions may lead to stabilization of G4 structures. To test this hypothesis, we assessed how G4 formation in cells is perturbed upon treatment with JQ1, which disrupts the binding of BRD4 to H3K27ac-marked enhancer chromatin and dissolves the phase-separated condensates comprised of mediators and RNA polymerase II at super-enhancers.[51] Immunofluorescence microscopy analysis with the use of BG4 antibody again revealed a significant diminution in BG4 foci in U2OS cells after JQ1 treatment (Figure 4.5). In this regard, we observed that a 12-hr treatment with 1.0 μM JQ-1 did not alter the

survival of U2OS cells (Figure 4.6). This result substantiated our conclusion that G4 structures in cells are stabilized in phase-separated condensates, though attenuated enhancer activity, arising from JQ-1 treatment, may also contribute in part to diminished G4 structures.

To further examine the roles of phase separation in stabilizing G4 structures in cells, we performed BG4-ChIP-seq experiment for U2OS cells that were untreated or treated with 1.5% 1,6-HD for 2 min.[30, 31] Our results from two biological replicates revealed a substantially diminished numbers of BG4 ChIP-seq peaks, namely, from 3507 peaks in control cells to 2269 peaks in 1,6-HD-treated cells. The majority of the peaks (58.9% and 61% in control and 1,6-HD-treated cells, respectively) are located in promoter regions (Figure 4.7), which is in agreement with the previously reported results[48]. We also observed that the peak intensities were substantially attenuated upon 1,6-HD treatment (Figure 4.7). It is worth noting that, among the 1369 overlapped peaks detected from control and 1,6-HD-treated cells, more than 65% of the peaks are located in promoter regions.

We further validated the BG4 ChIP-seq data by using ChIP-qPCR. We chose several genes, including *HYAL3, CCDC88A, PRA3* and *PRR14*, whose promoter regions exhibit high G4-forming potential and display diminished BG4 ChIP-seq peak intensities following 1,6-HD treatment (Figure 4.8-4.9). Consistent with the ChIP-seq results, we detected attenuated enrichment of G4 structures in the promoter regions of all four genes upon 1,6-HD treatment. Hence, we conclude that phase separation contributes to the stabilization of G4 structures in cells.

Because G4 structures were found to be binding hubs for transcription factors (TFs) in chromatin,[48] we next examined whether those loci with G4 structures being sensitive to 1,6-HD treatment are enriched with TF factor binding sites. To this end, we first divided those G4 loci commonly detected in control and 1,6-HD-treated cells into two groups based on IP enrichment scores, i.e., those G4 loci exhibiting decreased and increased BG4 ChIP signals after 1,6-HD treatment (i.e., labeled as 'Diminished G4' and 'Augmented G4', respectively, in Figure 4C). We subsequently calculated the overlapping percentage of BG4 peaks in each group with ChIP-seq results of TFs using publicly available data from ChIP-Atlas. Our results showed that those loci with diminished BG4 signal upon 1,6-HD treatment are more likely to co-localize with transcription factor binding sites than those with elevated BG4 signal after 1,6-HD treatment (Figure 4C). Along this line, it remains unclear how 1,6-HD treatment leads to elevated BG4 signal in some genomic regions, though we speculate that this may arise from augmented accessibilities of G4 structures in chromatin of these genomic regions. A comparison of BG4 ChIP-seq peaks detected in control and 1,6-HD-treated cells with respect to individual ChIP-seq data of TFs, including CTCF, MYC and ZBTB48, showed minimal or no effects of their overall genome-wide occupancy upon 1,6-HD treatment (Figure 4.8 and 4.10). We, however, observed pronounced diminutions in enrichment of these TFs at those binding sites enriched with G4 structures. These data suggest that G4-mediated LLPS is involved in regulating the chromatin occupancy of TFs.

4.4 Conclusion

In summary, we demonstrated that G4 DNA can undergo phase separation, and disruption of phase separation leads to genome-wide destabilization of G4s. Previous studies showed that DNA G4 structures promote phase separation through binding to phase-separated proteins.[52] Here we revealed another mechanism in modulating dynamic formation of DNA G4 structures, which may bear important implications in transcriptional regulation. In this vein, G4 structures are enriched in gene promoters,[5] and RNA polymerase II can partition into two distinct types of phase-separated condensates, where the form with the C-terminal domain being hypophosphorylated is accompanied with transcription initiation, but the hyperphosphorylated form is associated with splicing complex.[53] While more efforts are needed to understand further how phase separation of G4 DNA regulates biological processes, our studies suggest that G4 structures are hubs of not only the transcription machinery,[48] but perhaps also other molecular complexes involved in chromatin remodeling and long-range DNA interactions.[7, 10, 25]

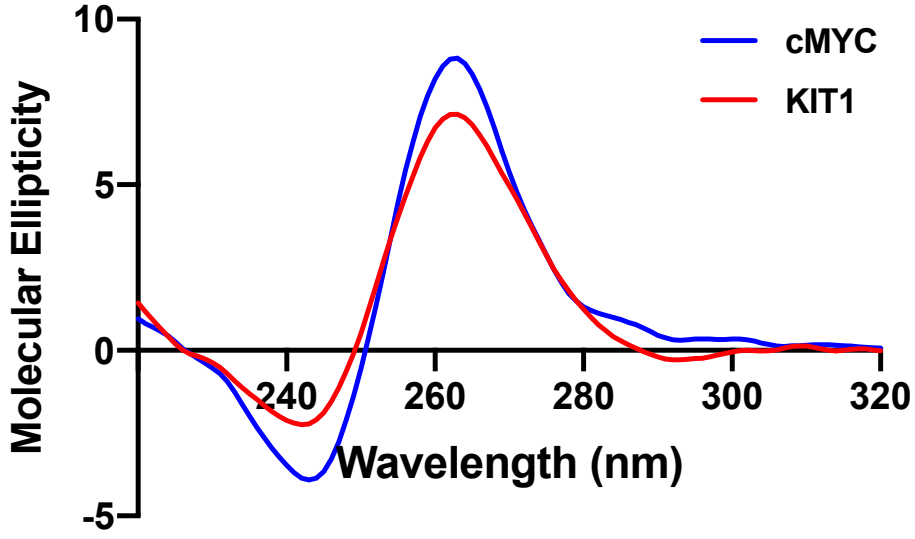

Figure 4.1. CD spectra of G4 DNA derived from the promoters of cMYC and KIT genes. The DNA probes were annealed in 10 mM KCl and diluted by 10-fold before measurement.

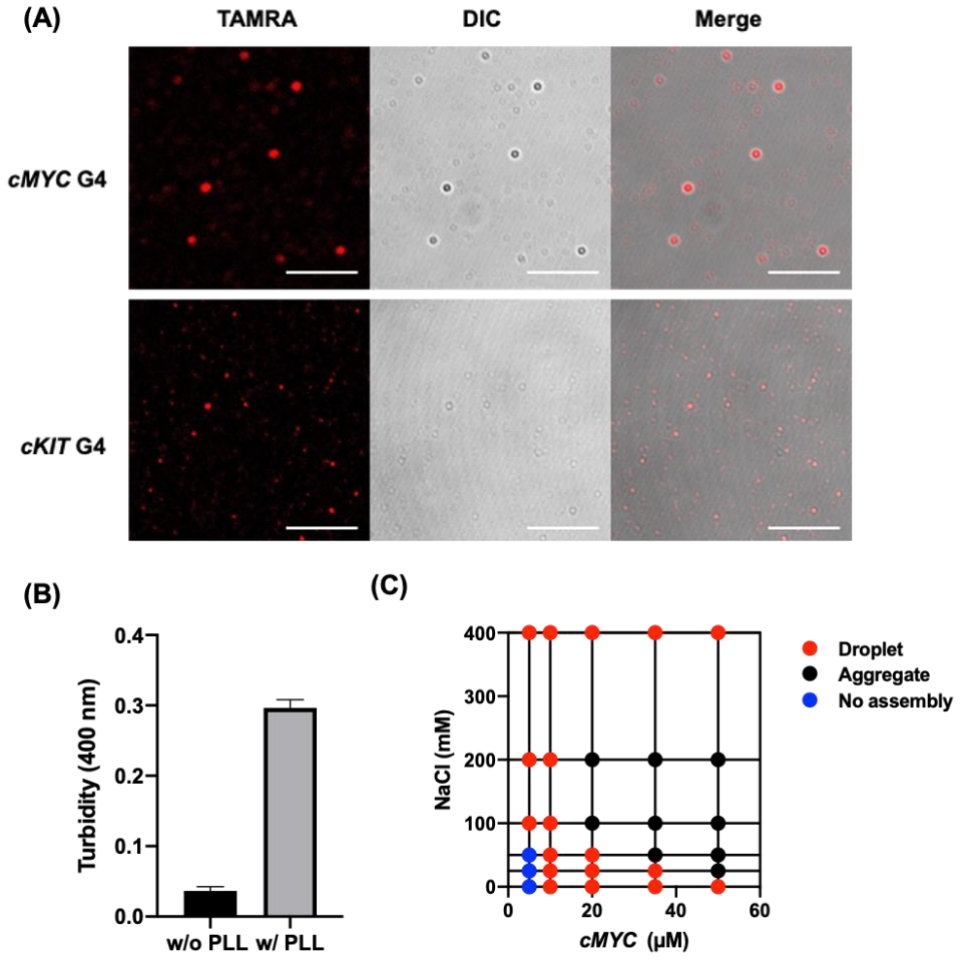

Figure 4.2. G4 DNA undergoes phase separation in vitro. (A) Images of 10 μM mixture of 5′-TAMRA-labeled and unlabeled (1:50) G4 oligodeoxynucleotides acquired from fluorescence microscopy and differential interference contrast (DIC) microscopy (scale bar = 10 μm). (B) Turbidity of 10 μM G4 DNA derived from cMYC promoter in the presence or absence of 2 μM PLL. (C) Matrix diagram showing the phase separation of different concentrations of cMYC G4 DNA in the presence of various concentrations of NaCl. Droplets are round shaped, whereas aggregates are of irregular shape (Figure S2).

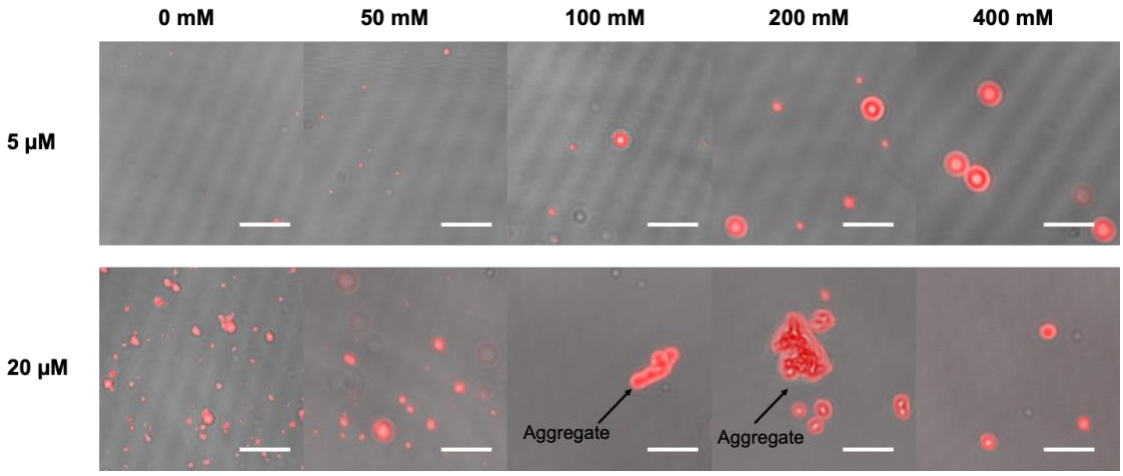

Figure 4.3: Droplet formation in solutions containing different concentrations of cMYC G4 DNA (labeled on the left) and NaCl (labeled on the top). Scale bar = 10 µm.

