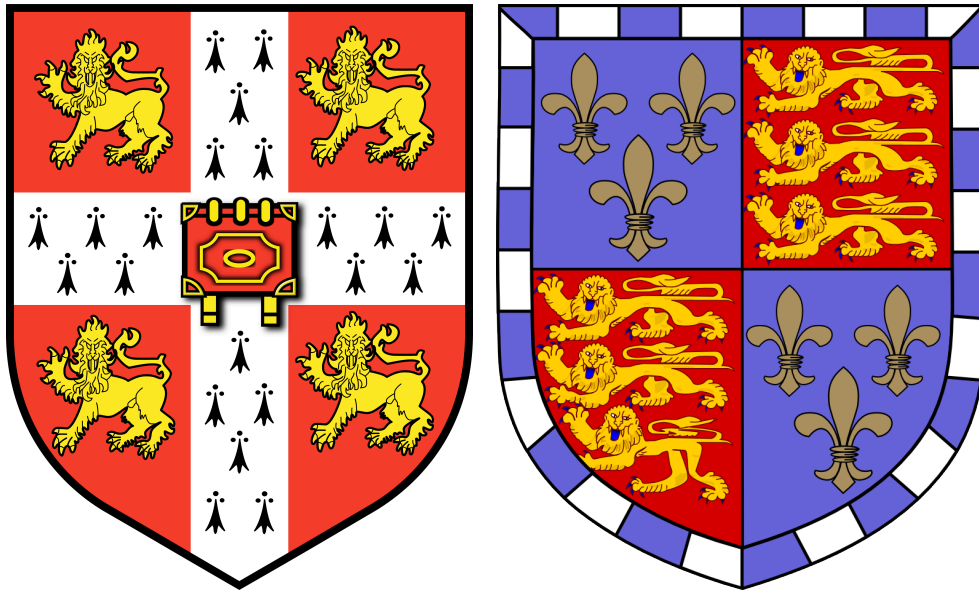# Exploring genetic interactions with G-quadruplex structures

Darcie Sinead Mulhearn

Cancer Research UK Cambridge Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Christ's College   August 2018

Preface

This dissertation is the result of my own work. Bioinformatic analysis of sequencing data was carried out in collaboration with Dr. Sergio Martinez-Cuesta. Performance and interpretation of the genome-wide screen (chapter 2) was conducted in collaboration with Dr. Katherine Zyner. All in-house-synthesis of PDS and PhenDC3, and CD spectra was performed by Dr. Marco Di Antonio. No other work was carried out in collaboration unless explicitly stated in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. This thesis does not exceed the word limit of 60,000 words.

Darcie Mulhearn

Christ's College

Cambridge

August 2018

Acknowledgements

Firstly, I would like to acknowledge Cancer Research UK for their funding which has allowed me the freedom and flexibility to perform my experiments. In particular I would like to thank my PhD supervisor Professor Sir Shankar Balasubramanian, for providing me with the intellectual and experimental freedom to follow my own ideas in the lab, allowing me to grow in independence and confidence on both a scientific and a personal level. Thank you for also teaching me that the best leader is not always the loudest voice in the room. I also express my sincere gratitude to Dr. David Tannahill, for his daily support and guidance, without which completing my PhD would have been nigh on impossible and Professor Greg Hannon, for his continued support throughout my studies. I also thank the Core Facilities, in particular Jane and Ian from Research Instrumentation and Bob and Manuela from Biorepository, without whom much of the work I have performed would not have been rapidly achievable and accessible.

Additionally, I am extremely grateful to all the amazing people I have worked with over the last four years. I thank Dr. Katie Zyner for her unending kindness and supervision; Dr. Jochen Spiegel for all the jokes, 'post-it' notes and Buzzfeed Fridays; Dr. Stefanie Lensing for showing me the ropes when I first started, for the Harry Potter audiobook sessions and making me laugh; Dr. Robyn Hardisty, Dr. Vicki Chambers and Judith Weber, for being the best Schwesters any Hoover could ask for; Dr. Giovanni "The Don" Marsico, Dr. Marco di Antonio; Dr. Clemens "Lukas" Mayer, Dr. Barbara Herdy, Dr. Robert "Bobbit" Hansel-Hertsch, Dr. Mike Gormally, Dr. Dhaval Varshney, Dr. Santosh Adhikari, Dr. Pierre Murat, Dr. Euni Ang-Raiber, Dr. Sergio Martinez-

For my Grandad Collins, my parents and, my niece and nephew;
my past, present and future inspirations.

Summary

G-quadruplexes are non-canonical nucleic acid secondary structures of increasing biological and medicinal interest due to their proposed physiological functions in transcription, replication, translation and telomere biology. Aberrant G4 formation and stabilisation have been linked to genome instability, cancer and other diseases. However, the specific genes and pathways involved are largely unknown, and the work within this thesis aims to investigate this. Stabilisation of G4s by small molecules can perturb G4-mediated processes and initial studies suggest that this approach has chemotherapeutic potential. I therefore also aimed to identify cell genotypes sensitive to G4-ligand treatment that may offer further therapeutic opportunities. To address these aims, I present the first unbiased genome-wide genetic screen in cells where genes were silenced via short-hairpin RNAs (shRNAs) whilst being treated with either PDS or PhenDC3, two independent G4-stabilising small molecules. I explored gene deficiencies that enhance cell death (sensitisation) or provide a growth advantage (resistance) in the presence of these G4-ligands. Additionally, I present a validation screen, comprising hits uncovered via genome-wide screening, and also the use of this in another cell line of different origin. Sensitivities were enriched in DNA replication, cell cycle, DNA damage repair, splicing and ubiquitin-mediated proteolysis proteins and pathways. Ultimately, I uncovered four synthetic lethalities *BRCA1, TOP1, DDX42, GAR1,* independent of cell line and ligand. These were validated with three G4-stabilising ligands (PDS, PhenDC3 and CX-5461) using an independent siRNA approach. The latter siRNA methodology was used to screen 12 PDS derivatives with improved

medicinal chemistry properties and ultimately identified SA-100-128, as a lead compound. The mechanism behind synthetic lethality with G4-stabilising ligands was explored further for DDX42, which I show has *in vitro* affinity for both RNA- and DNA-G4s and may represent a previously unknown G4-helicase. Also within this thesis, gene deficiencies that provided a growth advantage to PDS and/or PhenDC3 as uncovered by genome-wide and focused screening were explored. These showed enrichment in transcription, chromatin and lysosome-associated genes. The resistance phenotype of three gene deficiencies, *TAF1, DDX39A* and *ZNF217* was further supported by additional siRNA experiments. Overall, I satisfied the primary aims and established many novel synthetic lethal and resistance interactions that may represent new therapeutic possibilities. Additionally, the results expand our knowledge of G4-biology by identifying genes, functions and subcellular locations previously not known to involve or regulate G4s.

Table of Contents

Table of Figures

Abbreviations

(d)NTP	(deoxy)nucleoside triphosphate
2-Tet	2-tetrad antiparallel basket quadruplex structure
2'OH	2'hydroxyl
AAV	Adeno-associated virus
Ago	Argonaute
ALT	Alternative Lengthening of Telomeres
AML	Acute Myeloid Leukaemia
ATP	Adenosine Triphosphate
BCLAF1	BCL2 associated transcription factor 1
BLM	Bloom's syndrome protein
BRD	Bromodomain contain
Cas9	CRISPR associated protein 9
CD	Circular Dichroism
CDK	Cyclin Dependent Kinase
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
CFU	Colony Forming Units
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation-Sequencing methodology
CIN	Chromosome Instability
CLL	Chronic Lymphocytic Leukaemia
COSMIC	Catalogue of Somatic Mutations in Cancer
cPDS	Carboxy-Pyridostatin
CRISPR	Cluster Regularly Interspaced Short Palindromic Repeats

DAVID	Database for Annotation, Visualisation, and Integrated Discovery
DDR	DNA damage response
DDX/DHX-box	Asp-Glu-Ala-Asp-box helicase
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PK	DNA-protein kinase
DOG-1	Deletion of G-rich tracts 1
DSB	Double Strand Break
EASE	Expression Analysis Systematic Explorer
EDTA	Ethylenediaminetetraacetic Acid
ELISA	Enzyme-linked Immunosorbent Based Assay
EMT	Epithelial-Mesenchymal Transition
FA	Fanconi's anaemia
FACS	Fluorescent Activated Cell Sorting
FANC-J	Fanconi Anaemia group J protein
FBS	Foetal Bovine Serum
FDA	Food and Drug Administration
FDR	False Discovery Rate
Fe-S	Iron-Sulphur cluster domain family
FMRP	Fragile X Mental Retardation Protein/Syndrome
FRET	Fluorescence Resonance Energy Transfer
G-tract	Poly Guanine Tract
G4	G-quadruplex

G4-seq	High-throughput Sequencing of G-quadruplex structures
G4IPDB	G-quadruplex Interaction Protein Database
GARP	Golgi-associated retrograde protein
gDNA	Genomic DNA
GEMM	Genetically Engineered Mouse Model
GI_n	Concentration causing 'n'% Growth Inhibition
Glm	Generalised linear model
GO BP	Gene Ontology Biological Process
GO MF	Gene Ontology Molecular Function
GSK(i)	Glutathione Synthase Kinase (inhibitor)
GST	Glutathione Synthase Transferase
GTFs	General Transcription Factors
HBS buffer	HEPES-buffered Saline
HDAC	Histone Deacetylase
HEK	Human Embryonic Kidney
hnRNP	Heterogeneous nuclear ribonuclearproteins
hPIF1	Human Pif1
HR	Homologous recombination
HRP	Horseradish Peroxidase
IF	Immunofluorescence
IRES	Internal Ribosome Entry Site
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	Long non-coding RNA
Log_2FC	Log_2 Fold Change

LTR	Long Terminal Repeat
mAB	Monoclonal antibody
MDS	Multidimensional scaling
MEFs	Mouse Embryonic Fibroblasts
MeSH	Medical Subject Heading
MGP	Mouse Genetics Project
miRNA	Micro-RNA
MOI	Multiplicity of Infection
mRNA	Messenger RNA
MSCV	Murine Stem Cell Virus
N/A	Non-applicable
ncRNA	Non-coding RNA
NHE III	Nuclease hypersensitive element III
NHEJ	Non-homologous End Joining
NLS	Nuclear Localisation Sequence
NMM	N- methyl mesoporphyrin IX
NMR	Nuclear Magnetic Resonance
NPM1	Nucleophosmin
NS4A/NS5B	Non-structural Protein 4A/5B
NSCLC	Non-small Cell Lung Cancer
NT siRNA	Non-targeting small interfering siRNA
O/N	Overnight
ORC	Origin Recognition Complex
ORF	Open Reading Frame

PARP(i)	Poly-ADP ribose polymerase (inhibitor)
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PD	Population Doubling
PDAC	Pancreatic Ductal Adenocarcinoma
PDS	Pyridostatin
PGK promoter	Phosphoglycerate Kinase promoter
Plat-A	Platinum-A Retroviral Packaging Cell line
PMC	PubMed Central
PQS	Putative Quadruplex Sequence
PTM	Post Translational Modification
RFC	Replication Factor C
RIPA	Radioimmunoprecipitation Assay
RISC	RNA-induced Silencing Complex
RNA	Ribonucleic Acid
RNAi	RNA-interference
RNAP	RNA polymerase
RNR	Ribonucleotide Reductase
Rpm	Rotations per minute
RT	Room Temperature
scFV	Single-chain Variable Fragment
SD	Standard deviation
SEM	Standard Error of the Mean
sgRNA	Single guide RNA

shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SLC	Solute Carrier
snoRNA	Small nucleolar RNA
snoRNP	Small Nucleolar Ribonucleoprotein
ssDNA	Single-stranded DNA
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
t0	Initial/Reference time point
TAF	Transcription Associated Factor
TALENs	Transcription activator-like effector nucleases
TE	Tris-EDTA
TERRA	Telomeric Repeat containing RNA
tF	Final timepoint
TLS/FUS	Translocated in Sarcoma/Fused in Sarcoma
TMM	Trimmed mean of M-values
TRF1	Telomeric Repeat binding Factor 1
TRF2	Telomeric Repeat binding Factor 2
TSS	Transcription Start Site
UniprotKB	Universal Protein resource Knowledge Base
UTR	Untranslated Region
VSV-G	Vesicular Stomata Virus-Glycoprotein
WRN	Werner's syndrome protein
ZFN	Zinc Finger Nuclease

Chapter 1

General Introduction

Overview

This thesis describes investigations into a non-canonical four-stranded structure called the Guanine-quadruplex (G4), found ubiquitously in RNA and DNA. The purpose of these experiments relates to both G4s and to the small molecules that bind to and selectively stabilise these structures (referred to as G4-stabilising ligands henceforth). In particular, the work focuses on identifying genotypes that perturb G4 formation and/or function and thus alter G4-ligand sensitivity. Genes and pathways were identified via functional genomics and subsequent validation experiments, that when depleted via RNA-interference (RNAi), cause either decreased or increased growth rate in the presence of G4-stabilising ligands.

At the beginning of my studies and to date, such screening methodologies had yet to be explored with G4-stabilising ligands. The genes uncovered by such an approach have great potential to explore unknown clinical and basic biological areas regarding G4s (expanded in section 1.9). Firstly, the genes identified impart knowledge concerning the biological roles and locations of G4s, and the methods by which cells respond to and regulate G4-structures. From a clinical perspective, these experiments might help to define genotypes that are especially sensitive to G4-stabilising ligand treatment and highlight new therapeutic opportunities. Further, any identification of non-G4 associated resistance mechanisms will inform on areas for improving the

specificity of G4-ligands. Additionally, the tools and resources outlined within this thesis can be used to test and benchmark novel G4-stabilising ligands.

The first chapter of this thesis provides an introduction to G4s, including their cancer association and hypothesised endogenous roles, as well as the small molecules targeting G4-structures. Also discussed are the available functional genomic methods for pooled screening of the response of different cellular genotypes to pharmacological treatment. The experimental work is divided into four chapters. Initially, Chapter 2 describes the setup and outcome of a genome-wide screening approach and subsequent smaller, validation screens performed with two independent G4-stabilising ligands, PDS and PhenDC3 (see section 1.4.2.3). Chapter 2 focuses on gene-deficiencies that cause synthetic lethality with these molecules. Secondly in Chapter 3, I outline an independent approach to validate the top four sensitivities identified by genetic screening with PDS and PhenDC3. I also extend this validation to other G4-stabilising ligands. In Chapter 4, one of these key four sensitivities, DDX42, was selected for further investigation, and preliminary experiments to explore the possible mechanism of G4-ligand synthetic lethality with DDX42-deficiency are described. Then, I focus on the genetic screening outcome from a "resistance" perspective in Chapter 5, i.e. investigating gene deficiencies that provide a growth advantage to cells treated with G4-stabilising ligands. Finally, Chapter 6 provides an overall summary detailing how work within this thesis addresses the aims to define cellular genotypes that alter G4 formation and/or ligand accessibility from a biological, clinical and technical perspective.

1.1 G-quadruplex sequence and structures

Over 60 years ago, James Watson and Francis Crick suggested the first double helix model of DNA based on X-ray diffraction data performed by Rosalind Franklin and Maurice Wilkins (Watson & Crick, 1953; Franklin & Gosling, 1953; Wilkins *et al*, 1953). Since, it has been shown that DNA predominantly forms a B-duplex (Richmond & Davey, 2003). However, increasingly more studies have shown that non-canonical structures exist and that these configurations may influence a variety of biological processes in both RNA and DNA. Such secondary structures can occur, for example, when nucleotides interact via alternative hydrogen-bonding patterns to the classic Watson-Crick interaction between opposite strands of the DNA or RNA duplex (Murat & Balasubramanian, 2014; Wong & Huppert, 2009; Zhao *et al*, 2010). The subject of this thesis is one such non-canonical structure, the guanine-quadruplex (G4) which is formed when the guanines (Gs) interact by Hoogsteen base pairing (Hoogsteen, 1963). The existence of G4s was first suggested in 1910 following the observation that concentrated guanylic acid solutions formed a hydrogel (Bang, 1910), hypothesised to arise by guanine self-association. The existence of G4s were subsequently suggested by electrophoretic techniques whereby the formation of G4s from synthetic guanine (G)-rich DNA oligonucleotides derived from human (Sen & Gilbert, 1988) and ciliate *Tetrahymena* sequences (Williamson *et al*, 1989) caused their aggregation and retardation on the gel.

G4 structures form when four guanine tracts (G-tracts) self-associate via this Hoogsteen interface (Figure 1.1B) rather than interacting with cytosine via

3

Watson-Crick base pairing (Figure 1.1A). A consensus sequence for G4-formation, derived from biophysical structural and computational analysis, is shown below (Huppert & Balasubramanian, 2005; Todd *et al*, 2005) with G-tracts separated by nucleotide loops (N). More recent biophysical studies suggests that loops can be longer than seven nucleotides (Guédin *et al*, 2010; Mukundan & Phan, 2013) and *in vitro* spectroscopic analysis shows central loops of up to 21 nucleotides allow G4 formation (Bourdoncle *et al*, 2006).

$$G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3+}N_{1-7}G_{3-5}$$
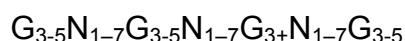
Four guanines (i.e. one from each tract) associate to form planar tetramolecular quartets (G-tetrads, Figure 1.1B). Stacking of these quartets via π - π interaction provides stability to the G4-structure (Figure 1.1C). Although generally consisting of a minimum of three quartets, two-tetrad stable structures have also been seen in DNA (Zhang *et al*, 2010; Lim *et al*, 2009). Monovalent cations through the G-quartet core electrostatically interact with electrons from O_6 from each guanine (see Figure 1.1B) providing further structural stability, with the strength of stabilisation in the order: K^+ > Na^+ > NH_4^+ > Li^+ (Davis, 2004; Pinnavaia *et al*, 1978; Sessler *et al*, 2000; Hud *et al*, 1999). More recently, RNA G4s have been biophysically investigated *in vitro* (Zhang *et al*, 2011).

A Watson and Crick base pairing

Cytosine (C)

Guanine (G)

B Hoogstein base pairing

Metal cation interaction with O_6 electrons

G-quartet

Guanine (G)

C Guanine Quadruplexes (G4)

strand orientation heterogeneity

Parallel

Mixed-type hybrid

Antiparallel

intramolecular G4 loop topologies

Bulges

Long loops
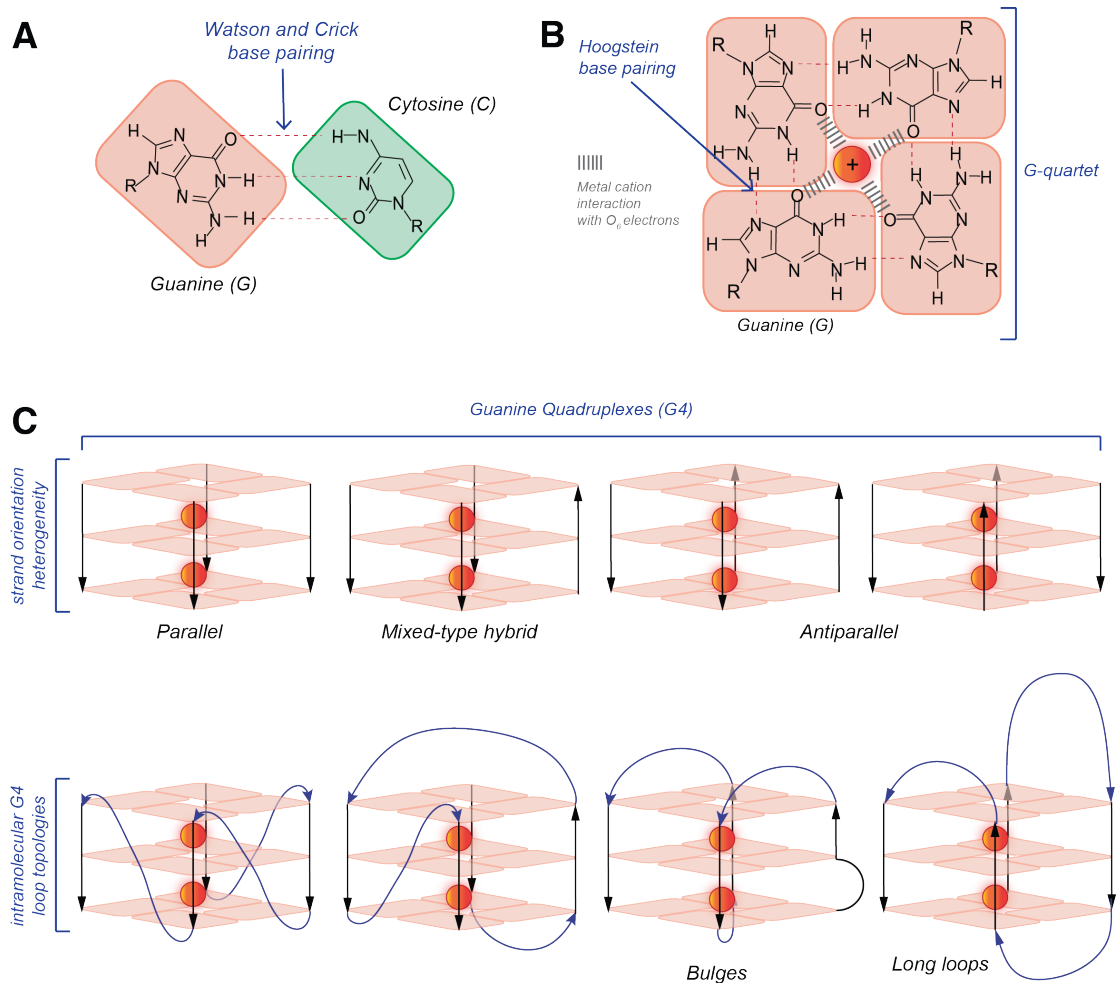
Figure 1.1. G-quadruplexes are diverse non-canonical nucleic acid secondary structures

(A) Watson-Crick hydrogen bonding between guanine and cytosine

(B) Schematic of a guanine (G)-quartet, in which four guanines interact via alternative Hoogstein hydrogen bonding, stabilised via a central monovalent cation interacting with electrons provided by the guanine O_6 atoms.

(C) Schematic of a series of G-quadruplexes formed from three tetrads, differing in strand orientation, which can influence loop topology in intramolecular G4-structures. Loops can also differ in length, and "bulges" can exist if guanines within the tract are non-consecutive.

5

1.2 G-quadruplexes are topologically heterogeneous

Although characterised by a common quartet arrangement, G4-structures as a family are diverse and some G4s are highly polymorphic, and both RNA and DNA have the propensity to fold into a variety of different G4-structures. One source of heterogeneity is the intervening loops (Figure 1.1C) differing in length, nucleotide composition and conformation (Neidle, 2010; Burge *et al*, 2006). Biophysical analysis of various G4 structures shows that the length and sequence of the intervening loops can greatly influence the structure and stability of intramolecular complexes (Risitano & Fox, 2004; Bugaut & Balasubramanian, 2008; Hazel *et al*, 2004). This was corroborated by *in vivo* investigations into loop length in yeast (Piazza *et al*, 2015), where short loops provided the most stable structures.

Structural polymorphism also arises from strand polarity: strands can be all parallel or have a mixture of parallel and antiparallel strands leading to additional diversity in loop topologies of G4 structures (Figure 1.1C). Both intra- and intermolecular G4s exist, the latter arising from interaction between two or four independent strands (Simonsson, 2001; Keniry; Burge *et al*, 2006; Schaffitzel *et al*, 2001). This includes hybrid DNA-RNA intermolecular G4s (Zhang *et al*, 2014). Such considerable polymorphism highlights that 'quadruplex' is an umbrella term for a diverse set of non-canonical structures.

1.3 Locations of G4s within the genome and the transcriptome

1.3.1 Bioinformatic analysis

Computational algorithms have been employed to identify Potential Quadruplex-forming Sequences (PQS) within the genome and transcriptome. One such algorithm is QuadParser (Huppert & Balasubramanian, 2005), which predicts ~376,000 PQS within the human genome adhering to the consensus $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3+}N_{1-7}G_{3-5}$ (Huppert & Balasubramanian, 2005; Todd *et al*, 2005). More recently the algorithm G4-hunter (Bedrat *et al*, 2016) suggests this number is 2-10 fold higher *in vitro* by additionally considering G-richness and G-skewness, however this algorithm specifies a very relaxed G4 definition thus encompasses significant false positives. Further extending the PQS definition to include longer loop lengths and bulges suggests over 716,000 potential genomic sites (Chambers *et al*, 2015; Schiavone *et al*, 2014). This may still be an underestimation of PQS based on the observation that repetitive DNA and RNA sequences, such as the repetitive G-rich sequences that form G4 structures, are often omitted from available genome databases (reviewed in Rhodes & Lipps, 2015). Many PQS, both RNA and DNA, are conserved at specific loci across species (König *et al*, 2010; Sahakyan *et al*, 2017; Capra *et al*, 2010) indicating a strong selection pressure for their retention in the genome and transcriptome. Such positive selection for PQS indicates important and evolutionarily constrained roles for the structures that they encode. These loci will be described in the remainder of section 1.3, and provide some preliminary indications of G4 functions.

Application of QuadParser to the human genome revealed that more than 40% of genes have a predicted PQS within 1 kb of the transcription start site (TSS), often overlapping with transcription factor binding sites (Neidle, 2010; Huppert & Balasubramanian, 2007, 2005). Promoters showing G4-enrichment include those at genes involved in immunoglobulin chain switching (Han & Hurley, 2000), cell development, signalling and growth (Eddy & Maizels, 2006). Of particular interest, PQS are enriched at oncogene promoters (Huppert & Balasubramanian, 2007) including MYC, BCL2 and VEGF (Sun *et al*, 2011) but depleted at tumour suppressor promoters (Eddy & Maizels, 2006). Additionally, genome-wide bioinformatic analysis of human replication origins suggest that over 90 % are in PQS proximity, with increased density at more frequently used origins (Cayrou *et al*, 2012; Besnard *et al*, 2012). That being said, replication origins are difficult to define in humans and show extensive variability (Leonard & Méchali, 2013), thus the relationship between G4s and replication origins are unclear. More concrete evidence beyond bioinformatic causation, could be provided by identifying synthetic lethality with replication-associated proteins. The other primary site of DNA-G4 enrichment is at telomeres (see section 1.3.3; Huppert & Balasubramanian, 2005; Blackburn, 1991; Simonsson, 2001). Additionally, DNA-G4 have been identified in the genomic DNA of viruses (Tlučková *et al*, 2013; Piekna-Przybylska *et al*, 2014) and bacteria (Beaume *et al*, 2013).

These bioinformatic analyses have recently been independently biophysically corroborated with two separate G4-mapping techniques, G4-seq and BG4-ChIPseq (Chambers *et al*, 2015; Hänsel-Hertsch *et al*, 2016). The former

mapped G4-structures in DNA isolated from primary human B-lymphocytes, combining polymerase stop assay and Next-generation sequencing techniques to compare sequencing readouts from conditions that promoted or discouraged G4-formation (Chambers *et al*, 2015). Conversely, BG4 ChIP-seq, uses the G4-specific antibody BG4 (see section 1.4.1) to perform chromatin Immunoprecipitation (ChIP) in the immortalised HaCaT keratinocyte cell line, and subsequent sequencing to identify more than 10,000 G4-peaks, which were preferentially formed in open chromatin (Hänsel-Hertsch *et al*, 2016).

1.3.2 Promoter quadruplexes

The human MYC gene illustrates the potential transcriptional regulatory role of promoter G4s. The MYC promoter contains a 27 bp, PQS-containing, nuclease hypersensitive element III (NHE III) (Simonsson *et al*, 1998). The NHE III element binds several transcription factors, controlling up to 90% of MYC transcription (González & Hurley, 2010). A synthetic oligonucleotide based on this sequence has been biophysically demonstrated to form two different G4-structures (Siddiqui-Jain *et al*, 2002), further corroborated by NMR visualisation of these structures in the human Myc sequence (Ambrus *et al*, 2005; Phan *et al*, 2004). These G4s act as hypothesised transcriptional repressor elements and destabilisation of these structures caused by NM23H2 binding causes increased MYC expression (Siddiqui-Jain *et al*, 2002).

1.3.3 Telomeres

DNA G4s are also enriched at telomeres, the protective nucleoprotein structures at chromosome ends. Human telomeres consist of up to 15 kilobases of TTAGGG tandem repeats (de Lange, 2009). This sequence adheres to the predicted G4 consensus sequence (Huppert & Balasubramanian, 2005). While predominantly double-stranded, the telomeres end in 50-500 basepairs of G-rich single-stranded DNA overhangs (de Lange, 2009; Makarov *et al*, 1997) that have the potential to form a polymorphic G4 array (Phan, 2010). Independent NMR spectroscopy and crystallographic studies with both Na+ and K+ cations suggest several stable telomeric DNA G4-conformations including parallel, antiparallel and 2-tetrad arrangements (reviewed in Bryan & Baumann, 2011) illustrating the topological heterogeneity of these structures.

1.3.4 Location of RNA quadruplexes

In RNA, more than 3,000 mRNAs have been bioinformatically and biophysically demonstrated to contain G4-structures (Bugaut & Balasubramanian, 2012; Kwok *et al*, 2016; Kwok & Balasubramanian, 2015), particularly enriched at 5'- and 3'-untranslated regions (UTRs). The latter was achieved via 'rG4-seq', an *in vitro* method that mapped K+-stabilised G4s in polyadenylated-enriched HeLa RNA, by coupling the ability of these structures to cause reverse transcriptase stalling with next-generation sequencing. Additionally viral mRNAs contain RNA G4 structures, such as that found in the Epstein-Barr virus encoded nuclear antigen 1 mRNA (Murat *et al*, 2014). PQS are further found in introns and near polyadenylation sites (Eddy & Maizels,

2008; Kikin *et al*, 2008; Chambers *et al*, 2015). RNA G4s are also predicted to form in non-coding (nc) RNA sequences, including precursor miRNAs (Mirihana Arachchilage *et al*, 2015) and long non-coding RNAs (lncRNAs) (Jayaraj *et al*, 2012) with the latter experimentally verified via G4-induced reverse transcriptase stalling in multiple lncRNAs (Kwok *et al*, 2016).

1.4 Tools to investigate G4 structures

Many hypothesised G4-roles arise from the biochemical inhibition that unresolved secondary structures can cause to processive RNA and DNA polymerases (Edenberg *et al*, 2014). This inhibition could provide an endogenous regulatory method for processes involving nucleic acids, however this has not been systematically evaluated. This may include modulation of transcription and replication (DNA G4s) and splicing and translation (RNA G4s) (Bochman *et al*, 2012; Cruz & Westhof, 2009; Murat & Balasubramanian, 2014; Balasubramanian & Neidle, 2009).

Several methods exist to explore the roles of G4s in cells (reviewed in Hänsel-Hertsch, Di Antonio, & Balasubramanian, 2017). One is the use of G4-specific antibodies. By experimentally exploiting the specific and high affinity interaction of antibodies, endogenous G4-structure forming sequences can be identified via ChIP or quantified and visualised within fixed cells, by coupling to fluoro- or chromophores. Alternatively, small molecule ligands that specifically bind to and stabilise G4-structures following their application to live cells, can be used. These two methods answer different interlinked questions. As antibodies predominantly bind pre-formed G4-structures, they

provide a global snapshot of the number and location of G4s within a cell. Further, antibodies are currently not able to image G4s and their dynamics in live cells. Conversely, small molecules actively stabilise and increase the lifetime of folded G4s in live cells. The phenotype induced by altering G4-dynamics with these small molecule ligands can inform on the roles that G4-structures perform and the cellular response to their increased levels and persistency. Both methods will be discussed below.

1.4.1 Development of G4-specific antibodies

The development of G4-specific antibodies has been instrumental in validating their *in vivo* existence. The first imaging of G4s in cells was provided by the antibody Sty49, generated via ribosome display and raised against telomeric DNA-G4s from the ciliate *Stylonychia lemnae* (Schaffitzel *et al*, 2001). *In situ* immunostaining of *Stylonichia* cells with Sty49 revealed telomeric G4 formation *in vivo* (Schaffitzel *et al*, 2001), later shown to depend on two telomeric binding proteins, TEBPα and TEBPβ (Paeschke *et al*, 2005). S-phase resolution of telomeric G4s required recruitment of the StyRecQL helicase as reflected by depletion in Sty49 staining (Postberg *et al*, 2012) indicating even at this early stage of G4 research, that cellular machinery exists to regulate these structures and that their resolution is necessary for accurate replication.

More recently developed antibodies allowed G4 detection in mammalian cells. For example, a single chain phage display (scFV) antibody BG4, was shown to recognise several G4-conformations with low nanomolar affinity (Biffi *et al*,

2013). BG4 immunofluorescence (IF) studies demonstrated the existence of DNA- (Biffi *et al*, 2013) and RNA- (Biffi *et al*, 2014a) G4s in fixed human cell lines and was subsequently used to detect G4-enrichment in human stomach and liver cancer tissues by immunohistochemistry compared to controls (Biffi *et al*, 2014b). Immunostaining with an independent monoclonal antibody 1H6, generated by immunising mice with stable G4-constructs also supports G4-existence in mammalian cells (Henderson *et al*, 2014). Furthermore, DNA Immunoprecipitation of genomic DNA from a human breast adenocarcinoma cell line, with the phage display antibody hf2 generated against a KIT promoter G4, showed PQS motif enrichment, suggesting that these sequences form G4 structures in genomic DNA (Lam *et al*, 2013; Fernando *et al*, 2008).

1.4.2 G4-stabilising ligands and therapeutic options

Many small molecules have been developed that bind to and stabilise G4 structures and have helped the investigation of G4 biology. G4 ligands are generally characterised by planar, aromatic structures allowing π-stacking onto the terminal G-quartet (Neidle, 2010). This is reflected in the structure of telomestatin, a naturally occurring telomeric G4 stabiliser (Figure 1.2A) isolated from *Streptomyces anulatus* (Shin-ya *et al*, 2001). This terminal quartet-binding model is supported for several G4-stabilising small molecules via NMR and crystallographic structural data (reviewed in Haider *et al*, 2011). G4-ligands are diverse but show similarities in their structures and hypothesised binding model to G4s. However, the literature reports phenotypic differences following their application to cells. To date however,

these observations arise from independent studies with different cell lines, G4-ligands and setups. To fully appreciate the distinct biological targets and efficacy of G4-stabilising ligands, a simultaneous, side-by-side comparison of the cellular response they induce is required. The work within this thesis makes a preliminary attempt at this goal. Below, the cellular response to G4-ligands will be generally discussed from three main perspectives: promoter G4-targeting, telomere G4-targeting and the role of G4s in replication and the DNA damage response (DDR). Next, the cellular phenotypes in response to two G4-stabilising ligands, PDS and PhenDC3 will be summarised, as these molecules have been extensively used for experiments outlined in this thesis.

Many of the documented G4-ligand treatment responses are hypothesised to arise from DNA-G4 stabilisation. Due to their later discovery in cells, RNA-G4s have been less well investigated as G4-ligand targets. Note that although many G4-ligands are biophysically characterised with DNA-G4 forming oligonucleotides, they may also bind RNA G4s. That being said, RNA specific G4-stabilising ligands have been developed (reviewed in Donlic & Hargrove, 2018), including carboxyPDS (cPDS; see section 1.4.2.4).
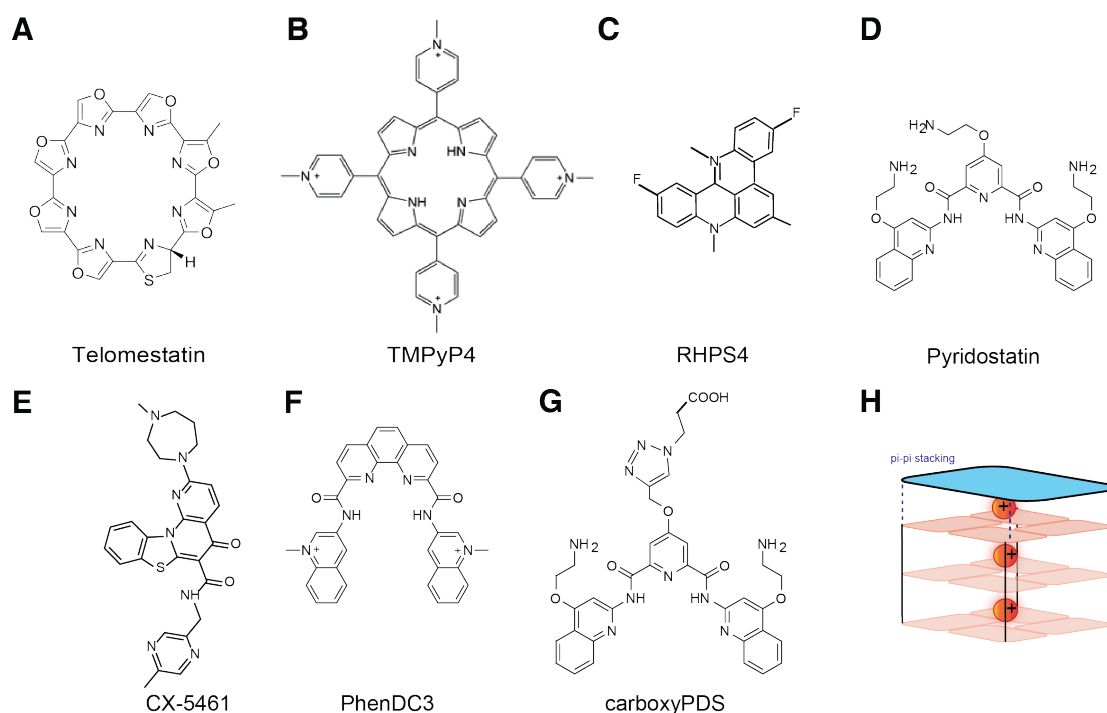
Figure 1.2. Examples of G4-stabilising ligands
(A) The naturally occurring G4-stabilising ligand telomestatin (Shin-ya et al., 2001), (B) cationic porphyrin TMPyP4 (Frank Xiaoguang Han et al, 1999), (C) pentacyclic acridine RHPS4 (Gowan et al, 2001), (D) Pyridostatin (PDS), (E) phase II clinical compound, quarfloxin derivative CX-5461 (Xu et al, 2017), (F) bisquinilium compound PhenDC3 (De Cian et al, 2007b), (G) carboxylated PDS derivative specifically stabilising RNA-G4, cPDS.(Di Antonio et al, 2012) (H) G4 ligands are hypothesised to stack onto the terminal G4 quartet via π-π interaction between planar aromatic rings.

1.4.2.1 G4-stabilising ligands targeting oncogene promoters

The bioinformatic and biophysical identification of predicted promoter G4s has gained experimental support from independent cell studies with various ligands inducing alterations in gene expression levels. This is postulated to arise from stabilisation of predicted promoter G4s and include modulation of oncogene expression. For example treatment with TMPyP4 (Figure 1.2B), actinomycin D and quinolone derivatives (Lagah *et al*, 2014; Thakur *et al*, 2009; Kang & Park, 2009; Ou *et al*, 2007) reduces *MYC* expression in several cell lines. Indeed telomestatin treatment reduces expression of *MYC* and *TERT* in medullablastoma and teratoid-rhabdoid cancer cell lines (Shalaby *et*

al, 2010) and *KRAS* in pancreatic cancer cell lines (Cogoi *et al*, 2009).

Acridinium derivatives have also been explored as G4-stabilisers, the most

effective of this class being RHPS4 (Figure 1.2C) (Gowan *et al*, 2001)

showing PQS-associated anti-proliferative effects in multiple glioblastoma cell

lines (Lagah *et al*, 2014). As many cancers are driven by deregulated

overexpression of various oncogenes, the ability to decrease their

transcription via treatment with G4-stabilising ligands provides an attractive

therapeutic prospect (Balasubramanian *et al*, 2011). However, as these

promoter G4s are also present in normal cells, in isolation this is perhaps

unlikely to provide adequate selectivity. Thus, there is merit in systematically

identifying and exploiting genotype synthetic lethality with G4-stabilising

ligands.

1.4.2.2 G4-stabilising ligands targeting telomeres

Supporting the existence of a tandem array of polymorphic DNA-G4s at the

telomere, application of G4-stabilising ligands to cells and *in vitro* induces a

plethora of telomere-associated phenotypes. These include telomerase

inhibition, telomere uncapping and subsequent senescence (Tahara *et al*,

2006; Gomez *et al*, 2006; Zahler *et al*, 1991; Zaug *et al*, 2005; Neidle, 2010).

For example, telomestatin inhibits telomerase by encouraging formation of,

and selectively binding to, intramolecular, antiparallel telomeric G4s (Kim *et*

al, 2002; Rezler *et al*, 2005). This may suggest that other telomeric G4

structures, such as parallel G4s, are not targeted by the natural compound, a

feature not shared by many synthetic G4-stabilising ligands which have been

shown to bind a range of G4-topologies (De Cian *et al*, 2007b; Le *et al*, 2015).

Treatment with low telomestatin concentrations, causes telomere attrition and apoptosis of several cancer cell lines whereas high doses induce telomere uncapping (Gomez *et al*, 2004b; Kim *et al*, 2003; Shammas *et al*, 2004; Tahara *et al*, 2006; Tauchi *et al*, 2003, 2006). As the majority of cancers reactivate telomerase and several have increased telomere lengths (Bailey & Murnane, 2006), the targeting of telomeric G4s represents a promising chemotherapeutic target (Han & Hurley, 2000; Neidle, 2010).

1.4.2.3 G4-stabilising ligands inducing a DDR

Beyond telomerase inhibition and transcriptional modulation, small molecule stabilisation of G4-structures is problematic during DNA replication and can induce DNA damage and activate a DDR, a mechanism hypothesised to arise via DNA polymerase inhibition. Such a response is observed following application of pyridostatin (PDS; Figure 1.2D, Rodriguez et al., 2008) to cell lines (section 1.4.2.3). This has been recently exploited in a synthetic lethality approach for treatment of BRCA-deficient tumours with the G4-stabilising ligand CX-5461 (Figure 1.2E), currently in Phase II clinical trials (Xu *et al*, 2017). Here the reduced ability to repair DNA-damage synergises with the induction of DNA damage via G4-stabilising ligand treatment. This will be discussed in greater detail in section 1.6.

1.4.2.4 PDS and PhenDC3

In many ways the G4-affinities of PDS, and the phenotypes attributed to PDS treatment of cells, reflect general attributes seen with other G4-ligands (sections 1.4.2.1-1.4.2.3). For example, PDS binds to promoter and telomeric

DNA-G4s and mRNA RNA-G4s *in vitro* (Murat & Balasubramanian, 2014; Kwok *et al*, 2016) and increases endogenous G4-formation as judged by BG4 immunofluorescence studies in fixed cells (Biffi et al., 2013). PDS-induced G4 stabilisation is accompanied by a strong DDR in transformed fibroblasts and colon cancer cell lines (Rodriguez *et al*, 2012). Subsequently, PDS treatment was shown to cause synthetic lethality in BRCA2-deficient cells, a gene involved in damage repair (McLuckie *et al*, 2013). Additionally, PDS-treated cells exhibit telomere dysfunction (Müller *et al*, 2012) and reduced expression of genes containing promoter PQS, including but not limited to oncogenes (Murat *et al*, 2013; Lam *et al*, 2013). As a consequence of the damage and telomere dysfunction that PDS induces, treatment causes long-term growth arrest and senescence in several human cancer cell lines (Müller *et al*, 2012).

The bisquinolinium compound PhenDC3 is large, planar, aromatic and is also thought to stabilise G4 via stacking onto the terminal quartet (Figure 1.2F) (De Cian *et al*, 2007b). PhenDC3 also modulates gene expression in cell lines; and can inhibit telomerase and replication fork progression *in vitro* (Halder *et al*, 2012; Madireddy *et al*, 2016; Castillo Bosch *et al*, 2014; De Cian *et al*, 2007b, 2007a). In contrast to PDS, the induction of a DDR following PhenDC3 treatment and whether this can be exploited in BRCA-deficient backgrounds has yet to be investigated.

PDS and PhenDC3 show different *in vitro* affinities for the 5'- and 3'- tetrads (Le *et al*, 2015). Biophysical G4-affinity differences are also reported for TMPyP4 and RHPS4: TMPyP4 has similar affinity for telomeric and promoter

G4s (Halder & Chowdhury, 2007; Lemarteleur *et al*, 2004) whereas RHPS4 mainly targets the telomere (Oganesian *et al*, 2006). Whether these *in vitro* differences translate to variation in the cellular phenotypes they induce, has yet to be systematically investigated in human cells.

1.4.2.5 carboxylated PDS – a step towards RNA-G4 specific ligands

Due to the transcriptional link between RNA and DNA, it is difficult to dissect RNA-G4 and DNA-G4 phenotypes following treatment of cell lines with pan-G4 ligands with broad specificity. There is an unmet need for the production of DNA and RNA specific G4-ligands, to more accurately segregate RNA and DNA effects. As a step towards this a carboxylated version of PDS (cPDS, Figure 1.2G) was developed, which selectively stabilised RNA-G4 over DNA-G4 *in vitro* (Di Antonio *et al*, 2012) and in cells (Biffi *et al*, 2014a). However this remains to be used in cellular phenotype experiments.

1.5 Hypothesised protein interactions and cellular roles of G4s

Increasingly, endogenous proteins are implicated in the modulation of cellular G4 landscape and play important roles via G4-interactions. Two broad classes of G4-binding proteins have been studied: G4-resolving proteins, helicases and nucleases, and non-enzymatic "docking" partners that are recruited by G4s to mediate various biological processes.

1.5.1 DNA G4-helicases and their regulatory roles

DNA G4s are thought to regulate transcription and replication. The proposed presence of dedicated G4-helicase machinery, to unwind these structures allows them to act as dynamic and regulatable switches (Patel & Donmez, 2006; Suhasini & Brosh, 2013). Artificial stabilisation of G4s by ligands compromises this switch-like role by antagonising helicase activity and encouraging the equilibrium towards the folded state. This may explain the transcriptional inhibition phenotypes imparted by multiple G4-ligands following their application to cells. The simplest model for this transcriptional inhibition is that DNA-G4s are RNA polymerase obstacles. This is supported by the observation that human fibroblast treatment with CX-4561 inhibits RNA polymerase 1, leading to activation of ATM/ATR DNA damage signalling (Quin *et al*, 2016). DNA replication studies in chicken DT40 cells offer an alternative explanation: G4 inhibition of the DNA polymerase results in the inheritance of incorrect histone epigenetic modifications and subsequent transcriptional silencing (Sarkies *et al*, 2010, 2012). G4 inhibition of DNA polymerase has also been linked to controlling replication, specifying replication origins and ensuring faithful and ordered DNA replication (reviewed in Rhodes & Lipps 2015). Concordant with this, PDS treatment induces cell cycle arrest (Rodriguez *et al*, 2012).

The clearest link between helicases and DNA G4s *in vivo* is provided by Pif1, a highly conserved DNA replicative helicase found in virtually all eukaryotes (Bochman *et al*, 2010). Both *S.cerevisiae* and human Pif1 unwind G4s *in vitro* (Paeschke *et al*, 2013; Ribeyre *et al*, 2009; Sanders, 2010) and yeast

genome-wide ChIP-Seq studies suggest that Pif1 binds to G4 motifs *in vivo* (Paeschke *et al*, 2011). Furthermore, in Pif1-deficient yeast, introducing artificial sequences encoding DNA-G4s induces replication stress and reduced cell growth (Paeschke *et al*, 2011). A functional role of human Pif1 (hPif1) is less understood. However supporting a role in DNA replication, hPif1 is heavily cell cycle regulated by proteosomal degradation, with peak abundance in G2-phase (Bochman *et al*, 2010). Further, siRNA-induced depletion of hPif1 slows replication fork progression and increases replication arrest in normal cells (Gagou *et al*, 2014). This is non-lethal to normal cells and not associated with double-strand break formation, but in several cancer cell lines where the replication stress high, hPif1 depletion induced apoptosis and reduced proliferation (Gagou *et al*, 2014). This highlights that cancer-associated genetic backgrounds exist that are sensitive to unresolved G4s, in this case due to hPif1-deficiency but also perhaps achievable by G4-stabilising ligand treatment. Exploring the latter, to define genotypes sensitive to G4-stabilising ligand treatment, is central to the aims of this thesis.

Another class of helicases with *in vitro* G4-unwinding activity are members of the Iron-Sulphur (Fe-S) cluster domain family (Wu & Brosh, 2012), including human FANC-J (Fanconi Anaemia Group J protein) and its *C.elegans* homologue DOG-1 (deletion of G-rich tracts) (London *et al*, 2008; Wu *et al*, 2008). FANC-J mutations cause the bone marrow failure disorder Fanconi's Anaemia (FA) and predisposition to hereditary breast and ovarian cancers (Seal *et al*, 2006; Levitus *et al*, 2005). FANC-J is proposed to promote DNA synthesis by G4 resolution, with deficiencies in FANC-J sensitise cells to G4-

stabilising ligands (Brosh & Cantor, 2014; Schwab *et al*, 2013; Bharti *et al*, 2013). Accordingly, G4 nuclear staining is increased in FANC-J deficient murine and human cells suggesting increased DNA-G4 levels (Henderson *et al*, 2014).

The transcriptional helicases XPB and XPD are also members of the Fe-S family (White, 2009); mutations in these proteins cause *Xeroderma pigmentosum*, a syndrome associated with melanoma predisposition (Kamileri *et al*, 2012). XPB and XPD mapping in the chromatin of human fibrosarcoma cells by ChIP-seq suggests that 40 % of binding sites overlap with a PQS (Gray *et al*, 2014). Furthermore, the regulator of telomere length (RTEL1), another Fe-S helicase (Wu & Brosh, 2012) is suggested to be responsible for the resolution of telomere G4s during S-phase, thus RTEL1 deficiencies are associated with telomere fragility (Vannier *et al*, 2012, 2014).

Likewise, the RecQ class of helicases bind and resolve DNA-G4s *in vitro* (Huber *et al*, 2006; Budhathoki *et al*, 2014). Members include the StyRecQL ciliate G4-helicase (Postberg *et al*, 2012) and Bloom's (BLM) (Sun *et al*, 1998), Werner (WRN) (Kamath-Loeb *et al*, 2001) and Rothmund-Thomson (RecQ4) (Hickson, 2003) syndrome proteins, mutations in which cause increased cancer susceptibility and premature aging (Suhasini & Brosh, 2013). Consistent with a role for RecQ helicase unwinding of cellular G4s, BLM- and WRN-deficient fibroblasts show down-regulation of genes with promoter PQS (Nguyen *et al*, 2014; Johnson *et al*, 2010) and an increase in nuclear G4 structures as judged by BG4 immunofluorescence (Drosopoulos

et al, 2015). Telomeric G4 resolution by WRN is reportedly essential to prevent replication fork stalling during lagging strand synthesis (Aggarwal *et al*, 2011), with telomere loss and premature senescence observed in WRN-depleted cells (Crabbe, 2004). In some ways, phenotypes associated with helicase deficiencies are similar to that following treatment of cells with G4-stabilising ligands. For example, fibroblasts derived from Werner and Bloom syndrome patients show altered gene expression at PQS sites (Brosh, 2013; Mendoza *et al*, 2016), similar to the transcriptional inhibition profile described above for several ligands in section 1.4.2.1. This may reflect a similar biological mechanism arising from unresolved G4-structures or may arise due to limited techniques to date available to investigate G4-associated phenotypes.

Although several G4-resolving helicases have been identified and studied, the full repertoire is unknown, thus there remains an unmet need to expand our knowledge of endogenous machinery responsible for regulating G4-formation. As helicase deficiencies may be expected to increase the numbers of G4s in cells, this may exacerbate sensitivity to G4-stabilisng ligands. Thus the genetic screening experiments performed within this thesis may uncover novel helicase candidates.

1.5.2 Other enzymatic interactors of DNA-G4s

In addition to helicases, nucleases are also emerging as important G4 regulators. The yeast endonuclease KEM1 can bind to and cleave at DNA-G4 sites (Liu & Gilbert, 1994) and XRN1, its murine and human homologue binds

both RNA- and DNA-G4s *in vitro* (Bashkirov *et al*, 1997)*.* In addition, viral mRNA G4 structures inhibit human XRN1 to prevent RNA decay of the viral genome (Charley *et al*, 2018). Other human nucleases DNA2, FEN1 and EXO1 cleave DNA G4s *in vitro* (Vallur & Maizels, 2008), and depletion of these genes cause telomere dysfunction (Saharia *et al*, 2008; León-Ortiz *et al*, 2014), suggesting that they may regulate telomeric G4 *in vivo*. Another putative enzymatic G4 interactor, is the translesion polymerase REV1, deficiencies in which cause G4-associated replication fork stalling in DT40 cells (Sarkies *et al*, 2010). Finally, BG4 staining of human HEK cells showed colocalisation of telomerase and a specific subset of telomeric G4, inferring a telomerase recruitment role by G4s during meiotic telomere extension (Moye *et al*, 2015).

1.5.3 DNA G4-docking partners

In addition, experimental evidence is increasingly suggesting that both RNA and DNA G4s can recruit protein partners to influence different aspects of genome and transcriptome function. Well-characterised examples of G4 binders are found at the telomere. For example, TLS/FUS can also bind both telomeric DNA-G4 and TERRA RNA-G4 (Takahama *et al*, 2013) and multiple shelterin components interact with G4s. G4-binding proteins are an ongoing area of investigation and there is an unmet need for its expansion.