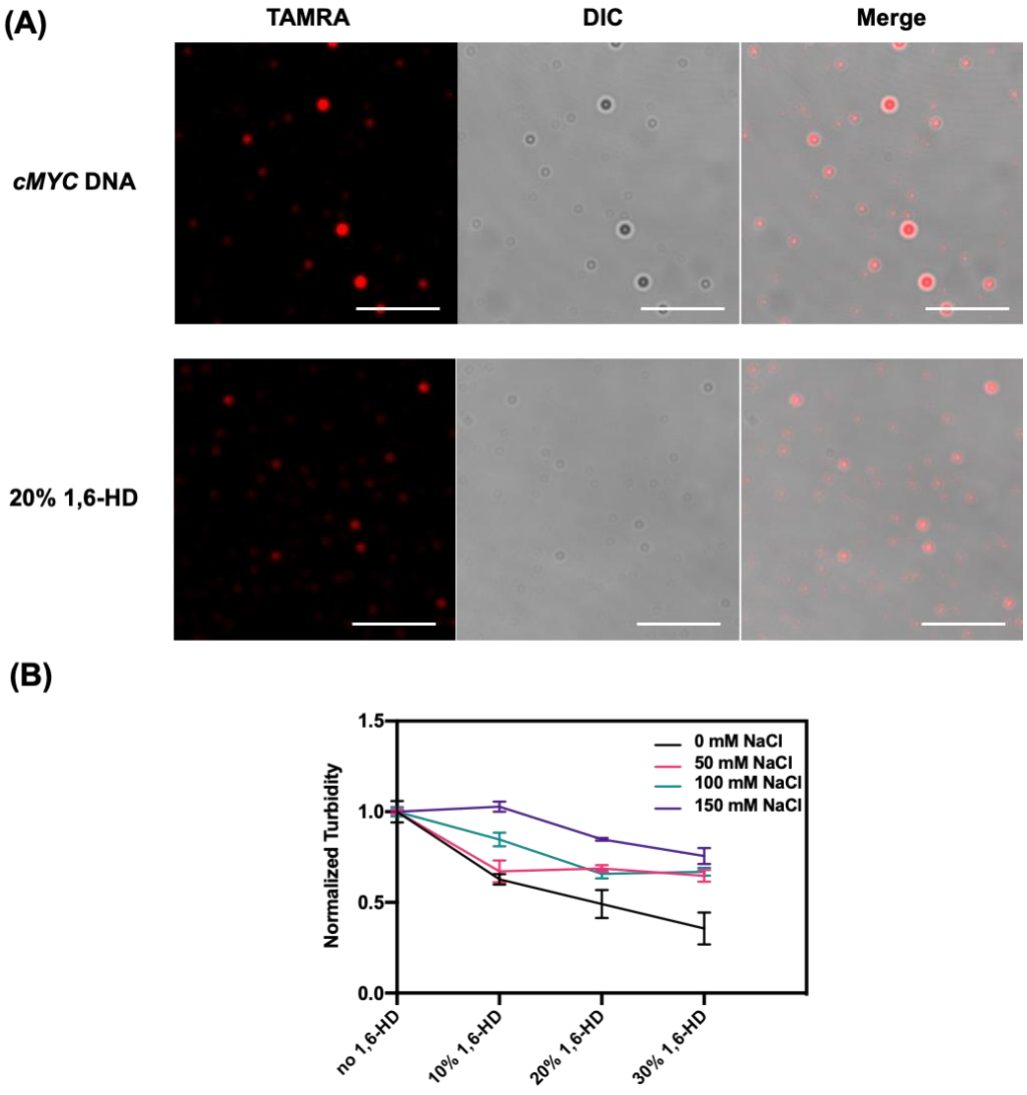

Figure 4.4. DNA G4 droplets are disrupted by 1,6-HD treatment. (A) Images of 10 μM 5′-TAMRA-labeled c*MYC* G4 DNA in a buffer containing 50 mM NaCl, and images of the corresponding samples after incubation with 20% 1,6-HD (scale bar = 10 μm). (B) Normalized turbidity of 10 μM c*MYC* DNA G4 in the presence of various concentrations of NaCl.

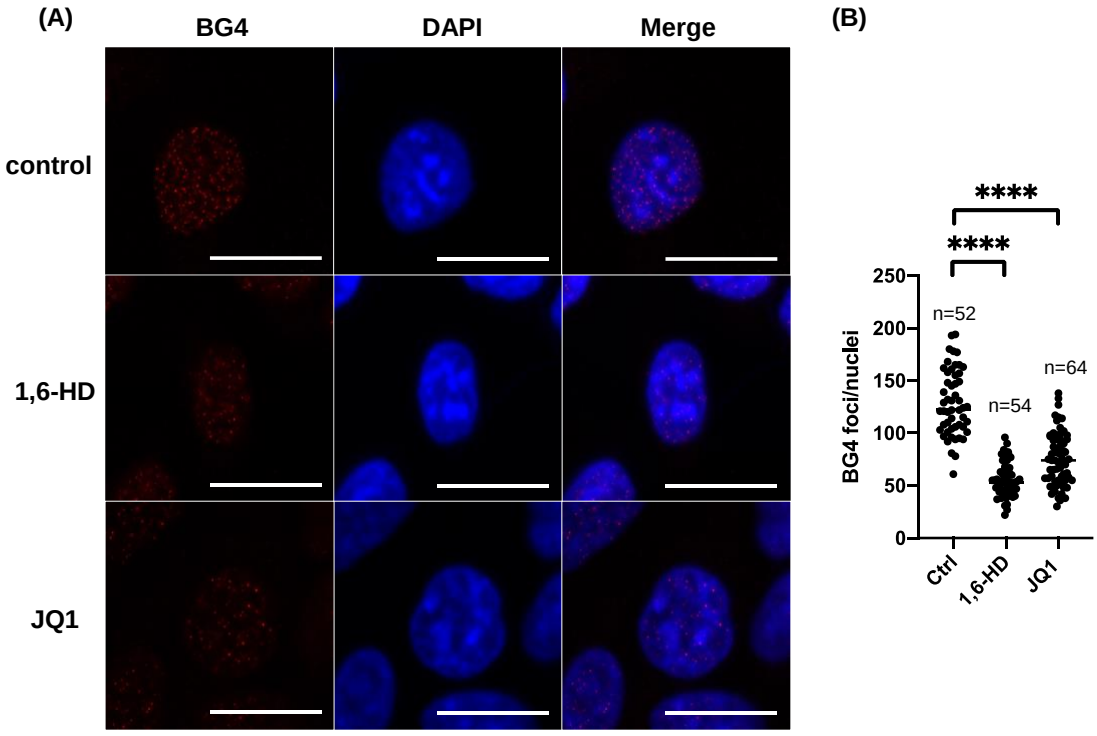

Figure 4.5. Disruption of phase separation led to diminished G4 structure foci in cells. (A) Immunofluorescence microscopy images showing the presence of G4 structure foci in control, as well as in 1,6-HD- and JQ1-treated U2OS cells. Nuclei were stained with DAPI; and G4 structures were monitored using BG4 antibody (scale bar = 20 μm). (B) Quantification of G4 structure foci after 1,6-HD and JQ1 treatment. The p values were calculated by using two-tailed, unpaired Student's t-test: ****, $p < 0.0001$.

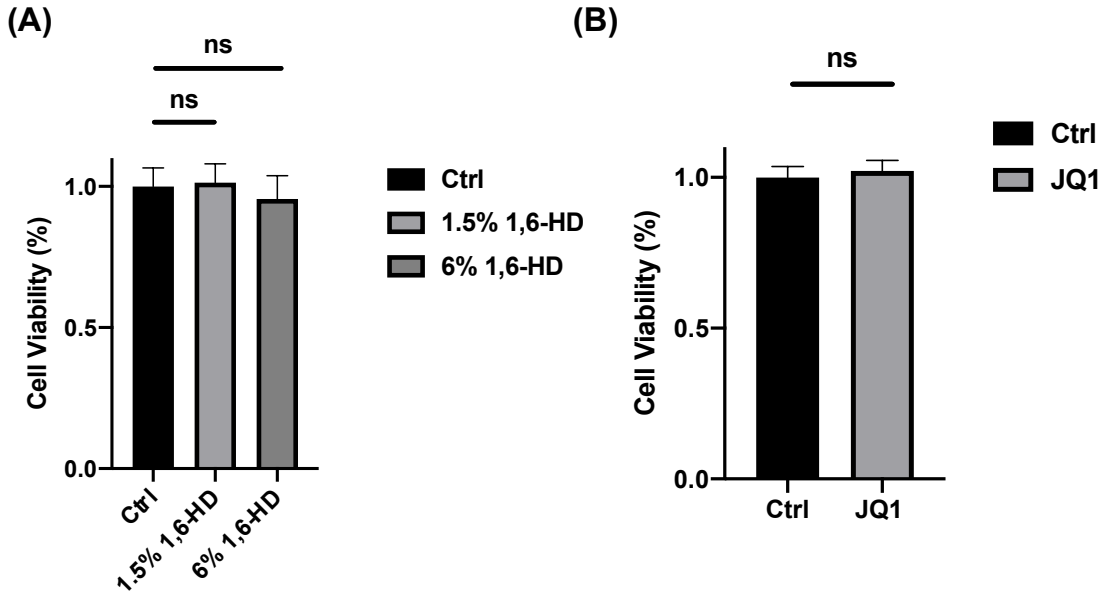

Figure 4.6. CCK8 cytotoxicity assay for assessing the effects of JQ1 and 1,6-HD treatment on survival of U2OS cells. (A) U2OS cells were treated with 1.5% 1,6-HD for 2 min and 6% 1,6-HD for 1 min. (B) Cells were treated with 1.0 μM JQ1 for 12 hr; and control cells was treated with 0.01% DMSO.

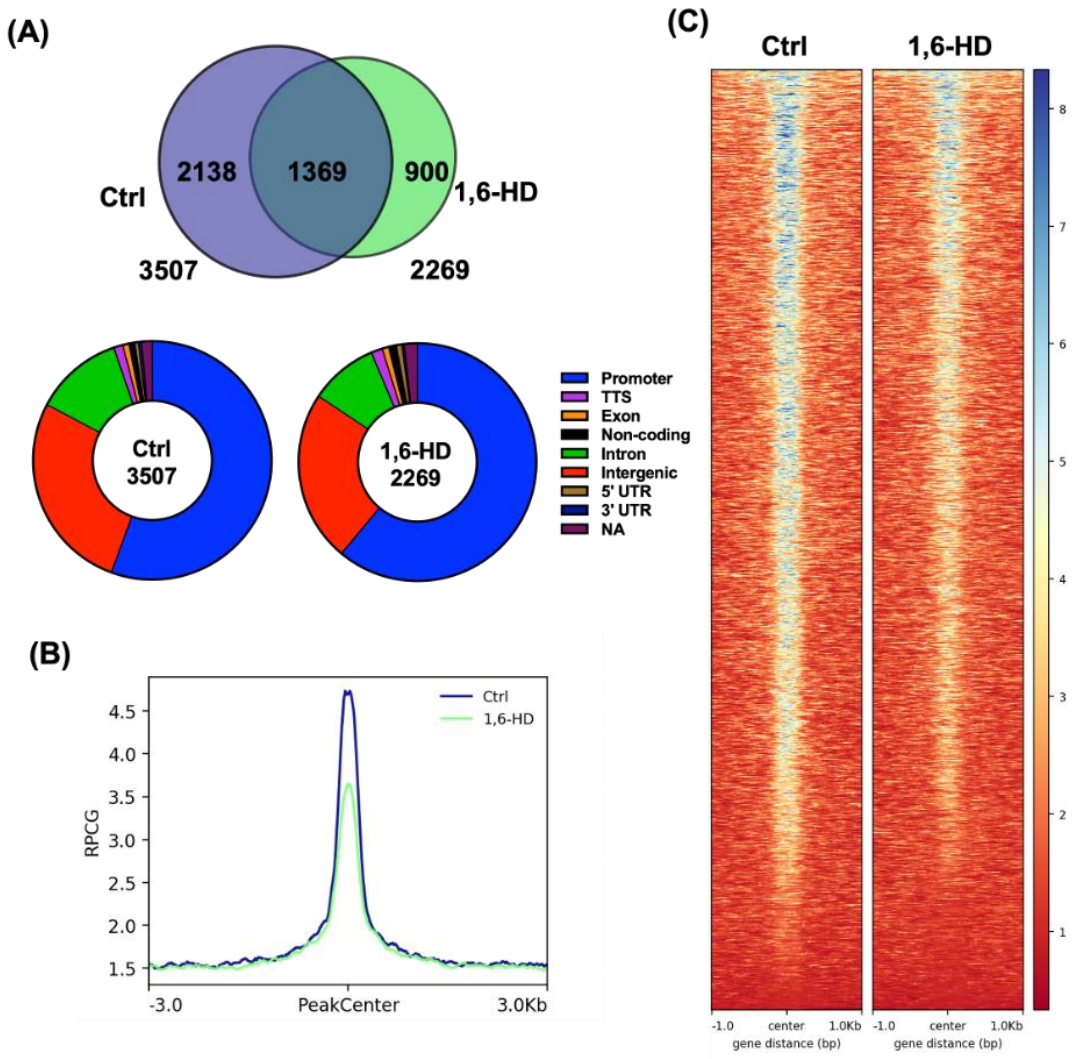

Figure 4.7. BG4 ChIP-seq results of U2OS cells with and without 1,6-HD treatment. (A) A Venn diagram depicting the numbers of significant BG4 ChIP-seq peaks in U2OS cells with (1,6-HD) or without (Ctrl) 1,6-HD treatment, and pie charts showing gene annotation of these peaks. (B) Genome-wide profiles of BG4 ChIP-seq peaks for control and 1,6-HD-treated cells showing diminished peak intensity after 1,6-HD treatment. (C) Heatmap of peak distribution and intensity in control and 1,6-HD-treated cells.

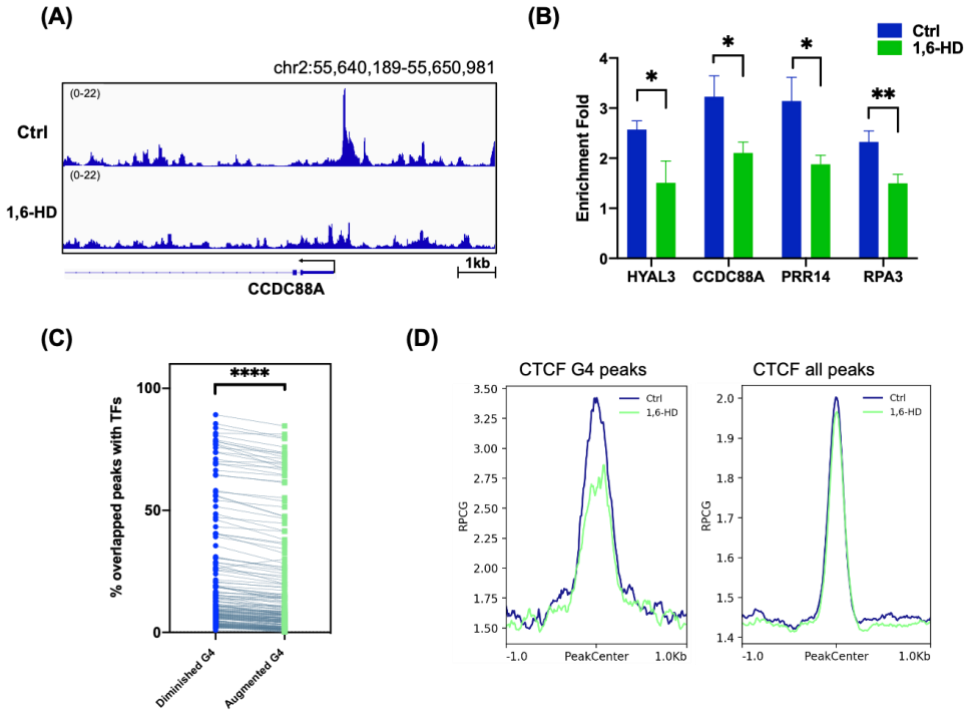

Figure 4.8. 1,6-HD treatment led to diminished G4 structures in cells. (A) ChIP-seq tracks showing the transcription start site (TSS) of CCDC88A gene in control and 1,6-HD-treated U2OS cells. (B) ChIP-qPCR results showing diminished enrichment of G4 structures in the promoter regions of *HYAL3*, *CCDC88A*, *PRR14*, and *RPA3* genes in U2OS cells upon 1,6-HD treatment. The enrichment was assessed using primers derived from the genomic regions around the TSS where the BG4 ChIP-seq peaks are located. *ESR1*, which has no G4 peak in the promoter region, served as a negative control to calculate the changes in enrichment fold. The data represent mean ± S.D. of results from three independent experiments. The *p* values were calculated by using two-tailed, unpaired Student's *t*-test: *, $p < 0.05$; **, $0.001 \leq p < 0.01$. (C) Comparison of overlapping percentage of transcription factors occupancy at G4 sites in control and 1,6-HD-treated cells. Dash lines connect the same TFs in both groups. The *p* values were calculated by using two-tailed, paired Student's *t*-test: ****, $p < 0.0001$. (D) Profiles of fold enrichment of G4 signals in control and 1,6-HD-treated cells at CTCF-binding sites with G4 structures and all CTCF binding sites, respectively.

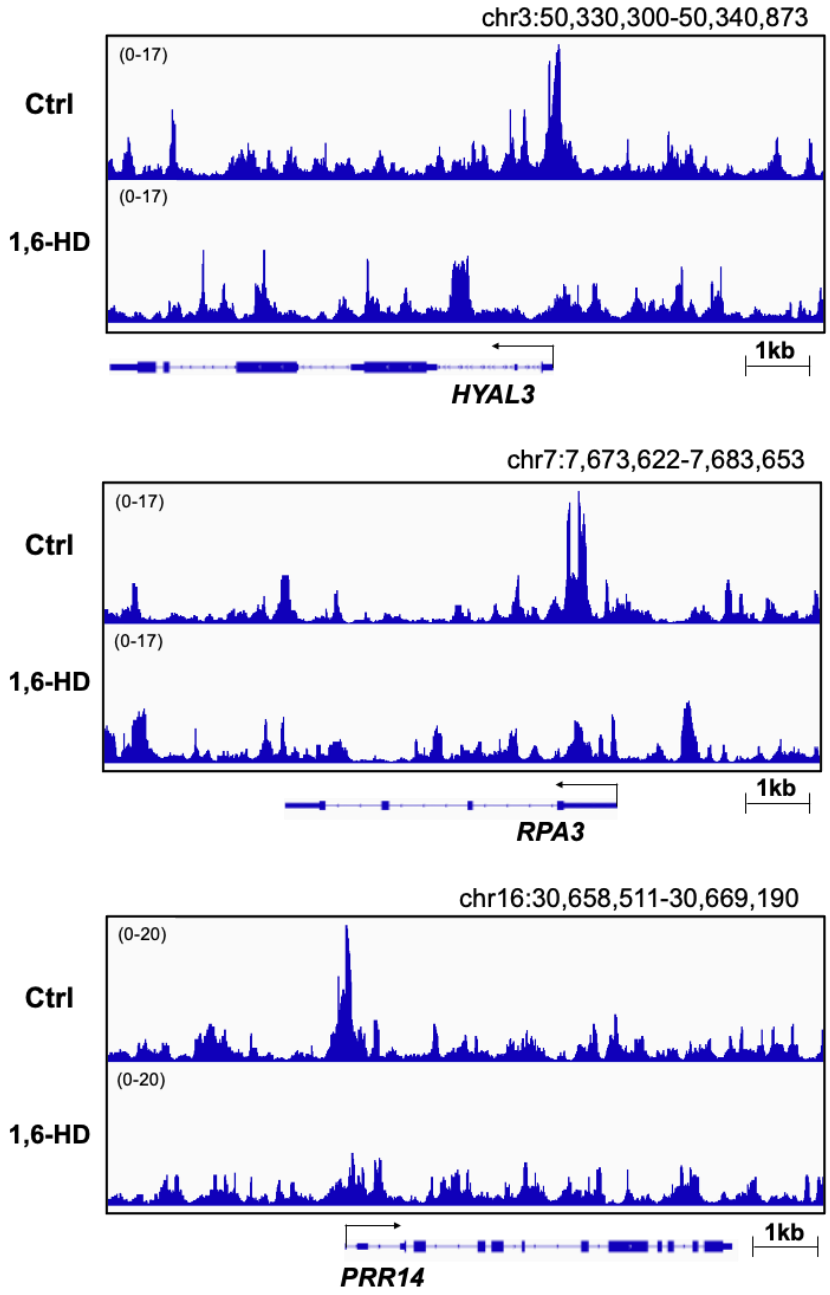

Figure 4.9: IGV plots showing the ChIP-seq signals near the TSS of HYAL3, RPA3 and PRR14 genes in control and 1,6-HD-treated U2OS cells.

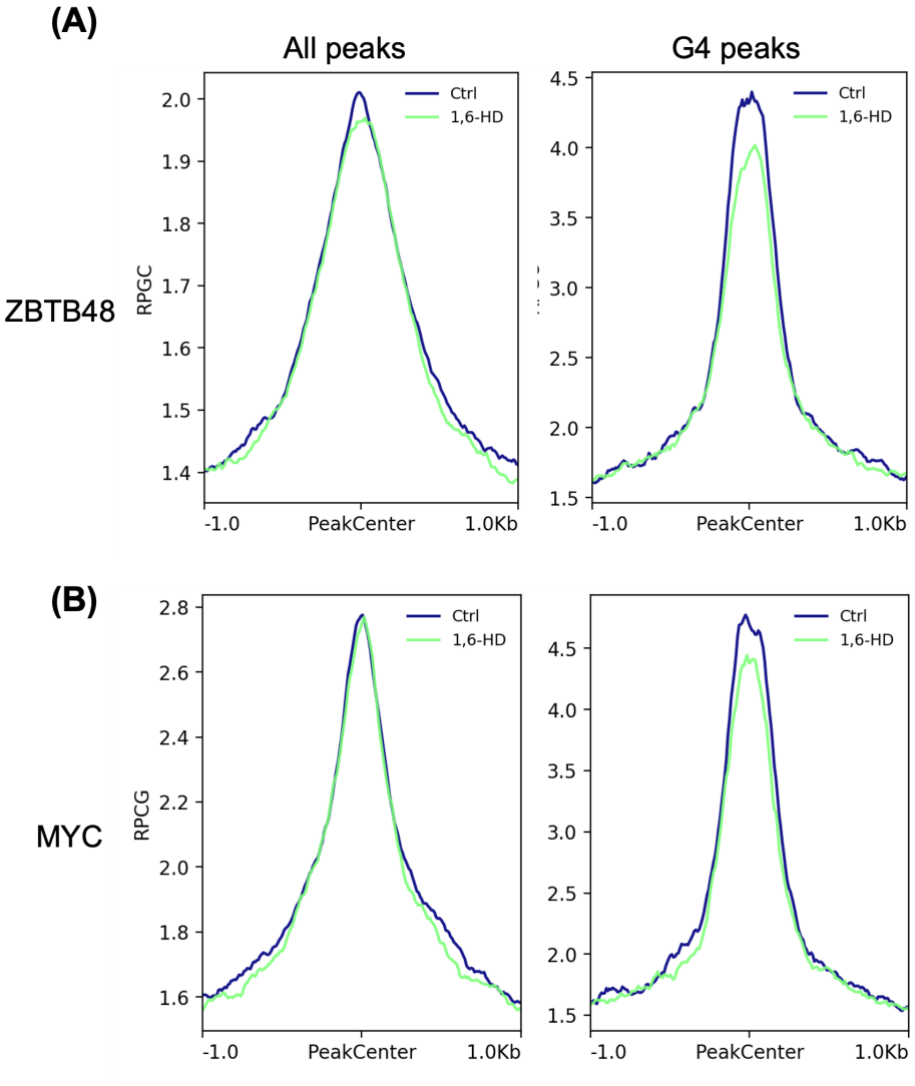

Figure 4.10: Profiles of fold enrichment of G4 signal in control and 1,6-HD-treated U2OS cells at all ZBTB48/MYC binding sites (All peaks) and ZBTB48/MYC sites with G4 structures (G4 peaks), respectively.

Name	DNA sequence	
KIT1 G4	5'-TAMRA-AGG GAG GGC GCT GGG AGG AGG G-3'	
cMYC G4	5'- TAMRA-TGA GGG TGG GGA GGG TGG GGA AGG-3'	
ChIP-qPCR primers		
Gene	Forward	Reverse
HYAL3	5'-GCC GAC TCA GTC TCT ACC CT-3'	5'-AAG CTG TGA CGC AAG GAG AA-3'
CCDC88A	5'-GAT ATC CTT CCG CCG ACT CC-3'	5'-CAG CGG TTT TCT TCT CCC AC-3'
PRR14	5'-AGC GTT GGG TAA GCT TGG TT-3'	5'-CTC TCC AGC CAC TCC TTG C-3'
RPA3	5'-CGG AAG TTG ACA GAT ACA GGG-3'	5'-GAT CGC AGA AAG GTA GTC TCA G-3'
ESR1	5'-GAA ACA GCC CCA AAT CTC AA-3'	5'-TTG TAG CCA GCA AGC AAA TG-3'

Table 4.1: A list of fluorescently labeled G4 DNA sequences and ChIP-qPCR primers used in this study.

References

(1) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **1987**, *51* (6), 899-908.

(2) Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **2005**, *33* (9), 2908-2916.

(3) Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **2005**, *33* (9), 2901-2907.

(4) Zheng, K. W.; Zhang, J. Y.; He, Y. D.; Gong, J. Y.; Wen, C. J.; Chen, J. N.; Hao, Y. H.; Zhao, Y.; Tan, Z. Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.* **2020**, *48* (20), 11706-11720.

(5) Bedrat, A.; Lacroix, L.; Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res* **2016**, *44* (4), 1746-1759.

(6) Di Antonio, M.; Ponjavic, A.; Radzevicius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L. M.; Lee, S. F.; Klenerman, D.; et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem* **2020**, *12* (9), 832-837.

(7) Hansel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48* (10), 1267-1272.

(8) Teng, Y. C.; Sundaresan, A.; O'Hara, R.; Gant, V. U.; Li, M.; Martire, S.; Warshaw, J. N.; Basu, A.; Banaszynski, L. A. ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nat. Commun.* **2021**, *12* (1), 3887.

(9) Eddy, S.; Ketkar, A.; Zafar, M. K.; Maddukuri, L.; Choi, J. Y.; Eoff, R. L. Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res.* **2014**, *42* (5), 3272-3285.

(10) Li, L.; Williams, P.; Ren, W.; Wang, M. Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **2021**, *17* (2), 161-168.

(11) Tikhonova, P.; Pavlova, I.; Isaakova, E.; Tsvetkov, V.; Bogomazova, A.; Vedekhina, T.; Luzhin, A. V.; Sultanov, R.; Severov, V.; Klimina, K.; et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. *Int. J. Mol. Sci.* **2021**, *22* (13).

(12) Wu, W. Q.; Hou, X. M.; Li, M.; Dou, S. X.; Xi, X. G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* **2015**, *43* (9), 4614-4626.

(13) Kamath-Loeb, A. S.; Loeb, L. A.; Johansson, E.; Burgers, P. M.; Fry, M. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. *J Biol Chem* **2001**, *276* (19), 16439-16446.

(14) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **2013**, *5* (3), 182-186.

(15) Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* **2022**.

(16) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L.; et al. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **2018**, *28* (6), 420-435.

(17) Ong, J. Y.; Torres, J. Z. Phase Separation in Cell Division. *Mol Cell* **2020**, *80* (1), 9-20.

(18) Shakya, A.; King, J. T. DNA Local-Flexibility-Dependent Assembly of Phase-Separated Liquid Droplets. *Biophys. J.* **2018**, *115* (10), 1840-1847.