1.5.3.1 Telomeric DNA-G4 interactors

The mammalian telomere is bound by shelterin, a heteromeric complex of six proteins: TRF1, TRF2, POT1, TPP1, TIN2 and Rap1 (Palm & de Lange,

2008). Shelterin provides a telomere protection role preventing recognition of the chromosome end as a double strand break (DSB), which would otherwise aberrantly activate the DDR. Replicative senescence occurs when telomeres are too short to interact with shelterin (Feuerhahn *et al*, 2015). The telomere G-rich overhang is mainly occupied by POT1 (protection of telomeres), while the double-stranded region is bound by TRF1 and TRF2, telomeric repeat binding factors 1 and 2, respectively, with other proteins performing bridging or stabilisation roles (Palm & de Lange, 2008).
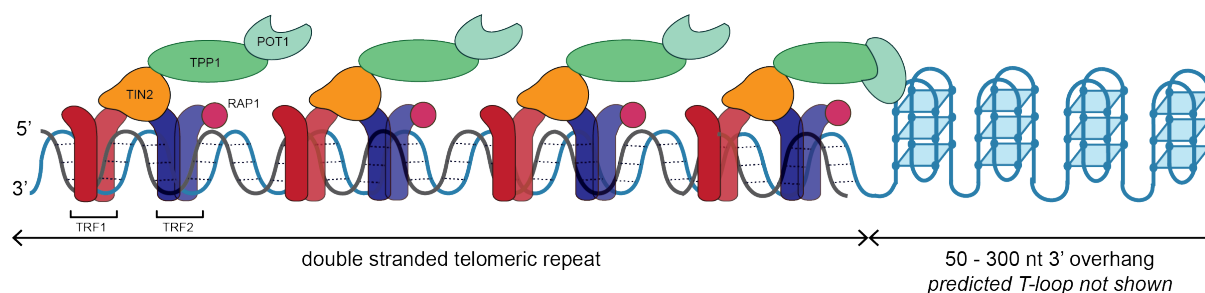
Figure 1.3. Telomere is protected by shelterin and tandem G4-structures
Schematic of the mammalian telomere. Telomeres consist of repetitive TTAGGG stretches forming up to 15 kb of double stranded structure and a G-rich overhang predicted to form a polymorphic array of quadruplexes. The telomere is protected by shelterin, a heterohexameric structure which binds to the telomere via three DNA binding proteins: TRF1, TRF2 and POT1

Consistent with telomeric DNA G4s contributing to telomere protection, TRF2 has *in vitro* telomeric DNA- and RNA-G4 (TERRA, see 1.5.4.2) affinity (Biffi *et al*, 2012). The latter interaction is likely to be structure rather than sequence specific, as TRF2 also has nanomolar affinity for other RNA-G4s, including those found in the 5'-untranslated regions (UTR) of BCL2 and NRAS mRNA (Biffi *et al*, 2012). Further *in vitro* data suggests TRF2 can promote both intra- and intermolecular DNA G4 formation (Pedroso *et al*, 2009). In keeping with a role for TRF2 interaction with G4s, TRF2 depletion in human fibroblasts

increases sensitivity to treatment with the G4-stabiliser RHPS4 (Salvati *et al*, 2015).

Shelterin also recruits the helicases hypothesised to resolve G4-structures: TRF1 recruits BLM (Sfeir *et al*, 2009) and TRF2 interacts with WRN (Opresko *et al*, 2002) and RTEL1 (Vannier *et al*, 2012; Sarek *et al*, 2015). Furthermore, FRET analysis shows the POT1-TPP1 complex can destabilise telomeric G4s *in vitro* (Hwang *et al*, 2012). Other proteins shown to interact with telomeric DNA-G4s include hnRNPA1 and hnRNPD (Krüger *et al*, 2010; Enokizono *et al*, 2003), UP1 (Hudson *et al*, 2014) and the single stranded binding protein RPA (Safa *et al*, 2014; Qureshi *et al*, 2012). This interaction with RPA shifts telomeric G4 equilibrium towards an unfolded state (Safa *et al*, 2014). The roles of these interactions beyond telomere protection are not fully characterised, however the yeast protein Rif-1 binds telomeric G4s and prevents replication at locations far from this telomeric binding site (Kanoh *et al*, 2015). The above studies firmly suggest the existence of telomeric G4-structures *in vivo* where they are thought to provide a structure specific 'docking' site for multiple proteins.

1.5.3.2 Non-telomeric DNA-G4 interactors

Several proteins are also hypothesised to bind promoter DNA-G4s. For example, the multifunctional phosphoprotein, nucleolin binds the MYC promoter G4 *in vitro* and in HeLa cells and inhibits transcription (González *et al*, 2009). Similarly, ChIP analysis shows that nucleolin also binds the oncogene VEGF promoter DNA-G4 (Sun *et al*, 2011), and may also perform a

transcription regulatory role. Likewise, the DNA-damage recognition protein PARP1 binds several promoter G4s, including *BCL-2, KRAS, MYB, KIT* and *VEGF* (Soldatenkov *et al*, 2008; Cogoi *et al*, 2010) an interaction proposed to promote PARP1 catalytic activity. Secondary to a DDR induction role, the docking of PARP with *KRAS* promoter DNA-G4s activates transcription when co-bound with the transcription factor MAZ1 and contributes to xenograft tumour growth of pancreatic cell lines injected into mice (Cogoi *et al*, 2010, 2013). Additionally a mutant form of p53 that causes a 'gain-of-function' phenotype in several cancers (Brázda, Hároníková, Liao, & Fojta, 2014) binds several promoter G4s *in vitro* acting as a transcriptional cofactor (Quante *et al*, 2012), providing a novel link between DNA-G4s and cancer via this p53 role.

Beyond promoter G4 interactions, the Origin Recognition Complex (ORC) is shown to bind both RNA and DNA-G4 *in vitro* via an RNA binding domain in ORC1, adjacent to the ATPase domain (Hoshina *et al*, 2013). This is consistent with the enrichment of PQS at replication origins (Besnard *et al*, 2012)*.* Further, nucleophosmin (NPM1), required for ribosomal RNA maturation and export, and frequently overexpressed in acute myeloid leukaemia (Falini *et al*, 2005) binds ribosomal DNA-G4, *in vivo* and *in vitro* (Chiarella *et al*, 2013). Moreover, DNA-G4s interact with several chromatin remodellers although the full repertoire is unknown. For example the SWI/SNF chromatic remodeller ATRX, binds GC rich sequences at the telomere and centrosome encouraging heterochromatin formation (De La Fuente *et al*, 2011), including multiple PQS (Clynes *et al*, 2013) and has DNA-

G4 selective affinity *in vitro* (Law *et al*, 2010). Synthetic lethality between ATRX-deficiencies and telemostatin treatment of cells has been reported (Watson *et al*, 2013), further supporting an endogenous DNA-G4 interaction. Such inactivation frequently occurs in glioma and ATRX-deficient neuroepithelial progenitor cells exhibit large-scale alterations in chromatin accessibility and transcription exclusively at normal ATRX binding sites (Danussi *et al*, 2018).

G4s may also prevent DNA-protein interactions. For instance, bioinformatic analysis suggests that DNA-G4s are nucleosome exclusion signals in yeast, nematode and mammalian cells (Halder *et al*, 2009; Wong & Huppert, 2009). This is experimentally supported by ChIP-seq data (see section 1.3.1) which indicates that G4s preferentially form in nucleosome free regions (Hänsel-Hertsch *et al*, 2016). Additionally, the preferential overlap of PQS with transcription factor binding sites (Huppert & Balasubramanian, 2005) coupled with the inverse correlation between gene expression levels and the number of promoter PQS (Balasubramanian *et al*, 2011), could suggest G4-formation prevents transcription factor interaction with the DNA-duplex. However, perhaps incompatible with this, BG4 ChIP-seq indicated that more transcriptionally active genes had the highest PQS density (Hänsel-Hertsch *et al*, 2016). Further, some transcription factors, such as SP1 bind G4s *in vitro* (Raiber *et al*, 2012)*,* suggesting potential mechanisms for G4s to also promote transcription.

Another hypothesised role for DNA G4 is in sister chromatid alignment during meiosis (Sen & Gilbert, 1988). Although there is no direct evidence for this in mammals, the yeast protein Hop1 was observed to link sister chromatids during synapsis via formation of intermolecular G4s (Anuradha & Muniyappa, 2004) and KEM1-depleted cells (yeast homologue of XRN1) were blocked in meiosis (Liu & Gilbert, 1994). Additionally, investigations in gram-negative bacteria, suggest a DNA-G4 is necessary at recombination sites for pilin antigenic variation (Cahoon & Seifert, 2009; Walia & Chaconas, 2013). Similar roles remain to be explored in human cells.

1.5.4 RNA G4 roles

Although less studied than DNA-G4s, RNA-G4s are emerging as equally important regulatory endogenous structures and are reportedly more stable and refractory to unfolding than DNA-G4, with direct evidence provided by NMR and X-Ray crystallography (reviewed in Cammas & Millevoi, 2016). The mostly single stranded nature of RNA may favour G4 formation compared to predominantly double-stranded DNA (reviewed in Brázda et al., 2014). The ribonucleotide 2'-OH group also contributes to this RNA-G4 rigidity and stability (Saccà *et al*, 2005).

1.5.4.1 Translation and mRNA processing

Multiple studies report that *in vitro* G4-stabilisation of RNA G4-structures can inhibit translation (Bugaut & Balasubramanian, 2012), first experimentally shown with the 5'UTR G4 in *NRAS* mRNA (Kumari *et al*, 2007). Subsequently, translationally repressive 5' UTR G4s have been identified in

several other mRNAs, including the tumour suppressor *BCL2* and the shelterin component *TRF2* (Arora *et al*, 2008; Balkwill *et al*, 2009; Gomez *et al*, 2010; Morris & Basu, 2009; Shahid *et al*, 2010; Beaudoin & Perreault, 2010). RNA-G4s can also activate translation by protein recruitment. For example, a G4 found in the IRES of the VEGF promotes translation by recruiting the ribosome (Morris *et al*, 2010). Similarly 5'UTR RNA-G4s increase eIF4A-dependent oncogene translation (Wolfe *et al*, 2014) and induce translation of the oncogenic inflammatory cytokine TGFβ (Agarwala, Pandey, Mapa, & Maiti, 2013). Additionally, RNA G4s can cause translational recoding by inducing +1 frameshifts, both *in vitro* and in HEK cells, and this is exacerbated by PhenDC3 treatment (Yu *et al*, 2014).

RNA G4s may also regulate mRNA processing, with roles ranging from splicing, as suggested for IGFII, TP53 and BACE1 (Christiansen *et al*, 1994; Fisette *et al*, 2012; Marcel *et al*, 2011) to alternative polyadenylation for 3' UTR G4s in LRP5 and FXR1 pre-mRNAs (Beaudoin & Perreault, 2013). As for DNA-G4s, such processes can be subverted by treatment of cell lines with G4-stabilising ligands, for example, applying the stabiliser compound 12459 to A549 cells prevented correct splicing of the TERT pre-mRNA (Gomez *et al*, 2004a). Finally, mRNA G4s also have a proposed role in localising mRNA to defined subcellular compartments in neurons including to the synapses, surface membrane and cytoskeleton, via interaction with COP-1 vesicle components (Todd *et al*, 2013; Subramanian *et al*, 2011).

1.5.4.2 Non-coding RNA G4 roles

RNA-G4 structures are also found in non-coding regions. One well characterised ncRNA G4 example, is Telomeric Repeat containing RNA (TERRA) (Schoeftner & Blasco, 2009; Azzalin & Lingner, 2014), transcribed from the C-rich strand of the telomere (Maicher *et al*, 2014). The resultant transcripts contain an average of 400 bp of UUAGGG sequence, though transcripts can also be several kilobases in length (Porro *et al*, 2010). Consequently TERRA is thought to form stable intramolecular tandem RNA-G4s, *in vitro* and when synthetic TERRA oligonucleotides are transfected in cells (Biffi *et al*, 2012; Xu *et al*, 2010; Martadinata *et al*, 2011)**.** TERRA G4 interacts with the telomere-binding protein TLS/FUS (translocated in liposarcoma) *in vitro* (Takahama *et al*, 2013). More recently, TLS/FUS has been to shown to simultaneously bind the TERRA RNA-G4 and telomeric DNA-G4, to form a ternary complex that is thought to be important for telomere length regulation (Kondo *et al*, 2018). Other TERRA functions include oncogene transcriptional suppression (Hirashima & Seimiya, 2015) and telomere heterochromatin maintenance (Wang *et al*; Schoeftner & Blasco, 2009), possibly mediated by forming DNA-RNA hybrid intermolecular G4s with telomeric DNA (Xu *et al*, 2012). At the telomere TRF2 has also been shown to bind DNA-RNA G4 hybrids (Pedroso *et al*, 2009) and these intermolecular structures are also observed in the mitochondria, where they are necessary for transcriptional termination (Zheng *et al*, 2014). Finally, ncRNA G4s formed in intron lariats are thought to regulate splicing and facilitate immunoglobulin class switching (Zheng *et al*, 2015).

1.5.5 RNA G4 hypothesised enzymatic and non-enzymatic protein interactions

1.5.5.1 RNA helicases

As for DNA-G4 structures, various helicases also unwind RNA G4s and RNA-DNA hybrid G4s (reviewed in Cammas & Millevoi, 2016). These include DEAH box RNA-helicases DHX36 and DHX9 (Tanner & Linder, 2001). DHX36 for example tightly binds and specifically unwinds RNA and DNA-G4 *in vitro* (Creacy *et al*, 2008; Giri *et al*, 2011; Chen *et al*, 2015) and in cells (Vaughn *et al*, 2005). This G4-resolving activity is required for gene expression regulation, including oncogenes such as YY1 (Huang *et al*, 2012) and the tumour suppressor PITX (Booy *et al*, 2014). For YY1 for example, this was explored via plasmid based luciferase assays in HEK-293T cells, where destabilising mutations of the 5'-UTR G4s were sufficient to increase luciferase expression, supporting the notion that the presence of RNA-G4s can inhibit translation. Similarly, DHX9 unwinds G4-structures in co-transcriptionally formed R-loops *in vitro* (Chakraborty & Grosse, 2011). Furthermore DHX9 interacts with BRCA1 (Anderson *et al*, 1998) and the DNA-G4 helicase WRN (Friedemann *et al*, 2005) suggesting a recruitment role to D-loop G4 in addition to the RNA G4 related roles, to repair and prevent DNA-G4 induced damage, respectively. A more specific RNA G4 resolving protein provided by the oncogene eukaryote initiation factor 4A (eIF4A) has been revealed by ribosomal footprinting to target 5'UTR RNA-G4s (Wolfe *et al*, 2014). During translation initiation, the ATP-dependent RNA helicase eIF4A enables ribosome translocation by removing mRNA secondary structures (Jackson *et al*, 2010), including but not limited to, G4s.

1.5.5.2 RNA binding partners

Due to the structural similarity between RNA- and DNA-G4, their binding partners show significant overlap and include nucleolin, TRF2, FUS/TLS and hnRNPA1 (reviewed in Brázda et al., 2014). For example TRF2 and FUS/TLS can bind both telomeric DNA- and TERRA RNA-G4 (Biffi *et al*, 2012; Takahama *et al*, 2013; Kondo *et al*, 2018). Specific RNA-G4 binders have also been identified including FMRP, which binds to multiple mRNAs with G4-forming potential, including its own mRNA in a feedback mechanism (Darnell *et al*, 2001; Brown *et al*, 2001; Schaeffer *et al*, 2001). As FMRP is found within polysomes, this interaction is thought to regulate translation (Darnell *et al*, 2001). Concordantly, FMRP overexpression increases oncogenic mRNA expression in melanoma cell lines (Zalfa *et al*, 2017). The anti-apoptotic protein, Aven also binds mRNA-G4 and activates translation of key leukaemia oncogenes *MLL1* and *MLL4*, which can be further advanced by cobinding of the helicase DHX36 (Thandapani *et al*, 2015).

RNA-G4s perform both inhibitory and activatory roles dependent on their binding partners. The increasing identification of RNA-G4 interacting partners will inform us on the full extent of RNA-G4 biology and regulation. A recent analysis used dimethyl sulfate (DMS) treatment coupled with high-throughput sequencing to map RNA G4s in yeast and mouse embryonic stem cells (Guo & Bartel, 2016). This study showed that G4 structures that formed *in vitro* did not protect the RNA from DMS methylation *in vivo,* which was interpreted as RNA-G4s being globally unfolded in eukaryotic cells. However, RNA G4 formation is dynamic (Harkness & Mittermaier, 2017) and recent work has

attempted to track these dynamics in real time in live cells via use of the G4-specific fluorescent probe QUMA-1 (Chen *et al*, 2018). Due to this dynamic nature, as soon as RNA-G4s unfold, the sequence is amenable to DMS methylation and is therefore 'trapped' in the unfolded state. Thus the experimental conditions of Bartel *et. al* may have promoted depletion of RNA-G4s and as such, the findings do not necessarily conflict with the RNA regulatory roles discussed here.

1.6 The link between DNA-quadruplexes and genome instability

1.6.1 Defining genome instability

Genome instability refers to mechanisms causing recurrent changes in genome structure and/or nucleotide composition, and is characteristic of many mammalian cancer genomes (Loeb, 1991; Hartwell, 1992; Hanahan & Weinberg, 2011). Instability is associated with problems in DNA replication, DDR activation and is increased in repetitive DNA and at sequences with the potential to form non-B DNA structures (Aguilera & García-Muse, 2013). Various types of genome instability exist: chromosomal instability (CIN), tandem repeat/microsatellite instability and small nucleotide alterations (Aguilera & García-Muse, 2013; Pikor *et al*, 2013). CIN refers to continuous changes in chromosome structure and integrity (Pikor *et al*, 2013) and is the most prevalent form of genomic instability in solid cancers (Lengauer *et al*, 1998). Repeat instability, prevalent in colon cancers (Thibodeau *et al*, 1993) is associated with expansion and retraction of 1-8bp microsatellite repeats

(Ellegren, 2004) due to DNA polymerase slippage (Pearson *et al*, 2005; McMurray, 2010). Finally, nucleotide instability is defined by increased deletions, insertions or substitutions of small nucleotide numbers (Pikor *et al*, 2013). This is associated with somatic mutation accumulation in cancers (Al-Tassan et al., 2002) and may be enhanced by non-canonical structure formation, such as G4s which can cause polymerase stalling and replication errors (Aguilera & García-Muse, 2013). Although G4-associated genome instability has been shown in yeast and *C.elegans* (see section 1.6.2), in humans it is not fully understood beyond BRCA deficiencies causing G4-ligand sensitivity (McLuckie *et al*, 2013; Zimmer *et al*, 2016; Xu *et al*, 2017).

1.6.2 Genome instability associated with deficiencies in G4-binding proteins

Fanconi Anaemia, Werner's and Bloom syndrome diseases are characterised by genetic instability (Bharti *et al*, 2013) with rearrangements in Werners encompassing chromosomal fusions (dicentric chromosomes and translocations) and deletions (Salk *et al*, 1981; Fukuchi *et al*, 1989; Gebhart *et al*, 1988) (see section 1.5.1). The proteins mutated in these syndromes are putative DNA-G4 helicases, suggesting that failure to resolve G4s contributes to an instability phenotype. A more direct link between helicases and G4-induced instability has been shown in Pif1-deficient yeast, in which break-points occur near both endogenous and experimentally inserted G4-forming sequences leading to CIN if unrepaired (Paeschke *et al*, 2011, 2013; Piazza *et al*, 2012). However, Pif1 is the only thoroughly characterised yeast G4-resolvase, whereas 31 putative DNA helicases have been identified in

humans (although not all are G4-specific) many of which are redundant (Brosh, 2013). This redundancy makes it difficult to draw direct parallels between findings in yeast and mammals, necessitating the systematic investigation of helicase deficiencies in human cells.

Yeast studies have also revealed that G4-induced instability can be exacerbated by non-helicase deficient genotypes such as the depletion of the telomere binding protein Cdc13 (Piazza *et al*, 2012), which performs the function of human POT1 (Churikov *et al*, 2006) and is known to destabilise G4-structures *in vitro*. This telomere instability phenotype could be rescued by overexpression of Stm1 (Hayashi & Murakami, 2002) or deletion of Cig21 (Downey *et al*, 2006) (both telomere binding proteins) but was antagonised by overexpression of the yeast RecQ G4-helicase Sgs1 (Huber *et al*, 2002; Hayashi & Murakami, 2002; Han *et al*, 2000). Yeast Stm1 mutants were resistant to the G4-stabilising compound NMM whereas Cig21-deficiencies correlated with increased NMM sensitivity (Hershman *et al*, 2008; Ren & Chaires, 1999; Kreig *et al*, 2015). This lead to the hypothesis that Stm1 overexpression and Cig21 deficiencies both stabilise telomeric DNA-G4s and inducing NMM sensitivity by increasing the available G4-ligand binding sites (reviewed in Johnson et al. 2008). This phenomenon of deficiencies in key G4-regulators inducing G4-ligand sensitivity or resistance, successfully highlighted here in yeast, is yet to be explored in human cells. In addition to being dependent on the yeast genotype the ability of G4-structures to induce

genomic instability, was subsequently shown to correlate with G4 thermal stability and number (Lopes *et al*, 2011; Piazza *et al*, 2010).

Helicase deficiencies in *C.elegans* also result in G4-associated instability, with DOG-1 depletion (homologue of human FANC-J) causing specific loss of G-rich tracts during replication (Cheung *et al*, 2002; Kruisselbrink *et al*, 2008; Tarailo-Graovac *et al*, 2015). More recently yeast G4-associated instability has been extended to topoisomerase I (TOP1), deficiencies of which promoted gross chromosomal rearrangements at a highly transcribed exogenous G4-forming sequence (Yadav *et al*, 2014). Whether TOP1 plays a similarly critical role at highly transcribed endogenous sequences in human cells remains to be explored.

1.6.3 Telomere instability and chromosomal fusions

The telomere forms tandem DNA-G4s and is bound by shelterin (see section 1.3.3). Shelterin prevents recognition of the telomeric end as a DSB, to prevent deleterious chromosomal fusions via aberrant DDR activation (Palm & de Lange, 2008; de Lange, 2009; O'Sullivan & Karlseder, 2010). Without proper regulation, the repetitive nature and the G4-forming ability render telomere sequences "hotspots" for instability. Such telomeric instability (a specific subset of CIN) and chromosomal fusions often contribute to cancer malignancy, for example, by placing oncogenes under the control of constitutively active promoters (Gisselsson *et al*, 2001; Rudolph *et al*, 2001;

Chang *et al*, 2001; Maser & DePinho, 2002). Whether the formation of telomeric G4s and their improper regulation can contribute to this telomeric instability remains to be systematically investigated in humans. Telomeric G4-repeats as sources of genome instability are not restricted to telomere fusion at chromosome ends. In ALT-positive cancers, telomeric sequences are aberrantly incorporated throughout the genome and encourages intrachromosomal fusions (Marzec *et al*, 2015). Non-telomeric G4-motifs are also enriched at chromosomal breakpoints in several human cancers and paediatric neurological diseases (Bose *et al*, 2014; De & Michor, 2011; Nambiar *et al*, 2011) suggesting a role for G4 structures in driving translocation events.

1.7 Quadruplexes and cancer

Supporting an association between G4 misregulation and cancer, more G4s are found in transformed keratinocytes and human stomach and liver cancer compared to controls (Biffi *et al*, 2014b; Hänsel-Hertsch *et al*, 2016). Similarly the use of naturally fluorescent G4-stabilising small molecules DAOTA-M2 and BMVC indicated greater G4 numbers in cancer cells versus their normal counterparts (Huang *et al*, 2015; Shivalingam *et al*, 2015). Concordantly, G4-ligand treatment causes senescence and/or apoptosis in several cancer cell lines (Neidle, 2017). More recent application of such ligands to mouse xenograft models can cause tumour shrinkage. For example, treatment with the G4-stabilising ligands MM41 (Ohnmacht *et al*, 2015) and CX-5461 (Xu *et al*, 2017) caused reduced pancreatic and BRCA-deficient breast tumour growth respectively.

Inferring a transcriptional role between G4s and cancer, bioinformatics, biophysical and G4 ChIP-seq data suggest that there is an enrichment of G4s in the promoters and 5' UTRs of cancer associated genes and copy number amplifications in human chromatin (Hänsel-Hertsch *et al*, 2016; Chambers *et al*, 2015; Huppert & Balasubramanian, 2005, 2007), which may allow transcriptional modulation of several oncogenes (see section 1.3). Additionally diseases deficient in DNA G4-helicases, as discussed for WRN and BLM, often show cancer predisposition (Brosh, 2013). Whether this cancer association extends to deficiencies in RNA G4 helicases has not yet been characterised.

Indirectly suggesting that telomeric G4s may contribute to carcinogenesis, shelterin mutations are associated with cancer and haematological malignancies. POT1 loss-of-function mutations are associated with a predisposition to familial melanoma (Robles-Espinoza *et al*, 2014; Shi *et al*, 2014), germ line glioma (Bainbridge *et al*, 2015) and chronic lymphocytic leukaemia (Ramsay *et al*, 2013). On the other hand deficiencies in TPP1 predispose to aplastic anaemia (Guo *et al*, 2014) and in TRF2 are found in melanoma (Aoude *et al*, 2015) and glioma (Vannier *et al*, 2014). RNA G4s have also been linked to the regulation of many cancer hallmarks outlined by Hanahan and Weinberg (Hanahan & Weinberg, 2011), due to their presence in the mRNAs of multiple oncogenes including but not limited to hTERT and TRF2 (required for immortality); VEGF and FGF2 (angiogenesis); MT3-MMP, ADAM10 and MST1R (induction of invasion and metastasis) (reviewed in Cammas & Millevoi, 2016).

1.8 Exploring synthetic lethality with G4-ligands

In this chapter I have discussed the multiple hypothetical roles of G4s including transcription, translation and replication regulation. Several of these roles are cancer related, suggesting that chemotherapeutic strategies with G4-stabilising ligands may exist. Exploitation with BRCA-deficient tumours has already been explored (section 1.4.2.1). Beyond BRCA-deficiencies, the understanding of the DNA damage proteins involved in repairing G4-induced damage is incomplete. Investigating DDR-deficient cancer genotypes susceptible to G4-targeting ligands may help identify these DDR proteins. Additionally, synthetic lethal strategies other than DNA damage deficiencies may exist. In support of this, in yeast, alterations in telomere G4-binding proteins resulted in NMM sensitivity (see section 1.6.2) and ATRX-deficient mouse embryonic fibroblasts are sensitive to telomestatin. Similar G4-ligand sensitive genotypes have yet to be characterised in human cells, and is one of the major aims of this thesis.

Methods for exploring such synthetic lethalities in a systematic and unbiased manner have been provided by the functional genomics field, where gene editing at both the RNA and DNA level has enabled screening of the pharmacological response of different cell genotypes.

1.8.1 Genetic screening as a step towards personalised medicine approaches

Extensive heterogeneity exists between individual responses to cancer treatment, which is often dependent on the genotype of the patient and/or

tumour cells, whether this be germline genetic polymorphisms or somatic mutation patterns which are dominant in the tumour (reviewed by Relling & Evans, 2015; Daly, 2017). Therefore, there is merit in understanding the genetic component dictating treatment efficacy. The sequencing of the human genome (Lander *et al*, 2001), was the first major step towards increasing the knowledge of our genetic underpinning and launched the pharmacogenomics field, to allow development of therapies that are specific to genetically identifiable subgroups within the population (i.e. personalised medicine). This 'subgroup' can either refer to people within a population, or to tumour cells within a 'normal' organ (i.e. genetic profiling of a specific cancer). As early as 1999, the genotype of lymphoblastic leukaemia cells was used to reliably dictate cancer aggressiveness and treatment intensity (Pui & Evans, 1998). As the functional genomics field expands and the genetic tools employed are improved, our ability to identify genetic backgrounds that are disease related and/or sensitive to certain chemotherapies has increased. This has yet to be fully and systematically applied to G4-stabilising ligands, and will be an important investigation in extending the knowledge of how these structures can be therapeutically exploited in a synthetic lethality approach.

Next, the available tools will be illustrated with specific examples from the literature, highlighting some of the key parameters for screening. For brevity, only discussed are pooled library techniques, that enable targeting of multiple genes within the same experiment on both genome-wide and specific subset scales. This is something I set out to achieve with G4-stabilising ligands. It is of note that the aim of this thesis differs somewhat from other screening

approaches with pharmacological ligands, in that I use synthetic lethality to G4-stabilising ligands as a proxy for investigating sensitivities to stabilisation of G4s.

1.8.2 Comparing functional genomic tools

To date, gene editing strategies fall in to two broad categories, either disrupting gene expression at the DNA level, or at the RNA level (RNA interference/RNAi). RNAi was discovered following the observation that injecting dsDNA into *C. elegans* could cause gene silencing, with the resultant phenotype indicating gene function (reverse genetics) (Fire *et al*, 1998). This was subsequently adapted for gene knockdowns in human cells (Elbashir *et al*, 2001). RNAi exploits an endogenous process by which small, genome-encoded double-stranded/short hairpin RNAs are cleaved via RNase III-like enzymes to form 21-23 bp small interfering RNAs (siRNAs) (reviewed in Perwitasari, Bakre, Tompkins, & Tripp, 2013). These RNA guides translationally silence genes by helping the RNA Induced Silencing Complex (RISC), including Argonaute (Ago) to target and cleave/translationally repress target mRNAs (Wilson & Doudna, 2013; Carthew & Sontheimer, 2009). By introducing synthetic small RNAs into cells, either by transient siRNA transfection or virally integrating a genome encoded short hairpin RNA (shRNA) (Mohr *et al*, 2014), this endogenous mechanism can be manipulated.

Alternatively, the identification and exploitation of sequence specific nucleases allows genomic DNA editing, initially using zinc-finger nucleases (ZFN), followed by transcription activator like effector nucleases (TALENs)

(Gaj *et al*, 2013). The modular structure of these enzymes enables 'swapping in' of synthetic DNA binding domains to guide the nuclease domain to target sequences and create gene knockouts via DSB-induced frameshift mutations (Sung *et al*, 2013). A more recent method adapts the bacterial CRISPR/Cas9 nuclease "immune" system (Barrangou *et al*, 2007). Modifying the Cas9 nuclease allows targeting of endogenous genomic loci, rather than native viral DNA substrates (Doudna & Charpentier, 2014). Whereas Cas9 uses a short guide RNA (sgRNA) for target site recognition, TALENs use the DNA binding domain amino acid sequence (Mali *et al*, 2013). Being a separate entity, the former is more easily modifiable and thus CRISPR/Cas systems have become the favoured method of DNA genome editing.

The use of such DNA-targeting nucleases provides a knockout while RNAi, where the genome copy of the gene remains intact and mRNA targeting is rarely complete, usually represents a gene knockdown. This can give rise to hypomorphic phenotypes (reviewed in Boettcher & McManus, 2015). Additionally, as RNAi acts on cytoplasm transcripts not genomic DNA, target accessibility is not influenced by the chromatin state (Kuscu *et al*, 2014; Wu *et al*, 2014). However, strictly nuclear RNA transcripts such as long-non coding RNAs, are therefore ineffectively silenced by RNAi (Derrien *et al*, 2012; Fatica & Bozzoni, 2014). As such, these approaches ask slightly different biological questions and the pros and cons of each will be more systematically discussed in chapter 2.

1.8.3 Synthetic Lethal Screening

One extensive use of CRISPR and RNAi is to screen synthetic lethal interactions in anticancer drug discovery (reviewed in Gerhards & Rottenberg 2018). This is particularly useful in the reprofiling of drugs for alternative targets, to circumvent the arduous process of novel drug development (Chong & Sullivan, 2007; DiMasi *et al*, 2003, 1995). Synthetic lethality involves genetic interactions, where two isolated mutations do not affect cell viability, but are lethal in combination (Figure 1.4A) (reviewed in Nijman, 2011). By identifying cancer specific mutations and pharmacologically mimicking a synergistic mutation, synthetic lethality approaches have been used to broaden the chemotherapeutic window of cancer versus normal cells (reviewed in Chan & Giaccia, 2011) (Figure 1.4B-C). Genome editing to identify drug sensitive backgrounds has been used as early as 1997 as proof-of-concept (Hartwell, 1992; Hartwell *et al*, 1997) in which FDA-approved drugs were screened against isogenic budding yeast DDR mutant strains, ultimately identifying different DDR synthetic lethalities with two cancer chemotherapeutics, cisplatin and mitoxantrone.
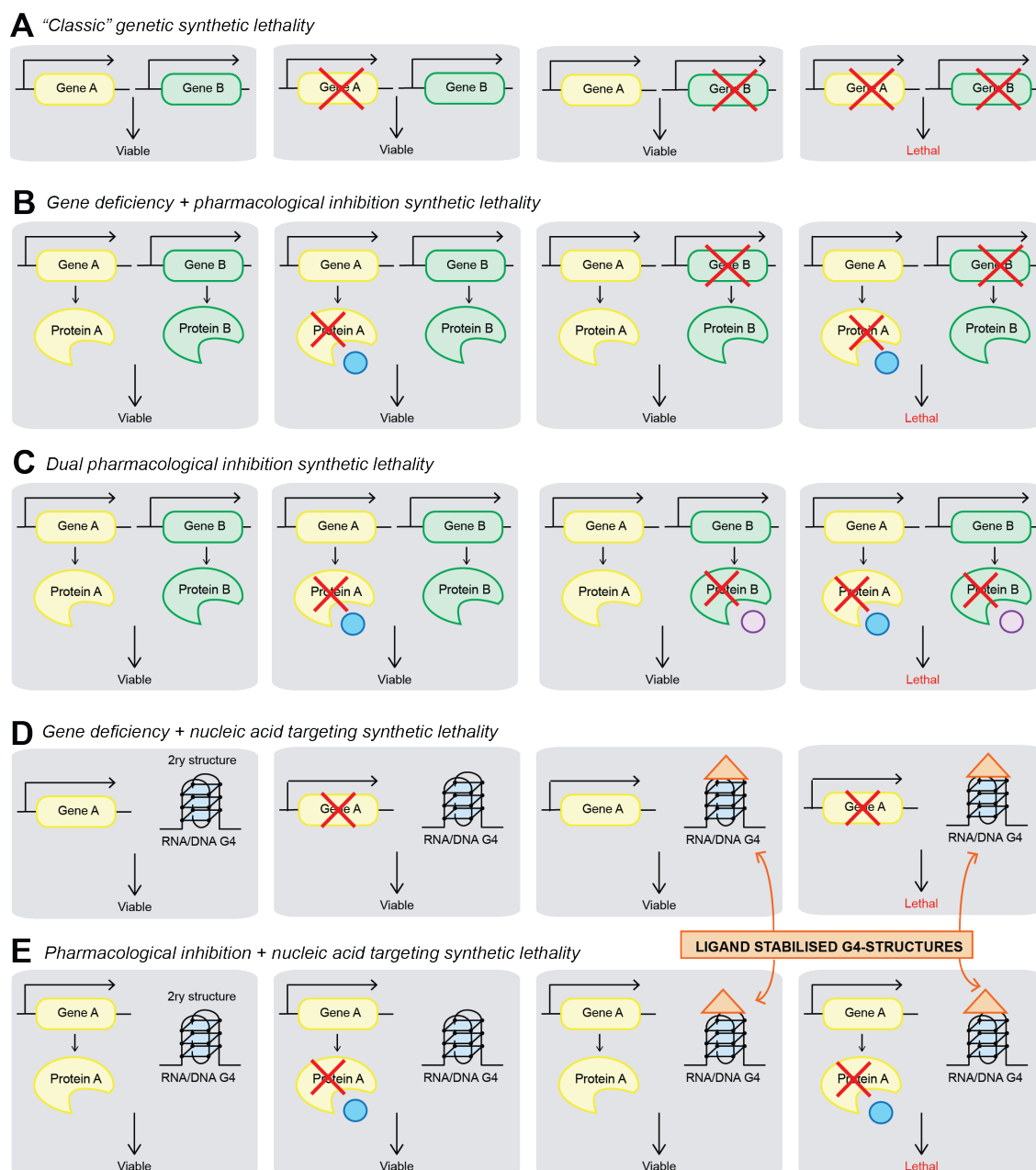
Figure 1.4. Synthetic lethality approaches to identify potential chemotherapeutic sensitive genotypes and/or combinatorial drug targets

(A) Inactivating mutations in gene A and gene B,

(B) Inactivating mutation in gene A and pharmacological inhibition of protein B,

(C) Pharmacological inhibition of gene A and B, are sublethal in isolation but lethal in combination

(D-E) Extension of synthetic lethality methodology to explore gene deficiencies to nucleic acid targeting ligands, for example G4-stabilising ligands: (D) Inactivating mutation in gene A or (E) Pharmacological inhibition of protein A is sublethal in isolation but sensitises cells to G4-stabilising ligands (depicted as an orange triangle).

Pooled screening development, where many siRNA/shRNA/sgRNA guides against multiple genes are combined and simultaneously evaluated, has allowed systematic identification of synthetic lethal interactions in mammalian cells. These are referred to as 'drop out' screens, as cells with a gene knockout (e.g. caused by RNAi or CRISPR) that reduces cell growth and/or viability will be negatively selected for compared to cells with phenotypically neutral gene deficiencies. Perhaps most famously is the synergistic use of several PARP inhibitors (PARPi) to cause selective lethality in BRCA1- and BRCA2-deficient cells (Bryant *et al*, 2005; Farmer *et al*, 2005) and subsequently breast and ovarian cancers (reviewed in Rouleau, Patel, Hendzel, Kaufmann, & Poirier, 2010). Combinatorial treatments have also been explored, where PARPi plus DNA-damaging drugs are used at doses sublethal in isolation but with dual lethality (reviewed in Chan & Giaccia, 2011)(Figure 1.4C). More recent synthetic lethal interactions have been identified between the DNA mismatch repair proteins (MLH1 and MSH2) and DNA polymerases (β and γ) (Martin *et al*, 2010, 2011), the latter being mutated in hereditary nonpolyposis colorectal carcinomas (Kinzler & Vogelstein, 1996), highlighting possible areas for pharmacological exploitation. This was verified by the selective efficacy of pol β inhibitors in such cancers (Nickoloff *et al*, 2017). Genotype sensitivities beyond DDR deficiencies include the sensitivity of VHL-deficient renal cancer cells to pharmacological inhibition of CDK4, CDK6 and autophagy proteins (Bommi-Reddy *et al*, 2008; Turcotte *et al*, 2008). Both CRISPR and RNAi techniques have been extensively used for such systematic cancer lethality investigations. Some of these applications will be discussed below, including

the parameters that have made them particularly robust techniques, such as the experiment duration, threshold for determining significant hits and drug concentrations used.

1.8.3.1 Identifying therapeutic synthetic lethalities via siRNA screening

Loss-of-function siRNA screens have been widely used for pharmacological target identification. For example, a focused kinase siRNA screen was successfully used to explore kinase deficiencies that sensitised pancreatic cancer cells to the nucleoside analogue gemcitabine, to identify potential combinatorial drug targets (Azorsa *et al*, 2009). This experiment was performed for 72 h at low 10-30 % growth inhibition concentrations (GI_{10-30}), ultimately identifying CHK1-deficiency as a potent gemcitabine sensitiser, verified by the increased gemcitabine efficacy in combination with two Chk1 inhibitors. Similarly, the application of siRNA screening to the HeLa cell line (Bartz *et al*, 2006) allowed identification of gene depletions that increase the potency of several established chemotherapeutics: gemicitabine, the DNA cross-linker cisplatin and paclitaxel, a mitotic spindle inhibitor. For this, a siRNA library targeting 20,000 genes, with 3 unique siRNAs per protein was used. Experiment duration was again limited to 72 h and GI_{10} drug concentrations.

Analogous siRNA screening in head and neck cancer identified deficiencies in fanconi-anaemia and BRCA DDR pathways as the main sensitisers to cisplatin treatment (Martens-de Kemp *et al*, 2017). Here a genome-wide targeting library, with 4 siRNAs per gene was used in combination with 72 h

cell treatment and a GI_{20} gemcitabine concentration. For this the VU-SCC-120 tongue squamous cell carcinoma cell line was selected for technical reasons, including cisplatin sensitivity, growth rate and amenability to siRNA transfection. A gene was designated significantly sensitised to gemcitabine treatment if at least 2 separate siRNAs were depleted (FDR ≤ 0.05). A smaller siRNA screen against 3,300 genes was used to investigate genes required for MYC-driven oncogenesis in human fibroblasts (Toyoshima *et al*, 2012). This ultimately identified 102 synthetic lethal interactions several of which were pharmacologically validated. These case studies reveal the success of RNAi screening to systematically identify genotypes that increase drug potency.

Beyond screening for genotypes that increase pharmacologically induced lethality, siRNA functional genomics have also been used to monitor other cellular features, such as genes required for epithelial migration (Simpson *et al*, 2008). For this a siRNA pool targeting 1,081 human genes (four siRNAs per gene) was screened for the ability to alter MCF-10A migration in a 24 h wound healing assay and identified 42 novel genes. Other siRNA screening applications in mammalian cells include identification of genes necessary for cell survival i.e. apoptotic resistance (MacKeigan *et al*, 2005) and the kinases required for endocytosis (Pelkmans, 2005). These prior siRNA screening studies reveal the importance of multiple siRNAs per genes, stringent thresholds for hit determination and the use of low ligand (often GI_{20-30}) doses to maximise treatment windows. Additionally, this reveals that many of the classical cancer chemotherapeutics are sensitised to check point and DDR deficiencies. However, such siRNA experiments are limited to 72 h, as dilution

and degradation of the siRNA below the critical value required to maintain gene depletion, means that such an approach is not suitable for longer time course.

1.8.3.2 Successful application of shRNAs for dropout screens

To combat this short time frame constraint, other functional genomics investigations have used genome encoded shRNA pools, allowing continuous shRNA production. For example, a shRNA lentiviral library targeting 1,028 genes (~5 unique shRNAs per gene) was successfully used to identify proliferation genes such as in HT29 colon cancer cells (Moffat *et al*, 2006). Here, a gene knockdown was considered a 'hit' if 2 or more shRNA knockdowns caused an increase in mitotic indices. A similar retroviral shRNA screening approach identified genes that when knocked down allowed bypass of p53-dependent cell cycle arrest (Berns *et al*, 2004): 23,742 shRNAs targeting approximately 8,000 human genes (3 independent knockdowns per gene). Both of these studies were performed for 96 h. A more recent, optimised shRNA screening approach was employed for the discovery of cancer proliferation genes over a longer time course (Schlabach *et al*, 2008). This study applied their screening technology of 8,203 shRNAs (2,924 genes; ~3 hairpins per gene) to four different human cell lines: DLD-1 and HCT116 colon cancer cells, the HCC1954 breast cancer cell line and normal human mammary epithelial cells and cultured the cells for 10 – 16 population doublings.

As for siRNA, shRNA approaches have also been used in conjunction with drug screening. For instance, a recent genome-wide shRNA screen identified synthetic lethal interactions with the PARPi olaparib (Bajrami *et al*, 2014). This ultimately uncovered expected deficiencies in DNA damage repair proteins, but also novel deficiencies in DNA cohesion and chromatin remodelling, including CDK12, a gene commonly mutated in high-grade serous ovarian cancer. For this study, they used olaparib resistant, BRCA1/2 wild-type MCF-7 breast tumour cells and a lentiviral pool containing 57,540 shRNAs (~17,000 genes; 3-4 hairpins per gene), a GI_{20} olaparib concentration and a duration of 10 population doublings. Significance was scored at $p < 0.05$ and focussed on genes that where 2 or greater shRNAs were depleted following olaparib treatment. These reflect common themes and parameters for successful screening approaches, including $p < 0.05$ to determine significant hits and use of multiple hairpins per gene.

1.8.3.3 *Identifying therapeutic synthetic lethalities via CRISPR screening*

CRISPR dropout screens have also been used to identify viability genes and pharmacological synthetic lethal interactions. As an example of the former, a 30 day lentiviral CRISPR knockout screen was applied to five leukaemia cell lines using a library of 90,709 different guide RNAs against 18,010 genes (~5 guides per gene), to identify a core set of genes that were critical for leukaemia viability (Tzelepis *et al*, 2016). Differences were considered significant $p < 0.01$ and log_2FC threshold < -1.5, with *KAT6A*, identified as a potential leukaemia therapeutic target. In a similar approach, a 14 day

CRISPR dropout screen was used to identify synthetic lethal drug target pairs in K562 chronic myeloid leukaemia cells, using 490,000 double-sgRNA against 21,321 pairs of drug targets (Han *et al*, 2017). This ultimately highlighted five strong, new drug target interactions including BCL2L1 and MCL1 inhibition as a strong synergistic combination.

These case studies highlight the power of genetic screening to obtain a global overview of gene and processes involved in a specific phenotype or in response to drug treatment. The advantages and disadvantages of these will be discussed in greater detail in chapter 2, where this approach is applied to G4-stabilising ligands. The rationale for this will be discussed in section 1.9.

1.9 Aims and rationale for the work described within this thesis

1.9.1 Expanding the biological knowledge of endogenous G4s

The functions, locations and importance of both RNA- and DNA-G4 structures are not fully understood. A major unanswered question within the field is whether G4 structures have any functional importance or if they are "nuisance" structures that need resolving. Additionally, identification of the plethora of proteins interacting with and regulating these structures has not been systematically investigated. Thus the first aim of this thesis is to expand our biological knowledge of both DNA and RNA G4-structures. The roles and interaction partners of RNA-G4s in particular are less understood relative to their DNA-G4 counterparts. With the recent suggestion that RNA-G4s are

globally unfolded in the cell (Guo & Bartel, 2016), it is also imperative to evaluate whether RNA-G4s have functional intracellular roles. If RNA-G4s do have pivotal roles, one would expect treatment of cells with G4-stabilising ligands to perturb these functions and the RNA-associated pathways in which RNA-G4s are involved, thus altering cellular growth rate.

I hypothesise that the identity and range of proteins uncovered from the functional genomic and follow-up studies within this thesis will provide insight into these outstanding questions. The application of genome-wide functional genomics in combination with G4-ligands to human cells should enable the identification of a wide-range of both known and novel G4-interacting proteins. In support of this, as discussed in section 1.6.2, yeast deficient in the G4-interacting proteins Sgs1 and Cig21 caused resistance and sensitivity to the G4-ligand NMM respectively (Hershman *et al*, 2008; Ren & Chaires, 1999; Kreig *et al*, 2015). For the majority of these investigations, PDS and PhenDC3 are employed as representative G4-stabilising ligands with broad *in vitro* G4-specifity, as discussed above and outlined further in section 2.1.

1.9.2 Identifying chemotherapeutically exploitable genotypes

As well as extending our knowledge of G4-biology, this project aimed to identify disease-associated, particularly cancer-related, genetic backgrounds that may be susceptible to G4-stabilising ligands and also elucidate G4-mehcanisms in cancer. To date, the latter has only been explored for cells deficient in three genes: *BRCA1, BRCA2* and *ATRX* (McLuckie *et al*, 2013; Xu *et al*, 2017; Watson *et al*, 2013). For BRCA1- and BRCA2-deficiencies

(deficiencies in homologous recombination (HR) repair), this synthetic lethality has been clinically exploited for BRCA-deficient tumours (clinical trial, NCT02719977), suggesting a key role for these proteins in the repair of G4-induced damage. It is currently unknown if other DNA damage repair proteins, including further members of the HR pathway or alternative repair mechanisms (e.g. non homologous end joining) also repair G4-associated DNA damage, and are thus similarly exploitable with G4-stabilising ligands. As different cancers are characterised by a multitude of distinct DDR-deficiencies, this will be of potential chemotherapeutic interest. Given the wide-ranging processes G4s are thought to contribute to, there is great potential for uncovering cancer genotypes beyond DDR-deficiencies with G4-stabilising ligands sensitivity.

There are two ways in which the gene sensitivities uncovered by genetic screening within this thesis, can potentially be chemotherapeutically exploited via synthetic lethality approaches. Firstly, for a cancer-associated gene deficiency that causes sensitivity to G4-stabilising ligands (Figure 1.4D), such ligands may be used as single agent therapies. Secondly, pharmacological inhibition of a critical oncogene that mimics the gene deficiencies identified here might be potentiated by combinatorial treatment with G4-stabilising ligands (Figure 1.4E). I envisage that the work within this thesis will be of increasing clinical importance as more disease-associated genotypes are discovered and as G4-stabilising ligands are pharmacologically improved

1.9.3 Understanding and improving G4-stabilising ligands

The final aim of this thesis is to gain further knowledge of G4-stabilising ligands. One aspect is to increase the clinical potential of G4-stabilising ligands, which complements the aim to identify cancer-associated genotypes. Identifying susceptible genotypes, in addition to creating more "drug-like" and pharmacokinetically tolerated G4-ligands brings us closer to their chemotherapeutic use. Within this thesis, I also provide a systematic investigation of the gene deficiencies causing resistance to the representative G4-stabilising ligands PDS and PhenDC3, providing insights into both G4-dependent and G4-independent resistance mechanisms. The latter may highlight possible areas for improving the endogenous efficacy of these molecules. Further, within this thesis, tools are provided to benchmark the synthetic lethalities engaged by novel G4-stabilising ligands against PDS and PhenDC3.

Secondly, it is unknown how structural differences of G4-ligands and the variation in their *in vitro* affinity for different G4-structures (Le *et al*, 2015) influence their cellular phenotypes. While multiple cellular studies with individual G4-ligands exist (see section 1.4.2), there is a lack of a systematic comparison of different ligands and their cellular responses. The side-by-side exploration of PDS and PhenDC3 within this thesis may provide insight into this. In particular, it would help in understanding whether different G4-ligands selectively stabilise a subset of G4 structures and subsequently elicit different and defined biological effects in cells.

To summarise, the genes, themes, tools and resources provided within this thesis attempt to answer biological and clinical queries regarding G4-structures and the G4-stabilising ligands targeting them. This thesis aims to solidify links between G4s and the genes and pathways with which they have previously been implicated. In addition it highlights new genes and areas of research into the regulation of G4-biology and explores the therapeutic potential of molecules targeting these structures.

Chapter 2

A genetic screening approach to uncover synthetic lethalities with G4-stabilising ligands

2.1 Background and Objectives

Given the identification of G-quadruplex (G4) structures in cells, I aimed to uncover new biological and potentially therapeutic roles for G4s. Key challenges for the G4-field are understanding and identifying:

1) Normal genes and pathways influenced by G4s, including the identification of novel G4-interacting proteins

2) The cellular response to G4 stabilisation by small molecule ligands, including the DNA-damage response involved

3) Genotypes susceptible to G4-stabilising ligand treatment for future chemotherapies in cancer and other disease-related backgrounds.

To investigate these challenges in an unbiased and systematic manner, a genome-wide genetic screening approach was selected in which silencing of the protein coding genome was performed in conjunction with G4-stabilising ligands. In this chapter, first the experimental design and rationale will be discussed, including the advantages of this method. Secondly I describe the preliminary experiments and pilot screen performed to establish critical parameters, including choice of cell line, drug concentration, experiment duration and the thresholds used to determine ligand-specific hits. The outcome of the genome-wide screen is then discussed. Next, I outline the

design and use of a custom focused screen comprising primary hits identified from the genome-wide screen, to gain information in two different cell lines. Finally, I discuss four key genes that, when depleted, imparted consistent synthetic lethality to G4-stabilising ligands. I finish by detailing how these results directly inform on the key challenges and additionally explain how the tools and resources developed here can be deployed to address unmet biological questions concerning G4s and their ligands.

2.2 Designing a shRNA screen to explore genotypes that are synthetic lethal with G4-stabilising ligands

2.2.1 Advantages of stable, retroviral shRNA genetic screening for investigating genome-wide G4-ligand sensitivities

There are several ways to achieve gene knockdowns (see section 1.8 and Figure 2.1). One method is CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-Crispr Associated protein 9), a technology based on the bacterial viral defense mechanism, in which synthetic sgRNAs (short guide RNA sequences) are used to guide the Cas9 nuclease to recognise and remove target genes (Doudna & Charpentier, 2014). Alternatively, RNA interference can be used, in which RNA guides promote the RNA Induced Silencing Complex (RISC) to cleave and/or translatonally repress a target mRNA (Mohr *et al*, 2014; Carthew & Sontheimer, 2009). While CRISPR cause knockouts via gene deletion, RNAi causes knockdowns, as not all the mRNAs transcribed from the target gene will be silenced. Both have been successfully employed as genetic screening tools (section 1.8.2).
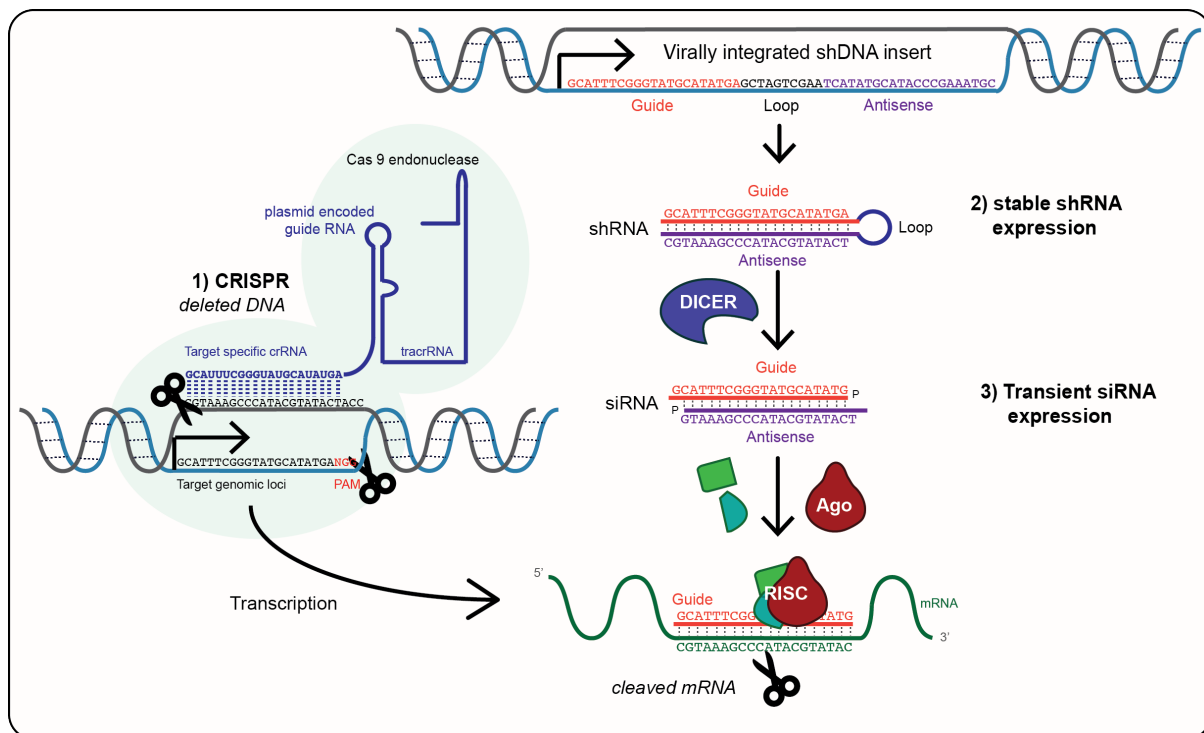
Figure 2.1. Outline of CRISPR, stable shRNA expression and transient siRNA expression for genome editing and functional genomics

The stronger perturbation provided by a gene knock-out is likely to be more deleterious to the cell than an RNAi-induced knockdown, particularly seeing as gene loss can induce compensatory mechanisms (El-Brolosy & Stainier, 2017). Therefore it was decided that RNAi will allow exploration of more viable phenotypes, potentially allowing detection of more hits. Also, with respect to disease related-phenotypes, including combinatorial opportunities with current chemotherapeutic drugs, RNAi was reasoned to be more physiologically relevant than a CRISPR knockout, as pharmacological inhibition is unlikely to remove all protein activity. Further, many cancer mutations are heterozygous and tumour supressor pathways are often downregulated rather than deleted, something mimicked more via RNAi than CRISPR. Additionally, many commonly used cancer cell lines exhibit polyploidy (Paulsson *et al*, 2013; Rondón-Lagos *et al*, 2014) making complete knockouts more technically

difficult. Therefore, RNAi methods may provide distinct advantages to a CRISPR approach.