(19) Gao, Z.; Williams, P.; Li, L.; Wang, Y. A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* **2021**, *20* (11), 4919-4924.

(20) Huang, Z. L.; Dai, J.; Luo, W. H.; Wang, X. G.; Tan, J. H.; Chen, S. B.; Huang, Z. S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J Am Chem Soc* **2018**, *140* (51), 17945-17955.

(21) Zhang, X.; Spiegel, J.; Martinez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **2021**, *13* (7), 626-633.

(22) Yagi, R.; Miyazaki, T.; Oyoshi, T. G-quadruplex binding ability of TLS/FUS depends on the beta-spiral structure of the RGG domain. *Nucleic Acids Res* **2018**, *46* (12), 5894-5901.

(23) Ghosh, M.; Singh, M. RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain. *Nucleic Acids Res* **2018**, *46* (19), 10246-10261.

(24) Wang, F.; Tang, M. L.; Zeng, Z. X.; Wu, R. Y.; Xue, Y.; Hao, Y. H.; Pang, D. W.; Zhao, Y.; Tan, Z. Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proc Natl Acad Sci U S A* **2012**, *109* (50), 20413-20418.

(25) Wang, W.; Qiao, S.; Li, G.; Cheng, J.; Yang, C.; Zhong, C.; Stovall, D. B.; Shi, J.; Teng, C.; Li, D.; et al. A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated enhancer clusters. *Nucleic Acids Res.* **2022**, *50* (9), 4917-4937.

(26) Mimura, M.; Tomita, S.; Shinkai, Y.; Hosokai, T.; Kumeta, H.; Saio, T.; Shiraki, K.; Kurita, R. Quadruplex Folding Promotes the Condensation of Linker Histones and DNAs via Liquid-Liquid Phase Separation. *J. Am. Chem. Soc.* **2021**, *143* (26), 9849-9857.

(27) Liu, X.; Xiong, Y.; Zhang, C.; Lai, R.; Liu, H.; Peng, R.; Fu, T.; Liu, Q.; Fang, X.; Mann, S.; et al. G-Quadruplex-Induced Liquid-Liquid Phase Separation in Biomimetic Protocells. *J. Am. Chem. Soc.* **2021**, *143* (29), 11036-11043.

(28) Liu, X.; Jiang, S.; Ma, L.; Qu, J.; Zhao, L.; Zhu, X.; Ding, J. Time-dependent effect of 1,6-hexanediol on biomolecular condensates and 3D chromatin organization. *Genome Biol.* **2021**, *22* (1), 230.

(29) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase separation drives heterochromatin domain formation. *Nature* **2017**, *547* (7662), 241-245.

(30) Hansel-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. *Nat Protoc* **2018**, *13* (3), 551-564.

(31) Tang, F.; Wang, Y.; Gao, Z.; Guo, S.; Wang, Y. Polymerase eta Recruits DHX9 Helicase to Promote Replication across Guanine Quadruplex Structures. *J. Am. Chem. Soc.* **2022**, *144* (31), 14016-14020.

(32) Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357-359.

(33) Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10* (2).

(34) Zhang, Y.; Liu, T.; Meyer, C. A.; Eeckhoute, J.; Johnson, D. S.; Bernstein, B. E.; Nusbaum, C.; Myers, R. M.; Brown, M.; Li, W.; et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9* (9), R137.

(35) Ramirez, F.; Ryan, D. P.; Gruning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A. S.; Heyne, S.; Dundar, F.; Manke, T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **2016**, *44* (W1), W160-165.

(36) Robinson, J. T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29* (1), 24-26.

(37) Quinlan, A. R.; Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26* (6), 841-842.

(38) Ducani, C.; Bernardinelli, G.; Hogberg, B.; Keppler, B. K.; Terenzi, A. Interplay of Three G-Quadruplex Units in the KIT Promoter. *J Am Chem Soc* **2019**, *141* (26), 10205-10213.

(39) Vieregg, J. R.; Lueckheide, M.; Marciel, A. B.; Leon, L.; Bologna, A. J.; Rivera, J. R.; Tirrell, M. V. Oligonucleotide-Peptide Complexes: Phase Control by Hybridization. *J. Am. Chem. Soc.* **2018**, *140* (5), 1632-1638.

(40) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419-434.

(41) Flock, S.; Labarbe, R.; Houssier, C. Dielectric constant and ionic strength effects on DNA precipitation. *Biophys J* **1996**, *70* (3), 1456-1465.

(42) Leng, M.; Felsenfeld, G. The preferential interactions of polylysine and polyarginine with specific base sequences in DNA. *Proc Natl Acad Sci U S A* **1966**, *56* (4), 1325-1332.

(43) Fleysher, L.; Oesingmann, N.; Brown, R.; Sodickson, D. K.; Wiggins, G. C.; Inglese, M. Noninvasive quantification of intracellular sodium in human brain using ultrahigh-field MRI. *NMR Biomed.* **2013**, *26* (1), 9-19.

(44) Daban, J. R. Physical constraints in the condensation of eukaryotic chromosomes. Local concentration of DNA versus linear packing ratio in higher order chromatin structures. *Biochemistry* **2000**, *39* (14), 3861-3866.

(45) Andre, A. A. M.; Spruijt, E. Liquid-liquid phase separation in crowded environments. *Int. J. Mol. Sci.* **2020**, *21* (16), 5908.

(46) Lin, Y.; Mori, E.; Kato, M.; Xiang, S.; Wu, L.; Kwon, I.; McKnight, S. L. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **2016**, *167* (3), 789-802 e712.

(47) Lago, S.; Nadai, M.; Cernilogar, F. M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S. N. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat.Commun.* **2021**, *12* (1), 3885.

(48) Spiegel, J.; Cuesta, S. M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **2021**, *22* (1), 117.

(49) Li, C.; Wang, H.; Yin, Z.; Fang, P.; Xiao, R.; Xiang, Y.; Wang, W.; Li, Q.; Huang, B.; Huang, J.; et al. Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res.* **2021**, *31* (9), 1546-1560.

(50) Sabari, B. R.; Dall'Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **2018**, *361* (6400).

(51) Cho, W. K.; Spille, J. H.; Hecht, M.; Lee, C.; Li, C.; Grube, V.; Cisse, II. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **2018**, *361* (6400), 412-415.

(52) Jack, A.; Kim, Y.; Strom, A. R.; Lee, D. S. W.; Williams, B.; Schaub, J. M.; Kellogg, E. H.; Finkelstein, I. J.; Ferro, L. S.; Yildiz, A.; et al. Compartmentalization of telomeres through DNA-scaffolded phase separation. *Dev. Cell* **2022**, *57* (2), 277-290 e279.

(53) Guo, Y. E.; Manteiga, J. C.; Henninger, J. E.; Sabari, B. R.; Dall'Agnese, A.; Hannett, N. M.; Spille, J. H.; Afeyan, L. K.; Zamudio, A. V.; Shrinivas, K.; et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **2019**, *572* (7770), 543-548.

Chapter 5: Targeted proteomic analysis of small GTPases in radioresistant breast cancer cells

5.1 Introduction

Small guanosine triphosphatases (small GTPases) are a superfamily of low-molecular-weight proteins that turn on their molecular functions through binding of GTP and turn off these functions through hydrolysis of the bound GTP to GDP.[1] Small GTPases are involved in many important cellular processes, including membrane trafficking, cell migration and cell cycle progression through modulating the relevant signaling pathways.[2-4] Many of them have been shown to promote cancer progression.[5-8]

In the latest report by the International Agency for Research on Cancer (IARC), breast cancer is the most commonly diagnosed cancer in the world.[9] Aside from surgery, treatment modalities for breast cancer include chemotherapy, hormone therapy and radiation.[10] Among them, radiation therapy is beneficial to cancer patients owing to its localized application and little effect to the rest of the body. More than 50% of cancer patients receive radiation therapy.[11] It is an effective way to cure and shrink the size of tumor, stop cancer recurrence, and it can be used to treat relapsed cancer. However, a significant portion of patients develop radioresistance.[12, 13] Therefore, to improve treatment efficacy and prognosis, it is important to understand the biological processes and molecular

mechanisms through which the sensitivity of cancer cells toward ionizing radiation (IR) is regulated.

An increasing number of studies revealed that small GTPases are involved in modulating radioresistance in cancer cells. For instance, down-regulation of RHOB in glioma cells reduces cancer cell survival after IR.[14] RAB27B, which is up-regulated in IR-exposed glioma cells, controls the proliferation of cancer cells through an epiregulin-mediated pathway.[15] RALA and RALB regulate colony formation, cell survival and DNA repair following IR exposure.[16] However, there is no systematic proteomic study on which small GTPases regulate radioresistance in cancer cells.

In this study, we employed an MRM-based targeted proteomic method, along with the use of stable isotope-labeled peptides, to examine the differential expression of small GTPases accompanied with the acquisition of radioresistance in two breast cancer cell lines. We identified several commonly altered small GTPases, and demonstrated that the diminished expression of one of them, i.e., ARFRP1, confers radioresistance in breast cancer cells.

5.2 Materials and Methods

Cell Culture and shRNA knockdown.

Radioresistant C5 and C6 cell lines were generated previously.[17-19] MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher) and 1% penicillin−streptomycin solution (PS, GE Healthcare). The

cells were maintained at 37°C in a humidified chamber supplemented with 5% CO_2. The shRNA stable knockdown cells were generated using pLKO.1-shRNA plasmids targeting ARFRP1 gene at the 3′-UTR and coding regions. Oligonucleotide sequences are listed in Table S1. Successful constructions of pLKO.1-shRNA plasmids were confirmed by Sanger sequencing.

Cell lysis and proteomic sample preparation.

Total protein lysates of MDA-MB-231/C5 and MCF7/C6 cells were prepared by using CelLytic M cell lysis reagent (Sigma) supplemented with a protease inhibitor cocktail (Sigma). Protein concentration was measured by using Quick Start Bradford Protein Assay (Bio-Rad). Fifty μg of total proteins in SDS-PAGE loading buffer were boiled for 10 min and loaded onto a 15% SDS-PAGE gel. The gel was subsequently stained with Coomassie Brilliant Blue R-250, destained, and proteins were digested in-gel as described previously.[20, 21] In brief, gel bands corresponding to a molecular weight range of 15−37 kDa were cut into 1 mm^3 cubes and destained sequentially with 25% and 50% CH_3CN in 50 mM ammonium bicarbonate (pH 7.8). Cysteine reduction and alkylation were performed by incubating the gel pieces in 10 mM dithiothreitol (DTT) at 37 °C for 1 h and 55 mM iodoacetamide at room temperature in the dark for 20 min, respectively. The proteins were digested with MS-grade trypsin (Pierce) at 37°C for 16 h. Peptides were eluted by shaking in a solution containing CH_3CN/H_2O/acetic acid (45/45/5, v/v) and dried by a Speed-Vac. The tryptic peptides were then desalted using C18 ZipTip (Agilent). Prior the LC-MRM analysis, each digestion mixture was spiked-in with 4 fmol each of synthetic

small GTPase peptides (New England Peptide, Inc.) with a C-terminal [$^{15}N_2$,$^{13}C_6$]-labeled lysine or [$^{15}N_4$,$^{13}C_6$]-labeled arginine.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

The MRM-based LC-MS/MS experiments were performed on a TSQ Altis triple-quadrupole mass spectrometer (Thermo Fisher) equipped with a Flex nanoelectrospray ion source (Thermo Fisher), where an UltiMate 3000 UPLC (Thermo Fisher) was employed for separation. The sample was loaded onto an in-house packed C18 (5 μm in particle size and 120 Å in pore size, Dr. Maisch GmbH HPLC) trapping column (150 μm i.d.) with buffer A, which contained 0.1% formic acid in water. The eluted peptides were loaded onto an analytical column (75 μm i.d.) packed in-house with C18 resin (3 μm in particle size and 120 Å in pore size, Dr. Maisch GmbH HPLC), using a 90-min gradient of 10–45% buffer B (80% acetonitrile in 0.1% formic acid). The peptides were ionized with a spray voltage of 2200 V, and the ion transport tube temperature was set at 325°C. The resolution of Q1 and Q3 was set at 0.7 Th full-width at half-maximum (FWHM). Fragmentation of precursor ions in Q2 was conducted with 1.5 mTorr argon. The collision energy was derived from default settings in Skyline (version 21.2).[22] The retention time of 10 tryptic peptides of BSA were monitored and used to derive the normalized retention time (iRT)–RT calibration curve and to generate the MRM method in Skyline. The mass spectrometer was scheduled to monitor the precursor to product ion transitions of 144 unique peptides of human small GTPases with a cycle time of 3 sec in a 4.5-min retention time (RT) window.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[23] via the PRIDE partner repository with the dataset identifier PXD034360.

LC-MRM data processing.

The acquired LC-MRM data were imported to Skyline. In Skyline, all peptides were filtered with dot-product (dotp) value of > 0.7, where the dotp scores were assigned by comparing the similarities of the observed relative abundances of fragment ions to those in the spectral library.[24] The potential interfering fragment ions that do not overlay with other fragment ions were manually excluded (i.e., processed data). The L/H ratios of each precursor ion were directly exported from Skyline. Detailed MRM quantification data are listed in Table S2.

Western blot

MDA-MB-231/C5 and MCF-7/C6 pairs of breast cancer cells were lysed with CelLytic M cell lysis reagent (Sigma) supplemented with 1% protease inhibitor cocktail, and denatured at 95 °C for 5-min in Laemmli loading buffer. Protein concentration was measured by Quick Start Bradford Protein Assay (Bio-Rad). Thirty μg proteins from the denatured lysates were separated using 15% SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was blocked with 5% milk in PBS-T (PBS with 0.1% Tween 20) for 1 hr, and then incubated with the corresponding primary antibodies, including human ARFRP1 (Proteintech, 17712-1-AP, 1:800 dilution), IFT27 (Proteintech, 17712-1-AP, 1:500 dilution) and anti-tubulin (Santa Cruz, SC-32293, 1:5000). The

secondary antibodies were donkey anti-rabbit secondary antibody (Sigma, A0545, 1:5000), or anti-mouse secondary antibody (Santa Cruz, m-IgGκ BP-HRP, 1:5000).

Total RNA extraction and real-time quantitative-PCR (RT-qPCR)

Total RNA was extracted with Total RNA Kit I (Omega) and purified with HiBind RNA mini columns (VWR). Two μg of total RNA was mixed with M-MLV Reverse Transcriptase (Promega, Madison, WI, USA) for cDNA synthesis. RT-qPCR was conducted as previously described,[25] with the use of Luna® Universal qPCR Master Mix (NEB) on a CFX96 RT-qPCR detection system (Bio-Rad).

Clonogenic Survival Assay

Clonogenic survival assay was performed as described previously.[26, 27] Briefly, control shRNA- and shARFRP1-treated MCF-7 and MDA-MB-231 cells were plated in triplicate in six-well plates at a concentration of 300 cells per well for 0 and 1.5 Gy treatment, and 600 cells per well for 3 and 5 Gy treatment. X-rays were delivered using a Rad Source RS-2000 cabinet irradiator (Rad Source Technologies, Buford, GA) followed by a 10-day incubation. Cell colonies were fixed and stained in an aqueous solution containing 6.0% glutaraldehyde and 0.5% crystal violet. The plates were then rinsed with water and dried at room temperature in air. Those colonies with at least 50 cells were counted. The survival fraction (SF) was calculated by equation:

$$SF = \frac{\text{no. of colonies formed for IR-treated cells}}{\text{no. of cells seeded for IR treatment} * \frac{\text{no. of colonies formed for control untreated cells}}{\text{no. of control untreated cells seeded}}}$$

5.3 Results

MRM-based Quantitative Profiling of Small GTPases in Modulating Radioresistance of Breast Cancer Cells

In this study, we compared the expression levels of small GTPases in two lines of breast cancers cells (i.e., MDA-MB-231 and MCF7) and their corresponding radioresistant C5 and C6 clones[17, 19] by applying a previously developed scheduled multiple-reaction monitoring (MRM)-based targeted quantitative proteomic approach, together with the use of synthetic stable isotope-labeled peptides as internal standards (Figure 5.1).[28, 29] We observed a number of small GTPases altered in radioresistant breast cancer cells relative to parental cells, including RHOB and RALB, which are known regulators of radiation sensitivity.[14, 16, 30] We also discovered novel candidate small GTPase regulators of radiation sensitivity in breast cancer cells, which provides potential targets for improving the efficacy of cancer radiotherapy.

To achieve high-throughput analysis of small GTPase proteins in breast cancer cells, we employed a previously developed scheduled LC-MRM method.[28, 29] In this regard, our MRM library consists of 144 unique peptides derived from 144 non-redundant small GTPases, representing 86% of the human small GTPase proteome with a total of 167 known proteins.[31] For MRM analysis, we chose three most abundant fragment ions observed in MS/MS acquired from shotgun proteomic analyses for each peptide. We conducted the experiments in three biological replicates, where samples from each replicate were analyzed by LC-MS/MS twice, and processed the data using Skyline.[22]

The LC-MRM data allowed us to quantify 82 and 68 proteins in the MCF7/C6 and MDA-MB-231/C5 pairs, respectively, with 62 small GTPases being quantified in both pairs (Figure 5.1). For some small GTPase peptides, we were able to detect the spiked-in heavy peptides, but not the respective endogenous peptides, which might be attributed to low levels of expression of the corresponding GTPase proteins in these breast cancer cells. On the other hand, post-translational modifications may shift molecular weights of some small GTPases out of the 15-37 kDa range (the portion of the gel that we employed for in-gel tryptic digestion and subsequent LC-MS/MS analysis), and/or introduce mass shifts of the tryptic peptides monitored in MRM, which may also contribute to failure in detecting some of the peptides.

We also performed hierarchical clustering analysis with Euclidean metric applied to the distance measurement (Figure 5.1). Such analysis illustrates the similarities and differences in the quantified levels of small GTPases in the radioresistant clones over the corresponding parental breast cancer cell lines. Many small GTPase proteins were commonly up- or down-regulated in the two radioresistant/parental pairs of breast cancer cells, while some exhibit different trends in the two pairs. This is not surprising on the grounds that MCF7 and MDA-MB-231 cells were derived from different breast cancer patients, where genetic heterogeneity may also contribute to differences in expression levels of small GTPases accompanied with the acquisition of radioresistance.

Validation of Differential Expression of Small GTPases in Parental/Radioresistant Breast Cancer Cells

We further categorized the quantified small GTPase proteins by imposing a cut-off of at least 1.5-fold difference in expression levels, which yielded 29 and 38 substantially changed proteins in MCF7/C6 and MDA-MB-231/C5 pairs, respectively (Figure 5.2). Among them, 7 proteins, including the known radioresistant regulators RALB and RHOB, were found to be commonly altered in both pairs with at least 1.5-fold changes (Figure 5.2). We further validated the differential expression of ARFRP1 and IFT27 proteins by Western blot analyses (Figures 5.3 and 5.4), suggesting the quantification accuracy of the MRM method.

We also compared our MRM quantification results for IFT27 and ARFRP1 proteins with their mRNA levels in parental breast cancer cells and the corresponding radioresistant cell lines (Figure 5.3 and 5.4). Consistent with the MRM and Western blot data, we observed higher levels of IFT27 mRNA and lower levels of ARFRP1 mRNA in C5 and C6 cell lines than the corresponding parental lines (Figure 5.5). These results suggest that the differential expression of these two small GTPase proteins arise from transcriptional regulation.

ARFRP1(ADP-ribosylation factor-related protein 1) was shown to be involved in trans-Golgi network through regulating ARL1-mediated Golgi recruitment.[32, 33] It was also found to be important for lipidation and assembly of lipoproteins.[34, 35] However, there are not many studies about its role in cancer.[36] Our MRM results prompted us to hypothesize that down-regulation of ARFRP1 may confer radioresistance in breast cancer cells.

ARFRP1 Knockdown Led to Increased Radioresistance in Breast Cancer Cells

To assess the role of ARFRP1 in modulating radiation sensitivity, we generated MCF7 and MDA-MB-231 cells with the *ARFRP1* gene being stably knocked down using shRNA. The knockdown efficiency was assessed by Western blot analyses (Figures 5.6 and 5.7). We then examined the cell survival rate after X-ray exposure.[26] The results from clonogenic survival assay showed pronounced increases in radioresistance after substantial knockdown of *ARFRP1* gene in both breast cancer cell lines (Figures 5.6 and 5.7). In particular, shARFRP1-1 and shARFRP1-3 led to 79.6% and 93.9% depletions, respectively, of ARFRP1 protein in MDA-MB-231 cells (Figure 5.6), and shARFRP1-2 and shARFRP1-3 led to 51.0% and 85.4% losses, respectively, of ARFRP1 protein in MCF7 cells (Figure 5.7). Our clonogenic survival assay results showed that shARFRP1-3 conferred a more pronounced elevation in radioresistance than shARFRP1-1 and shARFRP1-2, which is in agreement with the relative knockdown efficiencies of the three different sequences of shRNAs (Figures 5.6 and 5.7).

Using mRNA expression data of 25 breast carcinoma cell lines in the Cancer Cell Line Encyclopedia (CCLE) database, we compared the mRNA expression level of ARFRP1 to those of two known radioresistance regulators, RALB[16] and RAC1[37, 38], where RALB was also shown by our MRM results to be differentially expressed in the radioresistant/parental breast cancer cells (Figure 5.2). We found that the mRNA expression levels of *RALB* and *RAC1* genes are positively correlated with the mRNA levels of *ARFRP1* gene across the 25 breast cancer cell lines (Figure 5.8), further substantiating the role of ARFRP1 in modulating radioresistance.

5.4 Conclusion

In summary, we applied high-throughput scheduled LC-MRM analysis to explore the alterations in expression levels of small GTPase proteins accompanied with the development of radioresistance in breast cancer cells and to identify potential regulators of radioresistance. We were able to quantify 82 and 68 proteins in the MDA-MB-231/C5 and MCF-7/C6 pairs of breast cancer cells, respectively. Western blot analysis validated the MRM quantification results for two small GTPases, underscoring the quantification accuracy of the LC-MRM method. The LC-MRM analysis led to the discovery of 7 small GTPases that are commonly altered by at least 1.5-fold in the two pairs of breast cancer cells. These included two known radioresistance modulators (RHOB and RALB) and several novel candidate radioresistance regulators. In particular, we demonstrated, for the first time, that ARFRP1 is a regulator of radiation sensitivity, where its down-regulation in breast cancer cells conferred augmented radioresistance. Thus, our study also provides a new target for overcoming resistance in cancer radiotherapy and a novel protein biomarker for selecting radiotherapy for patients. In this regard, it is worth discussing a limitation of our work. In particular, owing to the limited availabilities of radioresistant breast cancer cells, we employed only two radioresistant breast cancer cell lines and the corresponding parental cell lines in the current study. It would be important to explore, in the future, if the findings made from these two cell lines can be expanded to other radioresistant breast cancer cell lines and radioresistant breast cancer tissues. Along this line, the quantitative proteomic method described in this study is also amenable to the quantification of small GTPase proteins in tissues. [28]

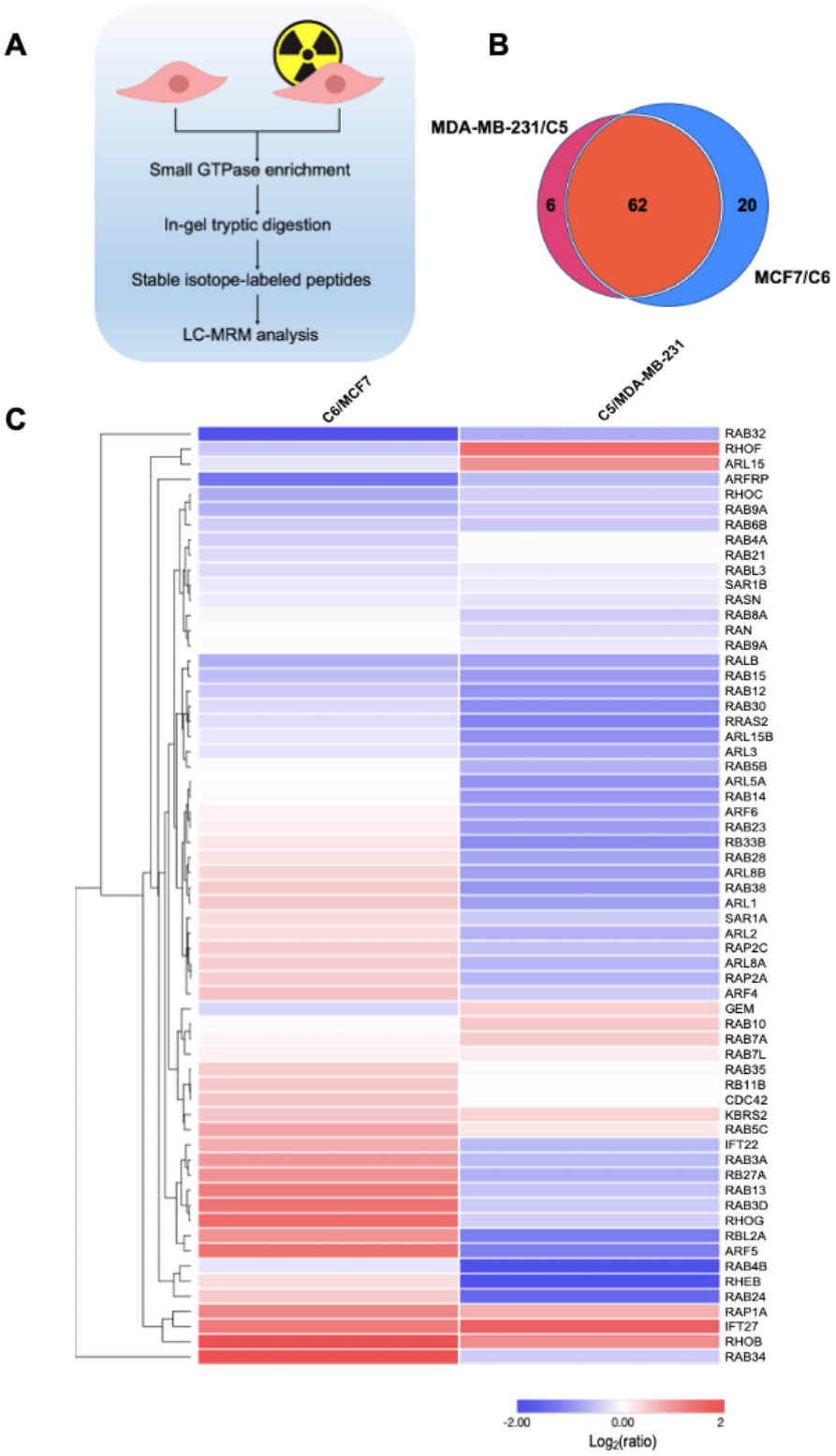
A
Small GTPase enrichment
In-gel tryptic digestion
Stable isotope-labeled peptides
LC-MRM analysis
B
MDA-MB-231/C5
MCF7/C6
6
62
20
C
C6/MCF7
C5/MDA-MB-231
RAB32
RHOF
ARL15
ARFRP
RHOC
RAB9A
RAB6B
RAB4A
RAB21
RABL3
SAR1B
RASN
RAB8A
RAN
RAB9A
RALB
RAB15
RAB12
RAB30
RRAS2
ARL15B
ARL3
RAB5B
ARL5A
RAB14
ARF6
RAB23
RB33B
RAB28
ARL8B
RAB38
ARL1
SAR1A
ARL2
RAP2C
ARL8A
RAP2A
ARF4
GEM
RAB10
RAB7A
RAB7L
RAB35
RB11B
CDC42
KBRS2
RAB5C
IFT22
RAB3A
RB27A
RAB13
RAB3D
RHOG
RBL2A
ARF5
RAB4B
RHEB
RAB24
RAP1A
IFT27
RHOB
RAB34
-2.00
0.00
2
Log2(ratio)