Generally siRNA approaches are restricted to 72-96 h, with a key limitting factor being siRNA degradation and dilution during cell division, below the critical threshold required for protein knockdown (Chiu & Rana, 2002; Singh *et al*, 2011). Due to this transiency, it is difficult to silence genes encoding long-lived proteins via siRNA. An additional complication is that non-physiological concentrations of commercial siRNAs may potentially trigger and exarcerbate off-target phenotypes (reviewed in Perwitasari, Bakre, Tompkins, & Tripp, 2013). For an unbiased and systematic screen, stable shRNA integration was chosen to allow silencing of proteins independent of half-life and over long time periods.

The most commonly used shRNA viral vectors are adeno-associated virus (AAV), retrovirus and lentivirus (Bukrinsky *et al*, 1992; Manjunath *et al*, 2009; Sliva & Schnierle, 2010). Lenti/retrovirus vectors result in low (~30 %) transduction efficiency, but show a high genome integration rate, stable throughout cellular passaging (reviewed in Perwitasari, Bakre, Tompkins, & Tripp, 2013). The near 100 % transduction efficiency with AAV infection does not normally allow chromosomal integration (Mitani & Kubo, 2002; Grimm *et al*, 2005). Therefore, for stable, long-term knockdown, retroviral/lentiviral vectors are advantageous, and several synthetic versions exist with different backbones (i.e. the plasmid into which the artificial shRNA is inserted and expressed). Several viral constuct variants place the shRNA within pol II

miRNA promoter contexts (shRNAmir vectors) rather than enforcing their pol III expression (Chung *et al*, 2006; Liu *et al*, 2008; Zeng *et al*, 2002). This prevents oversaturation of the endogenous pathway (Boudreau *et al*, 2008; Castanotto *et al*, 2007; McBride *et al*, 2008; Premsrirut *et al*, 2011).

For the unbiased genome-wide assessment of G4s, an shRNAmir vector was chosen using the latest generation shERWOOD-Ultramir shRNA pLMN human retroviral library (Transomic technologies). This system is designed by machine learning for effective engagement of the endogenous siRNA pathway, improved knock-down efficiency and reduced off-target effects (Knott *et al*, 2014; Fellmann *et al*, 2013, 2011; Auyeung *et al*, 2013). The development, design and advantages are summarised in section 2.2.2.

Figure 2.2 provides an overview of the synthetic lethality strategy. Briefly, shRNA knockdown was combined with G4-ligand treatment to stabilise genomic and RNA G4 structures. Two possible conceptual interpretations of an shRNA-induced phenotype are depicted: i) the gene is not required in a G4-associated process so there is no effect on viability (Gene A) or ii) loss of the gene results in cell death due to either removal of a protein interacting (e.g.binding or unwinding) with G4s or depletion of a G4-associated pathway e.g. DNA damage repair (Gene B). The synthetic lethal interactions in the 'Gene B' list will 1) help to inform on disease-associated genotypes amenable to G4-ligand targeting (blue box) and 2) identify a spectrum of biological pathways that can be perturbed by stabilised G4s (orange box), highlighting possible mechanistic features of G4-biology.
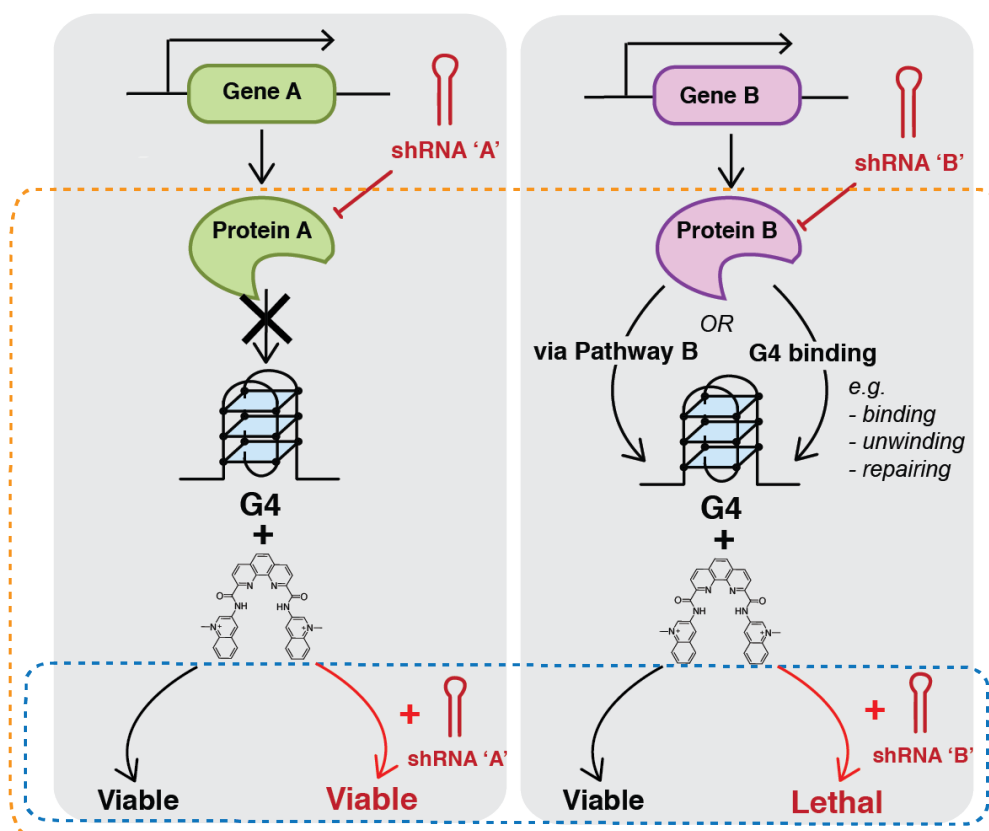
Figure 2.2. Overview of synthetic lethal strategy to identify G4 sensitiser genes
Gene A/Protein A is not, and Gene B/Protein B is, involved within a process or protein that interacts, either physically or mechanistically with G4 structures. Therefore in the presence of G4-ligands, viability unaffected by shRNA-induced knockdown of Gene A while Gene B deficiencies result in synthetic lethality.

2.2.2 Development and design of the shERWOOD-Ultramir shRNA pLMN human retroviral library

As the shERWOOD designed library was used for the G4-ligand screen, the technological developments that gave rise to the vector used are now described. Several considerations exist for effective flow through of the siRNA biogenesis pathway including recognition by processing enzymes (e.g. Drosha and Dicer) and components of the active RISC complex, such as Ago. For endogenous RNAi, cleavage of genome-encoded primary microRNAs by Drosha creates a shorter hairpin loop, called the pre-miRNA which is

recognised by Dicer (Denli *et al*, 2004; Lee *et al*, 2003). Artificial shRNAs, consist of synthetic guide and antisense sequences inserted into a 'backbone' vector (explained in section 2.2.1), which mimics this pre-miRNA substrate (Brummelkamp *et al*, 2002; Paddison *et al*, 2002). The sequence of the artificial shRNA heavily dictates knockdown potency (Chiu & Rana, 2002; Khvorova *et al*, 2003; Schwarz *et al*, 2003), but systematic exploration of optimal sequence and structure, beyond target mRNA complementarity has only recently been explored (Ameres & Zamore, 2013). Machine learning algorithms, such as BIOPREDSi and DSIR have been successfully implemented to discern potent siRNA, but not shRNA, features (Huesken *et al*, 2005; Vert *et al*, 2006). To address this Fellmann and colleagues used a fluorescent reporter assay to monitor the potency of 20,000 shRNAs against nine transcripts and train a machine learning algorithm (Fellmann *et al*, 2011). This identified potent shRNAs with rare features often missed by previous algorithms.

Despite the success of this 2011 algorithm and sensor assay, features of the shRNAmir construct and of the shRNA outside the mRNA targeting sequence had yet to be optimised. For example, most synthetic shRNAmir vectors, at single-copy expression, produced final siRNA levels much lower than their natural counterparts thus limiting their potency (Premsrirut *et al*, 2011). This low expression level arises from synthetic shRNAmir vectors lacking conserved backbone elements found in endogenous pri-miRNAs including bulges within the stem, alterations of bases flanking the loop and inclusion of restriction sites within conserved 3' regions. To investigate this 11 different

synthetic miR backbones were validated via the sensor assay, creating a more efficiently processed backbone with 10-30 fold higher mature siRNA levels with repositioned restriction sites (Fellmann *et al*, 2013).

Following optimisation of shRNA and shRNAmir construct criteria, Knott and colleagues generated 250,000 shRNAs which they interrogated via the sensor assay (Knott *et al*, 2014). The results were used to train an shRNA potency prediction machine-learning algorithm, named shERWOOD, which showed 180 % and 126 % increase in efficacy over existing siRNA- and shRNA-predicting algorithms respectively. This algorithm was benchmarked via shRNA screening against 2,200 cell growth and survival genes, comparing the efficacy of the top 10 shRNAs predicted by shERWOOD or DSIR. 40 % of shERWOOD shRNAs achieved significant depletion versus 31 % for DSIR constructs, with shERWOOD shRNAs giving greater depletion. The shERWOOD predicted shRNAmir backbones were further optimised by removing vector restriction sites, incorporating shRNAs via Gibson assembly and changing the first nucleotide of the shRNA sequence to U. Both features were validated to increase the potency of shERWOOD predicted shRNAs and used to generate human shRNA libraries targeting the protein coding genome. The library and its scaffold have undergone further iteration and are currently in the 7th generation, with 60 % of predicted hairpins causing strong depletion. This latest generation shRNAmir library was used for G4-ligand investigation.

The library consists of 12 randomised pools of ~10,000 hairpins targeting the protein coding genome (~20,000 genes), with an average of 6 optimised hairpins against each gene. This provides a total of 113,000 shRNAs. The vector is outlined below (Figure 2.3). The shRNA sequence (dark blue), consists of a 5' guide strand (targeting the mRNA) and a 3'antisense strand and loop (both required for hairpin formation and Drosha binding). A strong promoter isolated from Murine Stem Cell Retrovirus (MSCV) controls shRNA expression. Therefore, although the shDNA is randomly integrated into the genome, the promoter ensures comparable expression between cells. A phosphoglycerate kinase (PGK) promoter driven bicistronic construct in the vector encoding both neomycin resistance and a fluorescent protein, allows antibiotic and fluorescent selection of cells that have integrated the shRNA vector. The shRNAs are packaged into retrovirus particles, and introduced into the cell by viral infection, to create stable shRNA expressing cell lines.
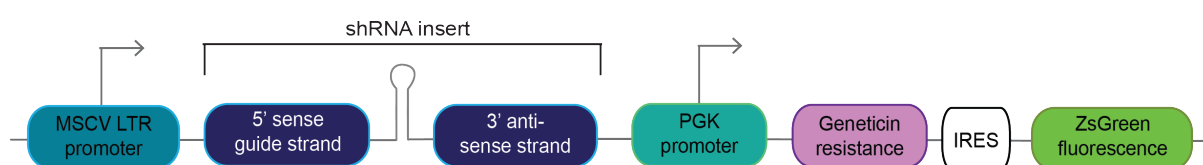
Figure 2.3. The shERWOOD retroviral shRNA construct for stable integration and knockdown of the protein coding genome.
The construct consists of 2 main components 1) shRNA cassette driven by the MSCV viral promoter and 2) a bicistronic "selection" cassette to allow identification of successful transformants via geneticin selection and ZsGreen fluorescent detection, driven by the constitutive PGK promoter.

2.2.3 PDS and PhenDC3 as representative G-quadruplex stabilising ligands

For this study the structurally distinct PDS and PhenDC3 were used (Rodriguez *et al*, 2012; De Cian *et al*, 2007b, structures shown in Figure 1.2D&F). Both ligands show relatively broad specificity for a wide range of G4s (see section 1.4.2.4) and are therefore suitable for probing genetic vulnerabilities to G4 stabilisation in cells.

To test and refine the experimental design, optimisation and pilot experiments were performed, using one pool chosen at random (pool 8) from the genome-wide set. This pool contains 9,600 individual shRNAs targeting 7,823 different protein-coding genes (1.23 hairpins per gene). In the entire library, each gene is targeted by an average of 5-6 hairpins per gene, randomised across the twelve pools. As such the pilot screen results are indicative rather than definitive. The design and outcome of these experiments are discussed below.

2.3 Experimental optimisation via pilot screening

2.3.1 Dose response curves for PDS and PhenDC3 treatment of three candidate cell types

Three cell lines were used for initial ligand sensitivity experiments. For this, technical considerations included adherancy, to minimise loss of cells during culture, fast doubling time and stable ploidy. The following adherent human cell lines were chosen: A375 (malignant melanoma; 20 h population doubling

time), HT1080 (malignant fibrosarcoma; 18.2 h doubling time) and MDA-MB-231 (metastatic breast cancer; 38 h doubling time). These have been used for previous functional genomic studies (Silva *et al*, 2008; Yang & Stockwell, 2008; Schlabach *et al*, 2008; Shalem *et al*, 2014).

First the PDS and PhenDC3 sensitivity of these cell lines was investigated. This is an important consideration as synthetic lethality is the readout of the screen. For example, too much ligand-induced cell death would reduce the available response window, while too little would not engage a strong enough ligand response. For this, cells were treated with serial dilutions of both ligands for 96 h and post-treatment viability determined by an ATP luminescence-based assay, that quantifies the number of metabolically active cells (see Methods for details). This was used to generate viability curves for three biological replicas (Figure 2.4A-F show example viability curves for one biological replica) from which average GI_{20} and GI_{50} values were calculated (Figure 2.4G). All cell lines were more sensitive to PDS versus PhenDC3, with HT1080 showing the greatest sensitivity (GI_{50} = 1.21 μM) and MDA-MB-231 the least sensitive, with a 10-fold higher average GI_{50} value (12.0 μM). For 96 h treatment, the three cell lines showed PhenDC3 sensitivity within the same order of magnitude.
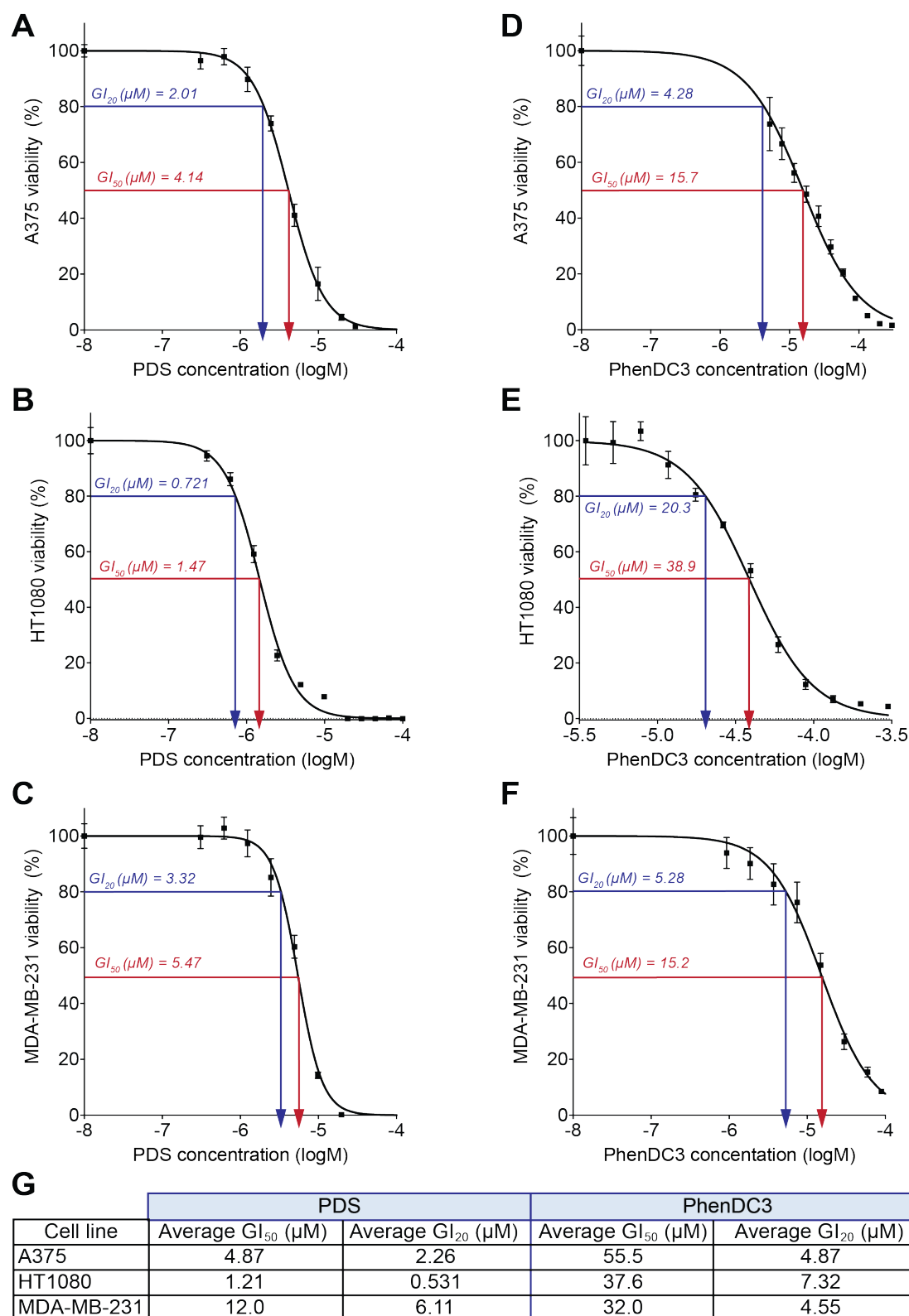
Figure 2.4. 96 h cellular viability assay reveals PDS and PhenDC3 sensitivity for the 3 candidate cell lines

(A-F) Representative viability curves for one of three biological replicas plotted as an average of four technical replicates using a non-linear regression model, following 96 h treatment with serial dilutions of (A-C) PDS and (D-F) PhenDC3 for A375, HT1080 and MDA-MB-231 as indicated. (G) Table summarising average GI_{20} and GI_{50} values for each cell line based on 96 h treatment with PDS or PhenDC3 for 3 biological replicas

Next, cell transfection efficiency with the pilot retrovirus pool was tested to explore whether sufficient hairpin representation was achievable. For this, virus was produced using the Platinum-A (Plat-A) packaging cell line, optimised for high viral titre production (see Methods for details). Plat-A cell transfection is pH sensitive, thus a range of buffers with pH 5.98-7.20, at 0.2 increments were tested for their ability to cause calcium phosphate precipitation by visual inspection. Buffers at pH 6.98 and 7.00 yielded fine precipitates (Figure 2.5A) and both pHs enabled successful hairpin integration in Plat-A in small-scale test transfections, as visualised by ZsGreen reporter gene expression (Figure 2.5B). Thus these buffers were used for virus production in the pilot and subsequent large-scale screens.
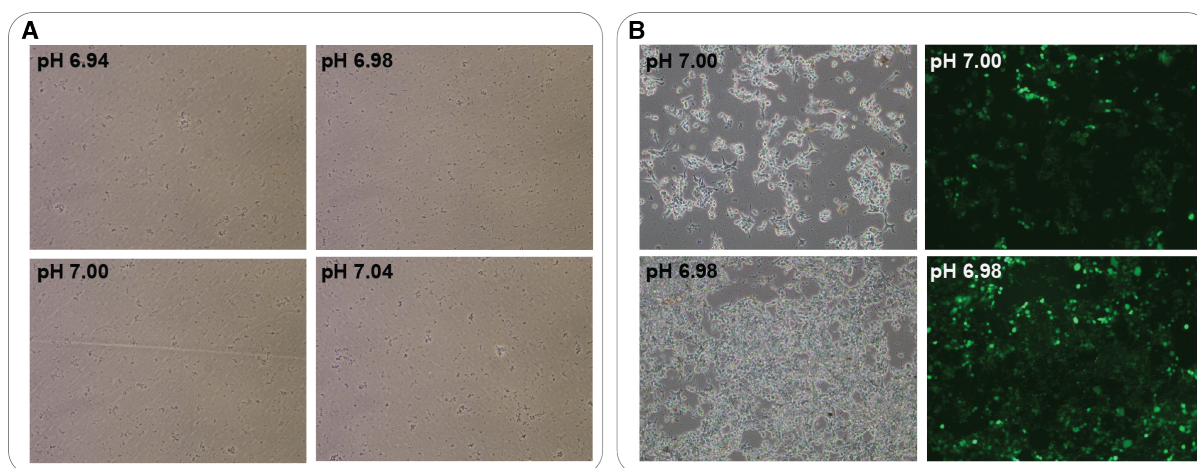
Figure 2.5. Optimising buffer pH for transfection of the Platinum-A packaging cell line
(A) Brightfield microcopy images comparing the calcium phosphate precipitation levels for HBS buffers at different pH values
(B) Phase contrast (left) and fluorescent (right) microscopy images of mock PlatA transfections with HBS buffers at pH 7.00 and 6.98 (as indicated). ZsGreen expression indicates successful transfection

Following successful virus production using both HBS buffers (pH 6.98 and pH 7.00), HT1080, A375 and MDA-MB-231 cells were transfected with 1:7, 5:14 and 4:7 dilutions of media containing the virus (viral media: normal

media) or media-only control. Plasmid integration was monitored by ZsGreen expression via flow cytometry, 48 h after transfection (Figure 2.6; see Methods for details). Based on these data, HT1080 > A375 >> MDA-MB-231 with respect to virus transfection efficiency. Confirming that cell autofluorescence does not influence fluorescent detection, in the absence of virus only 0.4 % of cells were denoted Zs-Green positive. Additionally, the data showed that viral titre was similar regardless of the HBS buffer pH used to transfect Plat-A cells.

	Percentage ZsGreen positive cells 48 h after infection (%)					
Cell line	HT1080		A375		MDA-MB-231	
Virus Dilution HBS buffer	pH 6.98	pH 7.00	pH 6.98	pH 7.00	pH 6.98	pH 7.00
0	0.48		0.39		0.41	
1:7	20.51	22.97	10.87	9.36	6.50	6.40
5:14	38.98	41.87	23.47	26.69	12.24	12.35
4:7	62.19	62.46	37.03	38.52	22.55	23.91

Figure 2.6. Comparing shRNA retroviral transfection and integration efficiency for three candidate cell lines.
HT1080, A375 and MDA-MB-231 cell lines were transfected with virus produced from Plat-A cell lines transfected with pool 8 shRNAs using either pH 6.98 or 7.00 HBS buffers. The following dilutions of virus were used 1:7, 5:14, 4:7 (virus:media) alongside a media-only control (0). 48 h after transfection, the percentage of each cell line that had integrated the shRNA construct was calculated via FACs detection of ZsGreen fluorescence.

This combined with the slower population doubling time, aneuploidy and reduced PDS sensitivity, suggested the MDA-MB-231 cell line was suboptimal. Therefore the HT1080 and A375 cell lines were selected for pilot shRNA screening.

2.3.2 A pilot shRNA screen using HT1080 and A375 cells

Next, a screening methodology designed to investigate synthetic lethality associated with G4-stabilising ligands was tested and refined using a pilot screen with HT1080 and A375 cells. This is outlined in the schematic below (Figure 2.7). Briefly, cells were infected with retrovirus shRNA vectors at a low titre to minimise multiple infections per cell. Following complete antibiotic selection (7-10 days), a reference sample was harvested (t0) and cells were then treated with a DMSO vehicle control or GI_{20} concentration of PDS and PhenDC3, for 'n' population doublings before harvesting the final timepoint (tF). Cells were split and replated in fresh media and ligand every 72 h. A GI_{20} dose was reasoned to be low enough to allow long-term cell culture while being sufficiently high to cause ligand-specific phenotypic effects. A low ligand concentration also allows detection of only the strongest synthetic lethal interactions with G4-stabilising ligands. To circumvent retrovirus cloning and infection efficiency limitations, cells were maintained at 1000-fold coverage (10×10^6 million cells). Additionally, logarithmic growth conditions were maintained by splitting at 70-80 % confluency to minimise shRNA representation changes due to localised restriction of cell growth.
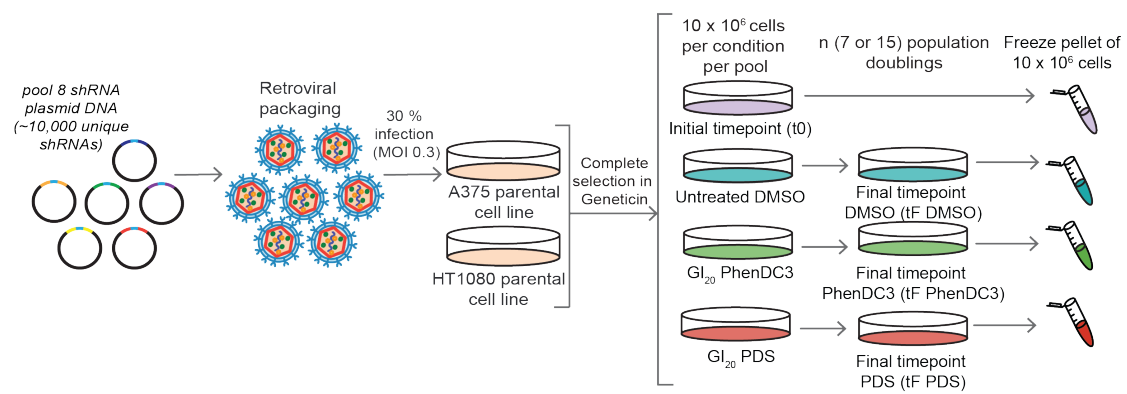
Figure 2.7. Outline of a pilot screening using shRNA retroviruses to uncover genetic vulnerabilities to G4-stabilising ligands
For the pilot screen pool 8 (~10,000 shRNAs) one of the 12 shRNA pools targeting the protein coding genome was used. Plasmids are retrovirally packaged and then A375 and HT1080 cells are infected at MOI 0.3 (30 %). Following complete antibiotic selection, for each cell line the initial time point (t0) was harvested and cells cultured for 'n' population doublings (7 or 15) in DMSO, PDS or PhenDC3 before the final time point was harvested (tF). The workflow shown above represents one biological replica. Three replicas were performed for each cell line (consisting of three independent transfections).

To determine differentially expressed shRNAs in tF versus t0, the unique 3'-antisense sequences were recovered by two rounds of PCR, using primers specific to common regions within the shRNAmir construct, and quantified by massively paralleled sequencing (Figure 2.8). Here, 10×10^6 reads (1000 x sequencing depth) per sample were sequenced to recover all shRNAs within each cell. If a gene knockdown compromises cell viability then the shRNA is depleted compared to those against genes without effect. Therefore, the count in tF is less than t0 and $\log_2(tF/t0)$ is negative.
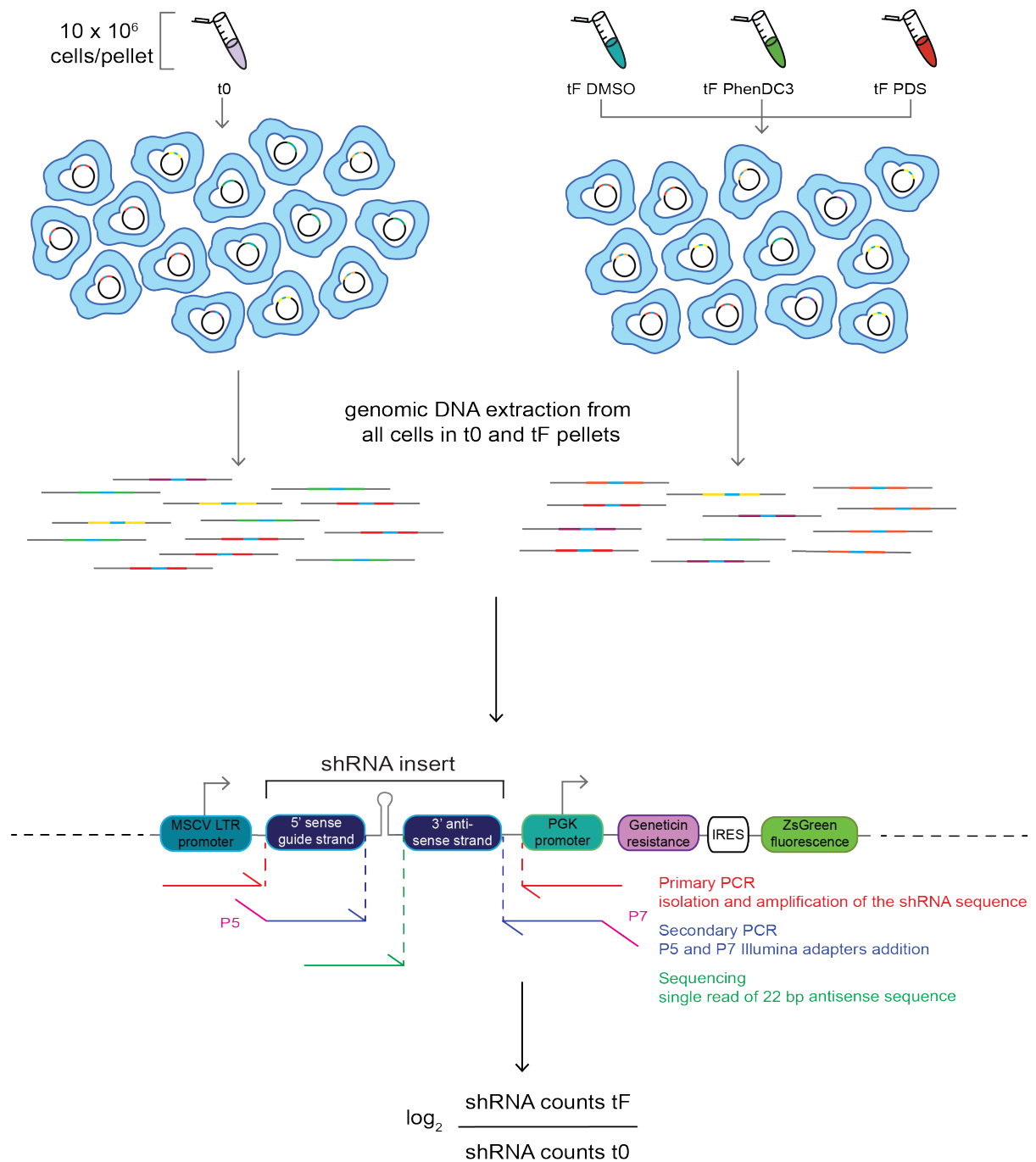
Figure 2.8. PCR and sequencing pipeline to identify differentially expressed shRNAs
Genomic DNA (gDNA) was extracted from pellets containing 10×10^6 harvested from the pilot screen corresponding to reference timepoints (t0) and final timepoints for each treatment condition (tF). The shRNA insert was recovered from surrounding gDNA via PCR (primary). The secondary PCR reaction added the P5 and P7 adaptors necessary for sequencing on the NextSeq platform. shRNAs were identified and quantified for each sample via a custom sequencing primer that gives a single read of the unique "barcode" 3' antisense sequence. To identify shRNA changes within the cell population after treatment, the $\log_2$ fold change was calculated for each shRNA.

The experimental design assumes that GI_{20} values calculated for 96 h are applicable for long-term culture and, that this methodology will detect ligand-specific synthetic lethal interactions. In addition to testing these assumptions, pilot screening allows for optimisation of two further parameters: experiment duration and suitable threshold cutoffs, i.e. false discovery rate and fold change (FC).

2.3.3 A375 cells more amenable than HT1080 for long-term culture with both PDS and PhenDC3

The pilot screen was performed as outlined in Figure 2.7, using the following ligand concentrations: 10 µM PhenDC3 for both cell lines and 0.5 µM and 2.5 µM PDS for A375 and HT1080 respectively, an approximate GI_{20} as determined by the 96 h viability assay. A375 cells were successfully cultured for 15 population doublings (PD) with PDS, PhenDC3 and DMSO (Figure 2.9A). Both G4-stabilising ligand treatments caused similar growth defects to the DMSO control and allowed continuous, linear growth for the experiment duration. HT1080 cells were also successfully cultured for 15 PD in PDS and DMSO. However, PDS induced growth inhibition was greater than expected for a PDS GI_{20} and reaching 15 PD necessitated culturing for 10 days longer than DMSO-treated cells (Figure 2.9B). For the first 120 h of PhenDC3 treatment, HT1080 growth inhibition was as expected. However, beyond this, cells were acutely sensitive to PhenDC3 and growth plateaued after ~ 3 population doublings. Based on these data, the A375 cell line, with 2.5 µM PDS and 10 µM PhenDC3 was chosen as the most suitable for performing the genome-wide genetic screen.

73
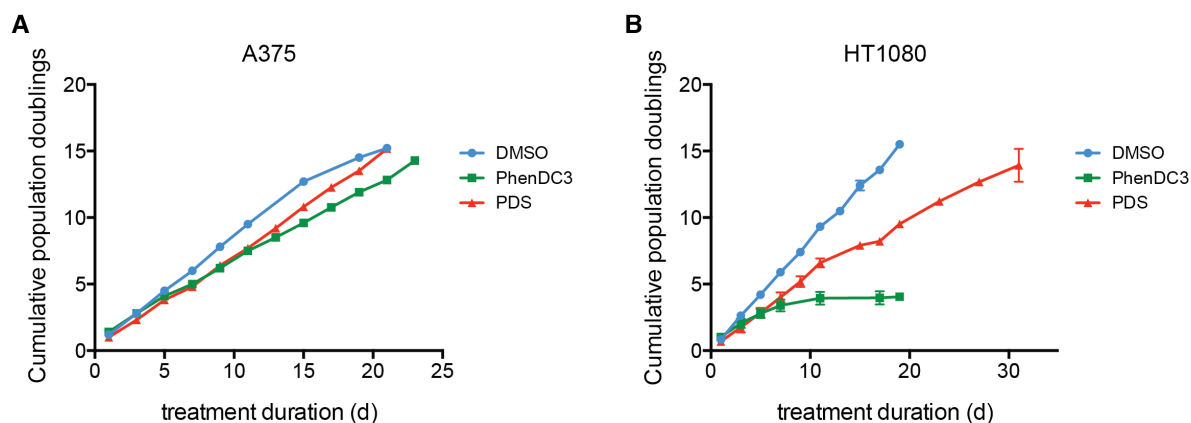
Figure 2.9. Pilot screen growth curves for A375 and HT1080 cells cultured in GI$_{20}$ ligand concentrations
Average cumulative population doublings were plotted for DMSO (blue), PhenDC3 (green) and PDS (red) treatment for the experiment duration for (A) A375 and (B) HT1080. Shown is the average of 3 biological replicates (mean ± standard deviation).

2.3.4 Determination of a suitable timepoint to uncover PDS and PhenDC3 synthetic lethalities

For A375 treated cells, two final timepoints were harvested for the pilot screen. An ideal end-point represents a balance between minimising experiment duration while allowing sufficient time to visualise significant phenotypic effects. Based on thresholds used in other screens, significance was set at FDR ≤ 0.05 (see introduction, section 1.8).

Cells were harvested after 7 and 15 population doublings (t7 and t15 respectively; 10 x 10^6 cell per conditions), and genomic DNA extraction, PCR and sequencing performed for t7, t15 and t0 samples. Then, average log$_2$FC (tF/t0) counts for each hairpin were plotted for PDS (Figure 2.10A) and PhenDC3 (Figure 2.10B) to provide a global overview of hairpin fold changes. Non-significant changes in shRNAs are shown in grey, while coloured points represent significant shRNAs. Blue indicates shRNAs which show a

significant $\log_2$FC for DMSO (untreated) while green and red represent ligand specific significant changes, for PhenDC3 and PDS respectively. Points below y=0 indicate a lower shRNA representation in tF versus t0, and therefore a gene knockdown causing sensitisation to the treatment. For t7, 13 and 12 shRNAs were significantly altered following PDS and PhenDC3 treatment respectively. Increasing this threshold ($\log_2$FC $\leq$ -1) to identify the strongest synthetic lethalities, revealed that only 6 (PDS) and 11 (PhenDC3) shRNA hairpins were altered for t7, suggesting that this is an insufficient time period to evaluate G4-ligand sensitivities. By contrast, at t15, more hairpins were significantly depleted following PDS and PhenDC3 treatment, 746 and 93 significant shRNAs respectively, excluding those also significantly changed in DMSO. Of these 322 (PDS) and 50 (PhenDC3) showed a $\log_2$FC $\leq$ -1, indicating a more appropriate final point.

Having obtained a global appreciation of shRNA representation changes, changes in individual hairpins were investigated. For this shRNAs targeting the known direct G4 binders XRN1 and hnRNPA2/B1 were explored (Figure 2.10C). It was hypothesised that synthetic lethality would arise from depletion of proteins that regulate processes via G4-interaction, in conjunction with disrupting pathway mechanics by ligand-induced G4 stabilisation. At t7, the XRN1 shRNA count was not significantly different in any treatment. However, at t15 this shRNA was significantly depleted in PDS and PhenDC3 but not DMSO treatments. For both ligands (but not DMSO), the $\log_2$FC was less than -1 (-1.51,PDS; -1.52,PhenDC3; -0.25 DMSO), indicating that both FDR $\leq$ 0.05 and $\log_2$FC $\leq$ -1 are critical for exploring G4-ligand synthetic lethality.

Similarly for hnRNPA2/B1, at t7 the hairpin did not show a significant fold change in any condition and did not pass the logFC threshold of < -1. This was also observed at t15 for PDS and DMSO treatment. However, for PhenDC3 treatment, despite at a global level having less overall significant sensitisers compared to PDS, the hnRNPA2/B1 shRNA exhibitted a $\log_2$FC of -1.18 at an FDR = 0.02. Perhaps surprisingly, a hairpin targeting BRCA1, a protein previously identified as synthetic lethal with G4-stabilising ligands, was not identified as significant (data not shown). This may reflect that several hairpins (of differing knock down potency) exist per gene and the relevant shRNAs targeting BRCA1 are not present within this pool.
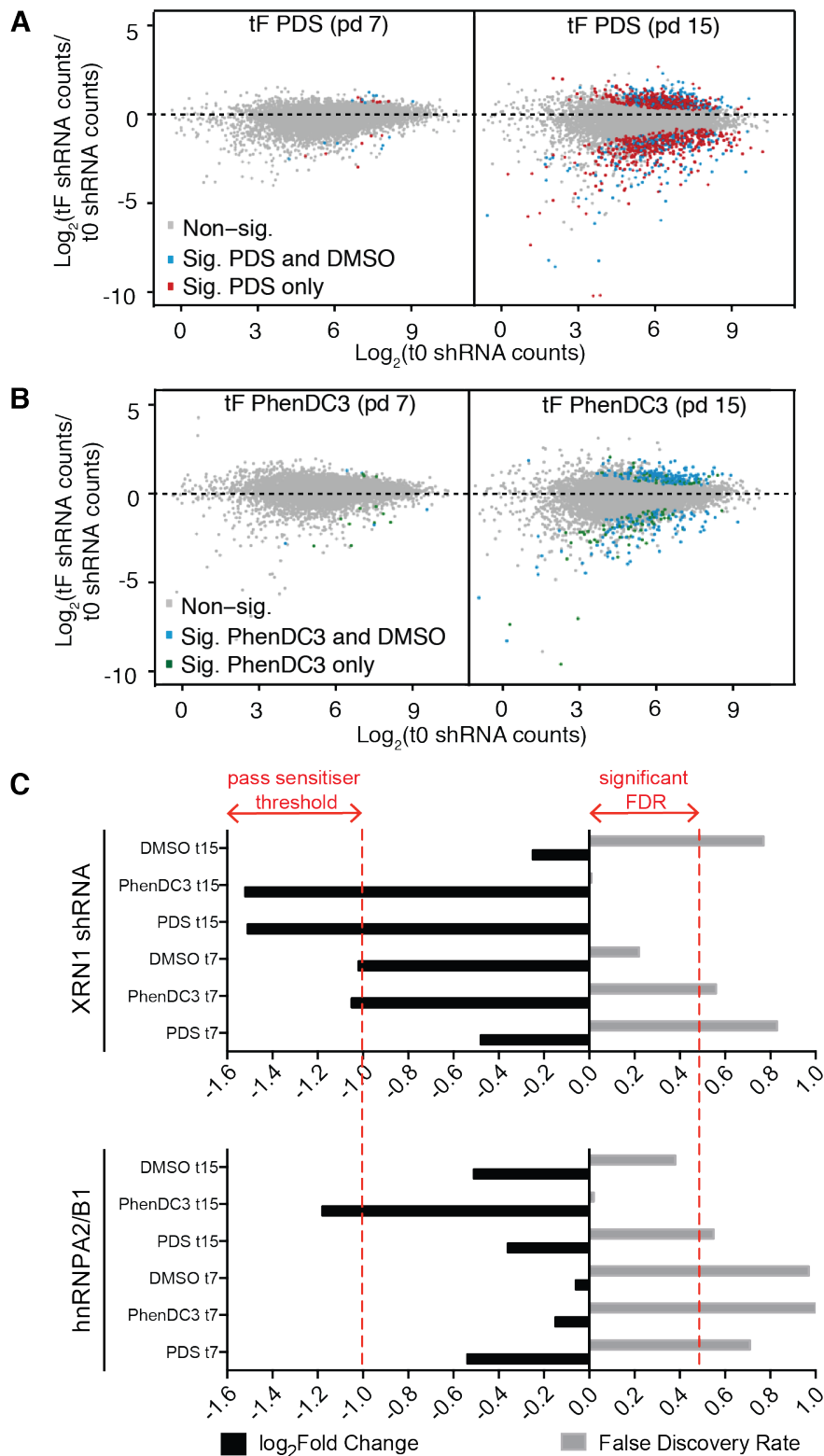
Figure 2.10. Pilot screening reveals t15 as suitable endpoint for uncovering G4-ligand specific hits.
(A-B) A comparison of significant hairpin hits (colours indicate significance at FDR ≤ 0.05) for two end points: population doubling 7 (left) and 15 (right) for (A) PDS and (B) PhenDC3
(C) A comparison of FDR and log2FC values for the XRN1 (top) and hnRNPA2/B1 (bottom) hairpins following DMSO, PDS or PhenDC3 treatment after 7 and 15 population doublings

The pilot screen was informative in several respects, foremost in identifying the A375 cell line as suitable for the screen, with the 96 h GI_{20} values for both drugs being applicable for the experiment duration. Additionally, it suggests a final timepoint after 15 PD; and FDR $\leq$ 0.05 and $\log_2 FC \leq -1$ thresholds as suitable parameters for identifying ligand specific synthetic lethalities. On a wider level, this pilot experiment suggested that the shRNA screening methodology outlined and developed here was capable of detecting synthetic lethalities to PDS and PhenDC3, and by proxy, stabilisation of G4-structures.

2.4 Genome-wide screening in A375 cells

2.4.1 PDS and PhenDC3 caused growth inhibition for all pools

Based on the success of the pilot screen, the pipeline was extended genome-wide, using 12 pools encompassing all shRNAs against the protein coding genome (Figure 2.11). The cumulative population doublings for each pool after treatment was monitored, to ensure linear growth inhibition for ligand-treated versus DMSO samples (Figure 2.12). Inter-pool differences in the degree of growth inhibition were evident. For PhenDC3 treatment, pool 4, 5, 8, 9, 10, 11 and 12 growth was reduced compared to DMSO for the experiment duration. Conversely, differences for pools 2, 3, 1 and 6 became apparent after treatment for 6, 12 and 15 days respectively. For PDS treatment, visible growth inhibition was seen for pool 8, 9, 11 and 12. For the remaining pools, the PDS growth overlapped with DMSO treatment.
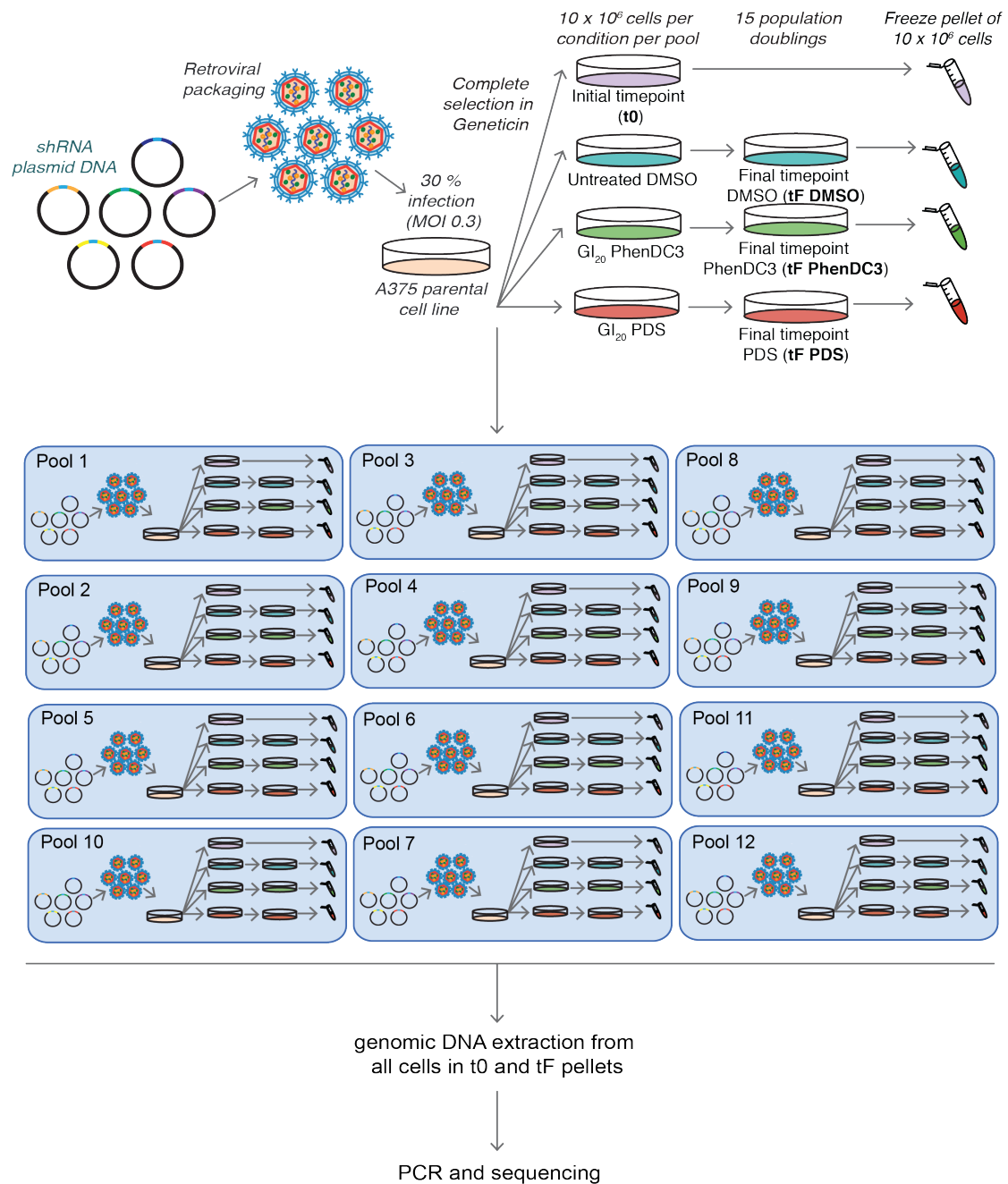
Figure 2.11. Outline of the genome-wide RNAi screen technology pipeline to uncover synthetic lethality with PDS and PhenDC3
For each pool (1-12), plasmids are retrovirally packaged and A375 cells are infected at MOI 0.3 (30 %). Following complete antibiotic selection, the initial time point (t0) was harvested and cells cultured for 15 population doublings in DMSO, PDS or PhenDC3 before the final time point was harvested (tF). Genomic DNA was extracted from all pellets and the antisense "barcode" of the shRNA extracted and amplified by PCR and quantified via sequencing on the NextSeq Illumina platform
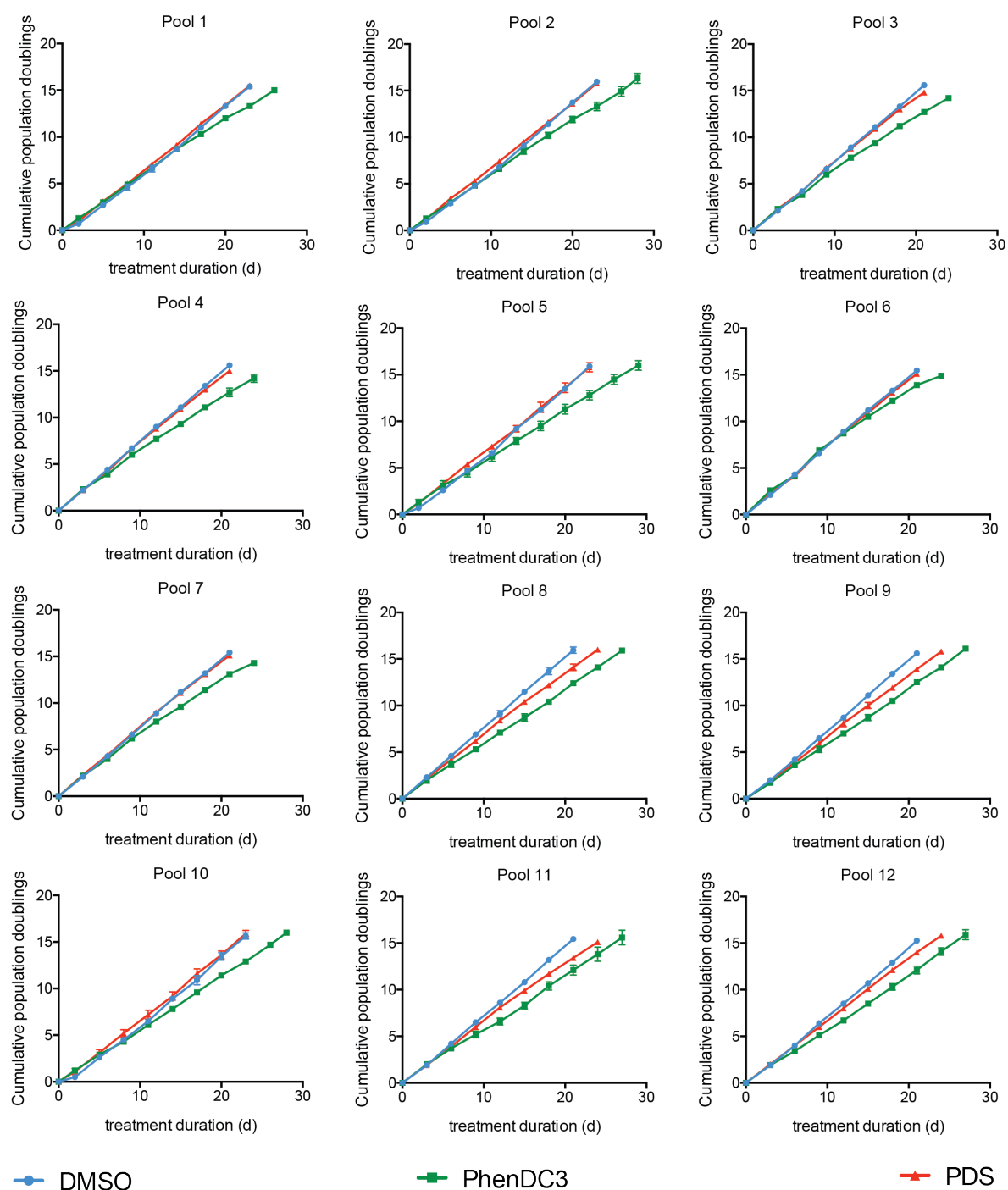
Figure 2.12. Inter-pool differences in growth inhibition following PDS and PhenDC3 treatment.

Average cumulative population doublings were plotted for DMSO (blue), PhenDC3 (green) and PDS (red) treatment for the experiment duration for the 12 pools as indicated. Shown is the average of 3 biological replicates (mean ± standard deviation).

2.4.2 Quality control check and sequencing results

Next t0 and tF samples were extracted and sequenced to monitor shRNA differential expression. As the NextSeq maximum output is 40 x 10^7 reads, t0 and tF samples were sequenced in batches of 2 (24 samples on the NextSeq, see Methods for details), to ensure sufficient sequencing depth per sample (8-10^6 per sample). Inter-pool shRNA contamination is a potential problem particularly at the PCR level, therefore following sequencing, reads were mapped according to their shRNA ID and reads belonging to another pool discarded before differential expression analysis (Figure 2.13). For some samples, shRNA contamination from other pools was particularly high. For contamination above 40 %, samples were resequenced and reads combined, to increase the number of pool specific reads (samples denoted with * and all pool 1 and 2 samples). A read count of 8 x 10^6 was considered sufficient for subsequent analysis (800 fold coverage, close to the original aim of 1000). Having achieved a minimum of 800-fold coverage for each pool, significantly differentially expressed hairpins for each pool were determined using the same parameters used for the pilot screen (Figure 2.14A).
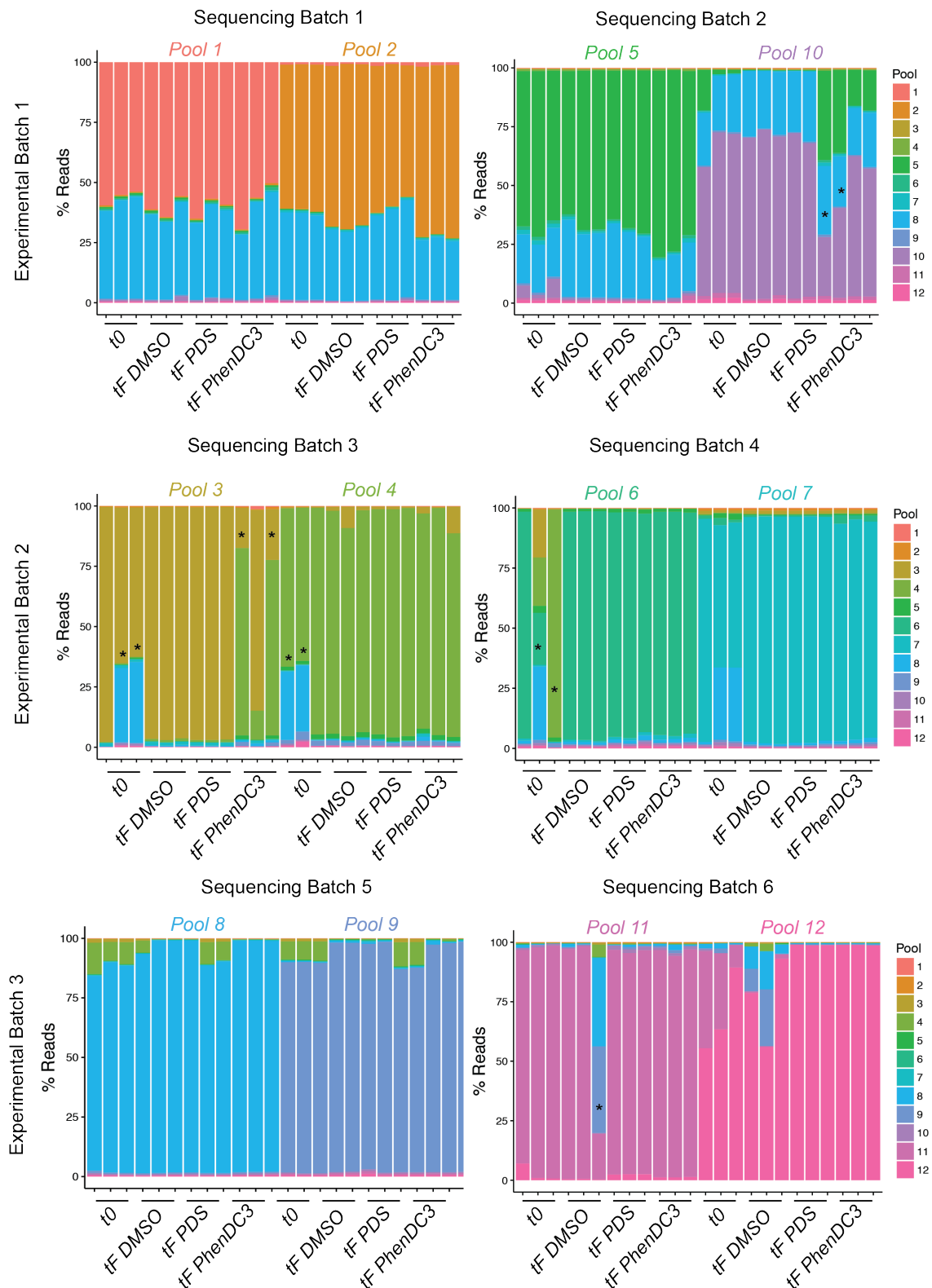
Figure 2.13. Number of pool-specific reads per sample following sequencing.
Screen was performed in three experimental batches consisting of four pools each for ease-of-handling: 1) pools 1,2,5,10; 2) pools 3,4,6,7; 3) pools 8,9,11,12. Sequencing was performed in 6 batches, with each run containing the samples from 2 pools (3 biological replicas per sample; 24 samples total) as indicated to ensure sufficient sequencing depth per sample. For each sample, reads were mapped to each pool to identify contamination from other pools.
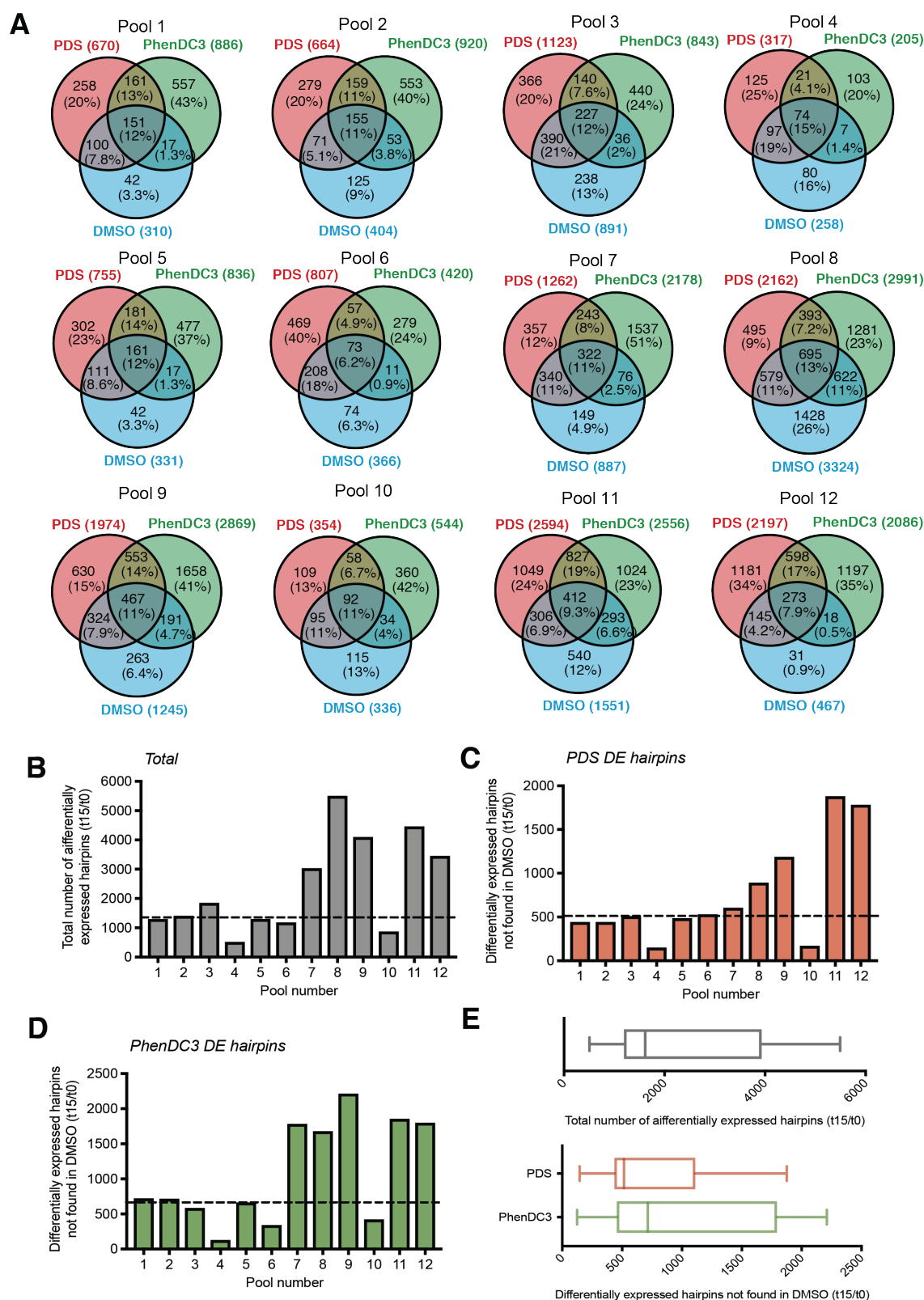
Figure 2.14. Genome-wide RNAi screening reveals ligand specific hairpin alterations.
(A) Venn diagrams showing the overlap between hairpins that were differentially expressed (FDR ≤ 0.05) following DMSO, PDS and PhenDC3 treatment, for the 12 pools. (B-E) comparison between numbers of significant hairpins for each pool for (B) total, (C) PDS excluding those also found in DMSO and (D) PhenDC3 excluding those also found in DMSO. Median is denoted by black dotted lines (E) Box and whisker plots for data in (B-D).

The number of shRNAs in each category and percentage that this represents of the total significantly differentially expressed shRNAs is indicated for each pool (2.14A). Interpool differences (referred to as batch effects henceforth) in the number of total (2.14B) and of ligand-specific hits (PDS 2.14C and PhenDC3 2.14D) were observed. This reflects the variation in PDS- and PhenDC3-induced growth inhibition (Figure 2.11). Across all conditions (DMSO, PDS and PhenDC3), a median of 1,616 hairpins per pool was differentially expressed, ranging from 507 (pool 4) to 5,493 (pool 8) shRNAs (Figure 2.14E). Excluding hairpins that were also significantly altered in DMSO conditions, following PDS and PhenDC3 treatment (i.e. ligand-specific), medians of 516 (ranging from 146 to 1,876) and 715 (from 124 to 2,211) differentially expressed hairpins were observed respectively. Despite this variation, the use of stringent thresholds and randomisation of gene hairpins between pools, gives confidence regarding drawing conclusions about PDS and PhenDC3 synthetic lethal interactions. These observed batch effects might arise because of technical factors, including cell status, absolute drug concentration differences and variation in DNA isolation and/or sequencing. However, they may reflect biological features – a pool may be particularly enriched in G4-ligand sensitive shRNAs and/or potent shRNAs causing the greatest knockdowns. I have assumed that to alter the drug concentrations on a pool-by-pool basis to achieve identical growth inhibition etc., would negatively impact the aim to conduct an unbiased screen.

2.4.3 The unbiased genome-wide screening methodology identifies known synthetic lethal interactions as top sensitisers

Having analysed each pool individually (section 2.4.2), all differentially expressed shRNA counts were combined, revealing 15,021 across all conditions (Figure 2.15A). As the aim is to investigate G4-ligand synthetic lethality, only shRNAs that were differentially expressed in ligand treatment (i.e. not also depleted in DMSO at FDR ≤ 0.05) were considered. This yielded 9,509 (8 % of total library) differentially expressed, ligand-specific hairpins (Figure 2.15B) with 2,561 PDS specific, 5,591 PhenDC3 specific and 1,357 common to both ligands.
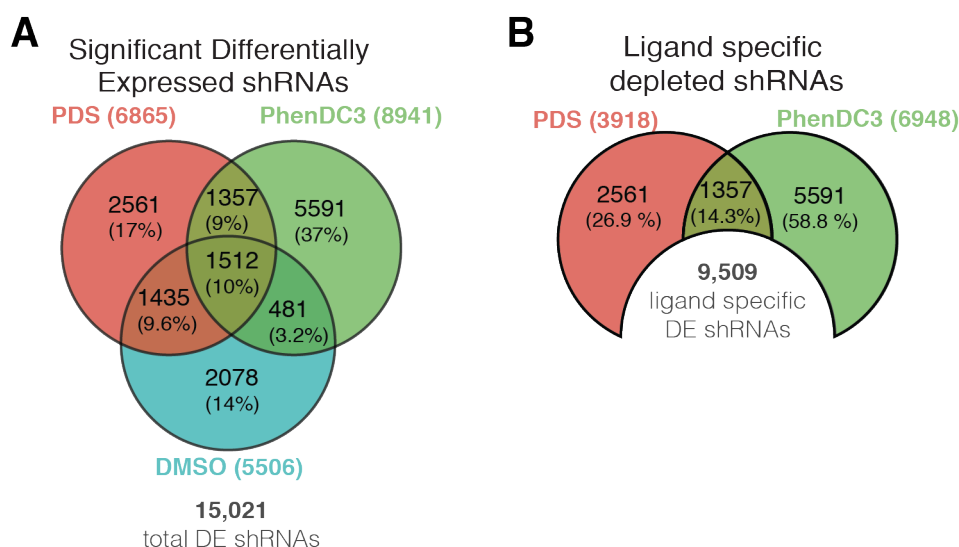
Figure 2.15. 9,509 hairpins show G4-stabilising ligand specific depletion
(A) Venn diagram showing the overlap between differentially expressed hairpins (FDR ≤ 0.05) following DMSO, PDS and PhenDC3 treatment, combined for all 12 pools
(B) Venn diagram showing ligand specific depleted shRNAs (FDR ≤ 0.05, not in DMSO, $log_2FC < 0$)

Although each gene is targeted by approximately 6 hairpins of varying knockdown efficiency (Figure 2.16A) some have many more hairpins (e.g. MYC which is targeted by 21 shRNAs). Therefore a gene was considered

differentially expressed if 50 % or 3 hairpins were significantly depleted, which resulted in 843 differentially expressed genes (Figure 2.16B). This threshold, while stringent, allows genes that are targeted by more or fewer hairpins than the average of 6 shRNAs per gene, to make the cut-off. To focus on top sensitisers an additional threshold of median $\log_2FC \leq -1$ was applied, which gave 758 genes (Figure 2.16C). These genes were denoted a preliminary list of G4-sensitisers, i.e. genes, that when knocked down, are synthetically lethal with G4-ligand treatment. This definition is maintained henceforth.
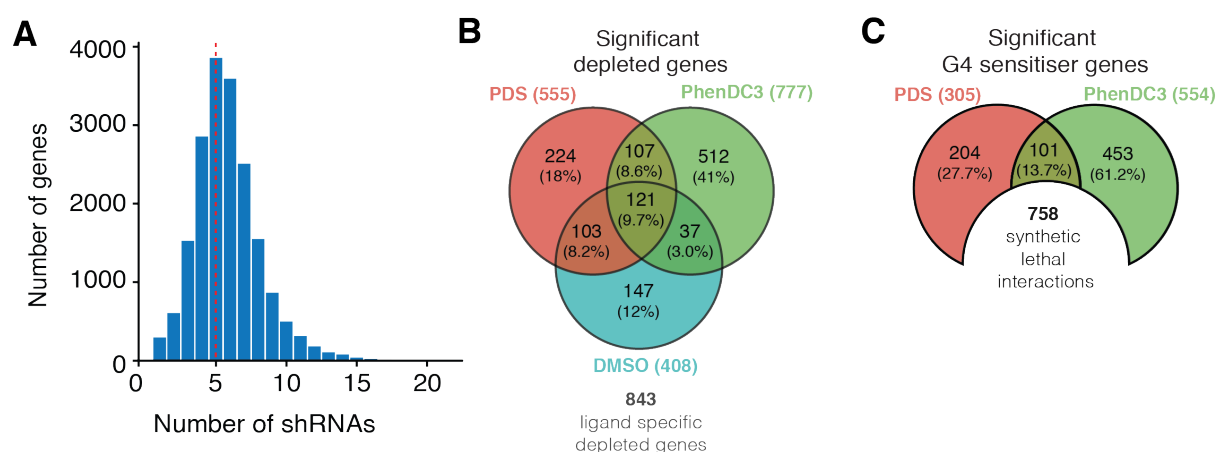
Figure 2.16. Genome-wide RNAi screening approach identifies 758 synthetic lethal interactions with PDS and PhenDC3 treatment in A375 cells
(A) Profile of the numbers of shRNAs targeting each gene, with the average indicated by a red dotted line
(B) Venn diagram showing significantly depleted genes (50 % or 3 hairpins, FDR ≤ 0.05, median $\log_2FC < 0$) following DMSO, PDS and PhenDC3 treatment
(C) Ligand sensitiser genes identified after applying a median $\log_2FC \leq -1$ cut off to PDS and PhenDC3 depleted genes.

G4-ligand synthetic lethality has been previously demonstrated for BRCA-deficiencies (PDS, CX-5461) and ATRX (telomestatin). BRCA1 and BRCA2 were independently uncovered as top sensitisers to both PDS and PhenDC3, and ATRX to PhenDC3, providing strong validation for the dropout screen capability to identify synthetic lethal interactions (Figure 2.17).

Gene	Hairpins in t0	Depleted hairpins (PDS)	Median log2FC (PDS)	Depleted hairpins (PhenDC3)	Median log2FC (PhenDC3)	PUBMED ID
ATRX	16	1	-1.6	5	-1.4	23563309
BRCA1	5	3	-2.8	3	-2.1	28211448, 26748828
BRCA2	3	3	-2.2	2	-1.8	28211448, 26748828, 23782415

Figure 2.17. Independent validation if previously identified genes known to be synthetic lethal with G4-stabilising ligands
Table showing the number of depleted hairpins and median log$_2$FC values for ATRX, BRCA1 and BRCA2, genes that have been previously been shown to be synthetic lethal with G4-stabilising ligands

2.4.4 The preliminary sensitiser list shows depletion in genes implicated in G4-biology

Genes in the 758-sensitiser list fall into three categories:

1) Genes with a known G4-relationship and previously shown to be synthetically lethal with G4-stabilising ligands (e.g. BRCA2)

2) Genes with a known G4-relationship (e.g. reported in the literature to bind and/or unwind G4-structures) but not known to be synthetically lethal with G4-stabilising ligands (e.g. DHX36)

3) Genes with neither prior known G4-relationship nor synthetic lethal interactions with G4-stabilising ligands

The first class of known synthetic lethal genes deficiencies have been described in section 2.4.3 and acted as positive controls. To investigate the second category the UniprotKB, Gene Ontology (GO) and G4 Interacting Proteins Database (G4IPDB) (Mishra *et al*, 2016) databases, were searched for genes annotated with G4-related terms that were also uncovered as sensitisers (see methods). This revealed nine sensitisers (*ADAR, ATRX,*

DHX36, DNA2, FUS, MCRS1, RECQL4, SF3B3 and *XRN1*) with a G4-relationship (Figure 2.18). Besides *ATRX* (Watson *et al*, 2013), this is the first report that these deficiencies cause G4-ligand sensitivity and provides strong support that these proteins are strongly linked to G4-structures. The observation of several G4-interacting proteins causing G4-ligand synthetic lethality, suggests that other unknown G4-binders may be present in the G4-

Uniprot KB, GO and G4IPDB search for G4-related genes

Gene	G4IPDB	GO term/ Uniprot KB G4 association	Hairpins in t0	Depleted hairpins (PDS)	Median log2FC (PDS)	Depleted hairpins (PhenDC3)	Median log2FC (PhenDC3)	PUBMED ID
ADAR1	YES	-	7	0	-	5	-3.9	24813121, 23381195
ATRX	YES	UniprotKB: P46100	16	1	-1.6	5	-1.4	23563309
DHX36	YES	GO: G4 DNA binding, G4 RNA binding	5	1	-1.9	4	-4.6	29269411, 28069994, 25653156, 25611385, 24151078, 22238380, 21149580, 18842585, 16150737
DNA2	-	GO: G4 DNA unwinding	5	3	-2.2	0	-	23604072
FUS	YES	-	3	2	-3.2	-	-	23521792
MCRS1	-	GO: G4 RNA binding	2	0	-	1	-1.7	16571602
RECQL4	YES	-	6	3	-1.8	2	-2.6	25336622
SF3B3	YES	-	4	0	-	3	-2.9	23381195
XRN1	-	GO: G4 DNA binding, G4 RNA binding	4	3	-1.1	3	-2.7	9049243

sensitiser list.

Figure 2.18. Proteins known to directly interact with a G4 structure identified as synthetic lethalities
Table showing the number of depleted hairpins and median log$_2$FC values for nine sensitisers annotated with a G4-associated term in the GO, UniprotKB or G4IPBD databases

It is notable that several known direct G4-interactors (e.g. binding and unwinding proteins) were not identified as causing synthetic lethality. This may reflect that these G4-interactions are redundant; less influenced by low ligand concentrations (gene C) and/or perform other integral roles, thus are

also lost following DMSO treatment (gene D). To evaluate this, for the 18 genes with a G4-associated UniprotKB/GO term, the significantly depleted hairpins and associated log_2FC values are summarised in Figure 2.19A. For example, despite the known *in vitro* ability to unwind G4s, PIF1, WRN and BLM knockdowns did not cause significant lethality in either DMSO or treated controls. This was reflected by all genes within the G4-interactor table, except hnRNPA2/B1 (Figure 2.19B; gene type C), which was identified as a sensitiser to both DMSO and PDS (3 hairpins, log_2FC -1.4 and 2 hairpins, log_2FC -1.8 respectively). This suggests that hnRNPA2/B1, despite the known *in vitro* G4-association, has an essential role in addition to stabilised G4s regulation (Figure 2.19B; gene type D). Using a screen readout of lethality and removing DMSO sensitisers, restricts the outcomes to the strongest ligand-specific gene knockout responses. This suggests that the nine direct interactors identified as sensitisers, are major players in the cellular response to stabilised G4-structures.

A

Gene	hairpins in t0	depleted hairpins (PDS)	median logFC (PDS)	depleted hairpins (PhenDC3)	median logFC (PhenDC3)	depleted hairpins (DMSO)	median logFC (DMSO)	Included in our sensitiser list?	Quadruplex GO term / UniprotKB association
ATRX	16	1	-1.6	5	-1.4	0	NA	yes	Quadruplex term present within UniprotKB:P46100 entry
DHX36	5	1	-1.9	4	-4.6	0	NA	yes	G-quadruplex DNA binding [GO:0051880]; G-quadruplex RNA binding [GO:0002151]
DNA2	5	3	-2.2	0	NA	0	NA	yes	G-quadruplex DNA unwinding [GO:0044806]
MCRS1	2	0	NA	1	-1.7	0	NA	yes	G-quadruplex RNA binding [GO:0002151]
XRN1	4	3	-1.1	3	-2.7	1	-1.6	yes	G-quadruplex DNA binding [GO:0051880];G-quadruplex RNA binding [GO:0002151]
AFF2	2	0	NA	0	NA	0	NA	no	G-quadruplex RNA binding [GO:0002151]
BLM	8	0	NA	2	-1.3	0	NA	no	G-quadruplex DNA binding [GO:0051880]; telomeric G-quadruplex DNA binding [GO:0061849]; G-quadruplex DNA unwinding [GO:0044806]
DDX11	7	1	-1.1	0	NA	1	-2.2	no	G-quadruplex DNA binding [GO:0051880]; G-quadruplex DNA unwinding [GO:0044806]
DDX17	6	1	-2.3	0	NA	2	-2.0	no	G-quadruplex term present within UniprotKB:Q92841 entry
FMR1	9	0	NA	1	-0.8	0	NA	no	G-quadruplex RNA binding [GO:0002151]
FXR1	5	0	NA	0	NA	0	NA	no	G-quadruplex RNA binding [GO:0002151]
HNRNPA2B1	3	3	-1.4	0	NA	2	-1.8	no	G-quadruplex DNA unwinding [GO:0044806]
HNRNPD	3	0	NA	0	NA	0	NA	no	quadruplex term present within UniprotKB:Q14103 entry
LIN28A	4	1	-4.8	0	NA	1	-7.5	no	G-quadruplex RNA binding [GO:0002151]
LONP1	5	0	NA	0	NA	0	NA	no	G-quadruplex DNA binding [GO:0051880]
PIF1	10	0	NA	1	-2.3	0	NA	no	G-quadruplex DNA unwinding [GO:0044806]
RAD50	8	1	-2.9	2	-2.1	1	-1.1	no	G-quadruplex DNA binding [GO:0051880]
WRN	6	0	NA	0	NA	0	NA	no	G-quadruplex DNA binding [GO:0051880]; telomeric G-quadruplex DNA binding [GO:0061849]; G-quadruplex DNA unwinding [GO:0044806]
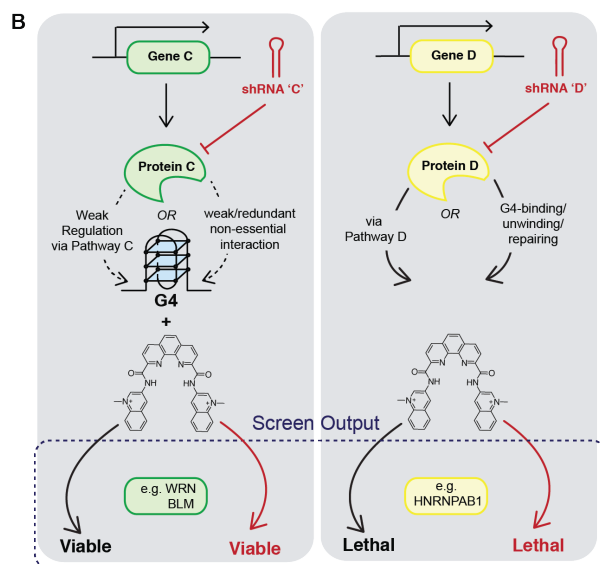
Figure 2.19. Several known direct G4-interactors are not identified as synthetic lethal with G4-stabilising ligand treatment
(A) Table showing the number of depleted hairpins and median log$_2$FC values for the 18 genes in the GO and UniprotKB database with a G4 relationship for PDS, PhenDC3 and DMSO treatment.
(B) Schematic explaining the absence of these genes from our sensitiser list. Either the G4-interaction (physical or mechanistic) is not strong enough in isolation to cause cell death following a single shRNA-induced gene knockdown (Gene C) or the gene also performs essential non-G4 roles in the cell and thus deficiency is also lethal in the DMSO control (Gene D).

Next, this G4-relationship was extended further by using G4 search terms and PolySearch2 (see methods; Liu et al., 2015) to scan PubMed abstracts (Figure 2.20). This identified 12 additional synthetic lethal interactions of G4-associated genes, covering categories including G4-helicases (*RTEL1, RECQL4*), DDR (*CHEK1, RAD17*), transcription (*POLR1A, CNBP*) and replication proteins (*ORC1, TOP1*). This is the first study to show that these gene deficiencies impart G4-ligand sensitivity. For those reported to be direct G4-binders, this extends the prior list (Figure 2.18 & 2.19). For genes with a G4-link unrelated to directly binding these structures, such as DDR (e.g. CHEK1) or transcription (POLR1A), their identification indicates that their G4 associated roles are strong and sufficiently essential, that their deficiency causes G4-stabilisation associated synthetic lethality.

Pubmed and PMC literature Search- manually curated top hits

Gene	PolySearch Score	Hairpins in t0	Depleted hairpins (PDS)	Median log2FC (PDS)	Depleted hairpins (PhenDC3)	Median log2FC (PhenDC3)	PUBMED ID	G-quadruplex Association
DHX36	855	5	1	-1.9	4	-4.6	29269411, 28069994, 25653156, 25611385, 24151078, 22238380, 21149580, 18842585, 16150737	DNA and RNA G4 helicase
ATRX	205	16	1	-1.6	5	-1.4	23563309	Chromatin remodeller and DNA G4 binder
RECQL4	130	6	3	-1.8	2	-2.6	25336622	Cooperates with BLM to unwind DNA G4
CNBP	115	5	1	-2.2	3	-5.5	23774591, 28329689,24594223, 26332732	Promotes formation of G4 at Myc promoter (NHE III region)
RTEL1	75	4	1	-3.0	2	-4.8	22579284, 24115439, 19596237	DNA G4 helicase at telomeres
DNA2	55	5	3	-2.2	0	-	18593712, 23604072	DNA G4 helicase/nuclease at telomeres
TFAM	55	5	2	-1.2	5	-3.0	28276514	DNA/RNA G4 binder in mitochondrial DNA
TOP1	45	8	3	-3.1	4	-3.9	11756434, 25473964	Prevents G4 formation
PCNA	35	6	3	-5.6	4	-2.6	24115439	Required for unwinding of telomeric G4 by RTEL1
PPM1D	35	6	5	-1.8	1	-1.9	23514618	DDR, response to G4 ligand
RAD17	35	5	4	-1.9	2	-2.6	17932567	DDR, response to G4 ligand
CHEK1	30	14	4	-3.2	7	-2.7	24441772, 23396447, 27597417, 24464582, 26845351	DDR, response to G4 ligand
EEF1A1	15	1	0	-	1	-5.1	27736771	Proteomic protein binder to RNA G4
ORC1	15	7	1	-2.7	4	-2.7	24003239	Contains G-rich RNA binding domain, proposed to preferentially bind G4
RPA3	15	13	3	-1.9	5	-3.1	16973897, 21772995, 24747047, 23708363, 27440048	RPA subunit: involved in disasociation of G4
POLR1A	10	4	2	-3.3	1	-5.1	27391441	G4 ligand treatment inhibits RNA pol 1

Figure 2.20. G4-associated proteins identified from the literature within the sensitiser list
Tables showing the number of depleted hairpins and median log_2FC values for 16 sensitisers identified as G4-related by text mining, with the associated PolySearch2 algorithm score and a brief description of G4 association.

2.4.5 Genome-wide sensitisers are enriched in a defined set of pathways

Next explored were genes for which both 1) G4-ligand synthetic lethality and 2) G4-structure relationship are largely unknown. For this, enriched KEGG pathways within the 758 G4-sensitiser gene list (i.e. gene depletions that caused sensitivity to PDS and/or PhenDC3) were investigated. This uncovered five significantly ($p < 0.05$) enriched clusters: 'cell cycle' ($p = 6.9 \times 10^{-7}$), 'ribosome' ($p = 3.1 \times 10^{-6}$), 'spliceosome' ($p = 1.5 \times 10^{-6}$), 'ubiquitin-mediated proteolysis' ($p = 1.3 \times 10^{-3}$) and 'DNA replication' ($p = 5.7 \times 10^{-3}$). PDS and PhenDC3 targeted common (blue) and unique (black) genes (Figure 2.20). The enrichment of transcription- and translation-associated genes and pathways is perhaps concordant with G4-structures having predicted genomic and transcriptomic roles. While predicted G4s are enriched at replication origins (Besnard *et al*, 2012), replication proteins themselves are currently unlinked to G4s and their deficiencies unexplored as synthetic lethal interactions. The ubiquitin pathway provides a novel link to G4-biology and ligand synthetic lethality.

Enriched KEGG pathways for all significant G4 sensitiser genes
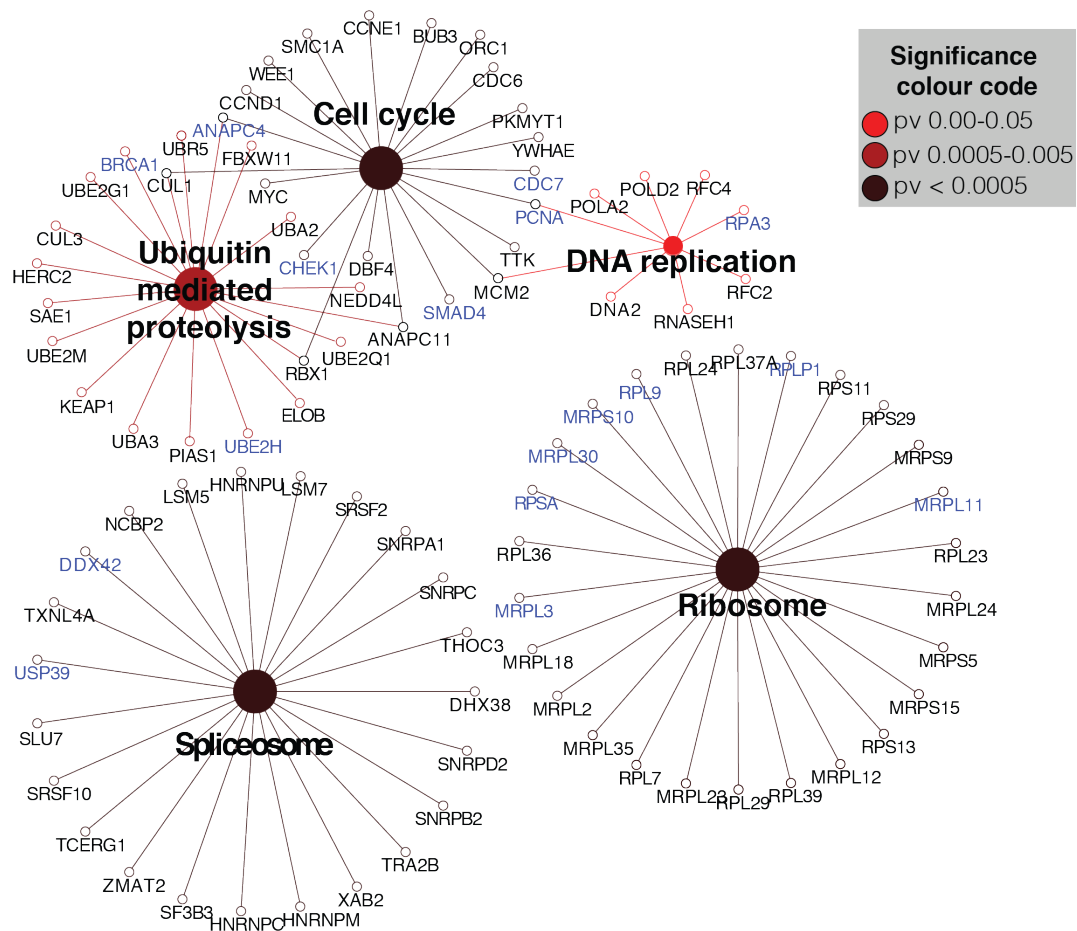
Figure 2.21. Genome-wide sensitiser genes enriched in five pathways including ubiquitin-mediated proteolysis
Enriched KEGG pathways for the 758 sensitiser genes (right-sided adjustment test based in hyper-geometric distribution, Bonferroni corrected p-value < 0.05). Node colour denotes significance and node size reflects the number of genes. Blue reflects gene deficiencies causing sensitivity to PDS and PhenDC3.

To gain possible functional insights into the G4-sensitisers, enriched GO Biological Process (BP) and Molecular Function (MF) terms were then determined (Figure 2.21A). The four MF terms and 20 out of 45 BP terms were directly related to DNA or RNA, consistent with PDS and PhenDC3 directly binding nucleic acid G4 targets. A search for enriched protein domains (Figure 2.21B) also revealed nucleic acid binding proteins such as helicase C-terminal; RNA recognition motifs including RRM, RBD and RNP; and DNA-binding domains including zinc fingers, bZIP motifs and HMG boxes. Also enriched were ubiquitin hydrolase and multifunctional ATPase domains, highlighting potential areas previously unknown to be affected by G4-intervention in mammalian cells. Additionally, the former is consistent with the observed ubiquitin-mediated proteolysis KEGG cluster (Figure 2.20).

Contributing to enriched helicase domains are several DEAD/DEAH-box helicases, including *DDX10, DDX42, DHX38* and *DHX36* (Figure 2.21C), the latter a confirmed RNA G4-helicase (Chen *et al*, 2015). Other DDX and DHX proteins were also uncovered as sensitisers, but not classified as enriched by the GO analysis parameters used (Figure 2.21D). With the exception of DHX36, putative novel G4-unwinders may also exist among these DDX/DHX proteins. The non-emergence of other helicases such as *WRN*, *BLM* and *PIF1* as sensitisers (section 2.4.4) suggests that helicases uncovered here may be key in the cellular response to stabilised G4-structures. Additionally, given the known synthetic lethality approaches with DNA-G4 (i.e. BRCA-deficiencies), it would be valuable to learn if RNA-G4 synthetic lethalities also exist, for example with RNA-helicase deficiencies.

A
Enriched GO terms for all significant G4 sensitiser genes

GO - Biological processes
GO - Molecular function

B
Enriched Protein Domains for all significant G4 sensitiser genes

C

Gene	hairpins in t0	depleted hairpins (PDS)	median logFC (PDS)	PDS sensitiser	depleted hairpins (PhenDC3)	median logFC (PhenDC3)	PhenDC3 sensitiser?
ATRX	16	1	-1.6		5	-1.4	✔
CHD8	10	0	NA		4	-2.3	✔
DDX10	8	3	-3.9	✔	2	-3.6	
DDX42	5	3	-3.0	✔	4	-4.5	✔
DDX6	5	2	-4.1		3	-2.9	✔
DHX36	5	1	-1.9		4	-4.6	✔
DHX38	8	1	-2.8		3	-2.2	✔
POLQ	8	5	-1.3	✔	1	-1.0	
RECQL4	6	3	-1.8	✔	4	-2.2	✔
SMARCA4	14	5	-1.8	✔	1	-1.5	✔
SRCAP	13	3	-2.3	✔	6	-2.2	✔

D

Gene	hairpins in t0	depleted hairpins (PDS)	median logFC (PDS)	PDS sensitiser	depleted hairpins (PhenDC3)	median logFC (PhenDC3)	PhenDC3 sensitiser?
DHX9	8	6	-4.9	✔	2	-3.0	
DHX30	7	4	-2.2	✔	3	-3.1	✔
DHX15	9	4	-4.7	✔	4	-4.7	✔
DDX5	7	3	-2.5	✔	3	-2.5	✔
DDX46	4	2	-3.6	✔	2	-3.6	✔
DDX3X	7	3	-1.9	✔	3	-1.9	✔

Figure 2.22. GO and protein domain analysis of genome-wide sensitisers reveals an enrichment in DNA, RNA and helicase associated genes
(A) Enriched Gene Ontology terms (GO Biological Processes and Molecular Functions) (Bonferroni corrected p-value < 0.05) and (B) Enriched protein domains (p ≤ 0.05) using GENE3D crystallographic data (black) and PFAM sequence information (grey) ordered by -Log10(EASE p-value) for the 758 sensitiser genes.
(C-D) Tables showing the number of depleted hairpins and median log_2FC values for selected genes within the sensitiser list:
(C) Contributing to the "Helicase conserved C-terminal domain" enrichment
(D) Additional DHX or DDX (DEAH and DEAD-box helicases) genes. Those highlighted in blue caused sensitivity to both PDS and PhenDC3 in the screen following shRNA induced depletion

2.4.6 Cancer associated gene depletions enhance sensitivity to G4-ligands

One aim was to identify disease-related genotypes where G4 ligands may be therapeutically exploited in the future. Multiple studies link G4s and cancer, including synthetic lethality of G4-stabilising molecules in BRCA-deficient tumours (Xu *et al*, 2017; McLuckie *et al*, 2013). Given the varied hypothesised roles of G4s, G4-stabilising molecules may offer synthetic lethality opportunities with other cancer genotypes. To explore this further, the COSMIC v83 database (Forbes *et al*, 2015) which describes genes with somatic mutations causally implicated in cancer, was used to investigate potential cancer-associated sensitisation within the genome-wide gene list (Figure 2.23).

A — 50 proteins (median log2FC≤ -1)

B — 52 proteins (median log2FC≤ 0)

C

All PhenDC3 sensitisers →
All PDS sensitisers →

Gene	total hairpins in t0	significant sensitiser hairpins in tF PDS	median logFC PDS
DDX10	8	3	-3.9
FIP1L1	4	2	-3.7
FUS	3	2	-3.2
TOP1	8	3	-3.1
CCND1	10	3	-3.1
BAP1	8	4	-3.1
MYC	8	6	-3.0
SPEN	4	3	-3.0
CCDC6	4	2	-2.9
BRCA1	5	3	-2.8
SMARCE1	16	9	-2.4
CLTC	5	3	-2.3
CBFB	4	2	-2.2
ARNT	5	3	-2.2
BRCA2	3	3	-2.2
SMAD4	4	2	-2.1
SMARCB1	8	4	-2.0
WHSC1	16	5	-1.9
RECQL4	6	3	-1.8
PPM1D	6	5	-1.8
SMARCA4	14	5	-1.8
CDK12	10	4	-1.7
NKX2-1	2	1	-1.6
DEK	7	3	-1.4
PAX7	4	2	-1.4
RAF1	7	3	-1.3
PALB2	5	3	-1.2
ZRSR2	8	4	-1.2
NONO	4	2	-0.9

Gene	total hairpins in t0	significant sensitiser hairpins in tF PhenDC3	median logFC PhenDC3
CNBP	5	3	-5.5
ALK	10	3	-4.7
TOP1	8	4	-3.9
ARNT	5	4	-3.7
FOXL2	3	2	-3.4
KEAP1	4	3	-3.3
RANBP2	11	3	-3.2
EWSR1	5	3	-3.2
FIP1L1	4	2	-3.0
DDX6	5	3	-2.9
CDK12	10	4	-2.8
FGFR3	10	3	-2.7
HSP90AB1	8	4	-2.4
CCDC6	4	3	-2.4
PPP6C	6	3	-2.4
SMAD4	4	3	-2.3
PLCG1	12	3	-2.2
BRCA1	5	3	-2.1
CARS	6	3	-2.0
CREB3L1	4	2	-1.9
WHSC1	16	4	-1.9
UBR5	3	3	-1.8
BRCA2	3	2	-1.8
SRSF2	7	3	-1.8
YWHAE	6	3	-1.5
NBN	6	3	-1.4
ATRX	16	5	-1.4
RUNX1	4	2	-1.4
DEK	7	4	-1.4
CCNE1	8	3	-1.3
NF2	4	2	-1.3
CYLD	8	3	-1.2
KAT6A	12	4	-0.9

Figure 2.23. Sensitivities uncovered from genome-wide screening are enriched in cancer-associated gene depletions

(A-B) STRING analysis for (A) 50 COSMIC proteins that overlapped with 'sensitiser genes' (50 % or 3 hairpins, FDR ≤ 0.05, median $\log_2$FC ≤ -1) and (B) 52 COSMIC proteins that overlapped with significantly depleted genes (50 % or 3 hairpins, FDR ≤ 0.05, median $\log_2$FC < 0). Shown are confidence interactions > 0.4 from co-expression and experimental data.

(C) Tables of the median $\log_2$FC and number of significantly depleted hairpins for the 53 depleted genes overlapping the COSMIC database for (left) PDS, (right) PhenDC3 and genes common to both drugs (red).

Within the 758 sensitisers, there was a 2-fold (p-value = 9.1×10^{-6}) enrichment for 50 cancer-associated genes. The enrichment increases to 3-fold (p-value = 2.5×10^{-3}) when considering 10 cancer-associated sensitisers common to both G4 ligands. The 50 sensitisers found in COSMIC were then explored via STRING network analysis to investigate functional interactions between them (Figure 2.22A). This revealed a cluster of DDR proteins including tumour suppressors PALB2 and BAP1 that physically interact with BRCA1 and BRCA2. Chromatin modifiers such as *SMARCA4*, *SMARCB1* and *SMARCE1* were also identified as sensitisers. Removing the $\log_2FC \leq -1$ threshold from the gene list identified, only adds an additional two proteins overlapping with the COSMIC database: NONO and KAT6A (both median $\log_2FC = 0.9$). This highlights how cancer-associated genes within significantly depleted genes of the screen show large fold changes ($\log_2FC$ -0.9 to -5.5; 46 % to 98 % depletion in tF versus t0), and their deletion can cause strong PDS and PhenDC3 sensitivity. When incorporated into the STRING analysis, *NONO* interacted with the splicing cluster and *KAT6A* was linked to the SMARC chromatin modifiers (Figure 2.22B). The 52 proteins, alongside their associated $\log_2FC$ for PDS or PhenDC3 treatment are shown in Figure 2.22C.

2.4.7 Establishment of a focused pool of potential G4-ligand sensitisers for validating and extending genome-wide findings

Using a genome-wide approach is necessary for an unbiased, systematic study but is labour-intensive. The approach is therefore difficult to scale to multiple ligands and cells. Additionally, as the genome-wide experiment is split between pools, all shRNAs are not simultaneously exposed to ligand and

all PCRs and sequencing runs are not performed concurrently. As revealed earlier this can result in batch effects for ligand response and sequencing quality/contamination (see section 2.4.2), which could compromise data interpretation.

To address these issues, a focused custom "quadruplex" pool was designed, derived from ligand-specific synthetic lethal hits uncovered in the A375 genome-wide screen. This was limited to 8,018 hairpins (1,363 genes), below the number within one pool of genome-wide screen (10,000 shRNAs). Therefore all shRNA-expressing cells can be cultured and sequenced simultaneously, minimising batch effects and allowing multiple comparisons to be simultaneously performed (Figure 2.24). Gene targets within this custom pool include 751 genome-wide G4-sensitisers (excluding 7 genes only targeted by 1 shRNA). Also targeted were 496 significantly upregulated genes (50 % or 3 hairpins significant, $log_2FC \geq 1$); 116 additional genes identified from the literature as potentially G4-associated (439 shRNAs) and shRNAs targeting 37 olfactory receptors as non-targeting controls (143 shRNAs). To reduce the total hairpin number, the number of shRNAs per gene was capped at seven.

A focused screen approach has its own limitations: the derivation from the PDS and PhenDC3 genome-wide screen in A375 cells creates an intrinsic bias to synthetic lethalities that can be uncovered with other ligands and/or cell lines. In light of this, the focused screen adds to but does not supersede or undermine the results of the genome-wide screen.
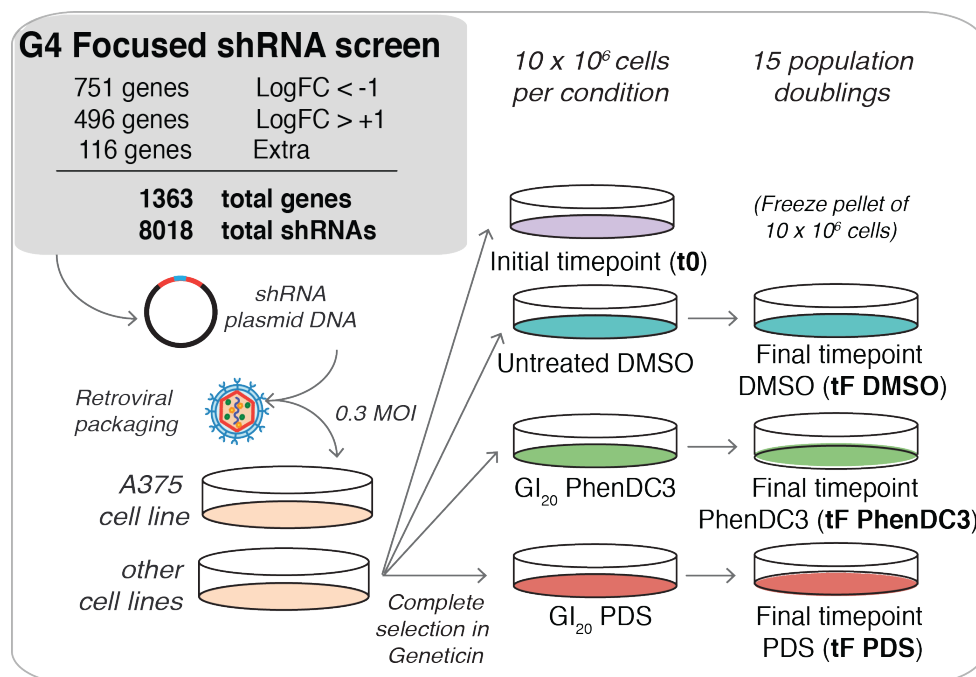
Figure 2.24. Development of a focused "quadruplex" pool strategy to circumvent batch effects
Schematic of the shRNAs in the custom pool, containing ~8000 hairpins and experimental outline for application to the A375 and other cell lines. Cells were infected at an MOI of 0.3 with retroviral particles. The reference time point (t0) was harvested following 100 % selection. Cells were cultured for 15 population doublings in DMSO, PDS or PhenDC3 and the final time point was harvested (tF).

2.4.8 Screening the A375 cell line with the focused library

Having designed the focused library, the A375 cell line was screened using the same experimental procedure, ligand concentration and parameters as for the genome-wide screen. Primarily, this was a proof-of-principle validation that this screening methodology could replicate genome-wide screen sensitivities. Secondly, as all shRNAs are simultaneously exposed to PDS and PhenDC3, this enables a stringent, comparative analysis of their sensitivities.

PDS and PhenDC3 caused continuous growth inhibition for the experiment duration (Figure 2.25A). Sequencing t0 confirmed that the maximum number

of shRNAs per gene was seven (Figure 2.25B). Depleted hairpins and sensitisers were defined as for the genome-wide screen, and recovered 2,236 depleted shRNAs (370 genes) across both ligands compared to only 221 shRNAs (10 genes) significantly depleted in DMSO (Figure 2.25C-D). 342 of the 360 ligand-specific depleted genes (95.0 %) were classed as sensitisers (Figure 2.25E), which represent 40.6 % of genes identified in the genome-wide screen (Figure 2.25F). The low number of differentially expressed genes in DMSO and high proportion of shRNA/genes depleted after G4 ligand-treatment provides confidence in the focused screen as a reliable, albeit more refined tool for investigating G4-sensitivities.
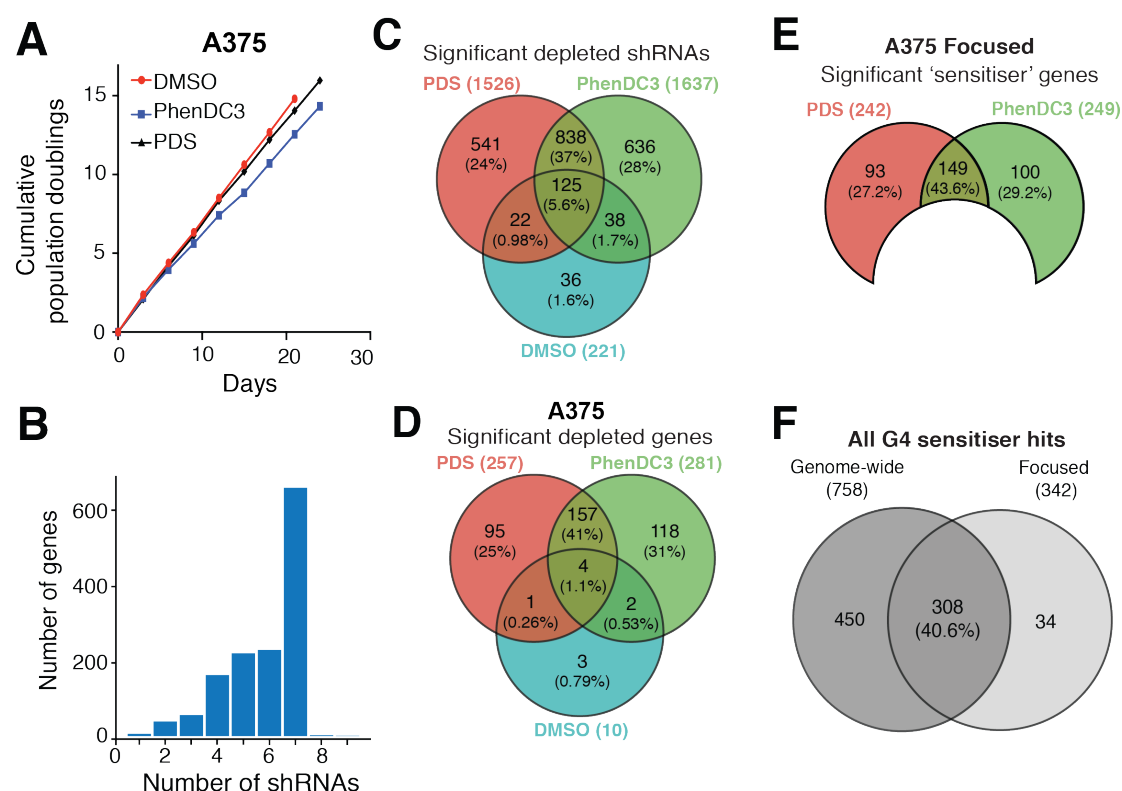
Figure 2.25. Focused screening in A375 reproduces 40 % of the genome-wide sensitisers

(A) Graph comparing cumulative population doublings for the experiment duration for the three conditions (DMSO, PDS and PhenDC3)

(B) Number of shRNAs targeting each gene, with the maximum capped at 7 per gene

(C-E) Venn diagrams showing (C) significantly depleted shRNAs in t15 v t0 (FDR ≤ 0.05) for DMSO, PDS and PhenDC3 treatment, (D) significantly depleted genes (minimum 50 % or 3 significant shRNAs), (E) Venn diagram of significant sensitiser genes from the A375 focused screen (50 % or 3 significantly depleted with a median $log_2FC \leq -1$).

(F) Venn diagram showing the overlap between combined PDS and PhenDC3 sensitisers from the genome-wide and A375 focused screen.