Figure 5.1: A scheduled LC-MRM method for interrogating the differential expression of small GTPases in breast cancer cells and the corresponding radio-resistant cell lines. (A) A schematic diagram illustrating the workflow of the LC-MRM method for discovering small GTPases that modulate radioresistance. (B) A Venn diagram showing the numbers of quantified small GTPases in MDA- MB-231/C5 and MCF7/C6 pairs of breast cancer cells. (C) Hierarchical clustering displaying the Log_2-transformed expression fold changes of small GTPases in radioresistant C5 and C6 cells relative to the corresponding parental MDA-MB-231 and MCF7 cells. Hierarchical clustering was generated using Perseus, where red and blue boxes designate proteins up- and down-regulated, respectively, in radioresistant breast cancer cells relative to the corresponding parental lines. Genes were clustered using Euclidean distance.

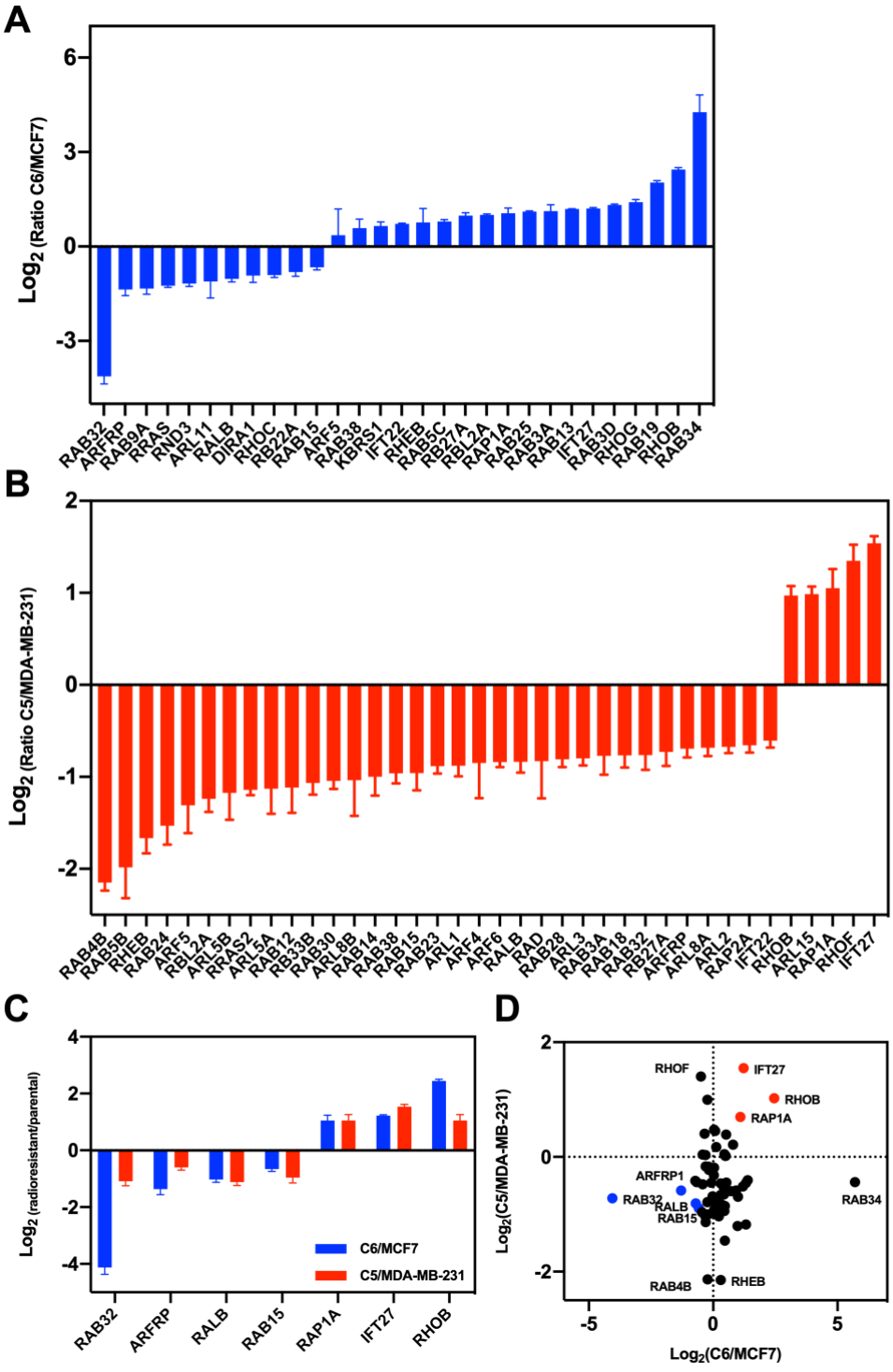
A
Log2 (Ratio C6/MCF7)
6
3
0
-3
RAB32
ARFRP
RAB9A
RRAS
RND3
ARL11
RALB
DIRA1
RHOC
RB22A
RAB15
ARF5
RAB38
KBRS1
IFT22
RHEB
RAB5C
RB27A
RBL2A
RAP1A
RAB25
RAB3A
RAB13
IFT27
RAB3D
RHOG
RAB19
RHOB
RAB34
B
Log2 (Ratio C5/MDA-MB-231)
2
1
0
-1
-2
RAB4B
RAB5B
RHEB
RAB24
ARF5
RBL2A
ARL5B
RRAS2
ARL5A
RAB12
RB33B
RAB30
ARL8B
RAB14
RAB38
RAB15
RAB23
ARL1
ARF4
ARF6
RALB
RAD
RAB28
ARL3
RAB3A
RAB18
RAB32
RB27A
ARFRP
ARL8A
ARL2
RAP2A
IFT22
RHOB
ARL15
RAP1A
RHOF
IFT27
C
Log2 (radioresistant/parental)
4
2
0
-2
-4
C6/MCF7
C5/MDA-MB-231
RAB32
ARFRP
RALB
RAB15
RAP1A
IFT27
RHOB
D
Log2(C5/MDA-MB-231)
2
0
-2
RHOF
IFT27
RHOB
RAP1A
ARFRP1
RAB32
RALB
RAB15
RAB34
RAB4B
RHEB
-5
0
5
Log2(C6/MCF7)

Figure 5.2. Quantification of differential expression of small GTPases in MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells. (A, B) Bar graphs showing the MRM quantification results of small GTPases with differences in expression being over 1.5-fold in radioresistant/parental pairs. (C) A bar graph illustrating the proteins commonly altered in the two pairs of radioresistant/parental breast cancer cell lines. Relative expression level is plotted as Log_2 ratio of radioresistant/parental cells. (D) A scatter plot illustrating significantly up- and down-regulated small GTPases in the two pairs of breast cancer cells. Seven commonly altered small GTPases are highlighted in red (for up-regulated small GTPases in radioresistant cell lines) and blue (for down-regulated small GTPases).

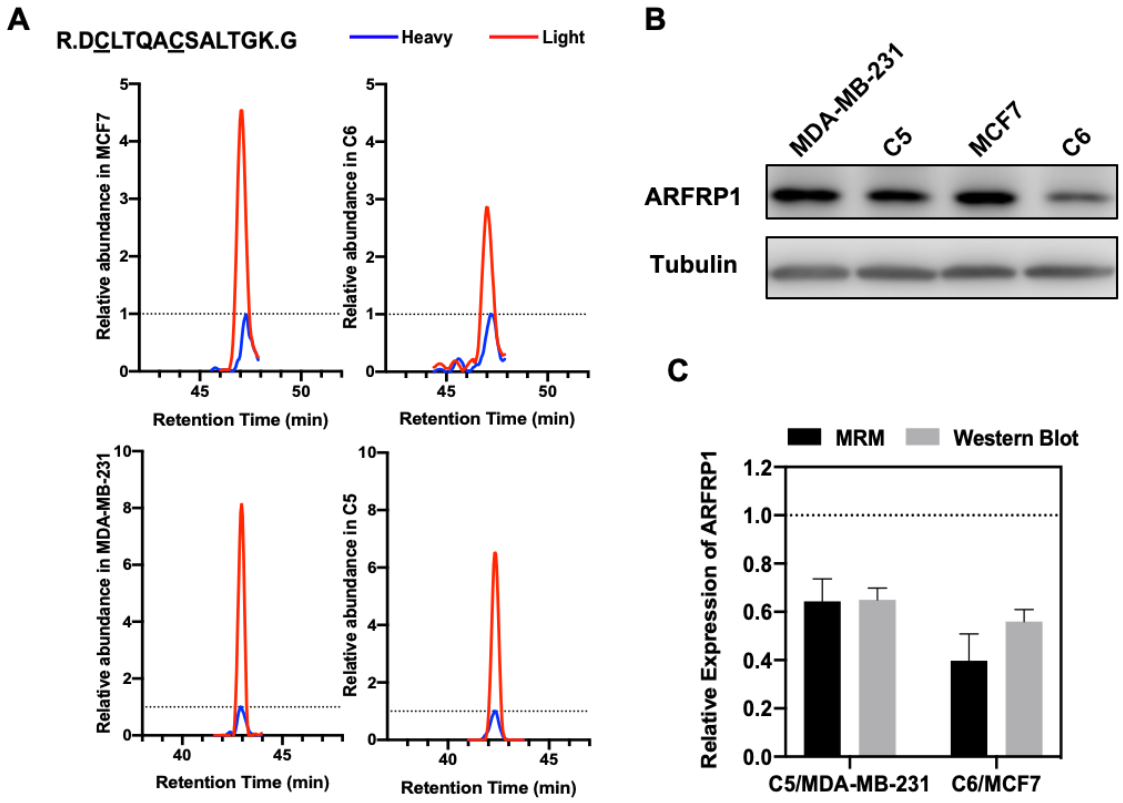

Figure 5.3. ARFRP1 is down-regulated in radioresistant lines of both MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells. (A) The MRM traces of a representative peptide (DCLTQACSALTGK, where underlined C represents carbamidomethylated cysteine) from ARFRP1 in MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells. The traces of the unlabeled peptide in parental and radioresistant cell lines are shown in red, and the spiked-in isotope-labeled peptide are depicted in blue. (B) Western blot for validating the MRM results for ARFRP1 in MDA-MB-231/C5 and MCF7/C6 cells. (C) Quantification results for the relative levels of expression of ARFRP1 protein in the two pairs of cell lines obtained from MRM and Western blot analyses. Error bars represent S.D. (n=3).

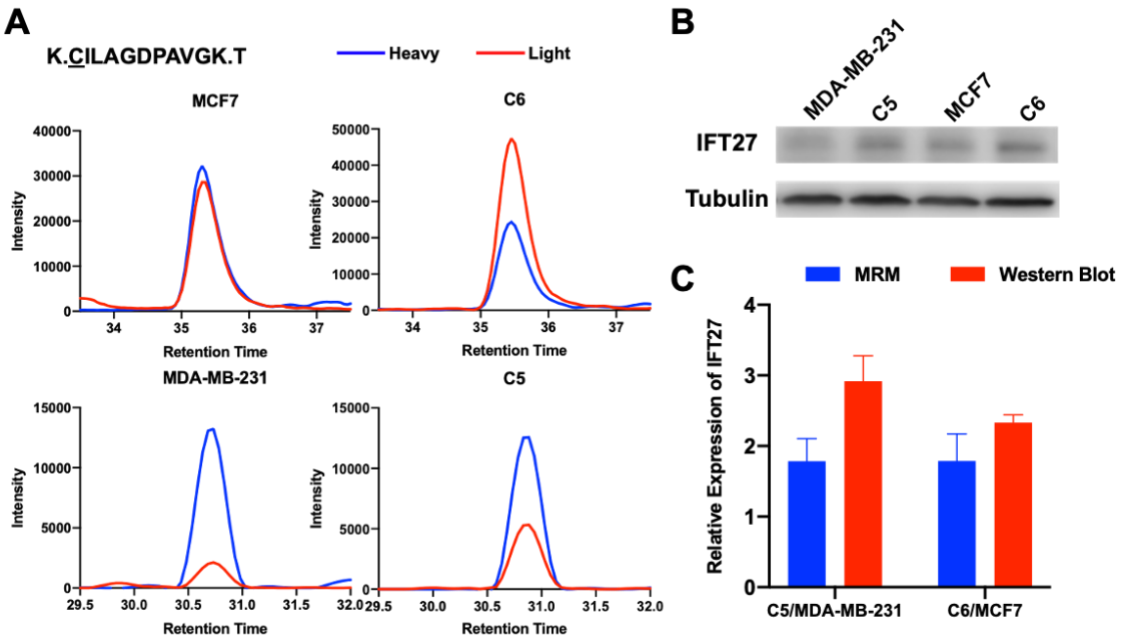

Figure 5.4. IFT27 is up-regulated in the radio-resistant lines of MDA-MB-231 and MCF7 cells. (A) MRM traces of a representative peptide (CILAGDPAVGK), where the underlined C represents carbamidomethylated cysteine, from IFT27 in MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells. (B) Western blot for validating MRM results for IFT27 in MDA-MB-231/C5 and MCF7/C6 pairs of cells. (C) Relative levels of expression of IFT27 protein in the two pairs of cell lines as obtained from MRM and Western blot analyses. Error bars represent S.D. (n=3).