Next, individual sensitivities to each molecule in the focused screen were determined using overlap with the genome-wide screen as a test of reproducibility (Figure 2.26A). This revealed 89 and 161 unique sensitivities for PDS and PhenDC3 respectively, and 40 genes common to both ligands (290 combined total). KEGG analysis revealed that although enriched pathways and contributing genes differ between PDS and PhenDC3 (Figure 2.26B-C), they can be broadly split into four categories: DNA-related (replication, HR and cell cycle); RNA-related (ribosome and spliceosome); mitochondrial oxidation pathway (TCA cycle and Huntington's disease) and ubiquitin-mediated proteolysis. Analysis of the GO Molecular Function and Biological Process terms (Figure 2.26D-E), revealed an enrichment in nucleic acid terms similar to those found for the genome-wide sensitisers, ranging from mRNA and RNA processing to DNA metabolism and replication, consistent with G4 targeting. Notably, 16 of 19 PhenDC3 terms were related to RNA and mRNA whereas PDS exhibited a DNA and chromatin bias (21 of 31 terms).
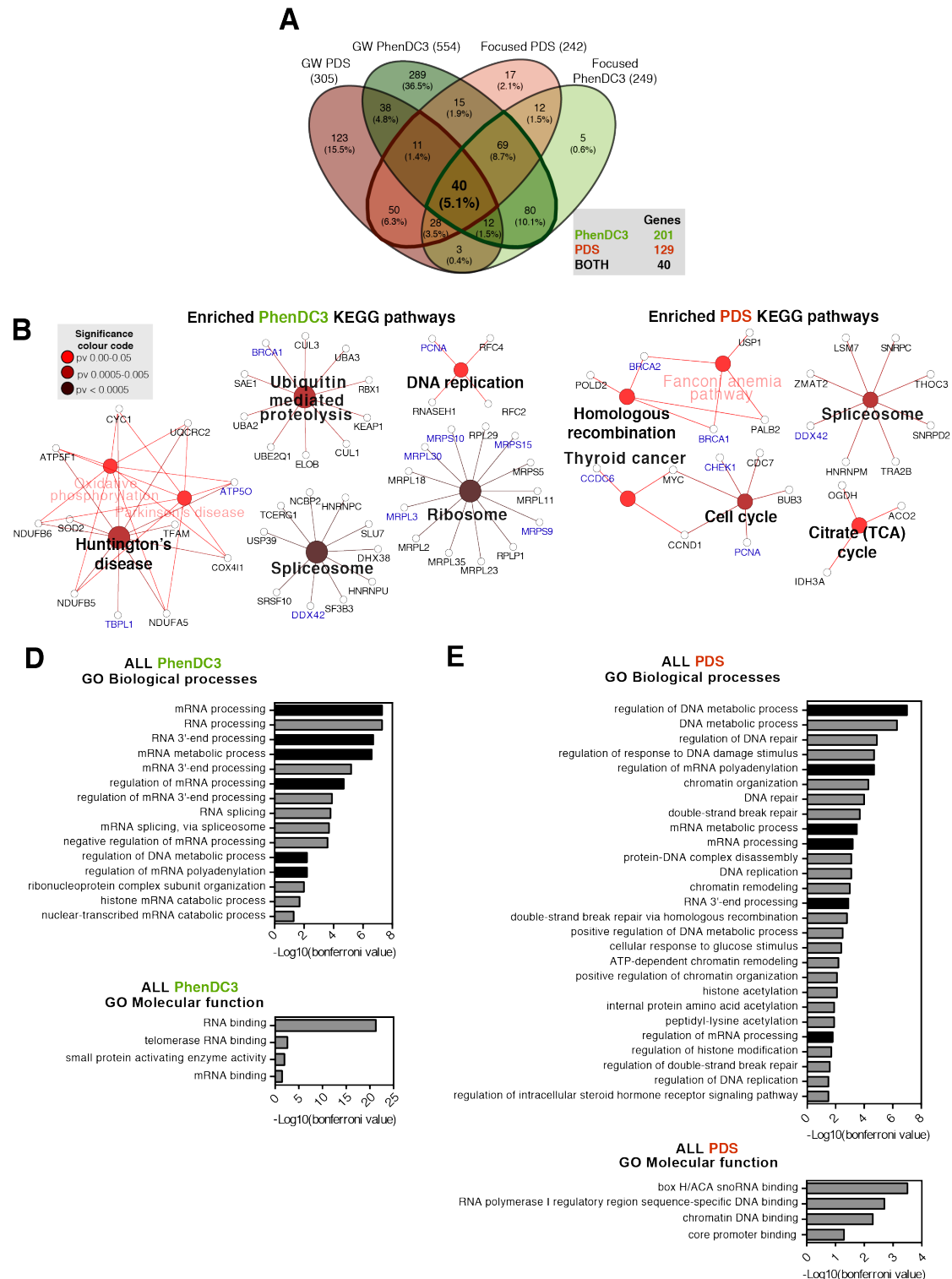
Figure 2.26. Focused screening in A375 cells reveals differences in PDS and PhenDC3 sensitivities alongside a core set of common sensitisers
(A) Venn diagrams overlapping genome-wide and A375 focused screens ligand-specific sensitisers: PDS – red (129 genes); PhenDC3 – green (201 genes), (B-C) Enriched KEGG pathways for (B) PhenDC3 and (C) PDS sensitisers in common between the genome-wide and focused screen (Bonferroni corrected p-value < 0.05). Node colour denotes significance as indicated and node size reflects the number of genes associated with the KEGG term, (E-F) Enriched Gene Ontology terms (GO Biological Processes and Molecular Functions) for (D) PhenDC3 and (E) PDS.

The 40 sensitiser genes common to PDS and PhenDC3 were designated as high-confidence as they arise from two independent ligands hitting the same target. These genes were designated into nucleic acid-related and 'other roles' (Figure 2.27A) using several databases (see Methods for details). Most of these genes were associated with DNA or RNA binding processes, such as transcription, replication and translation, consistent with direct targeting of G4-structures by the ligands. Many of the uncovered sensitisers revealed processes not previously recognised as 1) G4-ligand targets or 2) having G4 involvement, for example, the ubiquitin pathway (*PSMC4*, *RNF20*).
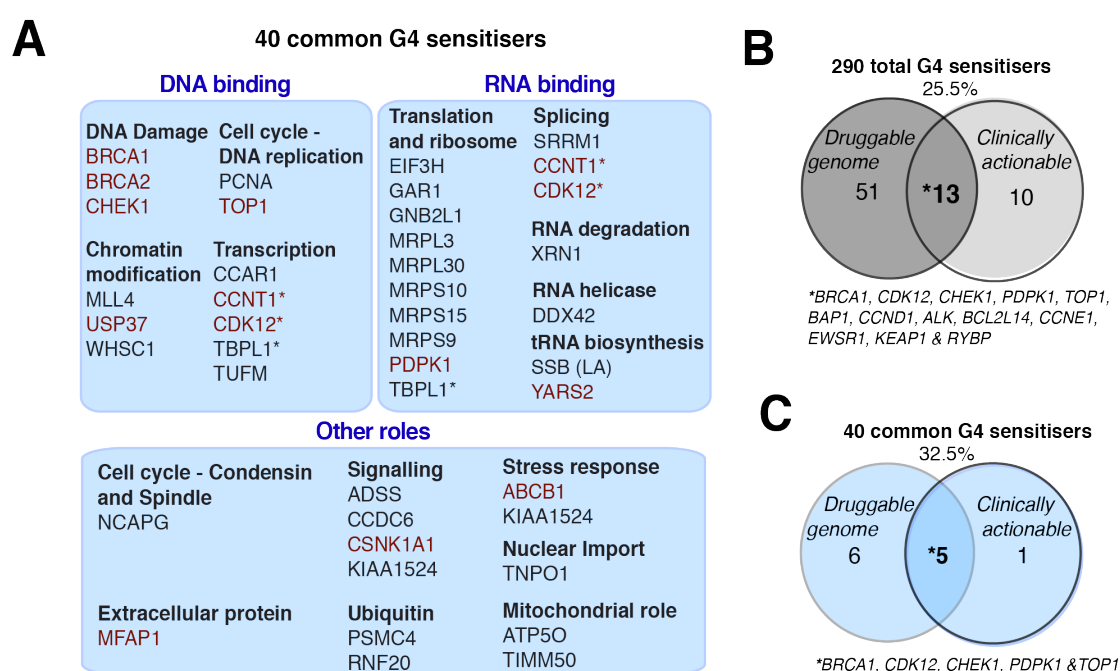
Figure 2.27. Focused A375 screening reveals 40 high confidence G4-ligand synthetic lethalities
*(A) DAVID, STRING (experimental data, co-expression, medium confidence (≥ 0.4)) interaction and UniprotKB data were used to categorize the 40 common hits between PDS and PhenDC3. Genes in red are found in the Drug Genome Interaction database (DGIdb 2.0). *genes in multiple categories.*
(B-C) Venn diagrams of sensitisers common to the A375 genome-wide and focused screens overlapping the DGIdb for (B) 290 PDS or PhenDC3 hits; and (C) 40 sensitisers common to both drugs and both screens. Druggable genome denotes genes with known or predicted drug interactions. Clinically actionable denotes genes actively used in targeted clinical sequencing panels for precision medicine in cancer. The genes found in the DGIdb are denoted as a percentage of the total.

2.4.9 Exploring therapeutic options within focused screen sensitisers

While the genome-wide screen provides a global appreciation of gene deficiencies sensitive to G4-ligand treatment, the overlap between this and the focused screen provides a set of reproducible hits to investigate therapeutic options (for discussion of therapeutic options for genome-wide sensitisers, section 2.5). For this, the druggable genome interaction database (DGIdb) (Griffith *et al*, 2013) was used to explore the potential for combinatorial pharmacological inhibition with G4-ligands. Two classifications were used: 'Druggable Genome' (genes with known or predicted drug interactions) and 'Clinically Actionable' (genes used in clinical cancer trials). Within the 290 genes that overlapped the A375 genome-wide and focused screen, 74 were identified as pharmacological targets, 13 of which were both druggable and clinically actionable (Figure 2.26B). This includes KEAP1, an E3 ubiquitin ligase adapter protein, bridging the therapeutic gap between ubiquitin-mediated proteolysis and G4-ligand synthetic lethality. For the 40 common hits between PDS and PhenDC3, 12 overlapped with DGIdb (Figure 2.27C), including five in both druggable and clinically actionable classifications (*BRCA1, CHEK1, CDK12, TOP1, PDKP1*).

2.4.10 Applying the focused screen approach to an independent cell line

Next the focused screen was applied to a second cell line to explore sensitivities common to distinct cell types. The HT1080 fibrosarcoma cell line was chosen for reasons discussed earlier (see section 2.3.1-2.3.2) and

additionally as it is derived from a distinct lineage (mesenchymal lineage) from the A375 melanoma cell line (neural crest lineage). However, the earlier pilot screen (section 2.3.3) suggested that 96 h drug treatment viability assays did not accurately predict a GI_{20} PhenDC3 concentration for the shRNA screen experiment duration (15 population doublings). To surmount this, several concentrations were trialled in a 15 day rather than 96 h experiment (data not shown), to determine a PhenDC3 concentration that gave the most similar growth inhibition to the 0.5 µM PDS treatment observed in the pilot screen. 1 µM PhenDC3 was therefore chosen and the focused shRNA library was applied to the HT1080 cell line. The growth inhibition for both PDS and PhenDC3 was seen to be linear for the experiment duration (Figure 2.28A) and the hairpin representation matched that of both the screen composition, and that of the A375 t0 (Figure 2.28B). Overall, the HT1080 screen revealed 121 ligand-specific depleted genes (838 hairpins, Figure 2.28C-D), all of which exhibited $log_2FC \leq -1$, and thus were considered sensitisers (Figure 2.28E).
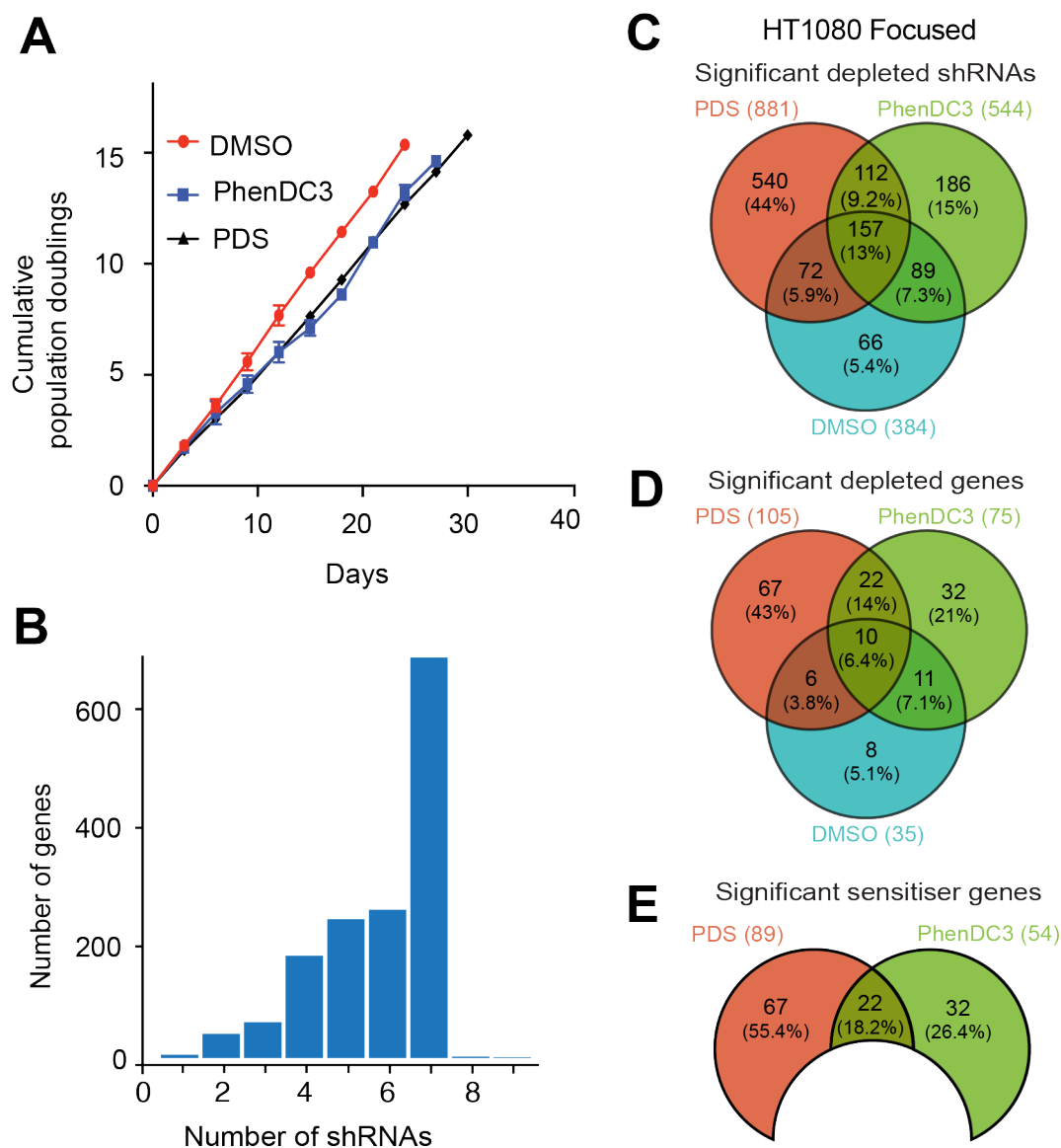
Figure 2.28. Successful application of the focused screen approach to the HT1080 cell line

(A) Graph comparing cumulative population doublings for the duration of the HT1080 experiment for the three conditions (DMSO, PDS and PhenDC3)

(B) Comparison between the numbers of shRNAs targeting each gene, with the maximum capped at 7 per gene

(C-E) Venn diagrams from the HT1080 focused screen showing (C) significantly depleted shRNAs in t15 v t0 (FDR ≤ 0.05) for DMSO, PDS and PhenDC3 treatment, (D) significantly depleted genes (minimum 50 % or 3 significant shRNAs; FDR ≤ 0.05) (E) sensitiser genes (significantly depleted in PDS or PhenDC3 but not DMSO and median $log_2FC ≤ -1$

2.4.11 G4-ligand synthetic lethalities common to two cell lines of different origins

In setting out to compare synthetic lethality between two different cell lines, it should be borne in mind that an inherent bias exists in the focused screen as it is primarily derived from A375 genome-wide hits. Thus some sensitivities in HT1080 may be missed. Sensitisers from the A375 genome-wide parental screen were compared to HT1080 sensitivities (Figure 2.29A). This revealed an overlap of 73 sensitisers, found reproducibly across PDS and PhenDC3 treatments in both cell lines which were enriched in spliceosome, HR and ubiquitin-mediated proteolysis processes ($p < 0.0005$) (Figure 2.29B). These data thus reveal a core set of G4-associated genes/pathways and synthetic lethalities common to the two cell lines. In particular this highlighted a central cluster of DNA damage proteins, including BRCA1 and BRCA2, extending PDS BRCA-deficient sensitivity to another two cell lines beyond HCT-116 and MEF cells (McLuckie *et al*, 2013; Zimmer *et al*, 2016) and revealing that PhenDC3 also has scope to be used in BRCA-deficiencies as for PDS and CX-5461. Additional DDR proteins were also highlighted in this comparison and notably several are deficient in cancers, including *BAP1* (uveal melanoma) and *PALB2* (breast and ovarian cancer), suggesting these malignancies may be similarly susceptible to G4-ligand treatment.
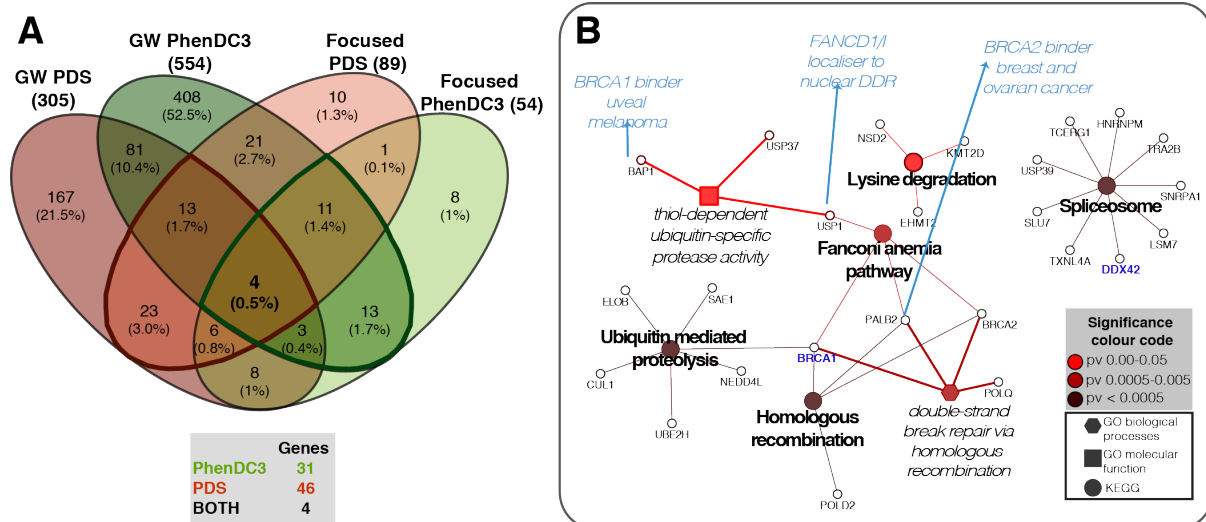
Figure 2.29. Successful application of the focused screening approach to the HT1080 cell line
(A) Venn diagram overlapping ligand-specific sensitisers between genome-wide A375 and HT1080 focused screens (50 % or 3 significantly depleted with a median log$_2$FC ≤ -1): PDS – red (46 genes); PhenDC3 – green (31 genes)
(B) Enriched KEGG and GO pathways of the 73 combined sensitisers (PDS or PhenDC3).

Next gene sensitivities, common across all three screens (i.e. A375 genome-wide, focused screens in A375 and HT1080 cells) were defined separately for PDS (29 genes, Figure 2.30A) and for PhenDC3 (22 genes, Figure 2.30B). These were categorised according to nucleic-acid related and "other" roles (Figure 2.30C-D). Lastly, *BRCA1, TOP1, DDX42* and *GAR1* were identified as key sensitisers common to both cell types and for both G4-ligands (Figure 2.29E; also highlighted in blue in Figure 2.29C-D). With the exception of these four genes PDS and PhenDC3 showed different consistent hits, however many performed similar roles, including transcription, splicing and ubiquitin mediated proteolysis. This lack of overlap of the strongest sensitivities could suggest that PDS and PhenDC3 target different genes within similar pathways.

A PDS sensitisers

GW (305)   A375 (242)

159 (36.1%) | 100 (22.7%) | 92 (20.9%)
17 (3.9%) | 29 (6.6%) | 21 (4.8%)
22 (5.0%)

HT1080 (89)

B PhenDC3 sensitisers

GW (554)   A375 (249)

344 (55.7%) | 179 (29.0%) | 41 (6.6%)
9 (1.5%) | 22 (3.6%) | 7 (1.1%)
16 (2.6%)

HT1080 (54)

C 22 consistent PhenDC3 sensitisers

DNA binding

Chromatin modification
WHSC1

Cell cycle - DNA replication
AHCYL1*
TOP1

Transcription
CNBP
HTATSF1
TCEB2
TCERG1

DNA Damage
BRCA1

RNA binding

Translation and ribosome
FASTKD2*
FCF1
GAR1

RNA helicase
DDX42

Splicing
AHCYL1*
SLU7

tRNA biosynthesis
NARS2

Other roles

Cell cycle - Condensin and Spindle
CUL1*
SUGT1
WEE1

Signalling
AHCYL1*

Mitochondrial role
FASTKD2*
PMPCB

Ubiquitin
CUL1*
SAE1

Vesicle/organelle transport
CAMSAP1L1

Unknown function
CBWD2
CBWD3

D 29 consistent PDS sensitisers

DNA binding

DNA Damage
BAP1
BRCA1
BRCA2
PALB2
POLD2
POLQ
USP1*

Chromatin modification
ANKRD11
HCFC1*
MLL4
UBE2H*

Cell cycle - DNA replication
HCFC1*
POLD2
POLQ*
TOP1
USP37*

Transcription
FUS*
HCFC1*
MYC

RNA binding

Translation and ribosome
FUS*
GAR1
MRPS10*

Splicing
HNRNPM
LSM7

RNA helicase
DDX10
DDX42

Other roles

Cell cycle - Condensin
NCAPG

Signalling
KIAA1524
TRA2B

Mitochondrial role
ABCB7
IDH3A
MRPS10*

Ubiquitin
NEDD4L
UBE2H*
USP1*
USP37*

Extracellular protein
SCGBL

Vesicle/organelle transport
CLTC

E

GW A375 shRNA screen					
Gene	hairpins in t0	Significant sensitiser hairpins (PhenDC3)	median logFC (PhenDC3)	Significant sensitiser hairpins (PDS)	median logFC (PDS)
BRCA1	5	3	-2.09	3	-2.80
DDX42	5	4	-4.51	3	-3.00
GAR1	7	5	-4.51	4	-3.63
TOP1	8	4	-3.92	3	-3.13

A375 Focused shRNA screen					
Gene	hairpins in t0	Significant sensitiser hairpins (PhenDC3)	median logFC (PhenDC3)	Significant sensitiser hairpins (PDS)	median logFC (PDS)
BRCA1	6	3	-3.56	4	-2.84
DDX42	7	7	-2.14	6	-2.16
GAR1	6	3	-3.79	3	-1.54
TOP1	7	3	-3.30	4	-4.34

HT1080 Focused shRNA screen					
Gene	hairpins in t0	Significant sensitiser hairpins (PhenDC3)	median logFC (PhenDC3)	Significant sensitiser hairpins (PDS)	median logFC (PDS)
BRCA1	6	3	-1.50	4	-2.21
DDX42	7	4	-2.91	5	-2.93
GAR1	6	4	-3.12	4	-2.76
TOP1	7	3	-3.10	4	-4.70

PDS & PhenDC3 sensitisers

GW (101)   A375 (149)

61 (28.1%) | 36 (16.6%) | 98 (45.2%)
0 (0%) | 4 (1.8%) | 11 (5.1%)
7 (3.2%)

HT1080 (22)

Figure 2.30. PDS and PhenDC3 sensitivities common across all screens
(A) Venn diagram overlapping ligand-specific sensitisers between genome-wide A375 and HT1080 focused screens (50 % or 3 significantly depleted with a median $\log_2 FC \leq -1$): PDS – red (46 genes); PhenDC3 – green (31 genes)
(B) Enriched KEGG and GO pathways of the 73 combined sensitisers (PDS or PhenDC3).
(C-D) Venn diagram of hits consistently identified as sensitisers across the three screens for C) PDS and D) PhenDC3
(E-F) DAVID, STRING (experimental data, co-expression, medium confidence ($\geq$ 0.4) interaction) and UniProtKB data was used to categorize consistent PDS and PhenDC3 hits. *genes in multiple categories. Sensitisers common to both drugs in blue.
(G) Venn diagram and tables showing the number of depleted hairpins and median $\log_2 FC$ values for four hits consistently sensitised to PDS&PhenDC3 across the three screens.

2.5 Discussion

2.5.1 Overall outcomes and observations

The overall aim of this chapter was to reveal new insights into G4-biology and identify new vulnerabilities for possible exploitation in a cancer context. G4 structures are emerging as promising clinical synthetic lethal targets in BRCA-deficient tumours (Xu *et al*, 2017; McLuckie *et al*, 2013). However, I reasoned that some cancer-associated genotypes that are synergistic with G4-ligand treatment were probably not yet defined. Below I will summarise the synthetic lethal findings uncovered here, including specific illustrative examples.

Genetic screening of the protein coding genome revealed over 700 genes that, when depleted, are synthetically lethal with the G4-ligands, PDS and/or PhenDC3. The majority of genes identified have no documented G4-biology link. Among these, 21 G4-associated proteins were identified via database searching and text mining, including direct G4-binders (*ADAR, ATRX, DHX36, DNA2, FUS, MCRS1, RECQL4, SF3B3* and *XRN1*). For *ATRX*, telomestatin sensitivity was recently reported for *ATRX*-null murine neuroprogenitor cells (Watson *et al*, 2013). This supports the PhenDC3 sensitivity seen here with ATRX-deficient A375 cells and establishes ATRX-deficiency as a key synthetic lethality for G4-stabilisation. Of the other 20 genes with prior links to G4-biology, this is the first time that synthetic lethality has been demonstrated with G4-stabilising ligands.

The genome-wide screen was key to providing a global overview of G4-ligand associated synthetic lethalities. However, interpool batch effects were

observed (section 2.4.1-2.4.2). To overcome this, a "quadruplex" shRNA pool of hits was developed that allowed simultaneous testing of component shRNAs. This was successfully applied to A375 cells, validating 290 genome-wide hits and revealing 40 high-confidence hits, common to both PDS and PhenDC3. Application to the HT1080 cell line identified 73 genes common across to both cell types. This could reflect that a core set of common G4s exists in both cell lines, despite HT1080 deriving from a different lineage to the A375 cell line (mesenchymal versus neural crest lineage). It remains to be tested if these synthetic lethalities are also common to other cell lines.

Ultimately, four gene deficiencies were designated as 'key' synthetic lethalities (*BRCA1*, *TOP1*, *DDX42*, and *GAR1)*. These genes reflect a range of hypothesised G4-related processes that comprise DDR (*BRCA1)* (McLuckie *et al*, 2013); relieving transcription and replication torsional stress (*TOP1)* (Wang, 2002); RNA unwinding (*DDX42)*, (Uhlmann-Schiffler *et al*, 2006); and rRNA processing and telomeric maintenance (*GAR1*) (Girard *et al*, 1992). These are investigated and discussed further in chapter 3.

2.5.2 Different synthetic lethalities for PDS and PhenDC3

Focused screening of the A375 cell line allowed investigation of recurring PDS and PhenDC3 sensitivities, and revealed that a large number of differences existed. Although largely considered as pan G4 stabilisers *in vitro* and capable of deregulating both RNA- and DNA-related roles in cells (see introduction section 1.4.2.3), there was a bias of RNA related terms for synthetic lethal interactions with PhenDC3-treated cells, while PDS

sensitivities were more often DNA-related. This may reflect an inherent variation in affinity for DNA over RNA-G4s or differences in ligand subcellular location when added to cells. For example, PDS may more easily enter the nucleus than PhenDC3, and thus more readily stabilise DNA-G4s. These biological response differences highlight the importance of cellular investigations to complement *in vitro* biophysical characterisation when designing and developing G4-stabilising ligands. Overall, beyond the RNA and DNA related term variations, PDS and PhenDC3 sensitivities were generally mediated by different proteins in similar pathways. These pathways are discussed below.

2.5.3 Sensitisers belong to several interlinked categories

Overall, the G4-sensitiser genes uncovered by genome-wide and focused screening uncovered several cellular pathways and proteins including DNA damage, helicases, transcription/chromatin interactors, spliceosome, cell cycle and ubiquitin-mediated proteolysis. These are individually discussed below in terms of possible mechanisms and cancer associations.

Within DNA damage-associated sensitivities, several homologous recombination (HR) proteins were identified as novel G4-sensitisers, including *PALB2* (a BRCA2 localiser), *BAP1* (BRCA1-binding partner and deubiquitinase for Histone 2A and tumour suppressor HCFC-1) and the deubiquitinase *USP1* required for FANCD2, FANCI and PCNA localisation to sites of DNA damage (reviewed in Harrigan et al., 2017). These findings extend our existing knowledge of the G4-associated DDR beyond BRCA1/2,

and highlights HR repair mechanisms as key cellular responses to unresolved G4s. Furthermore, *PALB2* inactivation causes predisposition to myeloid leukaemia, Wilm's tumour and Fanconi anaemia and *BAP1* deficiencies are found in uveal melanomas and mesotheliomas (reviewed in Carbone et al., 2013; Nepomuceno et al., 2017). This suggests new synthetic lethality opportunities for therapeutic G4-ligand intervention, beyond BRCA-deficiencies. One prospect would be using G4-ligands in situations of acquired resistance to conventional therapies that exploit DDR-deficiencies, such as cisplatin (general) and olaparib (specific for PARP inhibition). Supporting this, HR-defective olaparib-resistant cells remain PDS sensitive (Zimmer *et al*, 2016), suggesting that the engagement of independent pathways/genes underlies such synthetic lethality.

The enrichment of many helicases as sensitisers supports their mechanistic involvement with G4-structures. Helicase deficiencies are likely to reduce G4-structure resolution and shift the equilibrium from unfolded towards the G4-folded state. I propose that this increase is synergistic with G4-stabilising ligand treatment causing lethality. Sensitisers included both helicases previously implicated in G4 biology and those without prior G4 links. For the former, RECQL4 (Rothmund-Thomsun syndrome) and RTEL1 (Hoyeraal-Hreidarsson Syndrome), whose deficiencies impart increased risk of cancer, autoimmunity and premature aging (reviewed in Suhasini and Brosh, 2013) were uncovered and provides proof-of-principle of new synthetic lethality targets in these cancers. Knockdown of several known G4-helicases such as BLM, WRN, PIF1 and FANCJ (reviewed in Wu and Brosh, 2010) actually did

not cause G4-ligand sensitivity, and were also not lost in the DMSO control, suggesting non-essential roles. There could be several explanations, including redundancy (Spillare *et al*, 2006) and the use of low ligand concentrations . Several RNA helicases not previously associated with G4s or synthetic lethality with G4-ligands were also identified, including XRN2 and DHX38, DDX10 and DDX42. Notably, XRN2 is closely related to the 5'exoribonuclease XRN1, known to unwind RNA and DNA G4s (Zhang *et al*, 2002). As both were identified as sensitisers, this could suggest XRN1 also binds G4 structures, and that disruption of their mRNA decay roles are acutely sensitive to G4-stabilising ligand treatment. Furthermore, DHX38 inactivation causes splicing defects and retinitis pigmentosa (Ajmal *et al*, 2014). Although this does not directly offer therapeutic opportunities for retinitis pigmentosa treatment with G4-ligands, it does highlight possible G4 links to pathologies beyond cancer, and it would be interesting to explore the involvement of G4s in the progression of this disease.

Congruent with G4s as normal structural features of chromatin (Hänsel-Hertsch *et al*, 2016), numerous chromatin remodellers/transcriptional regulators were identified, most previously unassociated with G4 biology or G4 ligand synthetic lethality. These include the transcription factor *ANKRD11* a putative tumour suppressor in breast cancer (Noll *et al*, 2012; Lim *et al*, 2012; Neilsen *et al*, 2008); the cell cycle transcriptional coregulator HCF-1 (Wysocka *et al*, 2001; Reilly *et al*, 2002; Tyagi *et al*, 2007); *MLL4,* a lysine methyl transferase frequently inactivated in cancer (Rao & Dou, 2015; Froimchuk *et al*, 2017; Kandoth *et al*, 2013) and *WHSC1,* a histone

methyltransferase overexpressed in cancers including prostate and multiple myeloma (Bennett *et al*, 2017). Other notable G4-sensitisers, also without prior links to G4-biology, are the SWI/SNF proteins SMARCA4/B1/E1. These proteins are mutated in 20% of cancers (Shain & Pollack, 2013; Kadoch *et al*, 2013) and suggests an innovative synthetic lethal possibility whereby certain chromatin remodeller disease-associated genotypes may be G4-ligand sensitive. For example, loss of SMARCB1, SMARCA4, or KEAP1 is linked to malignancy of triple-negative breast cancer (Wijdeven *et al*, 2015), and deficiencies in these three genes were identified as sensitive to G4-ligand treatment.

The emergence of the ubiquitin-proteosome pathway, including ubiquitin-like modifications such as neddylation, represents a largely uninvestigated area with respect to G4s. Currently, the only other reported ubiquitin-G4 relationship is the yeast ubiquitin protein MMS1, recently identified as a possible G4 binding protein, although this interaction seems to be independent of a ubiquitin ligase role (Wanzek *et al*, 2017). Ubiquitin-G4-relationships identified within the genome-wide screen extend the full breadth of the proteosomal degradation pathway, including E1-ligases (*UBA3, UBA2, SAE1*), E2-ligases (*UBE2H*), E3-ligases (*NEDD4L, RBX1, CUL1, RNF20*), deubiquitinating enzymes (*USP1* and *USP37*) and proteasome components (*PSMC2)*. Their roles include DDR, cell cycle progression and proliferation. For example the E3-ligase *RNF20*, is involved in chromatin remodelling and the DDR (Moyal *et al*, 2011; Shema *et al*, 2008). As the screen also identified these processes, this links ubiquitination to DDR/chromatin remodeling via an

interaction with G4s. The ubiquitin-associated proteins identified in the screen, are generally deregulated in cancer (Senft *et al*, 2018; Harrigan *et al*, 2017; Liu *et al*, 2015a; Wei & Lin, 2012). For instance, the deubiquitiase *USP37* is upregulated in lung cancer (Pan *et al*, 2015) whilst reduced expression of the E3-ubiquitin ligase NEDD4L predicts poor prognoses in hepatocellular (Zhao *et al*, 2018) and gastric carcinoma (Gao *et al*, 2012).

Persistent G4s are known to be inhibitory to DNA replication (reviewed in Valton and Prioleau 2016) and the replication-associated gene deficiencies identified here strengthen this association. Furthermore, as DNA replication is tightly coordinated with the cell cycle, unsurprisingly a set of cell cycle G4-dependencies were also revealed. These include *WEE1, PCNA, CHEK1, CCND1, CDC7, RFC2* and *RFC4*. While such proteins are often up-regulated in cancer (Matheson *et al*, 2016), certain cancer sub-types show depletions. The opportunities to exploit and extend the use of both ubiquitin and cell cycle proteins as therapeutic targets with G4-ligands will be discussed in section 2.5.4

These results also reveal that unresolved G4s are problematic for splicing and that this is exarcerbated by splicing component deficiencies. For example, both SRRM1, an essential splicing factor (Meissner *et al*, 2003) and it's co-factor FUS, a known G4-interactor (Takahama *et al*, 2013) and liposarcoma oncogene (Crozat *et al*, 1993) were uncovered as G4-sensitisers. Several other cancer associated splicing factors were identified as G4 sensitisers, including SRSF10 and HNRNPM overexpressed in colon and breast cancers

respectively (Dvinge *et al*, 2016; Takahama *et al*, 2013; Crozat *et al*, 1993).

2.5.4 Exploring synthetic lethal targets for therapeutic application

Many of the sensitisers identified are deregulated in cancer. There are two approaches to exploit these genes therapeutically. Firstly, disease-associated backgrounds deficient in genes within the sensitiser list could be sensitive to G4-ligands in a clinical setting. Secondly, for cancers/diseases dependent on and/or driven by sensitiser genes, G4-ligand treatment may sensitise cells to pharmacological inhibition of these targets. For the latter, combinatorial treatment is often more efficacious than single treatments as cells are unlikely to simultaneously develop resistance against two drugs (reviewed in Chan & Giaccia, 2011). Further, as lower doses of each drug can be used, this increases the therapeutic window compared to normal cells, minimising adverse side effects. Such scenarios are discussed below, with particular focus on sensitisers within ubiquitin and spliceosome pathways, as these represent areas of increasing biological interest in cancer (Liu *et al*, 2015a; Lee & Abdel-Wahab, 2016) and are largely unexplored for G4 involvement. More generally, such lethal interactions described in this chapter may have future value when more cancer-associated genotypes are uncovered, and when new pharmacological inhibitors are developed for key oncogenic targets.

2.5.4.1 Synthetic lethal targets within the COSMIC database

Several proteins within the genome-wide sensitiser list were found within the COSMIC database of cancer-associated genes. Several are current direct or

indirect pharmacological targets, perhaps providing combinatorial opportunities with G4-ligands. For example, the CDK12 inhibitor dinaciclib (SCH77965) (Parry *et al*, 2010) was developed to treat high-grade serous ovarian cancer, which often exhibits a gain-of-function oncogenic CDK12 mutation (Bajrami *et al*, 2014). Similarly, AZD1775 inhibits WEE1 kinase (Richer *et al*, 2017), a crucial G2/M transition regulator overexpressed in several cancers (Matheson *et al*, 2016). Of note, some colon and non-small cell lung cancers (NSCLC) are WEE1-deficient (Backert *et al*, 1999; Yoshida *et al*, 2004). For this tumour subset, WEE-1 deficiency encourages early mitotic entry resulting in aneuploidy and DNA damage, due to incomplete DNA replication. This may synergise, and thus be exploitable with G4-associated DNA damage.

For KEAP1, an oxidative stress response protein, inactivating mutations prevent a degradatory interaction with Nrf2, which is consequently highly expressed in several cancers (Abed *et al*, 2015). Simllarly, CPUY192018 disrupts the KEAP1-Nrf2 interaction and peforms a cytoprotective role in NCM460 colonic cells (Lu *et al*, 2016), supporting that inactivating KEAP1 both genetically and pharmacologically can promote cancer cell survival. Recent analysis of common mutations within 'The Cancer Genome Atlas' database, also revealed prevalent KEAP1 deletions in several cancers including thoracic and endometrial (Sanchez-Vega *et al*, 2018). It would be interesting to explore the sensitivity of such cancer genotypes to G4-stabilising ligands. Conversely, several studies suggest an anti-tumorigenic role for Nrf2, with the KEAP1 inhibitor CDDO-Me, used to treat leukaemias

and solid tumours, in addition to chronic kidney diseases and diabetes mellitus (Wang *et al*, 2014).

SMAD4, is inactivated 50 % of pancreatic adenocarcinoma (PDAC) (Schutte *et al*, 1996; Hahn *et al*, 1996) and correlates with disease agressiveness (Tascilar *et al*, 2001). SMAD4 stability is negatively regulated by phosphorylation by glycogen synthase kinase (GSK), thus several GSK inhibitors (GSKi) have entered clinical trial (Demagny & De Robertis, 2016) including NCT01632306 (metastatic pancreatic cancer); NCT01214603 (acute leukaemia) and NCT01287520 (McCubrey *et al*, 2014). An alternative method, particularly in SMAD-deficient tumours could be G4-ligand treatment, either as the first-in-line treatment, or in GSKi resistant cancers. Similarly, CCDC6 a cell cycle regulator involved in apoptosis and DDR, is inactivated in thyroid (Puxeddu *et al*, 2005) and lung cancers (Morra *et al*, 2015). CCDC6-deficient lung cancers are cisplatin-resistant but are olaparib sensitive (Morra *et al*, 2015); the data presented here suggests G4-ligand treatment as an alternative to olaparib.

A less explored therapeutic target is oncogene DEK, overexpressed in colorectal cancer (Lin *et al*, 2014) and melanoma (Khodadoust *et al*, 2009) and may be interesting to explore with respect to G4-ligand sensitivity. This protein has multiple tumour facilitating roles including inflammation induction (Kavanaugh *et al*, 2011), a function additionally exarcerbated in inflammatory arthritis. For the latter, the use of SELEX-generated DNA aptamers to 'mop-up' excess DEK is under investigation (Mor-Vaknin *et al*, 2017) but has yet to

be explored from a cancer perspective, where combinatorial treatment with G4-ligands could benefit both DEK-driven diseases. Additionally, DEK binds cruciform secondary structures, another non-canonical DNA structure (Kavanaugh *et al*, 2011). Whether this affinity for non-canonical structures extends to G4s and whether this underlies the G4-ligand synthetic lethality associated with DEK-deficiency, should also be explored.

Similarly, XRN2 has been investigated as a chemotherapeutic target due to its upregulation in many cancers including prostate, brain, myeloma and lung (Lu *et al*, 2010) promoting EMT and metastasis in the latter (Zhang *et al*, 2017). XRN2 shRNA-induced depletion reduced migation and mesenchymal marker expression in multiple lung cancer cell lines (Fong *et al*, 2015) and within the screen, caused G4-ligand susceptibility. Together these experiments suggest a promising prospect for combinatorial XRN2 pharmacological inhibition and G4-ligand intervention. WHSC1, overexpressed in prostate cancer, multiple myeloma (Bennett *et al*, 2017), and mantle cell lymphoma (Beà *et al*, 2013), is an early experimental cancer target. Some success has been achieved with the antiparasitic, pan-methyltransferase inhibitor suramin (McGeary *et al*, 2008). More recently, the crystal structure of WHSC1 helped identify N-alkylsinefungin derivatives as more specific inhibitors (Tisi *et al*, 2016) and screening of 12,251 bioactive molecules, identified five further potent candidate inhibitors (Coussens *et al*, 2017). As the pharmacological exploration of these proteins become more advanced, the opportunities for combinatorial use of G4-stabilising ligands may be considered.

2.5.4.2 Therapeutic opportunities with ubiquitin proteins

Ubiquitin proteins were consistently enriched within the G4-sensitisers and, as they are often deregulated in cancers they are emerging as chemotherapetic targets (Huang & Dixit, 2016). For example, proteosome inhibitors have shown clinical success with multiple myeloma treatment including bortezomib (Chen *et al*, 2011a; Mattern *et al*, 2012), CEP187710 (Phase I), and carfizomib (Phase III against relapsed mutliple myeloma) (reviewed in Edelmann, Nicholson, & Kessler, 2011). Based on the ubiquitin-mediated proteolysis enrichment within the sensitisers, this proteosome inhibition could be potentiated by combinatorial G4-ligand treatment.

However, ubiquitin is not solely a degradative mark and can additionally alter protein localisation and/or function (Chen & Sun, 2009), necessitating the development of drugs targeting specific ubiquitin components, several of which are identified as sensitisers. One example is USP1, a DDR regulator in Fanconi's anaemia (Nijman *et al*, 2005) and translesion synthesis (Huang *et al*, 2006) pathways. USP1 mRNA overexpression is observed in melanoma, gastric, cervical and non-small cell lung (NSCLC) cancers while other malignancies show underexpression, including leukaemia and lymphoma (reviewed in García-Santisteban, Peters, Giovannetti, & Rodríguez, 2013). A high-throughput screen identified pimozide and GW7947 as potent USP1-targeting drugs (Chen *et al*, 2011b). Similarly, MLN4924 (Phase I clinical trials) targets UBA3, which is upregulated in AML and multiple solid cancers (Soucy *et al*, 2009), a protein identified as a G4-ligand sensitiser. As more ubiquitin components are linked to cancer and identified as candidate

chemotherapeutic targets, more therapeutic possibilities for G4-ligands may become increasingly evident, both as single agents for deficient cancers and combinatorial treatments in malignancies where they are overexpressed.

2.5.4.3 Therapeutic opportunities with splicing proteins

As reported here for PDS and PhenDC3, previous genetic screens with anti-cancer agents such as gemicitabine and cisplatin suggest that their efficacy can also be enhanced by deficiencies in HR, ubiquitin and replication pathways (Martens-de Kemp *et al*, 2017; Azorsa *et al*, 2009). Conversely, perhaps providing a novel and specific niche for G4-ligand based therapies, several cancer-associated splicing components were found in the G4-sensitisers, a pathway not enriched in these earlier genetic screens. One possible therapeutic avenue is via SRSF10, overexpressed in colon cancers (Zhou *et al*, 2014). Additionally, SRSF10 inhibition impairs HIV-1 replication, leading to the development of an inhibitor, 1C8, for HIV treatment (Shkreta *et al*, 2017; Cheung *et al*, 2016). While it remains unexplored whether 1C8 has similar efficacy in cancers, use of G4-ligands could sensitise cells to SRSF10 inhibition in both HIV and cancer settings. Rather than targeting specific splicing components, the anti-tumour drug E7107, a general splicing inhibitor which prevents spliceosome assembly (Kotake *et al*, 2007) may provide more viable combinatorial opportunities. As more dysregulated splicing genotypes are identified in a cancer context, such combinatorial opportunites will become more apparent.

2.5.5 Concluding remarks

In conclusion, I have addressed the aim of identifying genes and pathways involved with G4 structures and have provided insights into potential disease-related synthetic lethalities for G4-ligands, particularly focusing on cancer. These findings establish a unique and comprehensive resource that will give important insights for future experiments in G4 biology and therapeutic possibilities. Furthermore, the use of low ligand concentrations and lethality as a readout likely favours deficiencies in genes most sensitive to G4-ligand intervention. Such sensitivity suggests that these proteins are key players in synthetic lethality and G4 regulation, and will thus be of both clinical and basic biological significance, both within and beyond the G4-field. Finally, a focused "quadruplex" pool and pipeline was developed and tested in two cell lines, providing a tool to investigate the synthetic lethalities of other G4-stabilising ligands.

Chapter 3

An siRNA approach to validate and extend four key G4-stabilising ligand sensitivities uncovered by functional genetics

3.1 Objectives

In chapter 2, four key, reoccurring synthetic lethal interactions with the G4-stabilising ligands PDS and PhenDC3 (*BRCA1*, *TOP1*, *GAR1* and *DDX42*) were identified. These four genes are likely to have important roles in G4-biology and/or the response to stabilised G4-structures. Consequently, it was imperative to validate their PDS and PhenDC3 sensitivity via an independent methodology, and confirm that this phenotype correlated with protein knockdown. For this a siRNA approach was chosen, an alternative RNAi method (see section 2.2.1). Two possible scenarios for a validated sensitisation response are depicted (Figure 3.1). Briefly, protein depletion in the absence of G4-ligands (i.e. DMSO treatment) either does not alter cell growth (scenario 1) or causes slight growth inhibition (scenario 2) compared to cells transfected with a non-targeting control siRNA (blue and red solid lines respectively). Hence, for scenario 1, greater growth inhibition following G4-stabilising ligand treatment for the targeting siRNA than for the non-targeting control (blue and red dotted line respectively) indicates sensitisation. Conversely for scenario 2, protein depletion results in G4-ligand sensitivity if

growth inhibition compared to the non-targeting control, is greater than seen

following DMSO treatment.

Sensitiser - Scenario (1)

Sensitiser - Scenario (2)
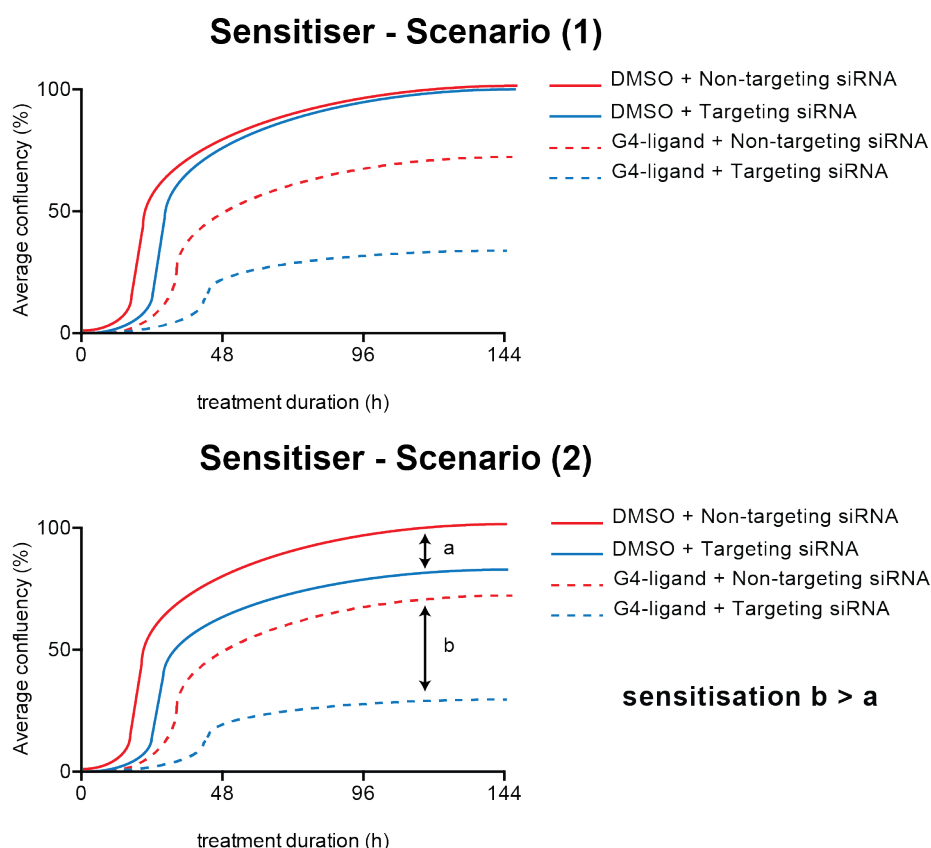
Figure 3.1. Schematic of the expected results for a validated sensitiser
Scenario (1) Depleting the protein of interest does not cause a growth defect compared to the non-targeting siRNA control upon DMSO treatment but does following G4-ligand treatment. Scenario (2) Protein depletion causes a growth defect compared to non-targeting siRNA control in the DMSO condition, but this growth defect is greater for G4-ligand treatment. Solid lines = DMSO treatment; Dotted lines = G4-ligand treatment; Red = non-targeting siRNA; Blue = Targeting siRNA.

Also within this chapter, to evaluate whether these four genes may represent

global synthetic lethalities to G4 stabilisation, also explored was the sensitivity

of the siRNA knockdowns to CX-5461 (see section 1.4.2.2; structure shown in

Figure 3.8A), a G4-ligand in Phase II clinical trials

(https://clinicaltrials.gov/ct2/show/NCT02719977; Xu et al., 2017). This

molecule was independently designed and is unrelated to PDS and PhenDC3.

Synthetic lethality was further probed with a carboxylated PDS derivative (cPDS). Unlike PDS, cPDS selectively binds RNA G4 *in vitro* (Di Antonio et al., 2012) and in cells, increasing cytoplasmic but not nuclear BG4 for the latter (Biffi *et al*, 2014a) (Figure 3.2B-C). This is the first time that the sensitivity of cells to cPDS, and in particular those of a certain genotype, has been investigated, and may go someway to evaluate whether observed phenotypes arise from either RNA-G4 or DNA-G4 stabilisation.
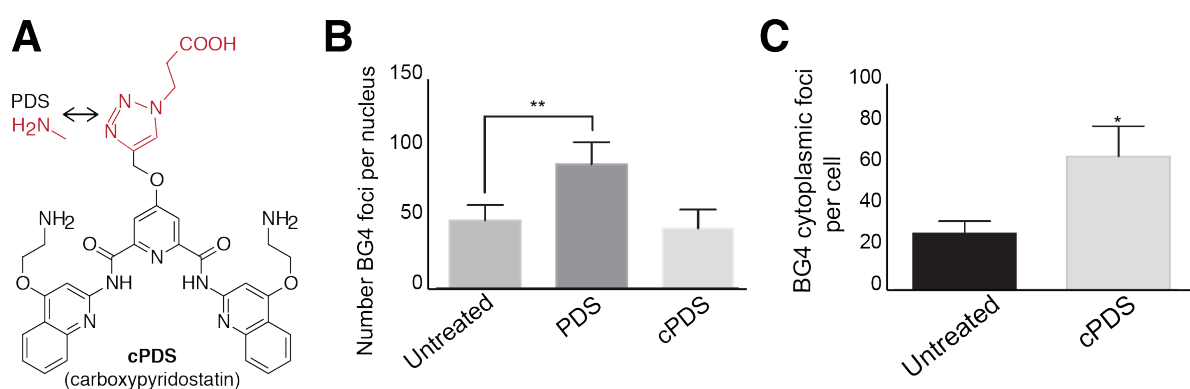
Figure 3.2. carboxy-PDS binds specifically to RNA G-quadruplexes in the cell
(A) Schematic of carboxyPDS (cPDS), derived from PDS; in red is the side chain that is altered between these two molecules, which contributes to specificity of the small molecule to RNA G4. (B-C) adapted Biffi et al., 2014 quantifying indirect immunofluorescence microscopy shows increased stabilisation of RNA but not DNA-G4 structures with cPDS in SV40-transformed MRC5 fibroblasts as revealed by BG4 staining (24 h treatment 2 µM cPDS). (B) Quantification of nuclear (DNA) foci and (C) quantification of cytoplasmic (RNA) foci.

The siRNA approach was then was used to screen the sensitivities of 12 G4-targeting PDS derivatives (synthesised by Dr. Santosh Adhikari and Mr. James Patterson, University of Cambridge). These derivatives were chosen as part of a development programme to explore whether ligands with improved clinically appropriate features, showed overlapping synthetic lethalities with PDS and PhenDC3. Good clinical properties are necessary to allow therapeutic exploitation of G4-stabilising ligands in specific cancer-

associated backgrounds (as discussed in the introduction and chapter 2). Unfortunately, PDS and PhenDC3 have multiple undesirable pharmacokinetic properties according to Lipinski's rule of five (Lipinski *et al*, 2001) used to estimate ligand solubility and cell permeability. The 12 molecules used within these investigations show improved drug-like properties (data not shown). Based on the resultant siRNA-induced sensitivities, one molecule was taken forward in a "quadruplex" focused shRNA screen (developed in chapter 2). The experiment performed to identify this derivative and subsequent screen outcome are also discussed.

Throughout this chapter, I aimed to establish these siRNA experiments as a rapid approach to 1) validate shRNA screen sensitivities, 2) extend and explore sensitivities with other established G4-ligands and 3) screen and help to refine novel G4-stabilising ligand candidates. For the latter, this siRNA approach enables synthetic lethality with these four genes as a pre-screen for novel G4-stabilising molecules before committing to a more detailed examination of sensitive genotypes.

3.2 Results

3.2.1. Short-term siRNA treatment validates sensitivity to PDS and PhenDC3 in *BRCA1, TOP1, DDX42* and *GAR1* deficient cells

To validate the shRNA screen results in an independent approach, a siRNA experiment was designed to explore the reproducibility of the PDS and PhenDC3 induced growth inhibition for cells deficient in four key sensitisers,

BRCA1, TOP1, GAR1 and DDX42. During the same experiment, protein-deficient cells were also treated with CX-5461 (discussed in section 3.2.2). Exploring synthetic lethalities to the three molecules within the same experiment, allows more meaningful comparisons between sensitivities. This siRNA experiment used a shorter timeframe than the shRNA screen (6 days versus ~28 days) and two concentrations per ligand (PDS: 0.25 µM and 0.5 µM HT1080, 5 µM and 10 µM A375; PhenDC3: 20 µM and 40 µM both cell lines). Both A375 and HT1080 cells were transfected with siRNAs targeting *BRCA1, TOP1, DDX42* or *GAR1* alongside non-targeting siRNA and non-transfected controls. Following 24 h, cells were treated with PDS, PhenDC3 or vehicle control for 144 h and cumulative cell growth monitored.

To confirm protein depletion following siRNA transfection, cell lysates were immunoblotted (see Methods) with the appropriate antibodies after 48 hours (Figure 3.3A). For both HT1080 and A375, protein depletion was evident for all targeting siRNAs (average 76-92 % knockdown HT1080; 41-69 % knockdown A375) (Figure 3.3B).

Figure 3.3. siRNA induced knockdown evident following 48 h post-transfection
(A) Representative western blot showing lysates immunoblotted for BRCA1, TOP1, DDX42, GAR1 and beta-actin loading control for non-transfected cells (U) and cells transfected with targeting (T) and non-targeting control (NT) siRNA, 48 h and 144 h after transfection as indicated for A375 (left) and HT1080 (right). (B) For both cell lines, BRCA1, TOP1, DDX42 and GAR1 proteins levels in targeting lysates were normalized to internally to the actin loading control and expressed as a percentage of the non-targeting control for three independent blots (mean ± standard deviation)

Cell confluency throughout the experiment was determined via Incucyte live cell scanning (see Methods for details) and growth curves plotted for cells transfected with siRNAs targeting the four proteins alongside the non-transfected (NT) and non-targeting controls (U) for both A375 and HT1080 (Figure 3.4 and 3.5). For HT1080 cells, non-targeting siRNA transfection caused slight growth inhibition compared to non-transfected controls (Figure 3.4) but gave comparable growth curves for A375 cells (Figure 3.5). For both cell lines, the non-targeting (NT) growth curve was used as a reference, allowing differentiation between siRNA sequence-specific growth phenotypes and non-specific effects from artificial siRNA transfection.

Figure 3.4. Short-term siRNA knockdowns in HT1080 of four key G4-sensitisers show dose dependent growth inhibition with G4-ligands

HT1080 cells were transfected with targeting siRNAs (orange) against BRCA1, TOP1, DDX42 or GAR1 for 24 h before treatment with PDS (0.25 μM and 0.5 μM), PhenDC3 (20 μM and 40 μM) or vehicle control (DMSO). For each knockdown, confluency over 144 h was monitored and plotted against confluency of non-transfected cells (blue) and cells transfected with a non-targeting control siRNA (green). Experiments were performed in triplicate and average confluency accumulation shown (mean ± standard deviation).
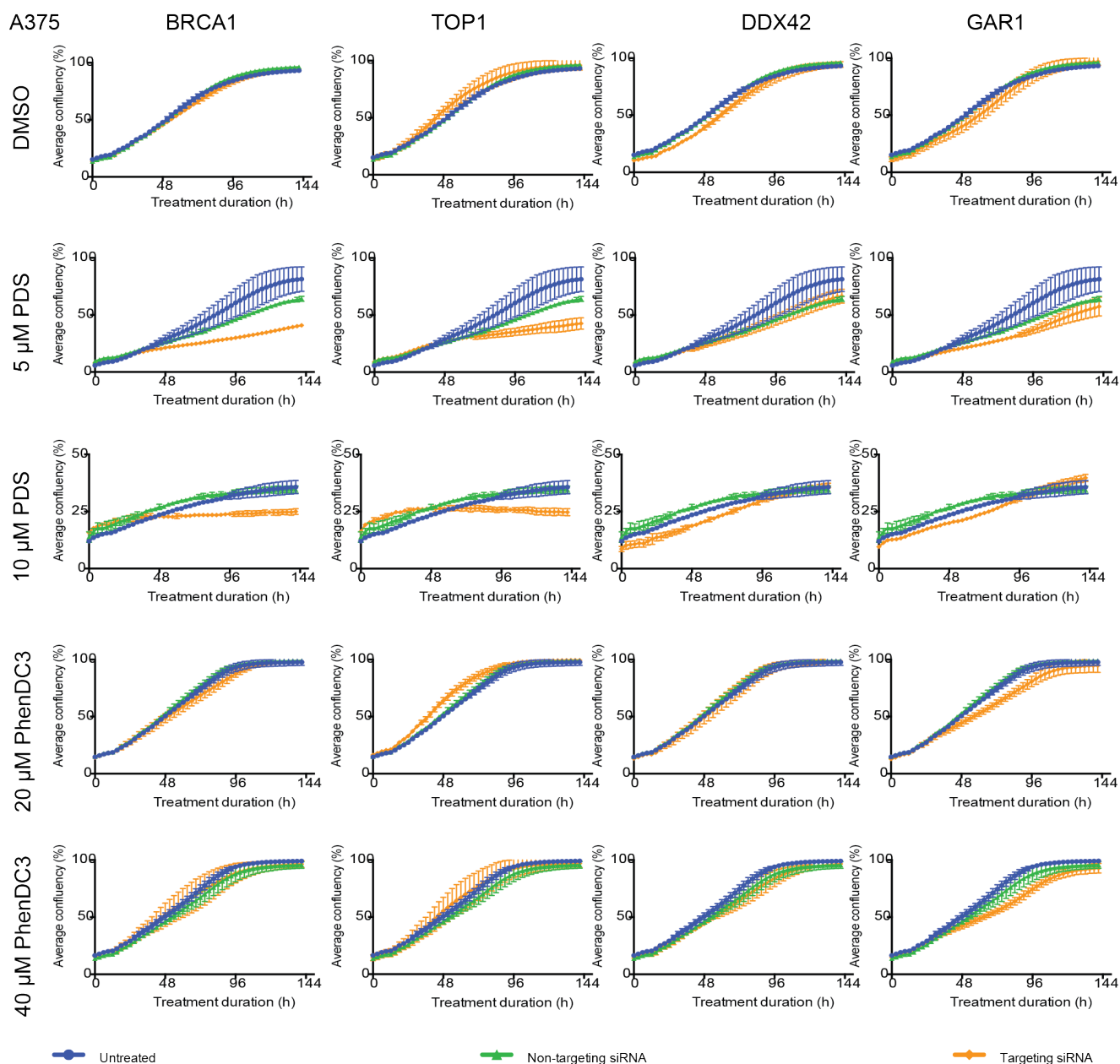
Figure 3.5. Short-term siRNA knockdowns in A375 of four key G4-sensitisers show dose dependent growth inhibition with G4-ligands
A375 cells were transfected with targeting siRNAs (orange) against BRCA1, TOP1, DDX42 or GAR1 for 24 h before treatment with PDS (5 µM and 10 µM), PhenDC3 (20 µM and 40 µM) or vehicle control (DMSO). For each knockdown, confluency over 144 h was monitored and plotted against confluency of non-transfected cells (blue) and cells transfected with a non-targeting control siRNA (green). Experiments were performed in triplicate and average confluency accumulation shown (mean ± standard deviation).

To investigate ligand-specific phenotypes, growth inhibition was compared for ligand and DMSO treated samples. For this, confluency differences for cells transfected with target siRNAs compared to non-targeting controls were plotted for each treatment (Figures 3.6A-D and Figures 3.7A-D) and differences at 72 h, 96 h and 120 h calculated (Figure 3.6E-H and 3.7E-H).

For HT1080 cells, all four siRNA knockdowns increased sensitivity to PDS and PhenDC3 versus DMSO, and this was greater and more statistically significant for later timepoints (Figure 3.6). The smallest and least significant phenotype was seen for DDX42 deficiencies (Figure 3.6C), however this may reflect the lower knockdown compared to the other proteins (~70 %; Figure 3.3). For DDX42, following 120 h treatment (Figure 3.6G), the median difference compared to DMSO for PDS was -28.9 % (0.25 µM, $p < 1 \times 10^{-4}$) and -32.4 % (0.5 µM, $p = 5 \times 10^{-3}$) and for PhenDC3, -15.5 % (20 µM, $p = 0.0271$) and -23.5 % (40 µM, $p = 0.0073$). For BRCA1 and TOP1 (Figure 3.6A-B and E-F) a lower PDS dose resulted in greater sensitisation. For example after 120 h, -69.3 % and -75.2 % respectively ($p < 1 \times 10^{-4}$) difference compared to DMSO was observed for 0.25 µM PDS versus -47.6 % and -56.7 % ($p < 1 \times 10^{-4}$), for 0.5 µM PDS treatment. Conversely for PhenDC3, growth inhibition compared to DMSO was similar between 20 µM and 40 µM concentrations: -15.0 % ($p = 0.0017$) and -20.4 % ($p = 0.0067$) respectively for BRCA1-deficient cells and -47.7 % and -45.4 % ($p < 1 \times 10^{-4}$) for TOP1 depletion. Overall, BRCA1- and TOP1-deficient HT1080 cells were more sensitive to PDS compared to PhenDC3. On the other hand, GAR1-deficient cells, as for DDX42 knockdowns, showed equivalent sensitisation to

both ligands (Figure 3.6 D and H), particularly evident at the 120 h timepoint, where PDS caused -51.6 % (0.25 μM, $p < 1 \times 10^{-4}$) and -44.2 % (0.5 μM, $p = 6 \times 10^{-4}$) and PhenDC3 induced -62.9 % (20 μM, $p < 1 \times 10^{-4}$) and -53.4 % (40 μM, $p < 1 \times 10^{-4}$) growth inhibition compared to DMSO treatment.
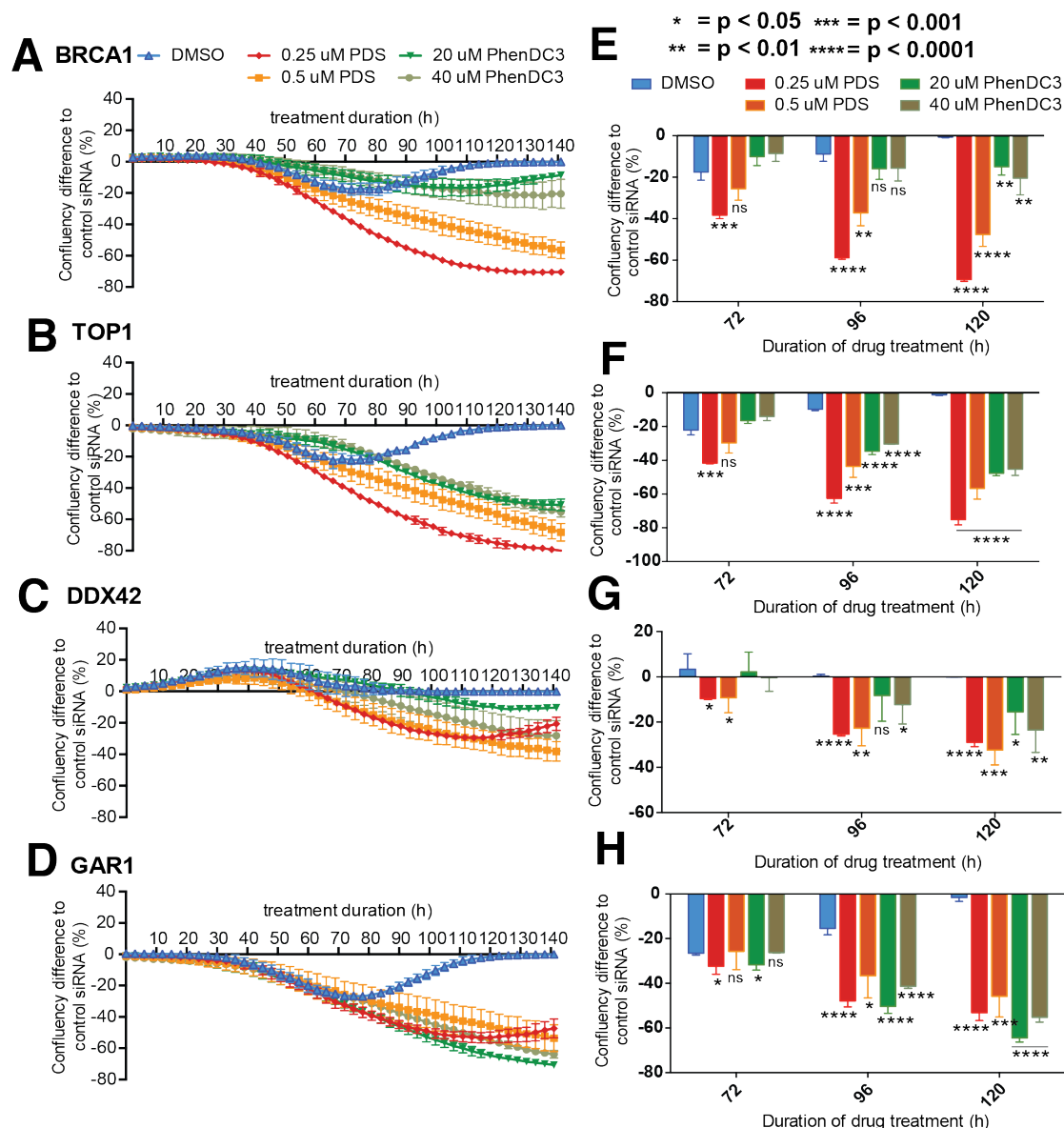
Figure 3.6. Short-term siRNA knockdowns validate 'top' G4-sensitisers identified by shRNA screening in HT1080

HT1080 cells were transfected with the targeting siRNAs for 24 h before PDS, PhenDC3 or DMSO treatment. Differences in percentage confluency for cells transfected with targeting siRNA to NT controls (mean ± standard deviation) were plotted across three replicates for A) BRCA1, B) TOP1, C) DDX42, D) GAR1. Confluency differences at 72, 96 and 120 h were plotted for comparison for E) BRCA1, F) TOP1, G) DDX42, H) GAR1. Significant confluency differences for G4-ligand versus DMSO treatment were determined using an unpaired parametric t-test, assuming equal standard deviation.

Generally, the magnitude and significance of the ligand sensitivities for A375 were less than for HT1080 cells (Figure 3.7) possibly reflecting that the siRNA-induced knockdowns were lower for each protein (Figure 3.3). Nonetheless, the sensitisation phenotypes show similar trends to that seen for HT1080 cells. BRCA1- and TOP1-deficiencies showed significant PDS sensitivity, with greater comparative growth inhibition at the lower dose (Figure 3.7C-H). Following 120 h treatment, for BRCA1 this growth difference was -20.0 % (p < 1 x 10^{-4}) and -7.74 % (p = 0.0019) for 5 μM and 10 μM PDS respectively, whereas TOP1-deficient cells showed -21.6 % (p = 0.0052) and -12.8 % (p = 0.0035) differences. BRCA1 and TOP1 knockdowns were also more sensitive to PDS than PhenDC3, whereby BRCA1-deficiency was only sensitive to 40 μM PhenDC3 (120 h; -3.34 % difference to DMSO treatment, p = 0.0307) and TOP1-deficiency was PhenDC3 insensitive. As for HT1080, GAR1 depletion caused sensitivity to both PDS and PhenDC3 in A375 cells (Figure 3.8F). For example, following 96 h treatment, GAR1-deficient cell growth was 10.6 % (p = 0.0212) and 18.0 % (p = 0.0017) less than for DMSO-treated, following 5 μM PDS and 40 μM PhenDC3 respectively. *DDX42*-deficient A375 cells did not replicate the ligand sensitivities of the shRNA screen (Figure 3.7E&I), perhaps arising from lower knockdown efficiency compared to the other three proteins (~40 %). Collectively, this independent siRNA assay largely substantiates that the four genes *BRCA1, TOP1, DDX42* and *GAR1* replicate the shRNA screen results.
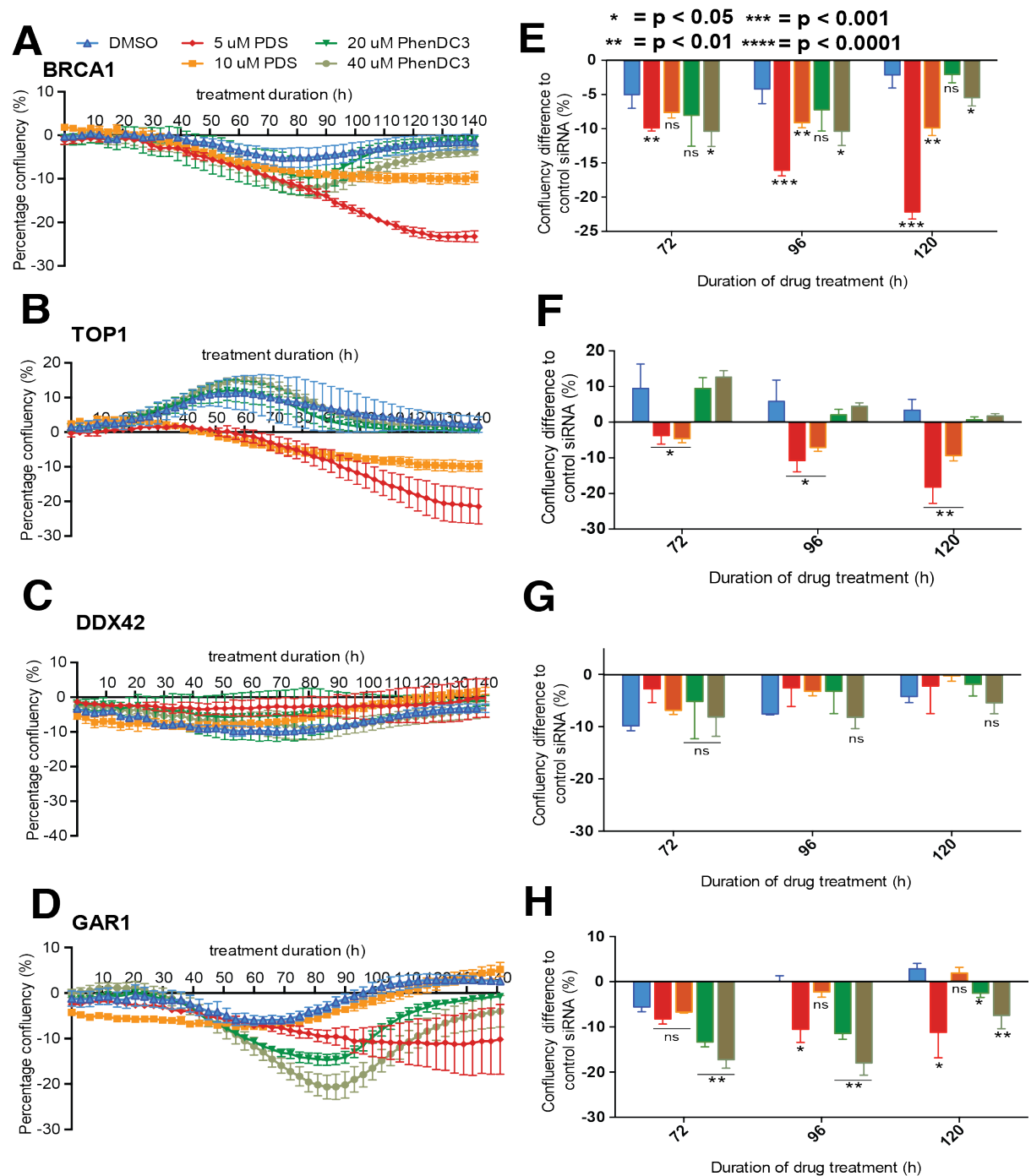
Figure 3.7. Short-term siRNA knockdowns reflect some of the growth inhibition of the 'top' G4-sensitisers identified by shRNA screening in A375

A375 cells were transfected with the targeting siRNAs for 24 h before PDS, PhenDC3 or DMSO treatment. Differences in percentage confluency for cells transfected with targeting siRNA to NT controls (mean ± standard deviation) were plotted across three replicates (mean ± standard deviation) for A) BRCA1, B) TOP1, C) DDX42, D) GAR1. Confluency differences at 72, 96 and 120 h were plotted for comparison for E) BRCA1, F) TOP1, G) DDX42, H) GAR1. Significant confluency differences for G4-ligand versus DMSO treatment were determined using an unpaired parametric t-test, assuming equal standard deviation.

3.2.2. The clinical G4-stabilising ligand CX-5461 is synthetically lethal with the four top sensitisers

The siRNA experimental design was also applied to CX-4561, a G4-ligand undergoing clinical testing for breast cancer treatment (Figure 3.8A). If the sensitivity is shared by CX-5461, this suggests that these four genes may represent universal synthetic lethalities to G4-stabilising ligands. Secondly, this provides a proof-of-principle experiment that the siRNA approach can identify the same PDS and PhenDC3 synthetic lethalities with other G4-ligands.
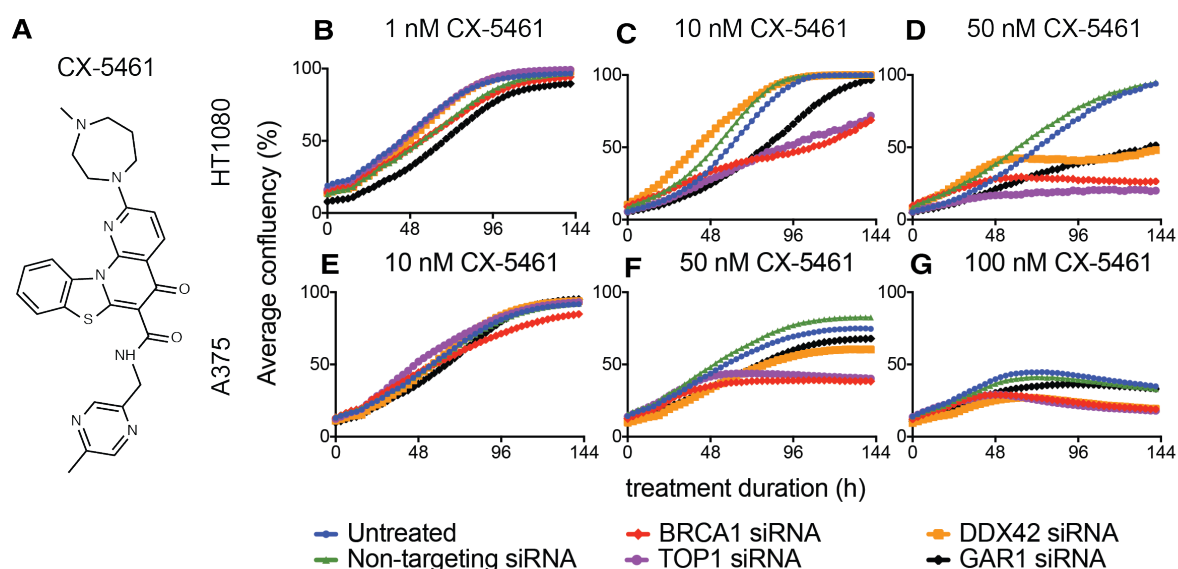
Figure 3.8. Investigating sensitivity of BRCA1, TOP1, DDX42 and GAR1-deficient HT1080 and A375 cells to CX-5461 treatment
(A) Structure of the G4-stabilising ligand CX-5461 currently in clinical trials. Both HT1080 (B-D) and A375 (E-G) siRNA induced BRCA1, TOP1, GAR1, DDX42 deficient cells were treated with 1 nM, 10 nM, 50 nM or 100 nM CX-5461 as indicated. Confluency over 144 h was then monitored and plotted against confluency of non-transfected cells (blue) and cells transfected with a non-targeting control siRNA (green). Shown are the average confluency accumulation for the two biological replicas (mean ± standard deviation).

The CX-5461 experiment was performed and analysed in A375 and HT1080 cells as for PDS and PhenDC3 siRNA experiments, thus the protein knockdowns are the same as in Figure 3.3. The absolute growth curves (Figure 3.8B) were used to investigate growth differences to the non-targeting control for the experiment duration (Figure 3.9A-D and Figure 3.9A-D) and individual timepoints (Figure 3.9E-H and Figure3.10E-H). BRCA1-deficiency induced CX-5461 sensitivity in both HT1080 (Figure 3.9A-B) and A375 cells (Figure 3.10A-B), with higher concentrations and later timepoints showing greater differences compared to DMSO treatment. This is exemplified by the 120 h timepoint, where treatment with 10 nM and 50 nM CX-5461 caused a median growth rate decrease of 44.0 % (p = 5 x 10^{-4}) and 61.7 % (p < 1 x 10^{-4}) compared to DMSO curves. TOP1-deficient HT1080 cells showed a similar profile to that seen for BRCA1 knockdown (Figure 3.9C-D) with increased sensitivity for CX-5461 concentrations above 10 nM. For example at 120 h, 50 nM CX-5461 treatment caused a -68.0 % (p < 1 x 10^{-4}) difference in growth rate for TOP1-deficient cells compared to DMSO controls. For both DDX42 and GAR1-deficient cells (Figure 3.9E-H), sensitivity was most noticeable at 50 nM CX-5461 (-45.7 % and -43.0 % difference respectively following 120 h treatment; p < 1 x 10^{-4}), with DDX42 deficiency insensitive to lower concentrations. Similarly, all knockdowns in A375 cells caused CX-5461 sensitivity compared to the DMSO control (Figure 3.10). Thus, these results confirm the BRCA1 G4-sensitivity to CX-5461 and suggests that TOP1, GAR1 and DDX42 deficiencies are also sensitive to this more clinically relevant ligand.
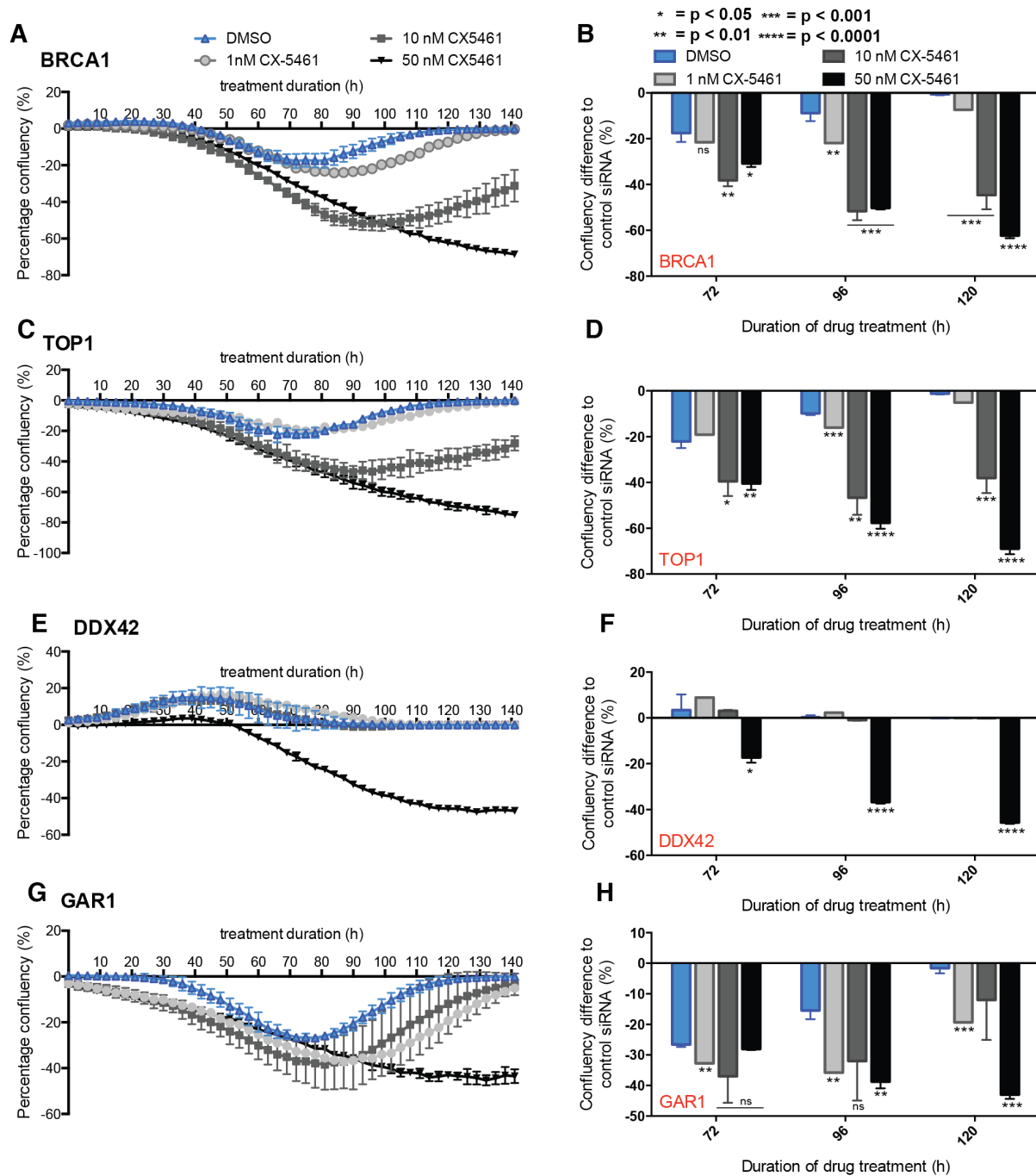
Figure 3.9. HT1080 cell deficient in BRCA1, TOP1, DDX42 and GAR1 are also sensitive to CX-5461.
HT1080 cells were transfected with the targeting siRNAs for 24 h before treatment with 1 nM, 10 nM or 50 nM CX-5461 or DMSO (vehicle control) and differences in confluency, compared to a non-targeting siRNA control were plotted for A) BRCA1, C) TOP1, E) DDX42, G) GAR1. Drug treatment was performed in duplicate, and average confluency differences (± standard deviation) are shown. The confluency differences at 72, 96 and 120 h for each of the drug conditions were plotted for comparison for B) BRCA1, D) TOP1, F) DDX42, H) GAR1. Significantly increased growth inhibition for G4-ligand versus DMSO treatment was determined using an unpaired parametric t-test, assuming equal standard deviation.
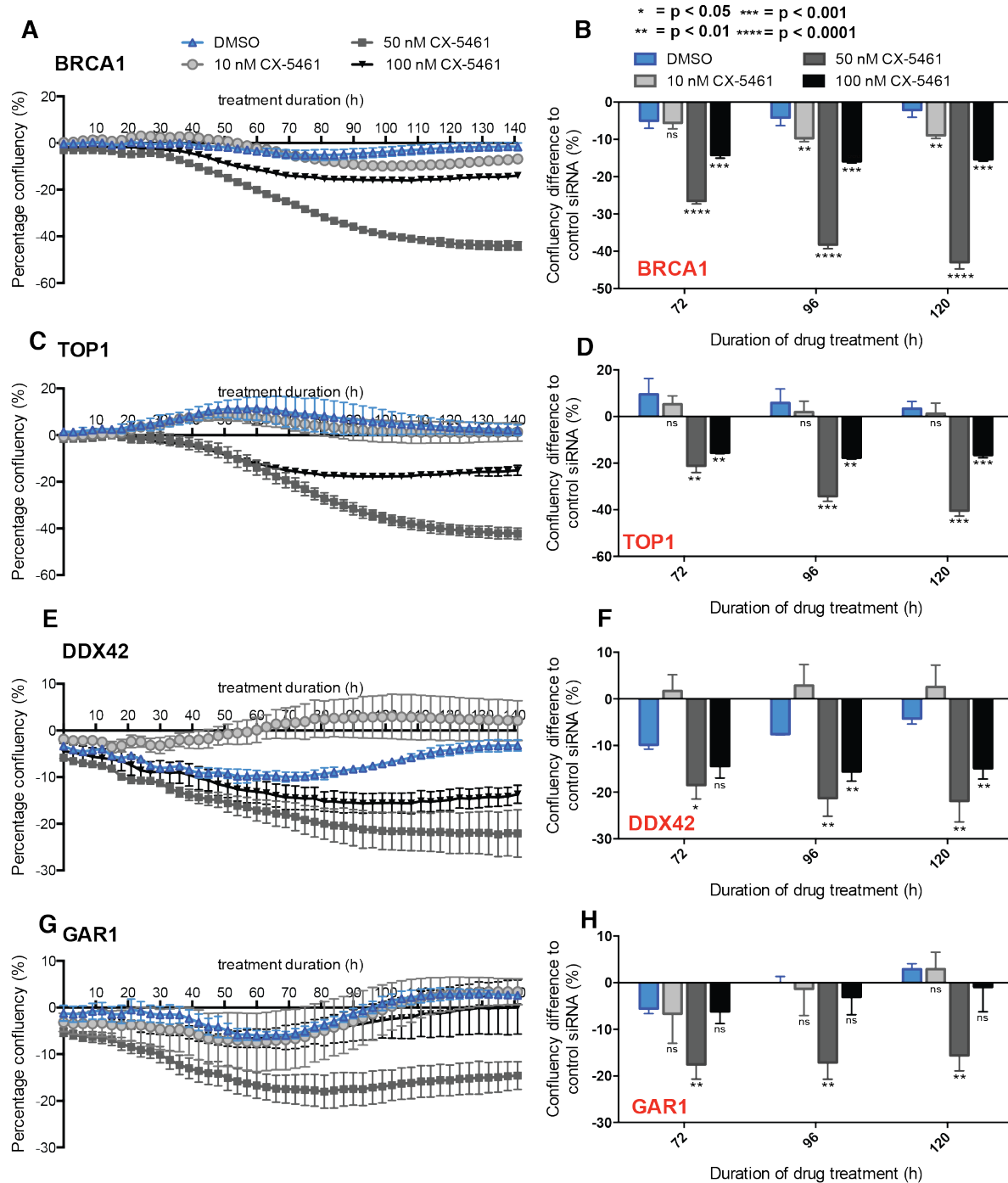
Figure 3.10. BRCA1, TOP1, DDX42 and GAR1 deficient A375 cells are sensitive to CX-5461

A375 cells were transfected with the targeting siRNAs for 24 h before treatment with 10 nM, 50 nM or 100 nM CX-5461, alongside a DMSO vehicle control and differences in confluency, compared to a non-targeting siRNA control were plotted for (A) BRCA1, (C) TOP1, (E) DDX42, (G) GAR1. Drug treatment was performed in duplicate, and average confluency differences (± standard deviation) are shown. The confluency differences at 72, 96 and 120 h for each of the drug conditions were plotted for comparison for (B) BRCA1, (D) TOP1, (F) DDX42, (H) GAR1. Significantly increased growth inhibition for G4-ligand versus DMSO treatment was determined using an unpaired parametric t-test, assuming equal standard deviation.

3.2.3 Deficiencies in the four key sensitisers show differences in cPDS sensitivity

The siRNA protocol was next applied to an RNA G4-selective PDS derivative, cPDS. For this the HT1080 cell line was used, as the siRNA-induced protein knockdowns and growth inhibition phenotypes were stronger compared to A375 (as discussed above). Cells were treated for five days with 5 µM and 10 µM cPDS alongside DMSO and 0.5 µM PDS (vehicle and positive controls, respectively). As this is independent data from the PDS, PhenDC3 and CX-5461 experiments, protein knockdown was first confirmed by immunoblotting (BRCA1 98 %, TOP1 82 %, GAR1 67 %; DDX42 49 %; Figure 3.11A&B) and similar growth rates between non-targeting and non-transfected controls (Figure 3.11C). The differences in confluency of cells transfected with targeting siRNAs compared to non-targeting controls were plotted as previously (Figure 3.12).
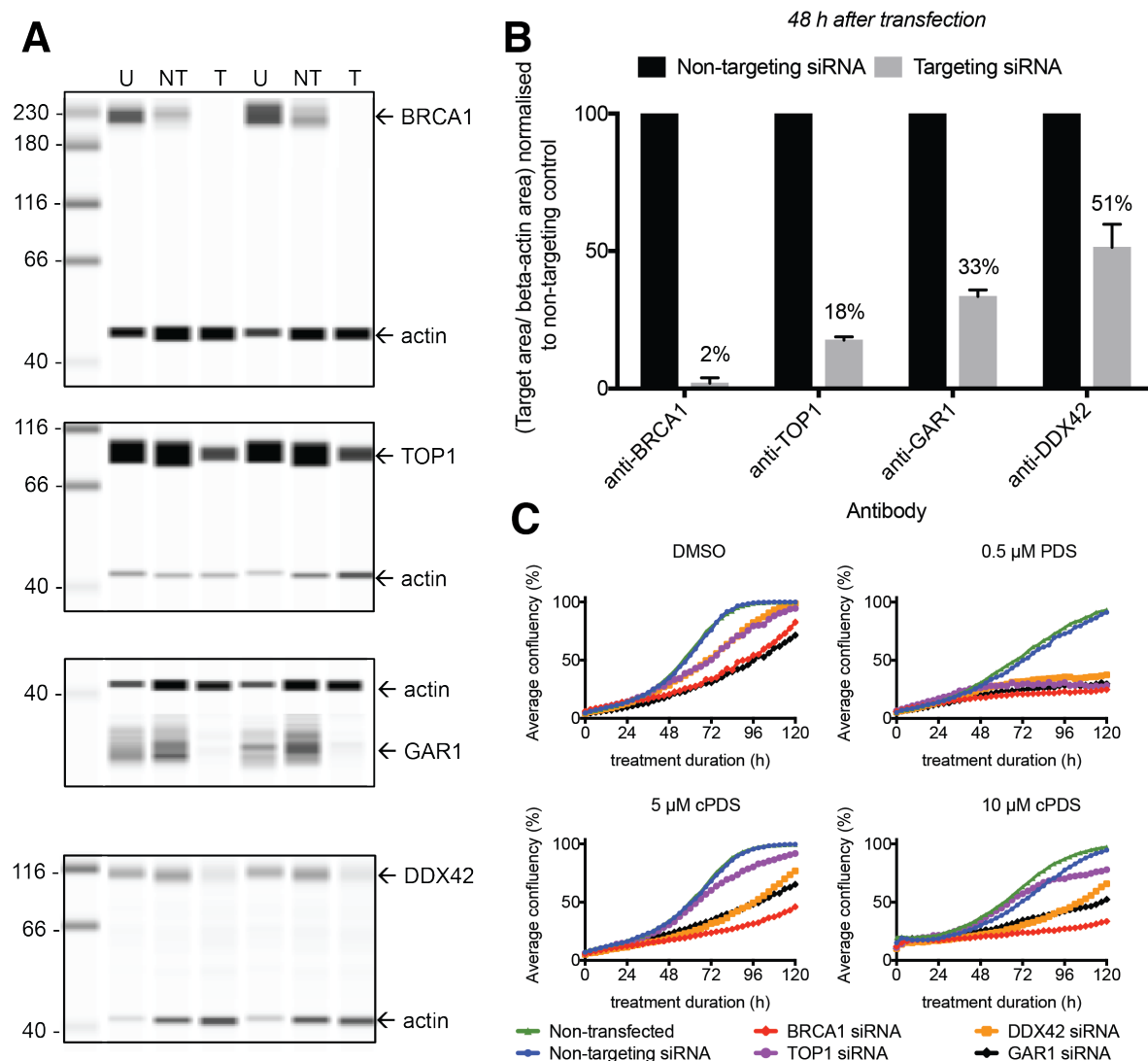
Figure 3.11. Knockdown of key sensitisers in HT1080 cells shows differential sensitivity responses to cPDS treatment
(A) Western blot showing HT1080 lysates immunoblotted for BRCA1, TOP1, DDX42, GAR1 and beta-actin loading control for untransfected cells (U) and cells transfected with targeting (T) and non-targeting control (NT) siRNA at 48hr post transfection for two biological replicas. (B) Graph showing BRCA1, TOP1, DDX42 and GAR1 proteins levels from siRNA targeting cell lysates normalized to the non-targeting control for the two replicas (mean ± standard deviation). (C) HT1080 cells either untransfected (blue line), or transfected with targeting siRNA (against BRCA1, TOP1, DDX42 or GAR1) or non-targetin control (green) were treated with 0.5 µM PDS, cPDS (5 µM or 10 µM) or vehicle control (DMSO). The average confluency accumulation over 120 h for the two biological replicas is shown (mean ± SEM).

Corroborating earlier observations (section 3.2.1) cells transfected with targeting siRNAs against the four proteins, all demonstrated growth sensitivity to 0.5 µM PDS treatment by 120 h (Figure 3.12C). For cPDS the greatest growth inhibition compared to DMSO treatment was seen for BRCA1- and DDX42-deficient cells, which was most significant at the 120 h timepoint for both 5 µM (-36.4 %, $p = 0.0063$; -21.4 %, $p = 0.0223$ respectively) and 10 µM cPDS (-43.7 %, $p = 0.0009$; -27.5 % $p = 0.0109$ respectively; Figure 3.12 A&D). Conversely, TOP1 and GAR1 knockdowns were relatively insensitive to cPDS treatment (unlike PDS, Figure 3.12B-C), only showing a significant difference to DMSO treatment at 120 h with 10 µM cPDS (-11.27 %, $p = 0.0317$; -13.93 %, $p = 0.0349$ respectively).

Figure 3.12. BRCA1 and DDX42 depletion causes sensitivity to cPDS treatment
For DMSO, PDS and cPDS treated HT1080 cells, confluency difference (mean ± standard deviation) compared to NT siRNA were plotted across two replicates for A) BRCA1, B) TOP1, C) GAR1, D) DDX42. Confluency differences at 72, 96 and 120 h were plotted for comparison for E) BRCA1, F) TOP1, G) GAR1, H) DDX42. Significant confluency differences for G4-ligand versus DMSO treatment were determined using an unpaired parametric t-test, assuming equal standard deviation.

3.2.4. Screening derivatives of PDS with improved medicinal chemistry properties via siRNA-induced sensitivity

Having shown that siRNA knockdown of *BRCA1, GAR1, DDX42* and *TOP1* shows sensitivity to three structurally independent G4-stabilising ligands, this approach was used to screen whether 12 PDS-analogues (structures not shown) with improved medicinal chemistry properties show similar sensitivities to the better characterised G4-ligands used elsewhere within this thesis.

For this HT1080 cells were used due to greater siRNA transfection efficiency, protein knockdown and PDS sensitivity compared to the A375 cell line as per previous experiments. As positive and negative controls 0.5 μM PDS and DMSO were included and the assay performed and analysed as previously, with two concentrations of each of the 12 molecules (0.5 μM and 2.5 μM). These concentrations were chosen as 0.5 μM matched that of PDS, while a concentration 5-fold higher (i.e. 2.5 μM) may ensure that any phenotypes are clear. The difference in cell confluency compared to the DMSO control is shown below, with the exception of molecule 11, which precipitated at 2.5 μM and thus was discarded from analysis (Figure 3.13). Molecule 8 (2.5 μM) was toxic to cells regardless of the siRNA treatment and was thus not included in the final assay. For easier visualisation, confluency differences for cells transfected with targeting siRNAs compared to the non-targeting control is shown for 48 h, 72 h and 96 h treatments (Figure 3.14).
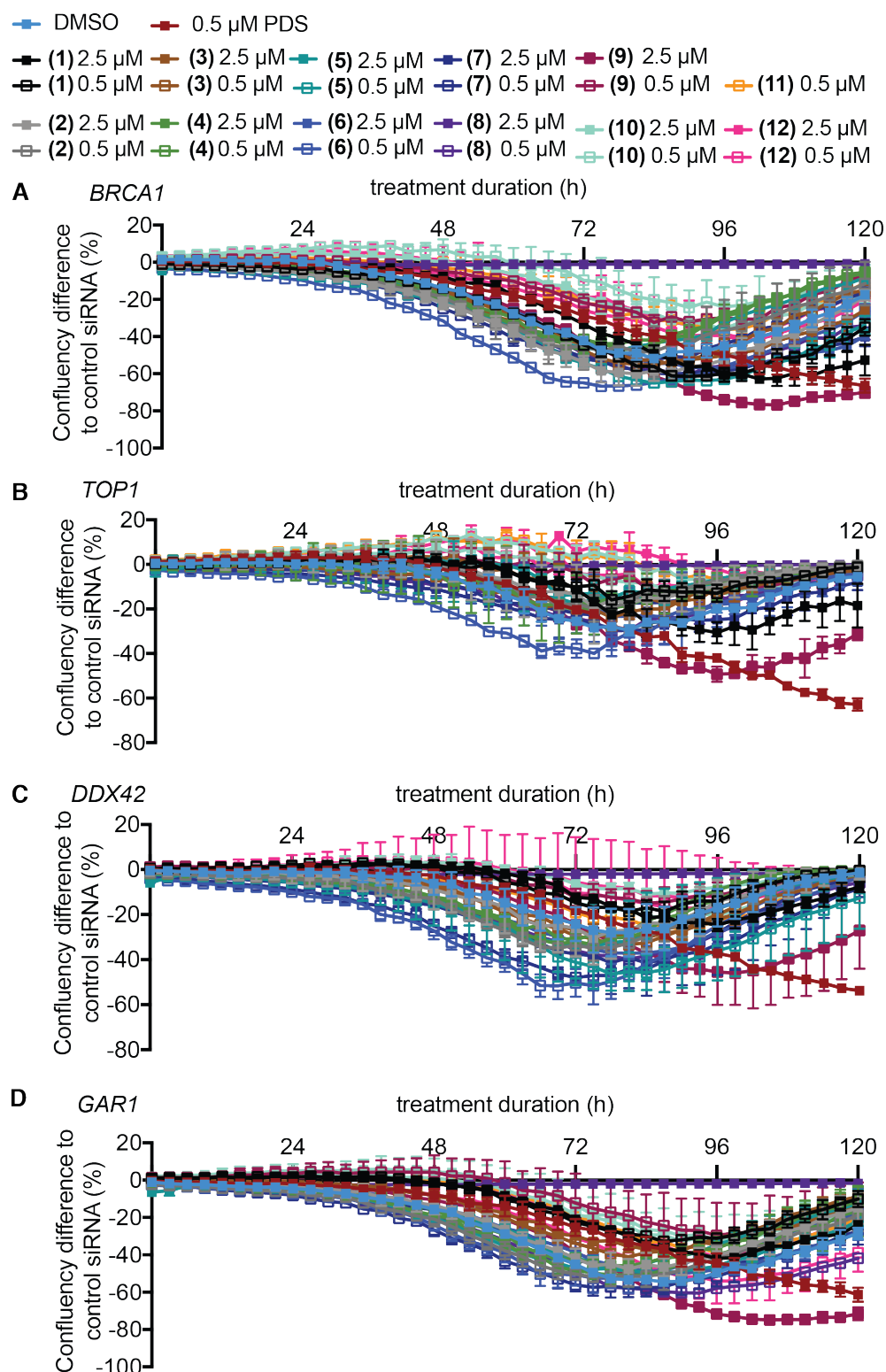
Figure 3.13. Screening clinically improved PDS derivatives via siRNA depletion of key sensitisers

HT1080 cells were transfected with targeting siRNAs alongside non-targeting controls for 24 h before treatment with 0.5 µM or 2.5 µM of 12 candidate G4-ligands derived from PDS, alongside 0.5 µM PDS and DMSO, positive and vehicle controls respectively, cultured for 120 h as previously. Confluency differences compared to a non-targeting siRNA control were plotted for (A) BRCA1, (B) TOP1, (C) DDX42, (D) GAR1, shown as an average of 2 biological replicas.

Five molecules gave promising growth inhibition phenotypes (Figure 3.14 and listed below). For these, confluency differences compared to DMSO and 0.5 µM PDS are depicted below (Figure 3.15). For the molecules that did not cause significant ligand-specific sensitivity, a higher concentration and/or a longer time period may be required.

BRCA1:   molecule 1 (0.5 and 2.5 µM), 4 (0.5 µM), 6 (2.5 µM), 7 (2.5 µM) and 9 (2.5 uM)

TOP1:   molecule 1 (2.5 µM) and 9 (2.5 µM)

DDX42:   molecule 4 (0.5 µM), 7 (2.5 µM), 9 (2.5 uM)

GAR1:   molecule 9 (2.5 µM)
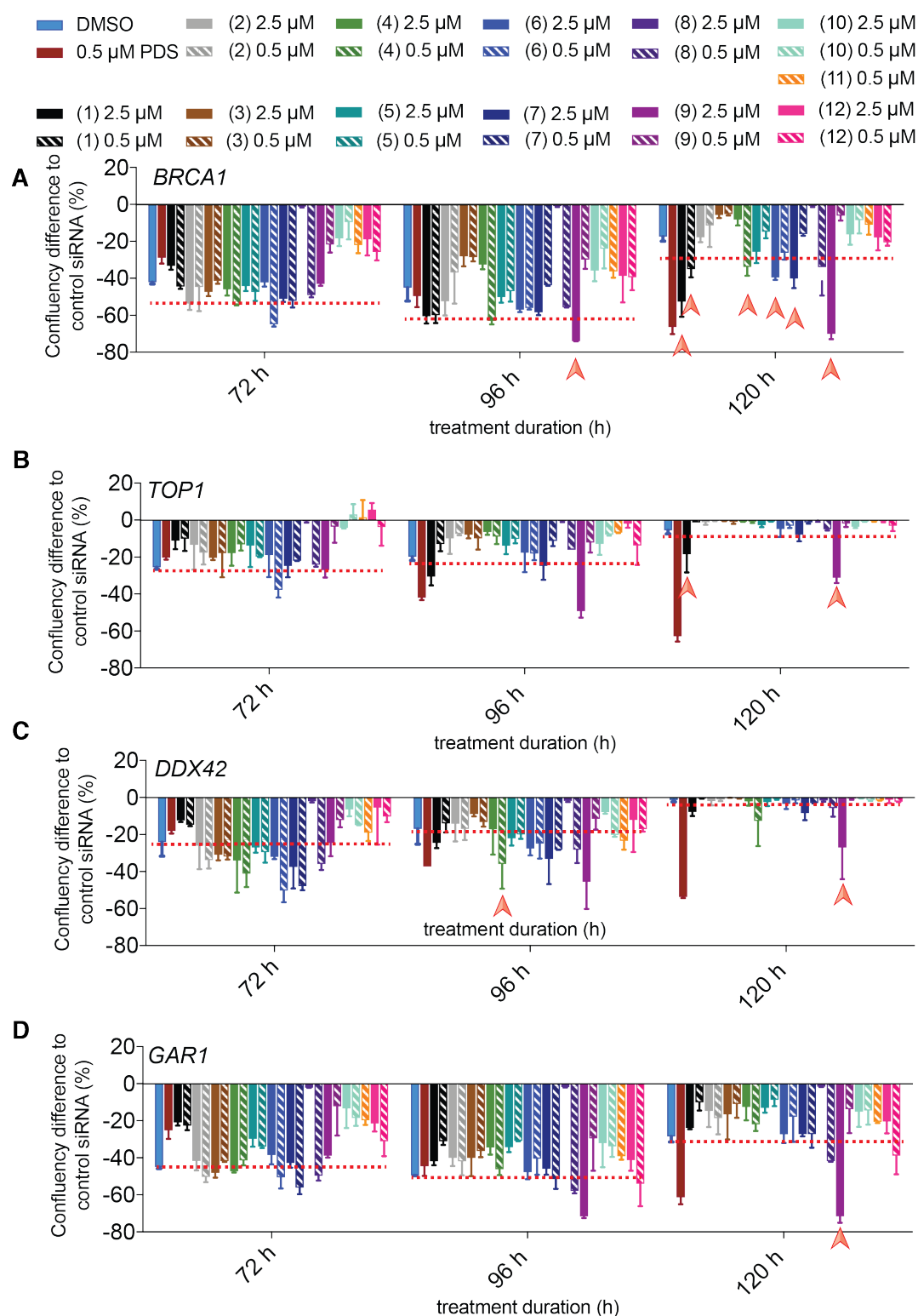
Figure 3.14. Five candidate molecules show sensitivity to siRNA-induced deficiencies in at least one of the key sensitivities

For the growth curves in figure 3.14, confluency differences at 72, 96 and 120 h for each drug conditions for (A) BRCA1, (B) TOP1, (C) DDX42, (D) GAR1, shown as an average of two biological replicas. Red dotted line denotes the growth inhibition caused by the protein knockdown compared to the non-targeting control for DMSO. Anything below this line denotes ligand specific sensitivity. A red arrow denotes the 5 molecules that gave the greatest sensitivities at the indicated timepoints.