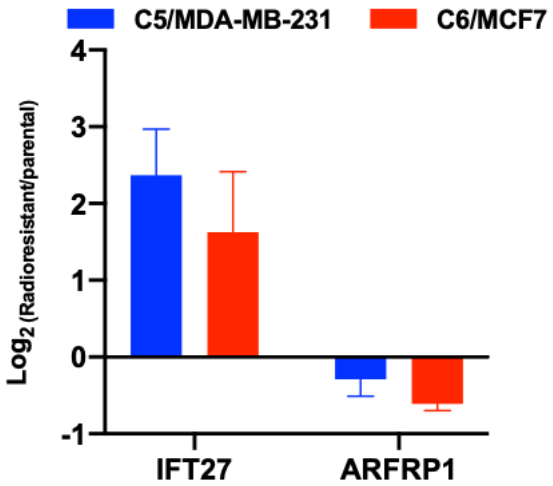

Figure 5.5. Real-time qPCR showing the mRNA expression levels of IFT27 and ARFRP1 genes in MDA-MB-231/C5 and MCF7/C6 pairs of breast cancer cells. Shown are Log_2 ratio of expression levels in radioresistant over parental cell lines. The data represent the mean ± S.D. of results obtained from three biological replicates.

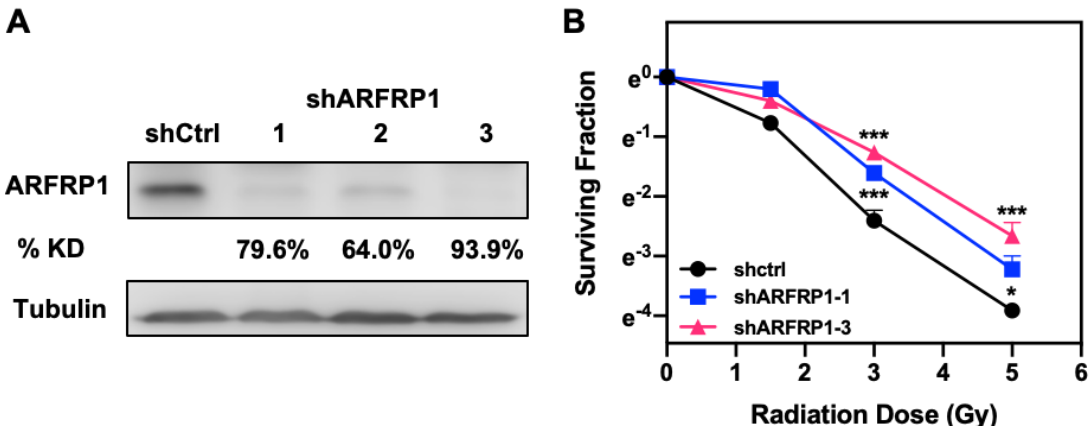

Figure 5.6. ARFRP1 modulates radioresistance in breast cancer cells. (A) Western blot for validating the knockdown efficiency of ARFRP1 in MDA-MB-231 cells. (B) Survival rates of MDA-MB-231 cells treated with control or ARFRP1 shRNA and exposed with the indicated doses of X rays. Error bars represent S.D. (n = 3). The *p* values were calculated using two-tailed, unpaired Student's *t*-test: *, $p < 0.05$; ***, $p < 0.001$.

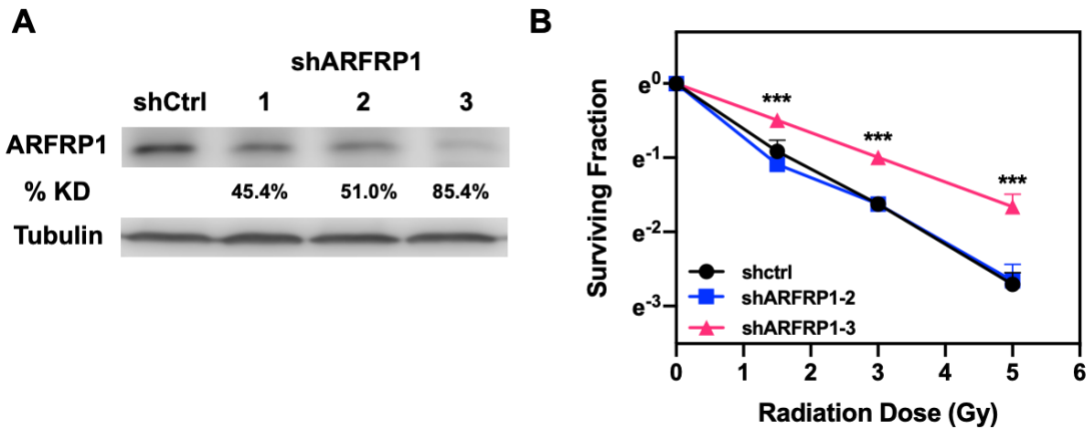

Figure 5.7. Down-regulation of ARFRP1 led to elevated radioresistance in MCF7 cells. (A) Validation of knockdown efficiency of ARFRP1 in MCF7 cells by Western blot analysis. (B) Survival rates of MCF7 cells treated with control or ARFRP1 shRNA and exposed with the indicated doses of γ rays. Error bars represent S.D. of results from three independent experiments. p values were calculated using unpaired, two-tailed Student's t-test: ***, $p < 0.001$.

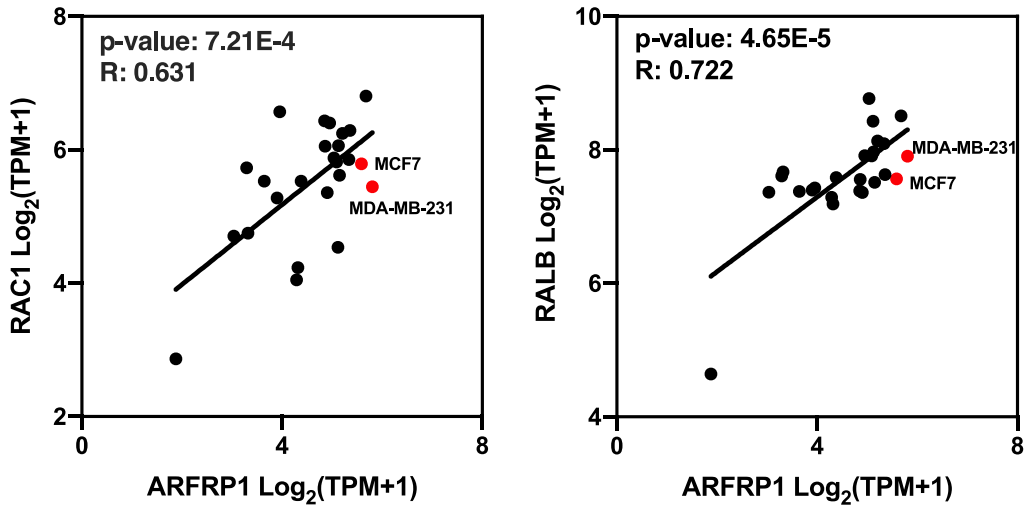

Figure 5.8. Scatter plots showing the correlation between mRNA expression level of *ARFRP1* gene and those of *RAC1*/*RALB* genes, which have been reported to be involved in regualting radioresistance in different breast cancer cell lines, including MDA-MB-231 and MCF7 cells highlighted in red.

Gene		**Oligonucleotides**
ARFRP1	shRNA-1	CGAAGACAAACTTTCCTCTAT
	shRNA-2	CCTCTCAATCCCTGACATCAA
	shRNA-3	GCAGTCTTTGTGGGACAAGTA
Genes	**qPCR Forward primers**	**qPCR Reverse primers**
ARFRP1	GTACAAGTACATGTTTCAGA	TCCTGGGCCTGGACAATGCT
IFT27	GCTCTTGTCCTCTGGGTACG	GTCTCCCGTGTCAGGAACTG

Table 5.1: Primers and Oligonucleotide Sequences

References

(1) Cherfils, J.; Zeghouf, M. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* **2013**, *93* (1), 269-309.

(2) Sancak, Y.; Peterson, T. R.; Shaul, Y. D.; Lindquist, R. A.; Thoreen, C. C.; Bar-Peled, L.; Sabatini, D. M. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **2008**, *320* (5882), 1496-1501.

(3) Stenmark, H. Rab GTPases as coordinators of vesicle traffic. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (8), 513-525.

(4) Guan, X.; Guan, X.; Dong, C.; Jiao, Z. Rho GTPases and related signaling complexes in cell migration and invasion. *Exp. Cell Res.* **2020**, *388* (1), 111824.

(5) Haga, R. B.; Ridley, A. J. Rho GTPases: Regulation and roles in cancer cell biology. *Small GTPases* **2016**, *7* (4), 207-221.

(6) Subramani, D.; Alahari, S. K. Integrin-mediated function of Rab GTPases in cancer progression. *Mol. Cancer* **2010**, *9*, 312.

(7) Vega, F. M.; Ridley, A. J. Rho GTPases in cancer cell biology. *FEBS Lett.* **2008**, *582* (14), 2093-2101.

(8) Cheng, K. W.; Lahad, J. P.; Gray, J. W.; Mills, G. B. Emerging role of RAB GTPases in cancer and human disease. *Cancer Res* **2005**, *65* (7), 2516-2519.

(9) Dyba, T.; Randi, G.; Bray, F.; Martos, C.; Giusti, F.; Nicholson, N.; Gavin, A.; Flego, M.; Neamtiu, L.; Dimitrova, N.; et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *Eur J Cancer* **2021**, *157*, 308-347.

(10) Loibl, S.; Poortmans, P.; Morrow, M.; Denkert, C.; Curigliano, G. Breast cancer. *Lancet* **2021**, *397* (10286), 1750-1769.

(11) Baskar, R.; Lee, K. A.; Yeo, R.; Yeoh, K. W. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci* **2012**, *9* (3), 193-199.

(12) Zhang, H.; Luo, H.; Jiang, Z.; Yue, J.; Hou, Q.; Xie, R.; Wu, S. Fractionated irradiation-induced EMT-like phenotype conferred radioresistance in esophageal squamous cell carcinoma. *J. Radiat. Res.* **2016**, *57* (4), 370-380.

(13) Koto, M.; Miyamoto, T.; Yamamoto, N.; Nishimura, H.; Yamada, S.; Tsujii, H. Local control and recurrence of stage I non-small cell lung cancer after carbon ion radiotherapy. *Radiother. Oncol.* **2004**, *71* (2), 147-156.

(14) Monferran, S.; Skuli, N.; Delmas, C.; Favre, G.; Bonnet, J.; Cohen-Jonathan-Moyal, E.; Toulas, C. αvβ3 and αvβ5 integrins control glioma cell response to ionising radiation through ILK and RhoB. *Int. J. Cancer.* **2008**, *123* (2), 357-364.

(15) Nishioka, S.; Wu, P. H.; Yakabe, T.; Giaccia, A. J.; Le, Q. T.; Aoyama, H.; Shimizu, S.; Shirato, H.; Onodera, Y.; Nam, J. M. Rab27b contributes to radioresistance and exerts a paracrine effect via epiregulin in glioblastoma. *Neurooncol. Adv.* **2020**, *2* (1), vdaa091.

(16) Kidd, A. R., 3rd; Snider, J. L.; Martin, T. D.; Graboski, S. F.; Der, C. J.; Cox, A. D. Ras-related small GTPases RalA and RalB regulate cellular survival after ionizing radiation. *Int. J. Radiat. Oncol. Biol. Phys.* **2010**, *78* (1), 205-212.

(17) Ahmed, K. M.; Dong, S.; Fan, M.; Li, J. J. Nuclear factor-kappaB p65 inhibits mitogen-activated protein kinase signaling pathway in radioresistant breast cancer cells. *Mol Cancer Res* **2006**, *4* (12), 945-955.

(18) Qi, T. F.; Miao, W.; Wang, Y. Targeted profiling of epitranscriptomic reader, writer, and eraser proteins accompanied with radioresistance in breast cancer cells. *Anal. Chem.* **2022**, *94* (3), 1525-1530.