150
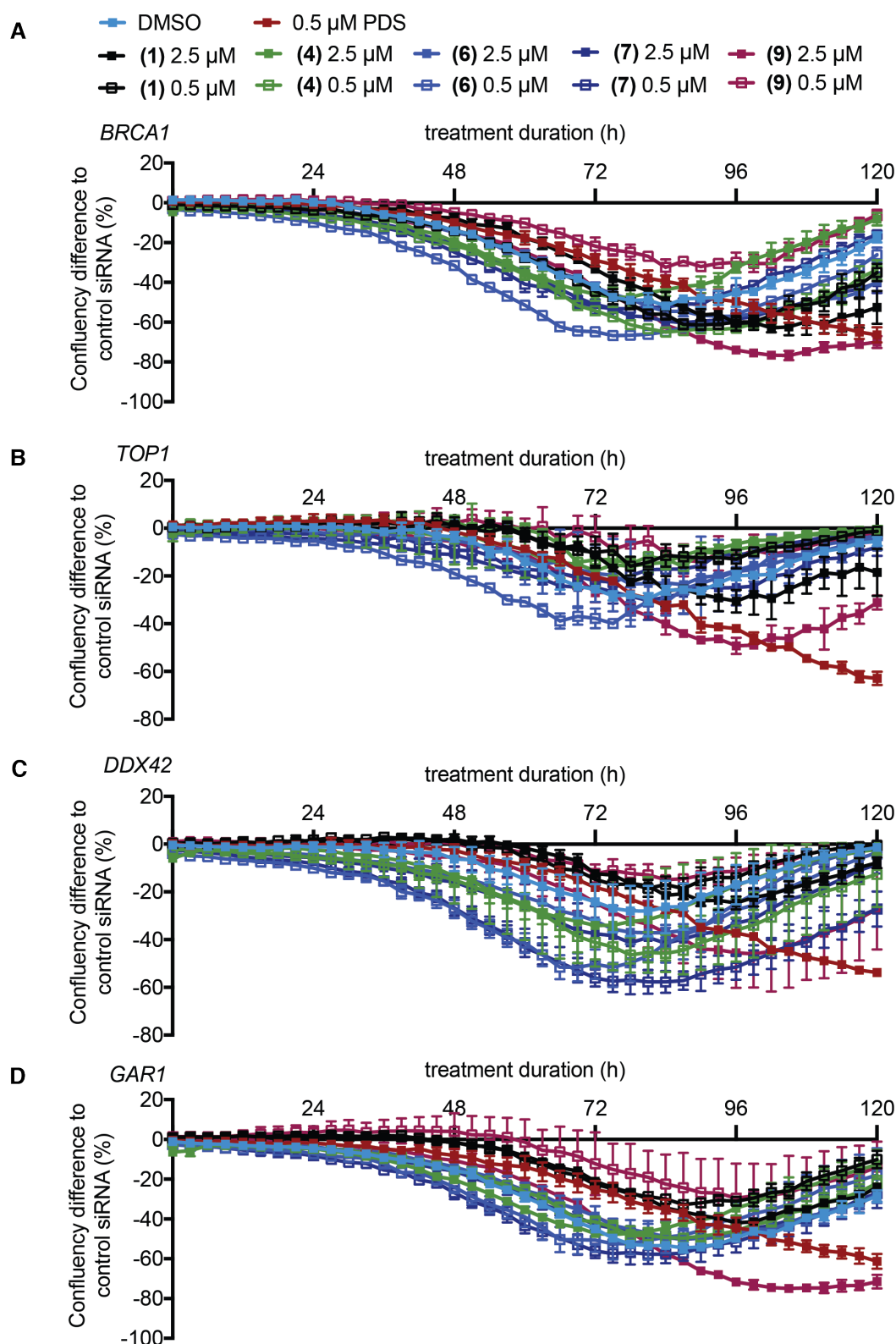
Figure 3.15. SA-100-128 (molecule 9) is synthetic lethal with BRCA1, TOP1,DDX42 and GAR1 siRNA knockdowns

For the 5 candidate molecules, DMSO and PDS, confluency differences compared to non-targeting siRNA controls following 120 h treatment were plotted for siRNA induced knockdown of (A) BRCA1, (B) TOP1, (C) DDX42 and (D) GAR1. One molecule **(9)** showed growth inhibition for all 4 knockdowns at 2.5 µM.

Molecule **9,** referred to as SA-100-128 henceforth, showed significant synthetic lethality with all four key protein knockdowns at 2.5 µM, with sensitivity comparable to 0.5 µM PDS treatment. This may suggest that SA-100-128 reflects a pan-G4 ligand with similar synthetic lethalities to PDS and PhenDC3, thus worthy of further exploration. SA-100-128 has improved physicochemical, adhering more to Lipinski's criteria compared to PDS (data not shown).

3.2.5. A focused "quadruplex" screen with SA-100-128

As outlined above SA-100-128 may present a promising "drug-like" PDS derivative: selective G4 stabilisation *in vitro;* improved pharmacokinetic properties and synthetic lethality with cells deficient in the top four key sensitisers (*BRCA1, TOP1, DDX42, GAR1*). Given its potential to be taken forward for future clinical development, it is of benefit to understand the spectrum of synthetic lethalities that this molecule could be applied to. As such, the focused G4 pool developed in chapter 2 was used to perform an shRNA screen with SA-100-128.

For consistency with the shRNA screens performed with PDS and PhenDC3 (chapter 2), cells were treated with a GI_{20} SA-100-128 concentration, pre-determined by a 96 h viability assay (see methods) (Figure 3.16A; 1 µM and 1.7 µM for A375 and HT1080, respectively), and the focused screen performed for 15 population doublings (t15) alongside a DMSO control as outlined previously. However, when upscaled for the long-term shRNA screen, this calculated value did not result in a growth rate difference

compared to the DMSO treatment for both A375 (Figure 3.16B) and HT1080 (Figure 3.16C). Increasing SA-100-128 concentrations caused binary outcomes, either complete death or no growth difference compared to DMSO treatment (data not shown). Therefore, t15 from cells treated with the 96 h GI_{20} concentrations (A375 1 µM and HT1080 1.7 µM, described above) were chosen for sequencing and analysis, to investigate how ligand treatment influenced the representation of individual shRNAs during the course of the experiment.

Using a hairpin threshold of FDR ≤ 0.05, resulted in 570 and 459 differentially expressed shRNAs for A375 and HT1080 respectively, of which 349 and 75 were ligand-specific i.e. not found in DMSO (Figure 3.18D-E). A gene was denoted significantly depleted if the median $\log_2FC < 0$ for 50 % or 3 hairpins. Overall, this gave 36 and 8 depleted genes for A375 and HT1080 cells respectively (Figure 3.16F-G). Considering the similarity between DMSO and ligand-treated growth curves, further applying a $\log_2FC \leq -1$ threshold, as for PDS and PhenDC3 screening, was deemed too stringent.
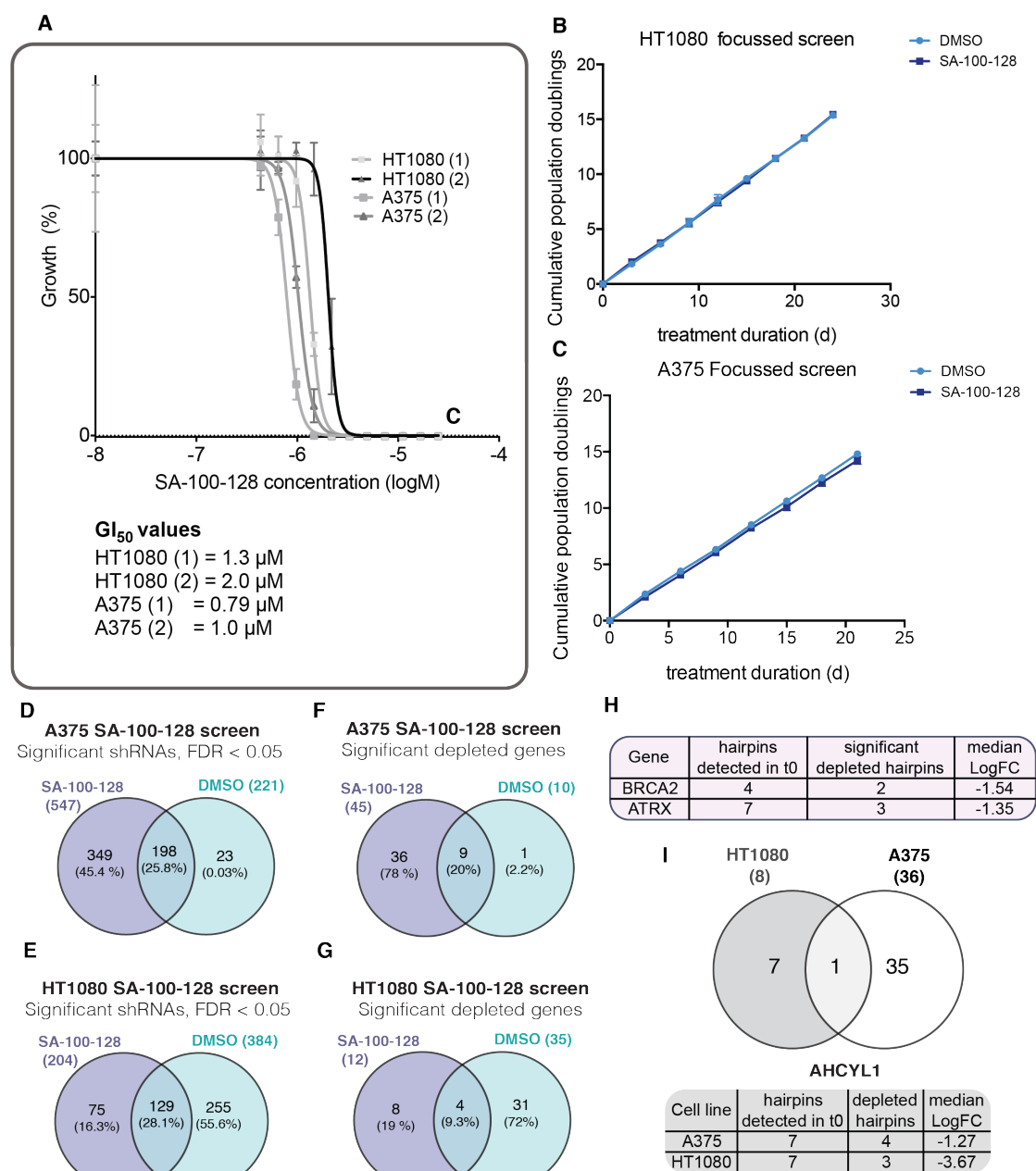
Figure 3.16. A focused 'quadruplex' screen performed with SA-100-128
(A) HT1080 and A375 cells at two different seeding densities were treated with serial dilutions of SA-100-128 for 96 h and viability recorded via Cell Titre-Glo viability luminescent assay. For each seeding density, curves were plotted averaging the four replicates using a non-linear regression model (± standard deviation) (B-C) cumulative population doubling graph for SA-100-128 and DMSO treated cells for experiment duration for (B) HT1080 and (C) A375 cells. (D-G) Venn diagrams overlapping, for DMSO and SA-100-128 treated cells, (D-E) significantly differentially expressed shRNAs, FDR ≤ 0.05 and (F-G) significantly depleted gene deficiencies (FDR ≤ 0.05; 50 % or 3 hairpins) for A375 (D&F) and HT1080 (E&G). (H) Table of median log_2FC and number of significantly depleted hairpins for "positive control" G4-ligand synthetic lethalities BRCA2 and ATRX that were uncovered as SA-100-128 specific hits in the A375 cell line. (I) Venn diagram overlapping the SA-100-128 specific depleted genes for both HT1080 and A375, revealing one common sensitivity (AHCYL1). For this gene, median log_2FC and number of significantly depleted hairpins are shown.

For A375 cells, among the gene deficiencies that showed SA-100-128 sensitivity are "positive controls" *ATRX* and *BRCA2* (Figure 3.16H) previously shown to be synthetic lethal with G4-stabilising ligands (McLuckie *et al*, 2013; Watson *et al*, 2013) and a known G4-interactor *HNRNPA2/B1.* This supports that SA-100-128 may stabilise endogenous G4-structures. Despite the low number of HT1080 hits, overlap with A375 significantly depleted genes revealed one common sensitivity to SA-100-128 treatment – AHCYL1 (Figure 3.16I).

Due to the low number of hits, HT1080 and A375 sensitivities were combined to provide 43 gene deficiencies synthetic lethal with SA-100-128 treatment. These were overlapped with the synthetic lethalities uncovered in the genome-wide and focused PDS and PhenDC3 screens to look for recurring sensitivities (Figure 3.17). Of the 19 genes common to all four screens, several components were found within the ubiquitin pathway: *CUL1, UBA5, USP1* and *USP37*. Additional ubiquitin components are also found in the overlap with the genome-wide and PhenDC3 focused screens, including *CAND1* and *USP9X*.
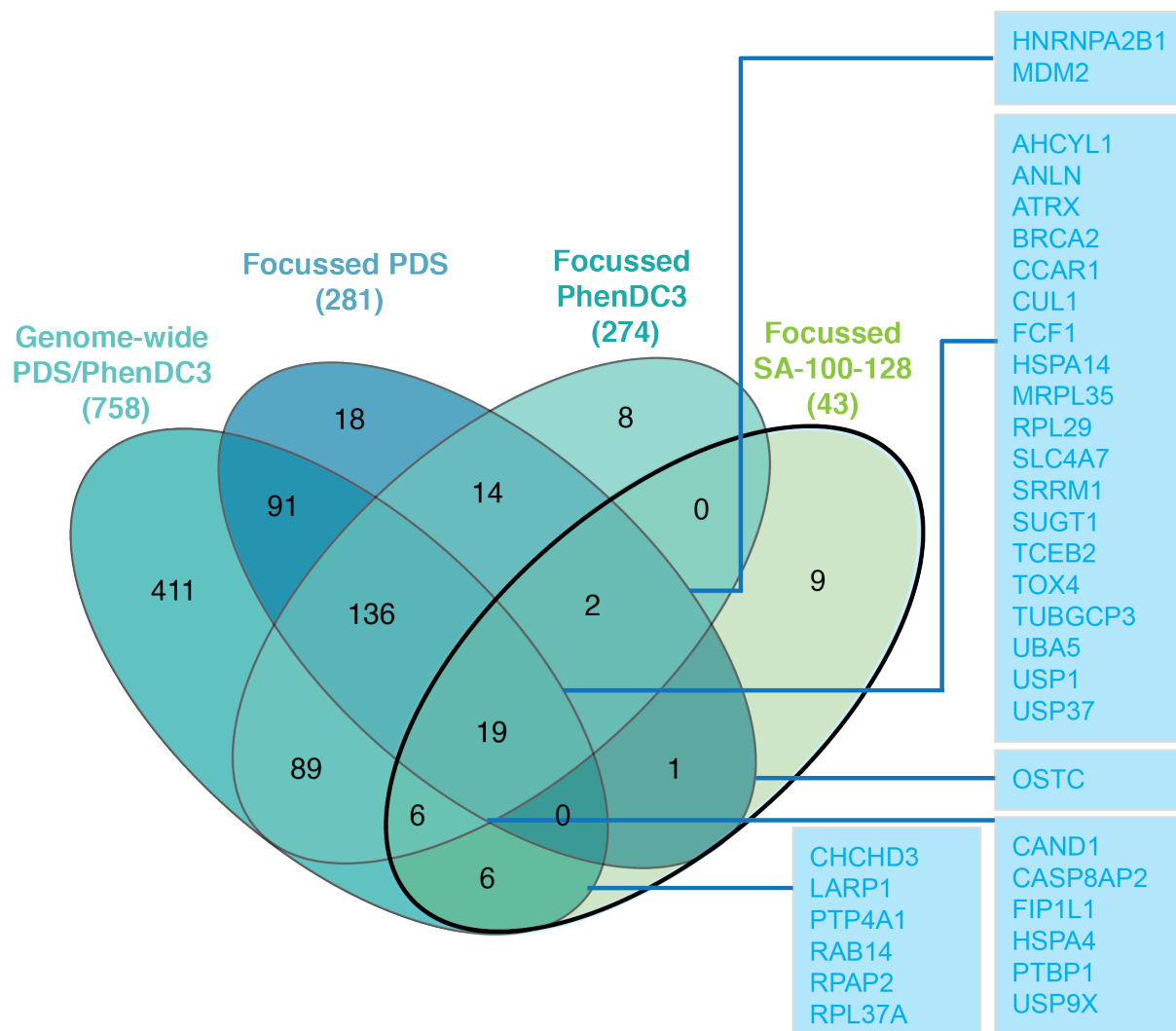
Figure 3.17. PDS and PhenDC3 sensitisers overlap with SA-100-128 depleted genes
Venn diagram showing the overlap between the 758 sensitisers identified in the genome-wide screen with PDS and PhenDC3, the 281 PDS sensitisers, 274 PhenDC3 sensitisers and 43 SA-100-128 upregulated genes uncovered in either the A375 or HT1080 focused screens. Significantly depleted genes are classified as 50 % or 3 hairpins significantly depleted at FDR ≤ 0.05. Additional thresholds of $log_2FC < -1$ were used to classify sensitisers throughout the screens.

3.3 Discussion

3.3.1 Validating the synthetic lethality of *BRCA, TOP1, GAR1* and *DDX42* with PDS and PhenDC3

Within this chapter, I set out to validate the four key sensitisers from the shRNA screen via an orthogonal short-term siRNA approach (5 days versus ~28 days for the shRNA screen). In the HT1080 cell line, reduced expression of all four genes resulted in PDS and PhenDC3 sensitivity, successfully recapitulating the shRNA screen results. Generally, the PDS and PhenDC3 sensitivities seen for cells deficient in the four proteins were greater for HT1080 cells versus A375 cells. This perhaps arises from the lower siRNA-induced depletion of all four proteins in A375 cells compared to the HT1080 cell line. Nonetheless, A375 BRCA1- and GAR1-deficient cells were PDS and PhenDC3 sensitive and TOP1-deficient cells were PDS sensitive and PhenDC3 insensitive.

For both cell lines, BRCA1- and TOP1-deficiencies caused comparatively greater PDS sensitivity compared to DDX42- and GAR1-deficient cells. As BRCA1 and TOP1 are primarily DNA-interacting proteins, one explanation could be that PDS has greater affinity for DNA-G4s over RNA-G4s. This hypothesis is in line with focused shRNA screen observations (chapter 2), where PDS hits were more enriched in DNA-related terms than for PhenDC3 hits. Alternatively, the DNA enrichment may be a reflection that although PDS stabilises both RNA- and DNA-G4 in cells, perturbation of the DNA-associated pathways resulted in a stronger phenotype than less essential RNA

dependent pathways. Perhaps TOP1 and BRCA1 represent key players in the cellular response to stabilised DNA-G4s: TOP1 prevents their formation while BRCA1 repairs G4-induced damage. For BRCA1 this hypothesis has been explored in the literature (McLuckie *et al*, 2013). For TOP1, this is the first study to show G4-stabilising ligand synthetic lethality, although human TOP1 has been shown to bind to G4s *in vitro* (Arimondo *et al*, 2000). This study also indicated TOP1 was able to encourage the formation of parallel G4-structures, whereas, another biophysical investigation indicated TOP1-inhibition by G4 structures (Ogloblina *et al*, 2015). More recently linked to suppression of DNA-damage at G4s in yeast (Yadav *et al*, 2014). Another explanation for *TOP1* synthetic lethality is that topoiosomerase I uses G4-structures as a docking site to bind and perform its superhelical strain relaxation role. In this scenario, TOP1-depletion potentially increases the quantity of "free" G4-structures for G4-stabilising ligands to bind.

The less PDS sensitive response of DDX42- and GAR1-deficiencies compared to TOP1- and BRCA1-knockdown may reflect less critical DNA-G4 roles and more of a possible RNA-G4 relationship. In support of an RNA (and perhaps RNA-G4) function, the strong PhenDC3 sensitisation following siRNA-induced GAR1 depletion corroborates the enrichment of RNA and mRNA related PhenDC3 sensitivities within the shRNA screen (chapter 2). For GAR1, the possible mechanisms of this sensitivity will be discussed in section 3.3.3. The mechanism of DDX42 sensitivity to G4-ligands will be further explored in chapter 4.

Of note, compared to non-targeting siRNA controls, cells deficient in the key four sensitisers were more sensitive to the lower PDS concentration. At higher concentrations, control cells may also show significant PDS-induced growth inhibition, thus limiting the window by which protein-deficiency induced sensitivity can be gauged. Both PhenDC3 concentrations (20 µM and 40 µM) were less toxic to normal cells than PDS. Therefore a similar sensitivity window exists for both concentrations, and revealed that 40 µM PhenDC3 caused greater growth inhibition than 20 µM. The PDS and PhenDC3 responses affirm that non-toxic ligand concentrations for normal cells may be the most useful, by widening the sensitivity window. As discussed in chapter 2, one way to achieve this is via combinatorial ligand treatments, a concept further explored in the overall discussion (chapter 6).

3.3.2 Extending validation of the four key sensitivities to CX-5461 and cPDS treatment

The synthetic lethality of the four genes was investigated with two further G4-ligands: CX-5461 (Xu *et al*, 2017) and the RNA-G4 specific cPDS (Biffi *et al*, 2014a; Di Antonio *et al*, 2012). CX-5461 showed synthetic lethality with all four proteins in both cell lines. The extension of this sensitivity to a further G4-ligand supports that these four proteins represent key players in G4-biology. It also suggests that GAR1-, TOP1- and DDX42-deficient backgrounds can be similarly therapeutically targeted by CX-5461 as for BRCA-deficiencies.

Deficiencies in BRCA1 and DDX42 were sensitive to cPDS whereas TOP1 and GAR1 were largely insensitive at the concentrations and time frame used.

For TOP1, this supports that this nuclear protein predominantly acts on DNA, perhaps involving DNA-G4 roles. For DDX42, a hypothesised RNA-helicase (see section 4.1.1), cPDS sensitivity supports the existence of RNA-G4 associated roles among its functions. This is not mutually exclusive with DDX42 performing DNA and perhaps DNA-G4 roles. The acute sensitivity of BRCA1-deficiency to cPDS is interesting, given that the primary association of BRCA1 with G4s is the repair of DNA-G4 associated damage. The cPDS sensitivity perhaps suggests an additional RNA-G4 associated BRCA1 role, and several studies in the literature link RNA to the DDR. For example, it was recently shown that small non-coding RNAs are transcribed at DNA DSB, and perform a recruitment role for an incompletely characterised complex of DNA damage proteins (Michelini *et al*, 2017). Of note, non-DDR BRCA1 roles include activatory and co-repressor transcriptional regulation, cell cycle control and ubiquitylation (Mullan *et al*, 2006), all of which were enriched sensitivities within the genetic screening results (chapter 2). In support of a possible RNA (and RNA-G4) role, BRCA1 shuttles between the nucleus and cytoplasm, enabling contact with RNA species (Fabbro & Henderson, 2003) and interacts with PABP1 (Poly(A)-binding protein). Further, BRCA1-depletion correlates with a global translation decrease in MCF7 cells (Dizin *et al*, 2006; Dacheux *et al*, 2013) possibly contributing to the cPDS sensitivity observed.

GAR1 cPDS insensitivity, but responsiveness to PDS, PhenDC3 and CX-5461, could suggest that while RNA-G4 stabilisation may contribute to GAR1 synthetic lethality, it is insufficient in isolation, and other G4-targets (perhaps DNA-G4) need to be targeted. However, having only tested one RNA-G4

ligand, this lack of response could be ligand-specific rather than due to a mechanistic/physical interaction of GAR1 with RNA-G4.

3.3.3 Possible mechanisms of GAR1 G4-ligand synthetic lethality

GAR1 is a member of the H/ACA small nucleolar ribonucleoproteins (snoRNPs), forming a complex with DKC1, NOP10 and NHP2 (reviewed in McMahon, Contreras, & Ruggero, 2015). DKC1 was also identified as a PDS and PhenDC3 sensitiser in the genome-wide screen (median $\log_2$FC -5.41 and -4.08 respectively), and a PDS sensitiser in the A375 focused screen (median $\log_2$FC -2.54). NHP2 was an A375 PDS sensitiser for both genome-wide and focused screens (median $\log_2$FC -2.1 and -2.0 respectively). This establishes that H/ACA protein deficiencies, and perhaps the deficiencies in the whole complex, as G4-ligand sensitivities.

The H/ACA snoRNPs are required for ribosomal RNA biosynthesis and telomere extension (Kiss, 2001; Pogacić *et al*, 2000) via interaction with an ACA RNA motif, within snoRNA guides (Ganot *et al*, 1997a; Ni *et al*, 1997) and the human telomerase RNA (hTR) component (Mitchell *et al*, 1999a, 1999b; Dragon *et al*, 2000; Pogacić *et al*, 2000) respectively. H/ACA protein depletion causes shortened telomeres and reduced telomerase and ribosome levels.

The structure of snoRNAs are well characterised, consisting of stem loops and conserved "pseudouridylation pockets" (Ganot *et al*, 1997b), but no G4-potential sequences. Thus synthetic lethality is unlikely to derive from GAR1

binding RNA G4s within the snoRNA. One explanation is that both G4-ligand treatment and GAR1-deficiency are problematic for translation, the former by inhibiting ribosomal scanning via mRNA 5' UTR G4-stabilisation and the latter by ribosome biogenesis reduction. Of note, the DNA G4-stabiliser, CX-5461 (Xu *et al*, 2017) was originally identified as an RNA polymerase inhibitor, specifically inhibiting rRNA synthesis (Drygin *et al*, 2011). Thus, given that GAR1 is required for ribosomal RNA maintenance, this may contribute to the observed CX-5461 sensitivity, independent of G4-stabilisation.

An alternative synthetic lethality mechanism is that GAR1-deficiency induced telomerase dysfunction may synergise with the telomere structure disruption caused by G4-ligand treatment (see section 1.3.3). A telomeric DNA-G4 associated role may explain why GAR1-deficient cells are PDS, PhenDC3 and CX-5461 sensitive but unaffected by cPDS treatment. Another possibility comes from recent Drosophila evidence suggesting that several snoRNA can associate with chromatin binding proteins, where they may perform chromatin-remodelling roles (Schubert *et al*, 2012). Chromatin alterations induced by GAR1 depletion may impact G4 formation and/or ligand accessibility and thus also contribute to G4-ligand sensitivity. In terms of therapeutic insights, several solid tumours and haematological malignancies show H/ACA snoRNA deregulation, although this often arises from alteration of specific snoRNA subsets, rather than changes in snoRNP levels themselves (McMahon *et al*, 2015). However, reduced GAR1 transcript levels are reported in chronic lymphocytic leukaemia cells compared to normal

clinical biopsies (Dos Santos *et al*, 2017), perhaps suggesting an area for G4-ligand therapeutic exploitation.

3.3.4 SA-100-128 as a novel G4-stabiliser with improved pharmacokinetic properties

Following siRNA screening of 12 PDS derivatives with improved medicinal chemistry properties, SA-100-128 showed synthetic lethality with the four siRNA knockdowns. The "quadruplex" focused screen (see chapter 2) was then employed to identify further SA-100-128 sensitivities. This screen revealed that, despite toxicity problems for the longer shRNA experiment, SA-100-128 shared some synthetic lethalities also seen with PDS and PhenDC3. As PDS and PhenDC3 were chosen as representative G4-stabilising ligands, perhaps SA-100-128 is a step closer to a promising "drug-like" molecule that shares the G4-stabilisation phenotypes experimentally reported for G4-ligands in the literature (see section 1.4.2).

Despite the siRNA knockdown of the key sensitivities (BRCA1, TOP1, DDX42 and GAR1) showing SA-100-128 synthetic lethality, these hits were not significantly depleted in the SA-100-128 shRNA screen. This suggests two main findings: 1) the SA-100-128 concentration used was insufficient to identify the complete spectrum of synthetic lethalities; 2) genes uncovered reflect strong sensitivities, more so than the BRCA1/TOP1 counterparts.

Overlapping the SA-100-128 depleted genes with sensitisers identified in PDS and PhenDC3 screens revealed positive controls *BRCA2*, *ATRX* and

HNRNPA1, and multiple components of the ubiquitin pathway: *CUL1, UBA5, USP1*, *USP37, CAND1* and *USP9X.* As discussed in chapter 2, several of these ubiquitin components are being explored as therapeutic targets due to their cancer deregulation. The sensitivity of cells deficient in these genes to SA-100-128, a more "drug-like" molecule than PDS, may bring us closer to the clinical development of G4-stabilising ligands in, for example, ubiquitin-deregulated malignancies. Another gene that showed sensitivity to PDS, PhenDC3 and SA-100-128 was *CCAR1* (cell cycle and apoptosis regulator) an oncogene overexpressed and required for tumour progression in gastric carcinoma (Chang *et al*, 2017). CCAR1 is currently being investigated as a therapeutic target, and such pharmacological inhibition could be potentiated by co-treatment with SA-100-128.

One gene deficiency, *AHCYL1* (S-adenosylhomocysteine hydrolase-like protein 1) caused sensitivity to SA-100-128, PDS and PhenDC3 in both HT1080 and A375 cells, perhaps indicating that this gene is integral in G4-biology regulation. AHCYL1 has several functions, including suppression of the inositol phospholipid pathway (Berridge *et al*, 2000; Ando *et al*, 2003, 2006) and regulation of several ion channels and transporters, among them the Cystic Fibrosis Transmembrane Receptor (CFTR) chloride channel (reviewed Ando, Kawaai, & Mikoshiba, 2014). More recently, AHCYL1 has been identified as a regulatory inhibitor of ribonucleotide reductase (RNR), the enzyme required for production and balance of dNTPs (Arnaoutov & Dasso, 2014). Compromising this nucleotide balance is problematic for replication and induces genomic instability (Pai & Kearsey, 2017). G4-ligand sensitivity

arising from AHCYL1 deficiencies may relate to cellular ionic composition deregulation, possibly favouring uptake of the positively charged G4-stabilising ligands and/or increase G4-stabilisation. Alternatively, AHCYL1-deficiencies could induce replicative damage at the level of RNR dysfunction, which is further exacerbated by treatment with G4-stabilising ligands.

From a possible therapeutic standpoint, AHCYL1 is a tumour suppressor in ovarian cancer with low expression levels correlating with poor prognosis (Jeong *et al*, 2012) and is often downregulated in malignant melanoma including those with acquired fotemustine and cisplatin resistance (Wittig *et al*, 2002). It would be interesting to investigate the response of these cancers to G4-ligands, as a single agent in ovarian cancers, and either as a combinatorial treatment for melanoma, or in isolation for melanoma tumours resistant to current standard-of-care drugs.

Overall, the siRNA investigation has been successful in validating the key four sensitisers (*BRCA1, TOP1, GAR1* and *DDX42*), and has provided hints at the mechanisms underlying this synthetic lethality. Additionally, they suggest SA-100-128 is a promising compound with potential for further development as a clinical G4-stabilising ligand.

Chapter 4

Characterising the mechanism of DDX42 synthetic lethality to G4-stabilising ligand treatment

4.1 Background and objectives

In addition to identifying synthetic lethal genotypes with G4-ligand treatment, another primary aim of this thesis was to gain insights into the regulation of G4-structures. Literature to date suggests that cells avoid deleterious effects of persistent G4s by rapidly repairing G4-induced damage and/or tightly regulating their formation by dedicated helicase machinery (see introduction 1.5.1 and 1.5.5). The full repertoire of these regulators is unknown and in part, could be expanded by the identification of helicases that specifically bind and unwind G4 structures. I reasoned that deficiencies in an important G4 regulator would impart consistent sensitivity, across cell lines and ligands, to G4-stabilisation by small molecule treatment. As shRNA-induced deficiencies in DDX42 caused synthetic lethality with four independent G4-stabilising ligands (PDS, PhenDC3, CX-5461 and cPDS) in both A375 and HT1080 cell lines (see chapters 2 and 3), I chose this protein for further investigation. DDX42 is a member of the Asp-Glu-Ala-Asp (DEAD) box protein family. Many of this family are RNA-helicases, with roles including RNA secondary structure regulation in translation initiation, splicing (nuclear and

mitochondrial), and ribosome and spliceosome assembly (Linder & Jankowsky, 2011). The majority of DDX proteins analysed to date have been shown to have a preference for RNA over a DNA substrates (Cordin *et al*, 2006), but in some cases show DNA helicase activity (Kikuma *et al*, 2004; Linder & Jankowsky, 2011). However, as DDX42 was identified as a putative splicing component (see section 4.1.1), the DNA affinity of this protein is largely uninvestigated. Given the results of the synthetic lethality screen and the roles of other helicases in G4 biology, DDX42 represents a promising novel G4-helicase candidate. Firstly I will summarise research to date for this largely uninvestigated gene before discussing my preliminary investigations into the relationship between DDX42 and G4-biology.

4.1.1 Discovery of DDX42 and its characterisation as a putative DEAD-box RNA helicase

DDX42 was first identified in 2000, as a highly abundant mRNA in the sera of insulin-dependent diabetes patients (Suk *et al*, 2000). In this study, ubiquitous DDX42 expression was identified in all tissues tested including liver, lung and pancreas. DDX42 was also independently uncovered by immunopurification and mass spectrometry as an interactor of the core splicing factor U2 snRNP protein (Will *et al*, 2002). Both studies identified DDX42 as a putative DEAD-box RNA helicase, based on peptide sequence motif and bioinformatics analyses respectively. Unusual for members of this family, DDX42 is encoded by two transcript variants with different 5'UTR introns, both lacking a canonical polyA box signal (Uhlmann-Schiffler *et al*, 2006) suggestive of translational regulation. Recombinant protein studies showed that *in vitro,*

DDX42 non-processively unwinds RNA strands in an ATP-bound form. This is increased in the presence of the single-stranded binding protein T4gp32. In an ADP-bound form, DDX42 function switches to promote association of synthetic complementary RNA strands *in vitro* (Uhlmann-Schiffler *et al*, 2006). Additionally, DDX42 interacts with and inhibits the apoptotic inducer ASPP2 (Uhlmann-Schiffler *et al*, 2009). ASPP2 is epigenetically silenced in several cancers, allowing apoptotic evasion (Song *et al*, 2015); DDX42 upregulation might provide an alternative tumorigenic mechanism. Consistent with an anti-apoptotic role for DDX42, Epstein-Barr Virus promotes cell death by downregulating mRNAs including DDX42 (Choi *et al*, 2013). The closest structurally related DEAD box proteins to DDX42 are DDX5 and DDX3, with amino acid conservation evident across the full length of the protein (~40 % homology) (Uhlmann-Schiffler *et al*, 2006; Suk *et al*, 2000). However, DDX5 and DDX3 are tumour suppressors, co-activating p53 and p21 respectively (Bates *et al*, 2005; Chao *et al*, 2006).

DDX42 has also been linked to the viral inflammatory immune response, interacting with the non-structural protein (NS4A) of the Japanese encephalitis virus *in vitro* (Lin *et al*, 2008). Consistent with a function for this *in vivo*, partial colocalisation of DDX42 and NS4A has also been reported in human medullablastoma cells (Lin *et al*, 2008). Based on these observations, DDX42 was hypothesised to detect foreign dsRNA, an interaction antagonized by NS4A. Evasion of the immune system and apoptosis are cited among the hallmarks of cancer (Hanahan & Weinberg, 2011). Thus, while DDX42 deficiencies are not fully characterised regarding their tumour forming

capabilities, with further research into this protein, a cancer association may be possible and perhaps exploitable in conjunction with G4-ligand treatment. Within this chapter, preliminary investigations into the mechanism of G4-ligand sensitivity in DDX42-deficient cells are described.

4.2 Results

4.2.1 DDX42 is a nuclear protein

As DDX42 is classified as an RNA-associated helicase, it is important to identify the subcellular localisation of this protein in order to understand the processes in which it could be involved. Within the cell, there are two defined areas of RNA localisation (and thus RNA-G4s). For example, lncRNAs, pre-miRNAs and pre-mRNAs (including nascent transcripts and R-loops) exist in the nucleus, where RNA G4s are involved in functions including transcriptional regulation and RNA processing. Conversely, mature mRNAs and miRNAs are cytoplasmic, with secondary structures contributing to post-transcriptional regulation e.g. stability, transport and translation.

To investigate which RNAs and therefore processes, DDX42 putatively regulates, subcellular fractionation was used to separate cytoplasmic and nuclear lysates from three independent cell lines, HT1080, HEK and HeLa cells, and these fractions probed via immunoblot for DDX42 (Figure 4.1A). This specificity of this antibody was validated in chapter 3, in which the single band detected at 120 kDa was depleted in lysates from cells transfected with an siRNA targeting DDX42 in two independent cell lines (see section 3.2.1).

To confirm efficient separation of the nuclear and cytoplasmic lysates, lysates were also probed with antibodies targeting laminB1 and GAPDH, nuclear and cytoplasmic proteins respectively (Figure 4.1B&C). DDX42 was found to be nuclear in all three cell lines (4-9-fold increase compared to cytoplasmic levels; figure 4.1D). The small amount of 'cytoplasmic' DDX42 protein is likely due to slight nucleoplasm leakage during fractionation, reflected by a low laminB1 in cytoplasmic fractions (Figure 4.1B). This suggests that the G4-associated roles of DDX42 are primarily nuclear, based on DDX42 abundance.
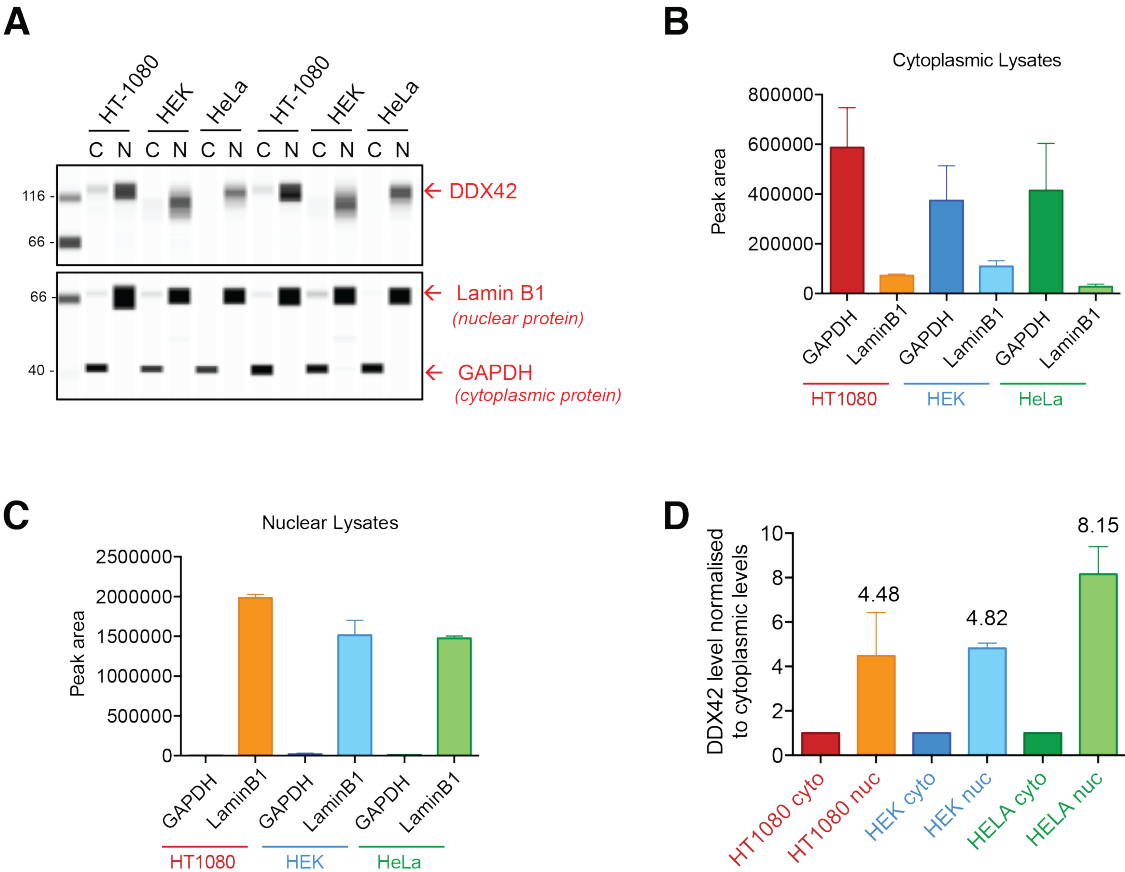
Figure 4.1. Subcellular location of DDX42 is exclusively nuclear
(A) Representative immunoblot showing cytoplasmic (C) and nuclear (N) lysates for HT1080, human embryonic kidney (HEK) and HeLa cells probed for DDX42 alongside laminB1 and GAPDH2 (exclusively nuclear and cytoplasmic expression respectively)
(B-C) Graph showing the average area for the GAPDH and laminB1 peaks for (B) cytoplasmic and (C) nuclear lysates for the three cell lines
(D) Graph showing area of DDX42 peak normalised to cytoplasmic levels (mean for 2 technical replicates ± standard deviation)

4.2.2 DDX42 specifically binds both RNA- and DNA-G4 oligonucleotides *in vitro*

As DDX42 is a predicted RNA helicase, its binding affinity for RNA-G4s was next investigated. For this, a G4 sequence derived from the NRAS 5'UTR, which forms a stable parallel G4, even in the absence of cations (Kumari *et al*, 2007) was used. For comparison the less thermodynamically stable RNA 5'UTR G4 from BCL2 mRNA was also employed (Shahid *et al*, 2010). Additionally the BCL2 sequence was modified replacing the terminal G-rich tract with adenines (i.e. GGG → AAA). This maintains the purine:pyramidine composition but abolishes formation of an intramolecular G4. However, the two remaining G-tracts permit intermolecular G4 formation. As controls, for both NRAS and BCL2 sequences, the central G in each G-rich tract was mutated (G → A), to prevent both intra- and intermolecular G4 formation.

These oligonucleotides were folded in either 100 mM KCl or LiCl, permitting strong and weak stabilisation of G4s respectively, and their resultant structures analysed by circular dichroism spectroscopy (CD; see Methods for details). In KCl, the CD spectrum of the NRAS and BCL2 intra- and intermolecular G4-sequences showed a positive peak at 263 nm and a negative peak at 241 nm, the characteristic signature for parallel G4 structures (Tang & Shafer, 2006; Balagurumoorthy *et al*, 1992). The NRAS G4 annealed under lithium conditions, also showed this signature, albeit weaker than its KCl counterpart, which is in keeping with its stability even in LiCl. For all other sequences, mutating the central G or annealing in LiCl prevented G4 formation as reflected by the altered CD signature (Figure 4.2).
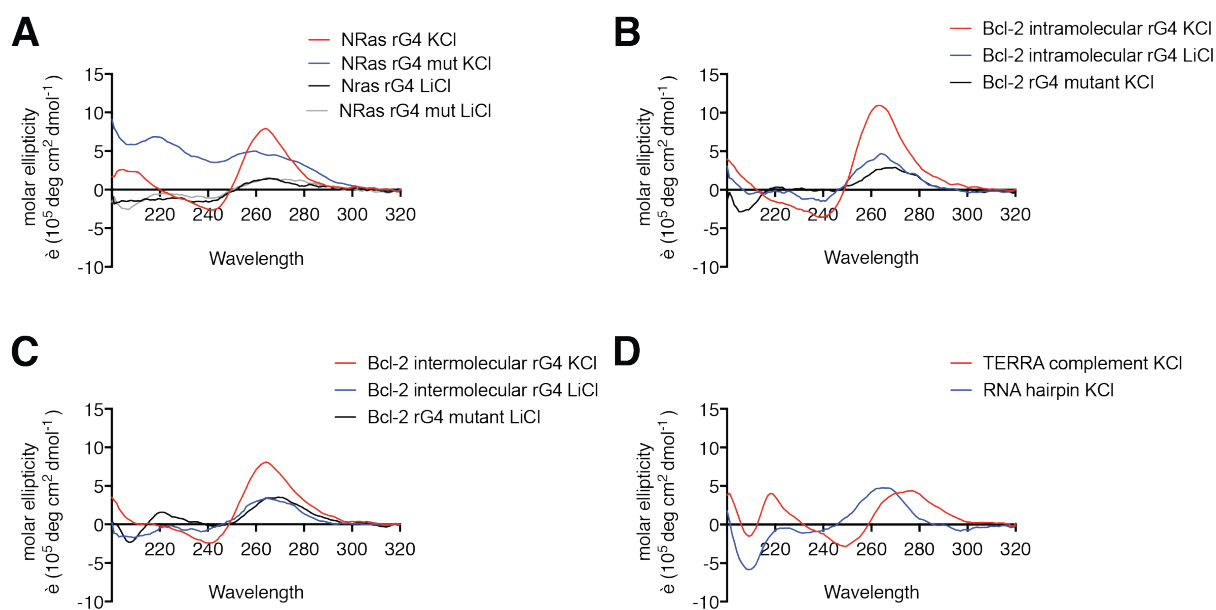
Figure 4.2. Structure of annealed RNA oligonucleotides
CD spectra for G4 forming RNA oligonucleotides (rG4) and controls (rG4 mut) folded either in 100 mM KCl or LiCl as indicated for sequences corresponding to (A) NRas 5'UTR G4 and controls; (B) Bcl-2 5'UTR RNA G4 and controls; (C) intermolecular G4 structure due to GGG → AAA substitution in one of the G-tracts; (D) single stranded/unstructured TERRA complement and RNA hairpin controls

Having confirmed that the various G4-sequences and mutants form the expected structures *in vitro*, the affinity of DDX42 for these structures was analysed via Enzyme Linked Immunosorbent Assay (ELISA; see Methods for details) and data modeled via a non-linear regression model, assuming one-site specific binding and saturation kinetics. With the exception of the binding curve for the KCl annealed NRAS G4, the curves did not reach saturation, a necessity to calculate an accurate K_d, therefore the values reported below reflect a best estimate for these values. DDX42 was found to bind both NRAS and BCL2 intramolecular G4s folded in KCl with nanomolar affinity, 46.15 ± 4.26 nM and 160.9 ± 21.3 nM respectively. The affinity difference may reflect the stability and/or persistence of the G4-structure. This is supported by the ~7-fold less affinity (approximate K_d = 333 nM based on non-saturated curve) of DDX42 for LiCl-annealed NRAS RNA-G4. Additionally, DDX42 also bound

the BCL2 intermolecular G4s, but with much lower affinity than for the intramolecular counterpart. Moreover, DDX42 did not bind detectably to mutant oligonucleotides or the BCL2 G4-forming oligonucleotides folded in LiCl, showing that the interaction is dependent on the presence of a G4-structure. The *in vitro* RNA-G4 specific affinity was further supported by the lack of DDX42 binding to an RNA hairpin or a single-stranded C-rich RNA sequence, derived from the complement of TERRA.
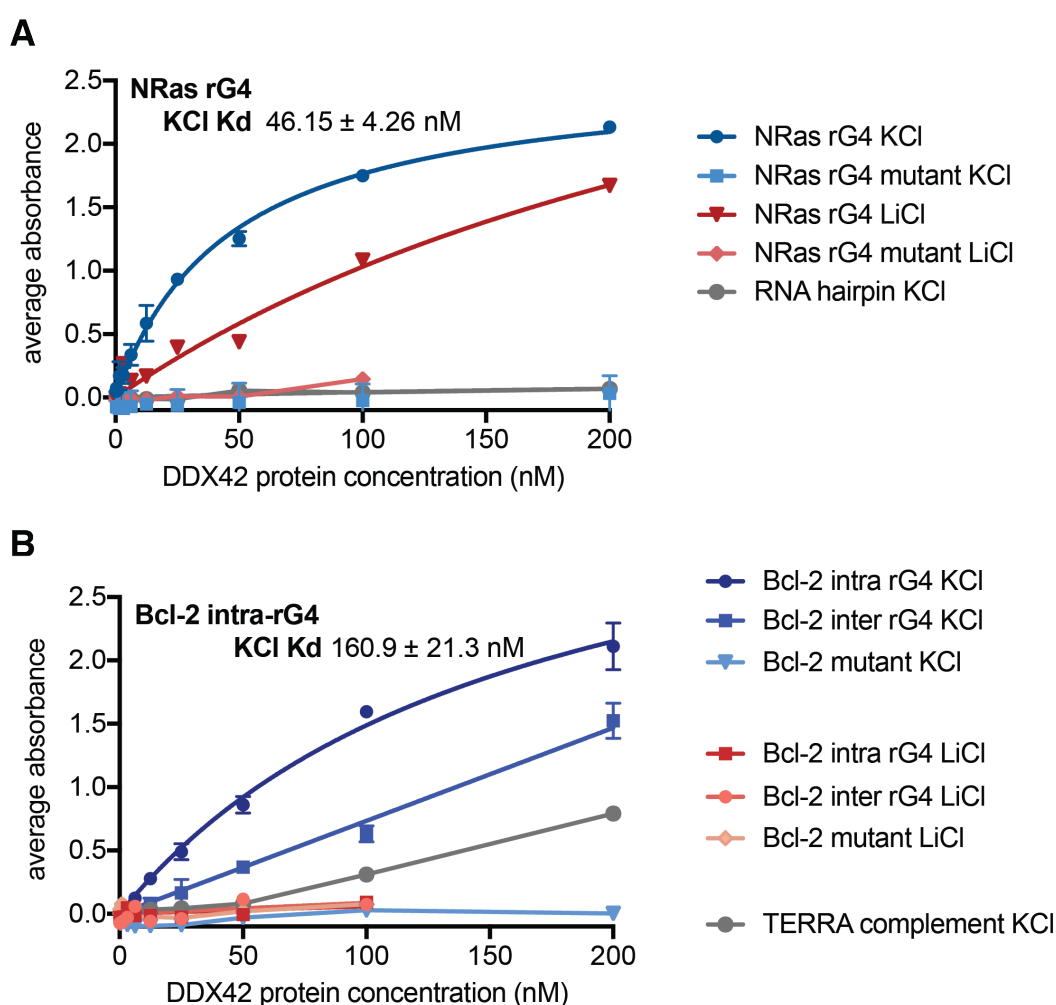
Figure 4.3. DDX42 specifically binds RNA-G4s in vitro
In vitro ELISA binding curves for a representative biological replicate (mean ± S.E.M) for the following oligonucleotides annealed in 100 mM KCl or LiCl as indicated (A) NRas 5' UTR RNA G4 and mutated G4 sequence alongside an RNA hairpin also annealed in 100 mM KCl and (B) Bcl-2 5' UTR intramolecular and intermolecular RNA-G4 and mutant G4 sequence alongside a single stranded RNA control corresponding to C-rich complement of TERRA. Values reflect the apparent K_d of two biological replicates, however as the model assumes saturations kinetics, these are only approximate values.

Given the nuclear localisation of DDX42 and as other members of the DDX family have been shown to have DNA helicase activity (Kikuma *et al*, 2004), the *in vitro* DNA-G4 affinity of DDX42 was also investigated. For this an oligonucleotide corresponding to the stable parallel G4 in the promoter of MYC (González & Hurley, 2010; Yang & Hurley, 2006) and a non-G4 forming control was used. These were folded in KCl and LiCl and analysed by CD (Figure 4.4A). In the presence of K^+ cations, MYC DNA-G4 showed positive and negative ellipticities at 263 and 241 nm, confirming parallel G4 formation. This CD spectroscopic signature was absent in mutant and LiCl-annealed oligonucleotides, affirning no G4-structure formation.

Next, DDX42 affinity for these annealed sequences was investigated (Figure 4.4B). This revealed that DDX42 also binds MYC DNA-G4 with comparable affinity to the BCL2 intramolecular RNA G4 (approximate K_d =159.2 ± 25.2 nM), but not to the mutant or controls folded in LiCl, suggesting a G4-structure specific interaction.
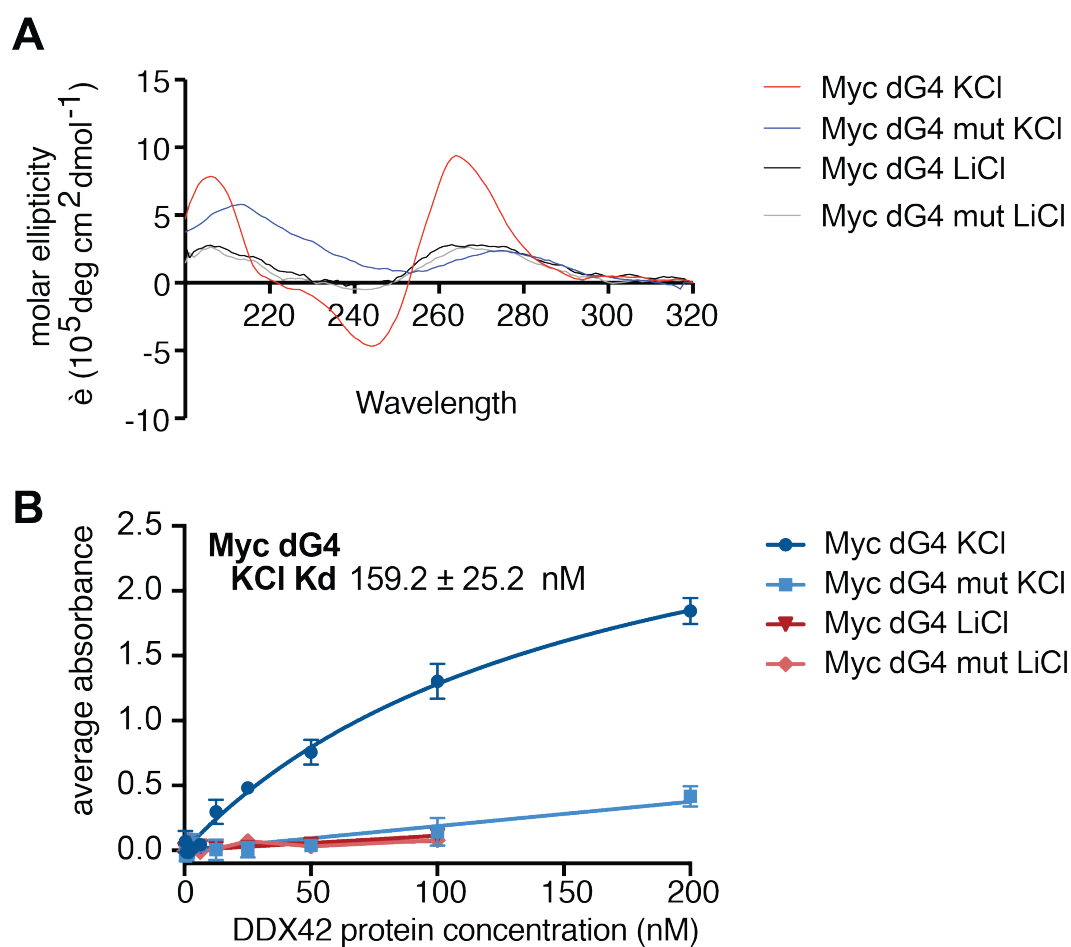
Figure 4.4. DDX42 bind a parallel DNA-G4 structure in vitro
A) CD spectra for G4 forming DNA oligonucleotide sequence corresponding to the cMyc promoter parallel G4 (dG4) and controls (dG4 mut) folded either in 100 mM KCl or LiCl as specified
B) In vitro ELISA binding curves for a representative biological replicate (mean ± S.E.M) for the Myc parallel promoter G4 and mutated G4 sequences annealed in 100 mM KCl or LiCl. The K_d value stated refers to the apparent K_d of two biological replicates.