(19) Cao, N.; Li, S.; Wang, Z.; Ahmed, K. M.; Degnan, M. E.; Fan, M.; Dynlacht, J. R.; Li, J. J. NF-kappaB-mediated HER2 overexpression in radiation-adaptive resistance. *Radiat Res* **2009**, *171* (1), 9-21.

(20) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856-2860.

(21) Yang, Y. Y.; Yu, K.; Li, L.; Huang, M.; Wang, Y. Proteome-wide interrogation of small GTPases regulated by N^6-methyladenosine modulators. *Anal. Chem.* **2020**, *92* (14), 10145-10152.

(22) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966-968.

(23) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; Garcia-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; et al. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2020**, *48* (D1), D1145-D1152.

(24) Kawahara, R.; Bollinger, J. G.; Rivera, C.; Ribeiro, A. C.; Brandao, T. B.; Paes Leme, A. F.; MacCoss, M. J. A targeted proteomic strategy for the measurement of oral cancer candidate biomarkers in human saliva. *Proteomics* **2016**, *16* (1), 159-173.

(25) Li, L.; Williams, P.; Gao, Z.; Wang, Y. VEZF1-guanine quadruplex DNA interaction regulates alternative polyadenylation and detyrosinase activity of VASH1. *Nucleic Acids Res.* **2020**, *48* (21), 11994-12003.

(26) Franken, N. A.; Rodermond, H. M.; Stap, J.; Haveman, J.; van Bree, C. Clonogenic assay of cells in vitro. *Nat. Protoc.* **2006**, *1* (5), 2315-2319.

(27) Miao, W.; Bade, D.; Wang, Y. Targeted proteomic analysis revealed kinome reprogramming during acquisition of radioresistance in breast cancer cells. *J Proteome Res* **2021**, *20* (5), 2830-2838.

(28) Huang, M.; Darvas, M.; Keene, C. D.; Wang, Y. Targeted quantitative proteomic approach for high-throughput quantitative profiling of small GTPases in brain tissues of Alzheimer's disease patients. *Anal. Chem.* **2019**, *91* (19), 12307-12314.

(29) Huang, M.; Qi, T. F.; Li, L.; Zhang, G.; Wang, Y. A targeted quantitative proteomic approach assesses the reprogramming of small GTPases during melanoma metastasis. *Cancer Res.* **2018**, *78* (18), 5431-5445.

(30) Ader, I.; Toulas, C.; Dalenc, F.; Delmas, C.; Bonnet, J.; Cohen-Jonathan, E.; Favre, G. RhoB controls the 24 kDa FGF-2-induced radioresistance in HeLa cells by preventing post-mitotic cell death. *Oncogene* **2002**, *21* (39), 5998-6006.

(31) Rojas, A. M.; Fuentes, G.; Rausell, A.; Valencia, A. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **2012**, *196* (2), 189-201.

(32) Jaschke, A.; Chung, B.; Hesse, D.; Kluge, R.; Zahn, C.; Moser, M.; Petzke, K. J.; Brigelius-Flohe, R.; Puchkov, D.; Koepsell, H.; et al. The GTPase ARFRP1 controls the lipidation of chylomicrons in the Golgi of the intestinal epithelium. *Hum. Mol. Genet.* **2012**, *21* (14), 3128-3142.

(33) Ishida, M.; Bonifacino, J. S. ARFRP1 functions upstream of ARL1 and ARL5 to coordinate recruitment of distinct tethering factors to the trans-Golgi network. *J. Cell Biol.* **2019**, *218* (11), 3681-3696.

(34) Hesse, D.; Radloff, K.; Jaschke, A.; Lagerpusch, M.; Chung, B.; Tailleux, A.; Staels, B.; Schurmann, A. Hepatic trans-Golgi action coordinated by the GTPase ARFRP1 is crucial for lipoprotein lipidation and assembly. *J. Lipid Res.* **2014**, *55* (1), 41-52.

(35) Hommel, A.; Hesse, D.; Volker, W.; Jaschke, A.; Moser, M.; Engel, T.; Bluher, M.; Zahn, C.; Chadt, A.; Ruschke, K.; et al. The ARF-like GTPase ARFRP1 is essential for lipid droplet growth and is involved in the regulation of lipolysis. *Mol. Cell Biol.* **2010**, *30* (5), 1231-1242.

(36) Starheim, K. K.; Kalvik, T. V.; Bjorkoy, G.; Arnesen, T. Depletion of the human N-terminal acetyltransferase hNaa30 disrupts Golgi integrity and ARFRP1 localization. *Biosci. Rep.* **2017**, *37* (2).

(37) Hein, A. L.; Post, C. M.; Sheinin, Y. M.; Lakshmanan, I.; Natarajan, A.; Enke, C. A.; Batra, S. K.; Ouellette, M. M.; Yan, Y. RAC1 GTPase promotes the survival of breast cancer cells in response to hyper-fractionated radiation treatment. *Oncogene* **2016**, *35* (49), 6319-6329.

(38) Tan, S.; Yi, P.; Wang, H.; Xia, L.; Han, Y.; Wang, H.; Zeng, B.; Tang, L.; Pan, Q.; Tian, Y.; et al. RAC1 Involves in the Radioresistance by Mediating Epithelial-Mesenchymal Transition in Lung Cancer. *Front Oncol* **2020**, *10*, 649.

Chapter 6: Concluding Remarks and Future Direction

In this dissertation, I examined comprehensively guanine quadruplex (G4) structures and their relevance to cellular processes and diseases. In particular, I focused on identifying and characterizing G4-binding proteins (G4BPs) to unravel the intricate protein interactome of G4 structures. I developed a novel quantitative proteomics approach using affinity pull-down coupled with LC-MS/MS analysis, which resulted in the identification of several dozen G4BPs. Among these proteins, SLIRP[1], YY1[2], GRSF1[3] and VEZF1[4], were validated, through *in vitro* biochemical assay, to be *bona fide* G4BPs.

I also utilized o-NBA-conjugated photo-crosslinking G4 probes, which facilitated the capture of weak and transient interactions, and, in conjunction with proteomic analysis, led to the identification of an expanded list of candidate G4-binding proteins. Notably, we uncovered HELLS as a novel G4BP using this approach. Furthermore, I examined comprehensively the functions of HELLS in its capacity as a G4 DNA helicase, specifically focusing on its role in unwinding G4 structures within promoter regions and its function in transcription regulation. Moreover, the investigation into phase separation phenomena in DNA G4 structures unveiled a previously unexplored mechanism of phase separation in modulating the stability and cellular functions of G4 structures. Collectively, these findings improved our understanding of G4 biology and open avenues for further research in the G4 field.

Throughout this dissertation, proteomic analysis played a pivotal role in identifying new G4-binding proteins, where discovery proteomics in data-dependent acquisition (DDA) mode has been extensively utilized. By employing this approach, a wide range of

proteins potentially involved in G4 biology have been discovered, offering valuable insights into the molecular interactions and functions of G4 structures. Additionally, targeted proteomics using multiple-reaction monitoring (MRM) has been employed as a powerful method to investigate biomarkers associated with radioresistance in breast cancer cells. The integration of these proteomics methodologies has greatly enhanced our understanding of G4 biology and provided valuable information for further research and clinical applications in cancer therapeutics.

Looking towards the future, there are exciting prospects for further investigation in the field of DNA G4s and their binding proteins. Building upon the current list of putative G4BPs, future studies can delve into roles of these proteins in modulating the biological functions G4 DNA in cells. It is evident that G4 structures have multi-faceted functions beyond their well-known roles in DNA replication, transcription, and telomere maintenance. Emerging research suggests their involvement in chromatin remodeling[5], epigenetic modification[6], ubiquitination[7], autophagy regulation[8], and stem cell pluripotency and differentiation[9], offering new avenues for exploring and expanding our understanding of G4 biology. Zyner et al.[9], using G4 ChIP-seq, demonstrated that G4 structures are highly prevalent and recurrent in the chromatin of human embryonic stem cells (hESCs), and are dynamically regulated during lineage differentiation. However, the specific proteins responsible for regulating the dynamic behavior of G4 structures during cell differentiation remain unidentified, presenting an intriguing area of research worthy of further investigation.

The therapeutic potential of G4 structures has garnered considerable attention in recent years, which emerged as promising targets for therapeutic interventions of various diseases. In cancer, G4 structures are being explored as targets to modulate gene expression by downregulating the transcription of oncogenes or blocking telomere elongation in cancer cells[10]. The development of small molecules specifically targeting G4 structures has shown promising results, with certain compounds, e.g., CX-5461, even progressing to clinical trials for patients with BRCA1/2-deficient tumors[11]. Furthermore, the involvement of G4 structures in neurodegenerative diseases, such as amyotrophic lateral sclerosis/frontotemporal dementia (ALS/FTD), holds significant potential for therapeutic interventions. Patients with ALS/FTD commonly exhibit hexanucleotide repeat expansions in the *C9ORF72* gene, which harbors G4-forming sequence of $(GGGGCC)_n$ in higher numbers compared to healthy controls[12]. The identification G4BPs involved in regulating the dynamics of G4 structures in disease contexts holds great promise for understanding the therapeutic implications of G4 structures. By unraveling the interactions between G4BPs and specific G4 structures in diverse disease contexts, we can gain insights into the underlying mechanisms and potentially develop targeted therapies with improved efficacy and specificity.

Regarding proteomics analysis, the current approach of photocrosslinking coupled with LC-MS/MS analysis has enabled the identification of a substantial number of putative G4BPs. However, one limitation of the current discovery proteomics method, data-dependent acquisition (DDA), resides in its reliance on selecting the most abundant precursor ions for subsequent MS/MS analysis. Consequently, proteins of low abundance

may escape detection. To address this limitation, utilization of data-independent acquisition (DIA) is a promising alternative. DIA captures information from all detectable peptides within a specified *m/z* range, including those of low abundance, thereby enabling a more comprehensive analysis of proteins captured by affinity pull-down [13]. By utilizing DIA, we can potentially uncover additional G4BPs and gain a more complete understanding of the protein interactome associated with G4 structures.

Furthermore, the integration of multi-omics approaches encompassing proteomics, transcriptomics, and epigenomics holds tremendous potential in elucidating the intricate interrelationships among G4 structures, G4BPs, and the broader cellular regulatory networks. By employing this integrated approach, researchers can achieve a comprehensive understanding of the molecular landscape underlying G4 biology. This method can be employed to screen for potential therapeutic targets using G4-binding small molecules and compare the profiles of different cell lines. By investigating these diverse cellular contexts, it might be possible to identify novel therapeutic targets and discover biomarkers for human diseases.

In summary, the findings presented in this dissertation contribute to the expanding field of G4 biology, highlighting the importance of G4BPs, stability modulation, and biomarker discovery. These advancements hold great promise for the development of targeted therapies. Continued research in this field will undoubtedly unravel further complexities of G4 structures and their functions, ultimately leading to improved diagnosis, treatment, and patient outcomes.

References

(1) Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J Am Chem Soc* **2017**, *139* (36), 12426-12429.

(2) Li, L.; Williams, P.; Ren, W.; Wang, M. Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **2021**, *17* (2), 161-168.

(3) Gao, Z.; Williams, P.; Li, L.; Wang, Y. A Quantitative Proteomic Approach for the Identification of DNA Guanine Quadruplex-Binding Proteins. *J Proteome Res* **2021**, *20* (11), 4919-4924.

(4) Li, L.; Williams, P.; Gao, Z.; Wang, Y. VEZF1-guanine quadruplex DNA interaction regulates alternative polyadenylation and detyrosinase activity of VASH1. *Nucleic Acids Res* **2020**, *48* (21), 11994-12003.

(5) Varizhuk, A.; Isaakova, E.; Pozmogova, G. DNA G-Quadruplexes (G4s) Modulate Epigenetic (Re)Programming and Chromatin Remodeling: Transient Genomic G4s Assist in the Establishment and Maintenance of Epigenetic Marks, While Persistent G4s May Erase Epigenetic Marks. *Bioessays* **2019**, *41* (9), e1900091.

(6) Reina, C.; Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int J Mol Sci* **2020**, *21* (11).

(7) Masud, T.; Soong, C.; Xu, H.; Biele, J.; Bjornson, S.; McKinney, S.; Aparicio, S. Ubiquitin-mediated DNA damage response is synthetic lethal with G-quadruplex stabilizer CX-5461. *Sci Rep* **2021**, *11* (1), 9812.

(8) Lejault, P.; Moruno-Manchon, J. F.; Vemu, S. M.; Honarpisheh, P.; Zhu, L.; Kim, N.; Urayama, A.; Monchaud, D.; McCullough, L. D.; Tsvetkov, A. S. Regulation of autophagy by DNA G-quadruplexes. *Autophagy* **2020**, *16* (12), 2252-2259.

(9) Zyner, K. G.; Simeone, A.; Flynn, S. M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S. G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **2022**, *13* (1), 142.

(10) Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* **2021**, *20* (1), 40.

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

(12) DeJesus-Hernandez, M.; Mackenzie, I. R.; Boeve, B. F.; Boxer, A. L.; Baker, M.; Rutherford, N. J.; Nicholson, A. M.; Finch, N. A.; Flynn, H.; Adamson, J.; et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **2011**, *72* (2), 245-256.

(13) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **2012**, *11* (6), O111 016717.