4.2.3 Investigation of whether DDX42 protein depletion alters nuclear BG4 foci

Deficiencies in helicases with G4-resolving activity may result in increased persistence of their G4-substrates, either globally at the majority of G4-forming sequences or for a select subset of G4s. Given that DDX42 is found in the nucleus and can specifically bind DNA and RNA-G4s *in vitro*, the effect of DDX42 depletion on nuclear G4 levels was next investigated in HT1080 cells.

First, it was confirmed via western blot that HT1080 cells transfected with a siRNA targeting DDX42 caused protein knockdown relative to the non-targeting (NT) control (Figure 4.5A; ~73 % knockdown, 72 h after transfection). Next, DDX42-depleted cells and non-targeting siRNAs were cultured with DMSO or 0.5 µM PDS for either 72 h or 120 h. This concentration of PDS was included to additionally investigate whether a global G4 increase at this low dose could explain the ligand sensitivity seen following siRNA-induced knockdown (chapter 3). Following treatment, cells were fixed, permeabilised and treated with the G4-specific FLAG-tagged BG4 single chain antibody (Biffi *et al.* 2013). BG4 signal was amplified via incubation with an anti-FLAG secondary antibody, which was subsequently recognised by a fluorophore-labeled tertiary antibody, to give nuclear foci corresponding to the detection of endogenous G4-structures (red foci; Figure 4.5B). For each condition the number of BG4 foci were quantified for 70-100 nuclei (Figure 4.6C-D). Additionally, no BG4 and no secondary antibody controls were

included for the 72 h timepoint, which confirmed the absence of non-specific binding by the secondary and tertiary antibodies respectively (Figure 4.5C).

After 72 h, no significant difference was seen between the average nuclear BG4-foci for any condition, suggesting neither the treatment with low concentration of PDS nor DDX42 depletion was sufficient to globally increase nuclear G4 levels, across two biological replicas (Figure 4.5C). Following 120 h, a significant increase was seen in the PDS-treated NT cells compared to DMSO vehicle control (Figure 4.5D), suggesting that 0.5 μM PDS can cause a small but significant increase in the nuclear G4-level in HT1080 for both biological replicas one ($p = 5 \times 10^{-4}$) and two ($p = 2.7 \times 10^{-3}$). However, no significant increase was seen in DDX42-depleted cells ($p = 0.2134$ and 0.1046 for biological replica 1 and 2 respectively), supporting that the single gene knockdown does not cause a detectable increase in nuclear G4 levels with this method.
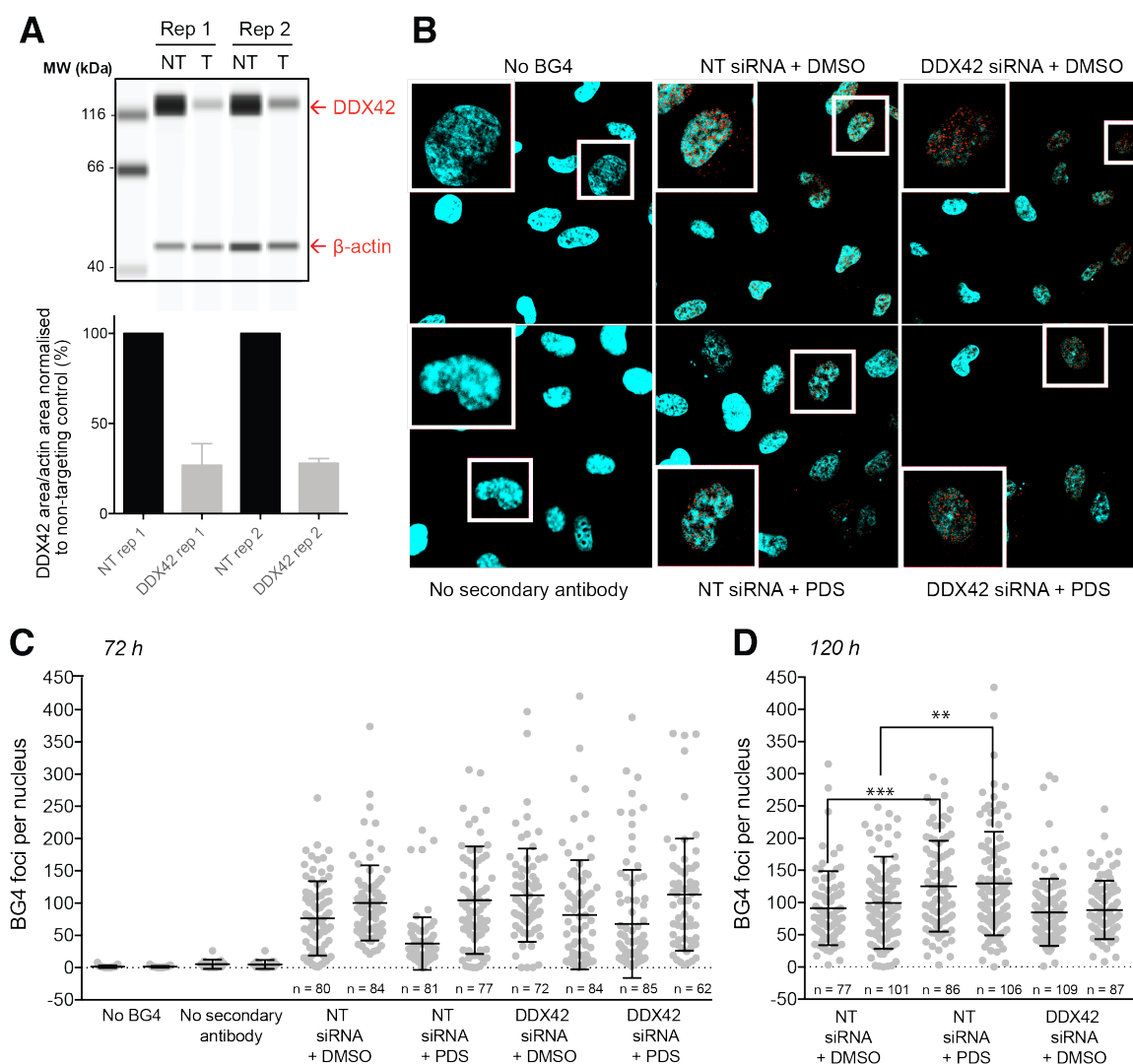
Figure 4.5. DDX42 deficient cells do not show a global increase in G4 structures
(A) Top - representative western blot showing lysates immunoblotted for DDX42 and beta-actin loading controls for cells transfected with a siRNA targeting DDX42 (T) or non-targeting control siRNA (NT), 72 h after transfection. Bottom - for all samples DDX42 protein levels were normalised internally to the beta actin control, and averaged for two technical replicates. For each biological replica DDX42 protein levels in targeting lysates were then expressed as a percentage of levels in the non-targeting control (technical replicate mean ± standard deviation).

(B-D) Indirect immunofluorescence microscopy to visualise endogenous G4-structures (B) Representative confocal microscopy images following 72 h treatment with either DMSO or PDS as indicated in conjunction with either NT or DDX42 siRNA transfection. No BG4 = NT cells + DMSO were incubated with secondary and tertiary fluorescent antibodies; No secondary antibody = NT cells + DMSO, incubated with BG4 and the tertiary fluorescent antibody but not the secondary anti-FLAG which binds to and amplifies the BG4 signal. Nuclei are counterstained with DAPI (blue) and BG4 foci are red (Alexafluor 594).

(C-D) Quantification of BG4 nuclear foci following (C) 72h and (D) 120 h siRNA transfection, for two biological replicas per condition. For each condition the number of nuclei counted (n) is indicated and the number of BG4 foci per nucleus plotted alongside mean ± standard deviation. Significance was determined via an unpaired parametric t-test, assuming equal standard deviation.

4.2.4 Investigating the effect of G4-ligand treatment on DDX42 levels

Previous RNA-seq data following 24 h treatment of MCF-7 breast cancer cell line with PDS had shown upregulation of DDX42 mRNA levels (personal communication Dr. Debbie Sanders). However, increased mRNA levels do not necessarily correlate with protein translation, and thus increased activity of a protein (Liu *et al*, 2016). Therefore, it was next investigated whether treatment of HT1080 with G4-stabilising ligands directly altered DDX42 endogenous protein levels, rather than drawing inferences from mRNA alterations.

First, to explore whether ligand-induced alterations in DDX42 expression contributed to its identification as a sensitiser in the shRNA screen (section 2.4.10), DDX42 protein levels in the initial reference (t0) and final timepoint (t15) samples (treated with either DMSO, 0.5 µM PDS and 1 µM PhenDC3 for 15 population doublings; ~25 days) from two biological replicates of the HT1080 focused screen were analysed via western blot. No significant difference was observed between t0 and t15 DDX42 levels (data not shown). It was hypothesised that the screen may represent too long a time period to evaluate translational changes in DDX42 levels, as the cells may have adapted to prolonged depletion of DDX42 and/or treatment with G4-stabilising ligands.

Therefore the influence of G4-ligand treatment on DDX42 levels in HT1080 cells for a shorter time period was evaluated. Based on short-term siRNA

validation assays (chapter 3), G4-ligand sensitivity was observed within 96 h after DDX42-knockdown. This could suggest that the G4-ligand response in DDX42-proficient cells may also occur within this time frame. Therefore, HT1080 cells were treated with G4-ligand concentrations previously optimised to cause greater growth inhbition in DDX42-deficient cells compared to non-targeting siRNA controls for 96 h (section 3.2.1). For this PDS (0.25 μM and 0.5 μM) and PhenDC3 (20 μM and 40 μM) were used. Cells were also treated with CX-5461 (10 nM and 50 nM) and SA-100-128 (0.5 μM), to compare responses.
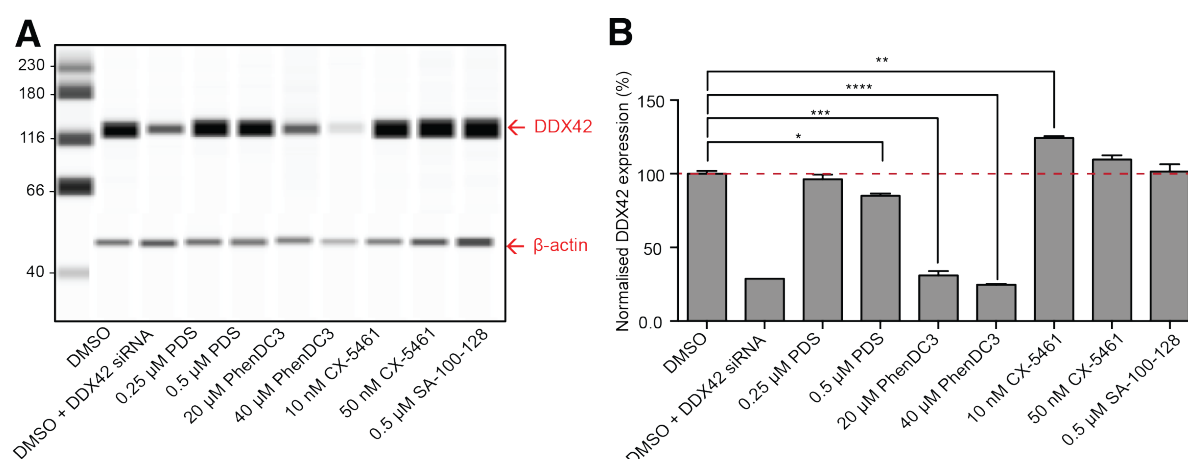
Figure 4.6. PDS and PhenDC3 treatment reduces DDX42 protein expression levels
(A) Representative western blot of one biological replica in which HT1080 lysates, treated with PDS, PhenDC3, CX-5461 and SA-100-128 for 96 h, alongside DMSO (vehicle) and anti-DDX42 siRNA treated controls were immunoblotted for DDX42 alongside beta-actin loading control.
(B) Graph showing the DDX42 level for each sample, normalized to the actin loading control and expressed as a percentage of the level found in DMSO treated cells for three biological replicates (mean ± standard deviation). Significance was determined by a parametric t-test assuming equal standard deviation (p < 0.05). * p<0.05; ** p< 0.01; *** p<0.001; **** p< 1 x 10^{-4}.

Following 96 h ligand or DMSO-vehicle control treatment, cell lysates were immunoblotted for DDX42, alongside DDX42-depleted control lysates for comparison (Figure 4.6A). Average DDX42 levels for two biological replicates

were expressed as a percentage of the levels in DMSO-treated cells for each condition (Figure 4.6B) and revealed that 0.25 μM PDS (p = 0.29), 50 nM CX-5461 (p = 0.08), SA-100-128 (p = 0.49) did not significantly alter DDX42 levels compared to the DMSO control at this time frame and ligand concentration. At 10 nM treatment CX-5461 induced a 24 % increase in DDX42 protein levels (p = 0.0014). For PhenDC3 treated cells the protein expression was similar to that caused by siRNA knockdown (28.8 %) at both 20 μM (31.1 %, p < 0.0001) and 40 μM (24.7 %, p < 0.0001) concentrations. Treatment with 0.5 μM PDS also caused a 15 % decrease relative to DMSO control (p = 0.011). Thus the DDX42 phenotypes at the protein level differed between ligand treatments, but what was apparent is that there is no major upregulation of the protein.

4.3 Discussion

In this chapter, preliminary investigations into the G4-*DDX42* relationship were performed, as this is a gene that when depleted caused robust sensitivity to several G4-ligands in A375 and HT1080 cells. As bioinformatic analysis indicates that DDX42 belongs to the DEAD-box RNA helicase family (Suk *et al*, 2000; Will *et al*, 2002) and other helicases have specific G4-resolving activity, I hypothesised that DDX42 may represent a novel G4-helicase.

DDX42 was exclusively nuclear in three separate cell lines. This location suggests that any RNA roles DDX42 performs are nuclear, such as splicing, and transcriptional regulation rather than cytoplasmic, e.g. translational

regulation at the ribosomal level. This is in keeping with its identification as an interactor with the U2 splicing component (Will *et al*, 2002). Although characterised as an RNA-helicase, given that several RNA-G4 binding proteins also show DNA-G4 affinity, for example TRF2 and XRN1 (see 1.5.5.2), a nuclear localisation also suggests the possibility of DNA-specific roles. In support of this, recombinant human DDX42 selectively bound several parallel G4-structures *in vitro* with nanomolar affinity: RNA-G4s from the 5'-UTR of NRAS and BCL2, and MYC promoter DNA-G4. For the RNA-G4s, this affinity correlated with G4 thermodynamic stability, as DDX42 bound the NRAS RNA-G4 with 4-fold greater affinity than the less stable BCL2-G4 (Kumari *et al*, 2007; Shahid *et al*, 2010). Additionally, DDX42 bound the NRAS RNA-G4 folded under lithium conditions, with 10-fold less affinity than the KCl folded counterpart. This may reflect the cationic stability dependence of G4s (Huppert & Balasubramanian, 2005; Davis, 2004). In addition to binding the BCL2 intramolecular G4, DDX42 could also bind the intermolecular structure, albeit with a much lower affinity. This could reflect a biological phenomenon that DDX42 binds intermolecular structures more weakly than intramolecular G4s. More likely the observation reflects a kinetic property, with intermolecular G4 formation being less likely and more transiently to form due to the loss of entropy of two molecules becoming one (Van Rysselberghe, 1935). Having confirmed that DDX42 could bind G4s *in vitro*, the next step would be to investigate whether DDX42 can also unwind these structures, to delineate whether DDX42 is a G4-specific helicase.

The effect of DDX42-deficiency on HT1080 nuclear G4 levels was also investigated. It was hypothesised that G4-ligand synthetic lethality arises from an inability to resolve either RNA/DNA-G4s causing translation, splicing or transcriptional blockages either at a global scale or at a small subset of G4s. Global BG4-foci levels were found unaltered upon DDX42 depletion, suggesting that DDX42 is not a universal G4-resolvase. This may make evolutionary sense, since if one protein was responsible for unwinding a significant proportion of the prevalent G4 structures, the cell would be acutely susceptible to mutations in this gene. Thus a universal resolvase would be incompatible with the 'genetic redundancy' theory in which genes with overlapping functions exist to allow genetic buffering and minimise adverse phenotypes (Kafri *et al*, 2009). In the nucleus, the majority of BG4 signal is derived from genomic DNA-G4 which can be largely abolished via DNAse treatment (Biffi *et al*, 2013). Therefore, an alternative explanation is that, although able to bind RNA and DNA-G4s *in vitro*, within the nucleus DDX42 only binds and/or unwinds RNA structures. Perhaps an increase in nuclear RNA-G4 levels is obscured by the large DNA-signal, thus is an artifact of the sensitivity limit of this method. Further, although it is hypothesised that the BG4 foci represent G4s, it is unknown if these represent a cluster of G4s or only detect G4s at specific loci. Therefore, even if DDX42 depletion does increase DNA- and/or RNA-G4 levels, this may not correlate with increased nuclear BG4 foci. Therefore, a more appropriate technique to detect alterations in DNA-G4 number following DDX42-depletion would be via G4-ChiP-seq (see section 1.3.1). Alternatively, it would be of interest to map the binding site of DDX42 via ChIP-seq, to investigate whether DNA-G4s are

genome-wide targets of DDX42, in an analogous manner to the mapping of the transcriptional helicases XPB and XPD to G4-motifs at transcription start sites (Gray *et al*, 2014).

If DDX42 is not a global G4-resolvase, DDX42 may bind to and/or unwind a specific subset of G4 structures. Given the presence of DDX42 in spliceosome complexes (Will *et al*, 2002), the most likely candidate substrates are mRNA G4s. In the absence of DDX42, perhaps these mRNAs are improperly processed, a defect exacerbated by G4-stabilising ligands, in turn inhibiting the growth rate of cells. Although the effect of DDX42 depletion on processing of specific transcripts has not been investigated, this mechanism has been explored for other essential splicing factors including SRSF10 and HNRNPM, genes also identified as G4-ligand sensitisers in the genetic screen (see section 2.4.5). SRSF10 was shown to be a key pre-mRNA splicing regulator of BCLAF1 (Bcl-2-associated transcription factor 1). SRSF10-depletion promoted exon5a inclusion of BCLAF1 which was sufficient and necessary to increase colon cancer growth rate (Zhou *et al*, 2014). Conversely, hnRNPM-depletion inhibits breast cancer growth and metastasis in mouse models specifically due to incorrect alternative splicing of CD44 (Xu *et al*, 2014b). These case studies highlight that dysregulation of a single pre-mRNA substrate can significantly impact cellular growth rate, a mechanism that may also be shared by DDX42. Thus for future investigation it would be of merit to perform RNA immunoprecipitation to identify DDX42 associated RNAs, and analyse the PQS prevalence in the enriched sequences. Additionally, transcriptome-wide analysis of alternative splicing differences

between wild-type and DDX42-deficient cells could be performed, as successfully achieved for SRSF10 knockout cells (Zhou *et al*, 2014). For a more focused approach, several mRNAs contain G4s that are thought to contribute to their alternative splicing: IGF-II (Christiansen *et al*, 1994), TP53 (Marcel *et al*, 2011), BACE1 (Fisette *et al*, 2012) and CD44 (Xu *et al*, 2014). Thus it is of merit to investigate the splicing alteration of these mRNAs following DDX42 depletion.

Finally, the effect of G4-ligand treatment on DDX42 expression levels was investigated. Such regulation at the gene expression level would be compatible with the gene and mRNA structure of DDX42 (Uhlmann-Schiffler *et al*, 2006a, section 4.1.1). DDX42 protein levels were decreased following treatment with PDS and PhenDC3 at concentrations used to validate synthetic lethality in HT1080 DDX42-deficient cell lines (section 3.2.1). This decrease was particularly apparent for PhenDC3 treatment where DDX42 protein expression was comparable to that seen following siRNA-induced knockdown. A decrease in DDX42 was not seen following CX-5461 and SA-100-128 treatment, although DDX42-deficient cells are also sensitive to these ligands. Thus although DDX42 was downregulated following PhenDC3 treatment, this does not seem to reflect a universal response to G4-ligand treatment and is unlikely to be the only mechanism behind the synthetic lethality observed.

On a wider level, the identification of DDX42 as a novel G4-binding protein *in vitro* also supports the concept that other specific interactors may exist within the sensitiser list of over 700 proteins (chapter 2). The lack of alteration in

nuclear G4 levels following DDX42 depletion may highlight that synthetic lethality with low-level G4-ligand treatment can arise from slight alterations rather than large-scale changes in endogenous regulatory mechanisms. More concrete support for this will arise based on the outcome of the further experiments I suggest above. Additionally, differences in the ability of the different G4-ligands to alter DDX42 expression levels, reflects a recurring theme within the G4-field and also this thesis. Despite their hypothesised common *in vitro* G4-targets, cellular treatment with G4-ligands can give differential phenotypes, which must be carefully considered when designing and adapting these molecules for future clinical use.

Chapter 5

Identifying gene deficiencies that cause resistance to G4-ligand induced cell death

5.1 Objectives

In chapter 2, a genetic shRNA screen approach (genome-wide and focused) was described for identifying synthetic lethal interactions with the G4 ligands PDS and PhenDC3, using low ligand doses and cell death as an experimental readout (Figure 5.1, gene C). Within this dataset, gene deficiencies also resulted in a growth advantage in the presence of these molecules compared to DMSO treatment alone (Figure 5.1, gene A). For this chapter, these genes will be investigated to give insight into the possible resistance mechanisms cells can develop in response to treatment with G4-stabilising ligands. Furthermore, a subset of these identified resistance genes will be supplemented with a secondary siRNA experiment to provide further validation. The use of a low GI_{20} ligand concentration, used to identify the most sensitive vulnerabilities, potentially does not provide a sufficiently strong driving pressure to engage all resistance mechanisms, compared to a classic 'positive' selection screen where most cells die, and 'surviving' cells are expanded (reviewed in Miles, Garippa, & Poirier, 2016). However, gene knockdowns that consistently cause resistance across the genome-wide and focused screens are unlikely to be false positives and worthy of exploration as important in G4-ligand induced resistance. Overall, this chapter aims to

provide further insights into 1) the cellular response to G4 stabilisation and 2) undesirable off-target effects for G4-ligands in order to lay the foundations to develop more targeted G4-stabilising ligands. Minimising such non-specific toxicities could widen the sensitivity window between normal cells and those with backgrounds acutely sensitive to G4-stabilisation. Additionally, this investigation may reveal key genes and/or pathways that when depleted reduce the number of intracellular G4s accessible to PDS and PhenDC3, complementing the synthetic lethality investigations outlined in chapter 2, and revealing genotypes that may not be applicable to G4-ligand treatment.
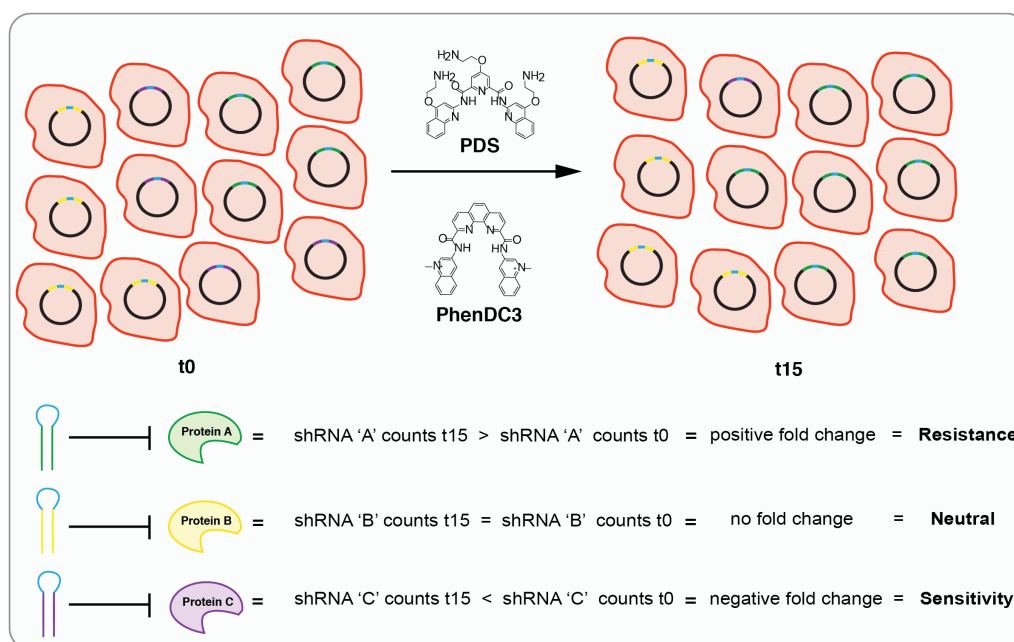
Figure 5.1. Identifying gene deficiencies that cause resistance to G4-stabilising ligands via genetic shRNA screening
Cells expressing genome encoded synthetic shRNAs against protein coding genes were cultured with G4-stablising ligands for 15 population doublings (t15), and shRNA counts were quantified and compared for t0 and t15. shRNAs can be classified into three categories based on their differential expression in t15 versus t0: 1) shRNA is present at greater levels in t15 versus t0 suggesting that the shRNA-induced protein knockdown offers the cells ligand specific-growth advantage, referred to as 'resistance' (green shRNA targeting protein A); 2) shRNA representation is unaltered, suggesting that growth rate was unaffected by the protein depletion, i.e. "neutral" (yellow shRNA targeting protein B); 3) shRNA is depleted in t15 versus t0, suggesting this protein deficiency causes sensitivity to G4-stabilising ligands (purple shRNA, protein C). Scenario 3 has been evaluated from a synthetic lethality perspective in chapter 2. Within this chapter resistance is analysed by investigating the genes and shRNAs that contribute to scenario 1.

5.2 Results

5.2.1 Defining a list of resistance genes from across the genome-wide and focused screens

For the A375 genome-wide screen, following ligand or DMSO treatment, 15,015 hairpins were differentially expressed across all three treatments (Figure 5.2A). From these, a gene knockdown was classified as causing significant growth advantage following ligand treatment, if 50 % or 3 hairpins had a positive $\log_2$FC with an FDR ≤ 0.05, in PDS or PhenDC3 but not in DMSO vehicle control (713 genes, Figure 5.2B). Of these, 510 exhibited a $\log_2$FC ≥ 1 (i.e. a minimum 2-fold increase) and were denoted G4 'resistance' genes (Figure 5.2C).

Of the genome-wide resistance genes, 464 (91 %) were included in the custom focused pool alongside the 751 sensitiser genes (discussed in chapter 2). The remaining 46 genes were excluded as they were only targeted by one hairpin (see Methods for further details). As mentioned previously, this focused screen was applied to the HT1080 and A375 cell lines and significantly upregulated shRNAs (Figure 5.2D&G) and genes (Figure 5.2E&H) classified using the genome-wide parameters. Further applying $\log_2$FC ≥ 1 threshold revealed that 207 (A375; Figure 5.2F) and 165 (HT1080; Figure 5.2G) gene knockdowns caused ligand specific resistance. Giving confidence that the majority of identified resistance genes are not false positives, only a small number of the genome-wide resistance genes were uncovered in the DMSO control during focused screening (A375 23 genes; 4.9 % and HT1080 43 genes; 9.3 %) (Figure 5.2E&H). Also considered within

our analysis of G4-ligand induced resistance, were genes uncovered with the SA-100-128 focused screen (Figure 5.2J-N, see also chapter 3). For this, 20 ligand-specific upregulated genes were observed in A375 despite the lack of growth rate difference between DMSO and SA-100-128 treated samples (Figure 5.2J-L). For HT1080, no resistance genes were identified, perhaps suggesting that the SA-100-128 concentration was too low in an already restricted screen to induce resistance mechanisms (Figure 5.2M&N). An alternative biological explanation could be that there are no resistance genes in the HT1080 cell line, however this warrants further investigation.

A **A375 Genome wide screen**
Significant shRNAs, FDR < 0.05
PDS (6865)   PhenDC3 (8941)
2561 (17%), 1357 (9%), 5591 (37%), 1512 (10%), 1435 (9.6%), 481 (3.2%), 2078 (14%), DMSO (5506)

B **A375 Genome wide screen**
upregulated genes
PDS (447)   PhenDC3 (537)
235 (24%), 67 (6.7%), 411 (41%), 36 (3.6%), 109 (11%), 23 (2.3%), 116 (12%), DMSO (284)

C **A375 Genome wide screen**
G4 resistance genes
PDS (209)   PhenDC3 (338)
172 (33%), 37 (7.3%), 301 (59%), 510 GW resistance genes

D **A375 focused screen**
Significant shRNAs, FDR < 0.05
PDS (1526)   PhenDC3 (1637)
541 (24%), 838 (37%), 636 (28%), 125 (5.6%), 22 (0.98%), 38 (1.7%), 36 (1.6%), DMSO (221)

E **A375 focused screen**
upregulated genes
PDS (225)   PhenDC3 (253)
99 (27%), 107 (29%), 135 (37%), 9 (2.5%), 10 (2.7%), 2 (0.55%), 2 (0.55%), DMSO (23)

F **A375 focused screen**
G4 resistance genes
PDS (89)   PhenDC3 (154)
53 (25%), 36 (17%), 118 (57%), 207 A375 resistance genes

G **HT1080 focused screen**
Significant shRNAs, FDR < 0.05
PDS (881)   PhenDC3 (544)
540 (44%), 112 (9.2%), 186 (15%), 157 (13%), 22 (0.98%), 89 (7.3%), 66 (5.4%), DMSO (384)

H **HT1080 focused screen**
upregulated genes
PDS (185)   PhenDC3 (52)
142 (68%), 12 (5.8%), 11 (5.3%), 24 (12%), 7 (3.4%), 5 (2.4%), 7 (3.4%), DMSO (43)

I **HT1080 focused screen**
HT1080 G4 resistance genes
PDS (154)   PhenDC3 (23)
142 (25%), 12 (17%), 11 (57%), 165 HT1080 resistance genes

J **A375 SA-100-128 screen**
Significant shRNAs, FDR < 0.05
SA-100-128 (547)   DMSO (221)
349 (45.4 %), 198 (25.8%), 23 (0.03%)

K **A375 SA-100-128 screen**
upregulated genes
SA-100-128 (33)   DMSO (23)
20 (47 %), 13 (30%), 10 (23%)

L

gene	hairpins detected in t0	significant shRNAs	median logFC
HMGB1	3	2	1.31
AFAP1	4	2	1.22
BEND3	5	3	1.07
SPTAN1	5	3	1.06
ZBTB7A	7	3	1.00
MYL1	4	2	0.99
HOXA11	5	3	0.93
BRP44L	5	3	0.91
GRIPAP1	7	3	0.88
SPACA5	6	3	0.87
ZNF268	4	2	0.85
GPR114	7	3	0.83
MUC5B	3	2	0.80
AGPAT3	7	3	0.80
SPINT1	6	3	0.80
ZNF562	7	3	0.80
C2orf51	2	1	0.77
PRDM2	7	3	0.76
PPCDC	7	4	0.70
DHX29	6	3	0.65

M **HT1080 SA-100-128 screen**
Significant shRNAs, FDR < 0.05
SA-100-128 (204)   DMSO (384)
75 (16.3%), 129 (28.1%), 255 (55.6%)

N **HT1080 SA-100-128 screen**
upregulated genes
SA-100-128 (12)   DMSO (43)
12 (28%), 31 (72%)

Figure 5.2. Gene sets that when knocked down cause resistance to G4-stabilising ligand treatment revealed by genome-wide and focused shRNA screening (see overleaf for figure legend description)

(A-C) Venn diagrams from the genome-wide A375 screen showing (A) significantly differentially expressed shRNAs (FDR <0.05), (B) significantly upregulated genes and (C) ligand specific resistance genes.

(D-F) As (A-C) for focused screening in A375 cells; (G-I) As (D-F) for HT1080 cells. (J-L) SA-100-128 A375 focused screen (J) significant differentially expressed shRNAs, (K) significant upregulated genes, (L) table showing the 20 ligand specific upregulated genes, with their associated significant shRNAs and median log FC. (M-N) SA-100-128 HT1080 focused screen (M) significant differentially expressed shRNAs, (N) significant upregulated genes.

The 510 resistance genes identified in the genome-wide screen was used to globally investigate gene deficiencies associated with G4-ligand resistance (section 5.2.2). Additionally, the resistance genes outlined above for PDS and PhenDC3 treatment in the focused screens were overlapped with the parent screen to allow 1) PDS and PhenDC3 resistance comparisons and 2) identify consistent, resistance genes common to both cell lines (Figure 5.3). For A375, this revealed 104 and 44 reproducible PhenDC3 and PDS hits (145 total, Figure 5.3A). Comparing the genome-wide and HT1080 focused screens revealed 6 and 41 PhenDC3 and PDS resistance genes respectively (45 total, Figure 5.3B). Also compared were focused screen specific resistance genes (Figure 5.3C) to reveal five PhenDC3 and 30 PDS common hits. The identified genes in the different screens and with different ligands (outlined in Figure 5.2 and 5.3) will now be evaluated.

A A375 PDS and PhenDC3 resistance hits

B HT1080 PDS and PhenDC3 resistance hits

C Focussed screen specific PDS and PhenDC3 resistance hits
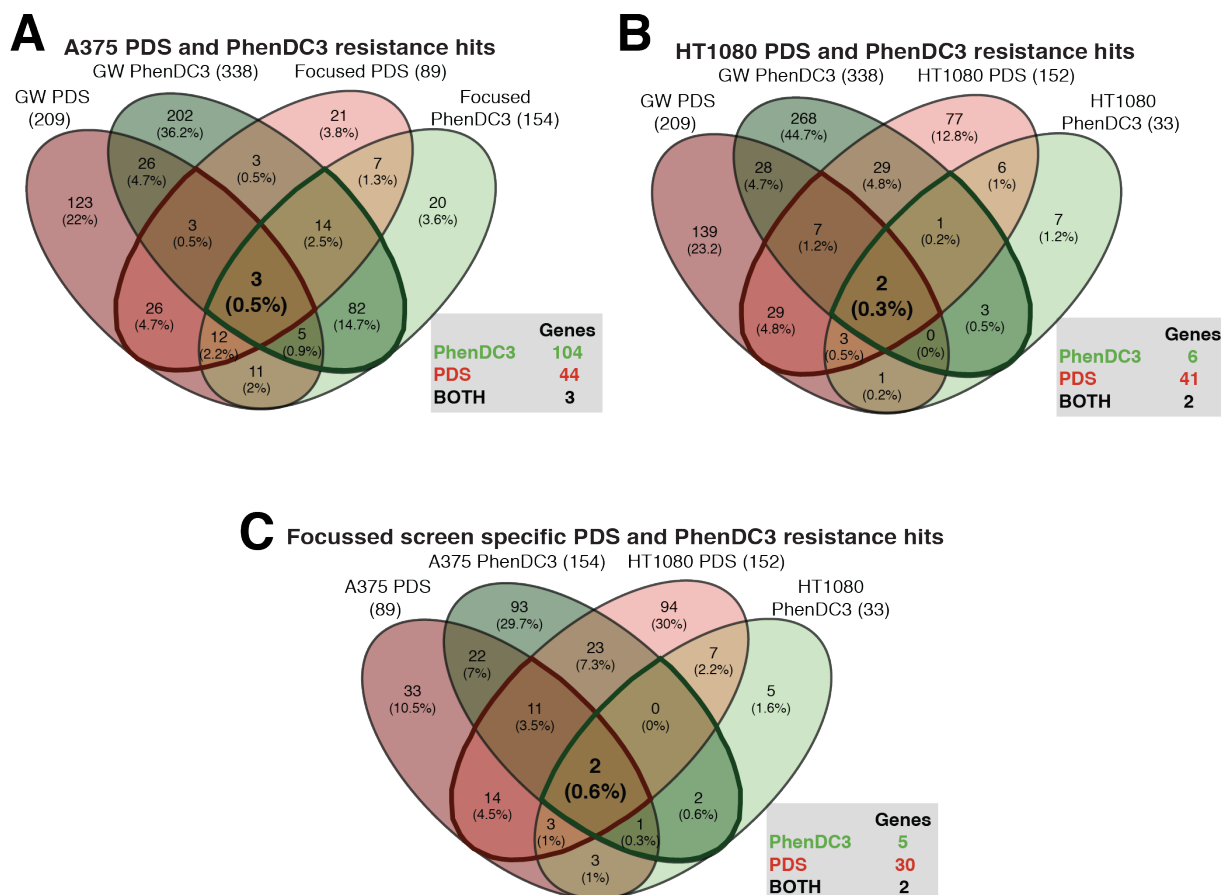
Figure 5.3. Focused screening reveals high confidence resistance hits
Venn diagrams showing the overlap of PDS and PhenDC3 resistance genes for (A) genome-wide and A375 focused screens, (B) genome-wide and HT1080 focused screens, (C) A375 and HT1080 focused screens.

5.2.2 Resistance genes uncovered by genome-wide screening provides a global view of G4-ligand resistance

First, to gain unbiased and systematic insights into possible resistance mechanisms to G4-ligand treatment, genome-wide resistance genes (i.e. gene deficiencies that caused resistance to either PDS and/or PhenDC3 in the genome-wide screen) were analysed. Inter-gene interactions were evaluated via STRING, performed with experimental evidence and coexpression data parameters and a high confidence threshold (0.7, Figure 5.4). This revealed a transcription factor cluster, including core transcriptional machinery components alongside tumour suppressors (e.g. *TP53, RB1* and

TP53IL) and a suite of interacting genes (Figure 5.4A). Smaller clusters of vacuolar protein sorting genes (*VPS53, VPS52* and *VPS54*) and lysosomal ATPase components (*ATP6V1C1, ATP6V0D1, ATP6V0D2, ATP6V1E1* and *ATP6V1G1*) were also seen (Figure 5.4C). For RB1, TP53 and TP53IL1 (an early transcriptional TP53 target) shRNA-enrichment following ligand treatment was high (Figure 5.4D). For example, RB1 shRNA-induced knockdown resulted in a log$_2$FC of 3.9 in PDS and 3.5 in PhenDC3 (~14-fold increase). This suggests that these genes are key in the cellular response to G4-ligand treatment, which may arise from an increase in G4-structure stabilisation.
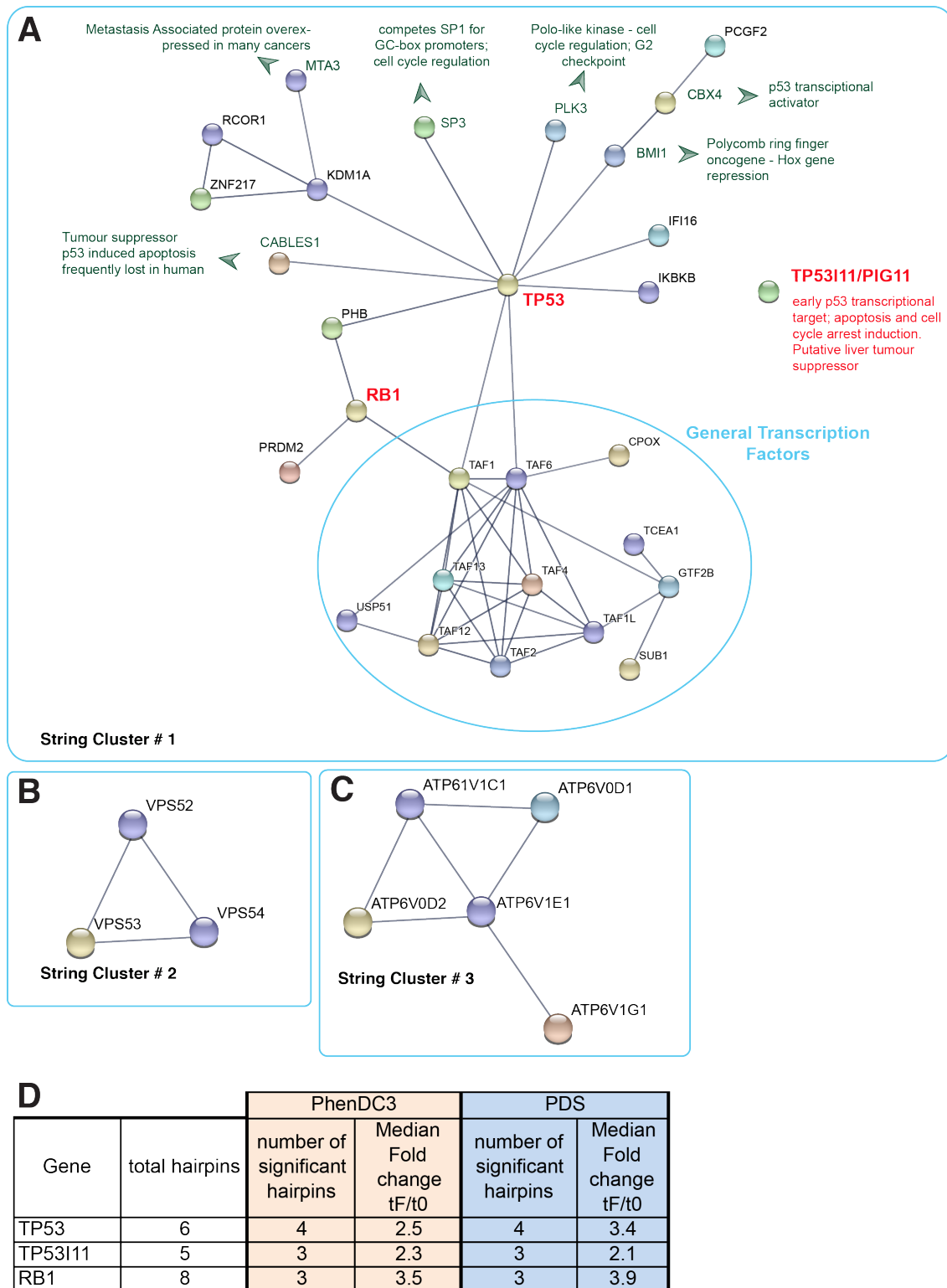
Figure 5.4. Genome-wide resistance genes show a transcription related interacting network including cancer related genes

(A-C) STRING (v2) network analysis using experimental evidence and coexpression data and a high confidence interaction threshold (0.7) of the 510 genome-wide resistance hits (A) cluster 1, largest cluster of transcription factors and associated proteins with the cancer association highlighted for selected genes, (B) cluster 2 of vacuolar proteins, (C) cluster 3 of ATPase components, (D) table showing the number of significant hairpins and associated $\log_2 FC$ for the tumour suppressors within cluster 1: TP53, TP53I11 and RB1.

195

To investigate the pathways involved in resistance to G4-ligands further, Cytoscape was used to analyse enriched KEGG and GO pathways for genome-wide hits (Figure 5.5). These can be broadly separated into two categories, 1) genes involved in drug uptake and processing (energy, transport, secretion, lysosome) and 2) nucleic-acid related genes involved in transcription and covalent modification. For genes identified via Cytoscape, enriched GO and KEGG terms are shown below in Figure 5.6.
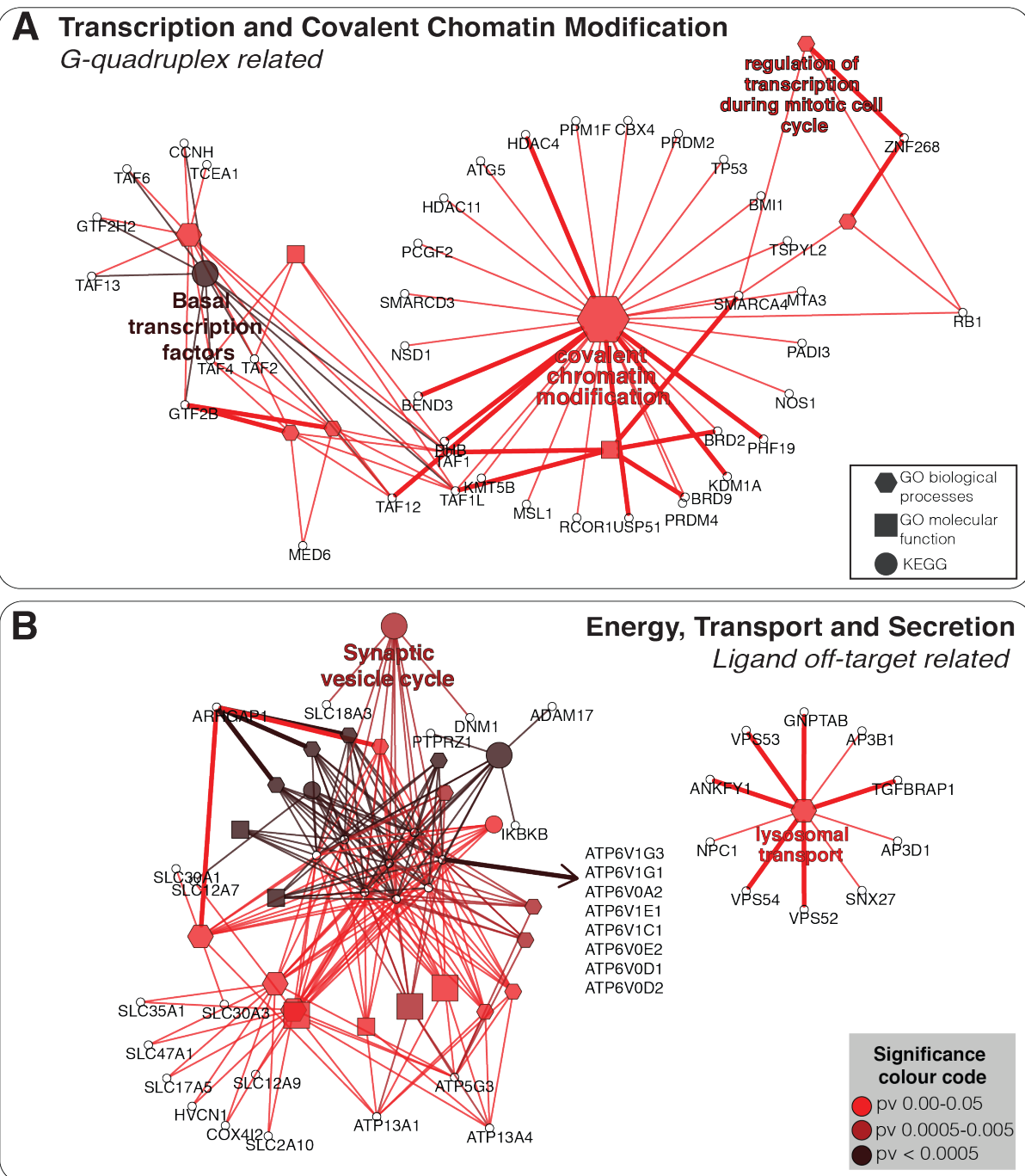
Figure 5.5. Genome-wide resistance genes are enriched in transcription, chromatin modification and lysosome associated terms
Cytoscape analysis for enriched KEGG and GO terms for the 510 genome-wide resistance genes showed two main categories of genes: (A) transcription and chromatin modification and (B) energy transport and secretion. Significance was determined using Bonferroni adjustment (p < 0.05) and p-value correction for multiple hypothesis testing. Node shape indicates database and colour represents significance (as outlined in the figure). Node size reflects the number of genes contributing to the term.
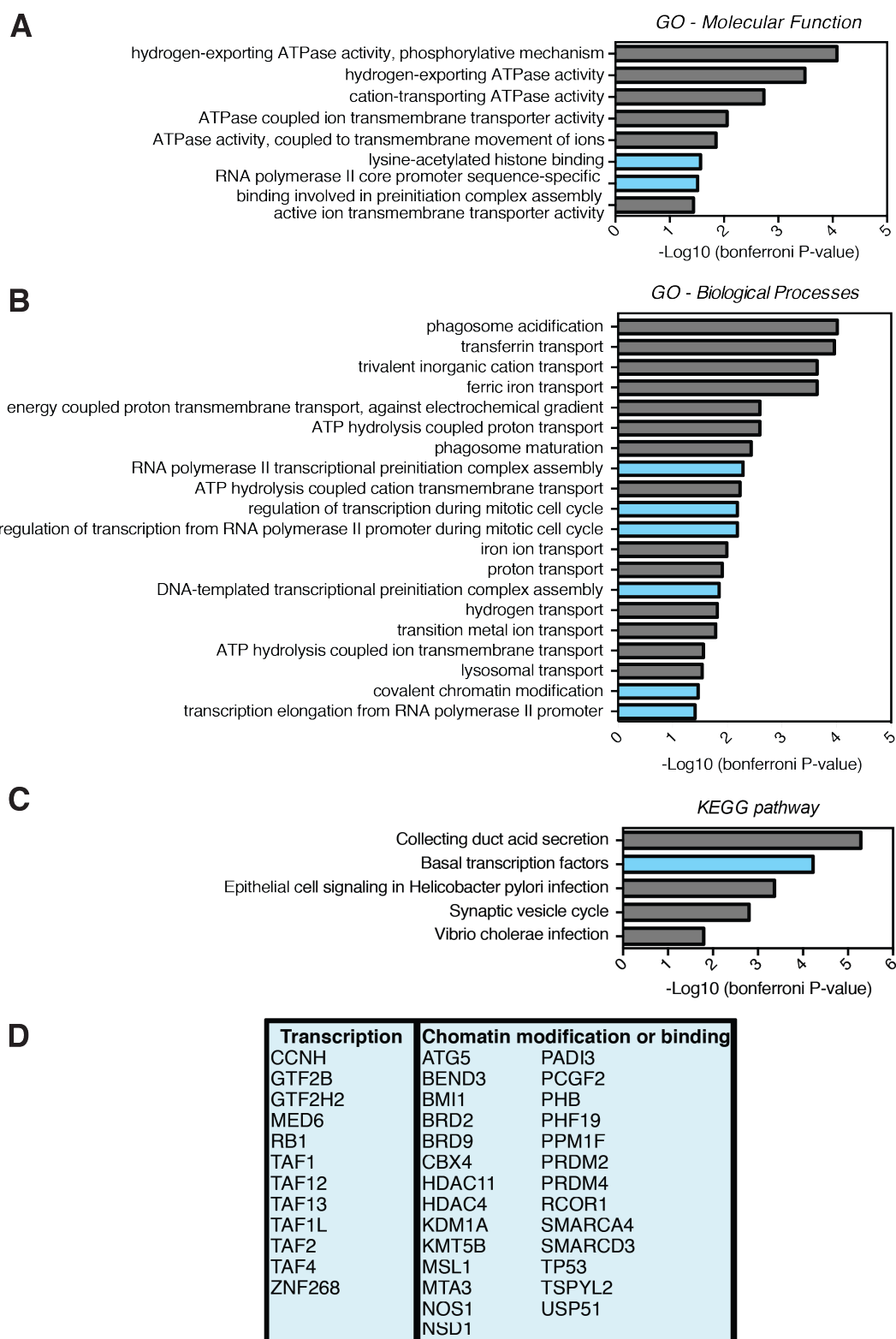
Figure 5.6 Genome-wide resistance genes show enrichment of general transcription factors and chromatin remodellers

(A-C) Enriched terms ordered by −log10 P-value (most to least significant); significance determine p < 0.05 for (A) GO Molecular Function terms, (B) GO Biological Processes and (C) KEGG pathway. Blue indicates transcription/chromatin modification related terms. (D) Table showing the genes enriched in chromatin and transcription associated terms for the genome-wide sensitisers.

5.2.3 Identifying gene deficiencies resulting in specific ligand resistance

To gain insights into ligand resistance effects specific to each molecule focused screen resistance genes were next explored. For the 145 resistance genes overlapping the genome-wide and A375 focused screens, chromatin/ histone associated gene enrichment in addition to lysosome/ligand-related terms were again observed (Figure 5.7A-E). Of the 39 genes that contributed the transcription and chromatin enrichment within the genome-wide screen, 15 reemerged in the focused screen (Figure 5.7E&F).

Specifically focusing on the 104 recurring PhenDC3 resistance genes, similar KEGG and GO enrichment terms are revealed (Figure 5.7G-I). These include several PhenDC3-specific transcription associated genes such as HOXA1, HOXA11, HOXB4 and SP3, a transcription factor that regulates DNMT1 expression (Kishikawa *et al*, 2002), a protein recently investigated in our lab to have DNA-G4 affinity (S. Mao *et al.* in press 2018) (Figure 5.7J). For PDS, there were too few resistance genes to analyse enrichment (44 genes causing PDS resistance in both the A375 genome-wide and focused screen). However, within this list, were also chromatin and transcription associated genes. The former comprises enzymes involved in post-translational modification of histones including ubiquitnation (*BMI-1*) and lysine methylation (*SUV420H1*). For the latter, deficiencies in the oncogenic leucine zipper, *MAF1* and three zinc finger transcription factors, *PHF6, ZNF268* and *ZNF217* were uncovered as inducing PDS resistance. Only one of these deficiencies, ZNF217, was also found to cause PhenDC3 resistance (see below).
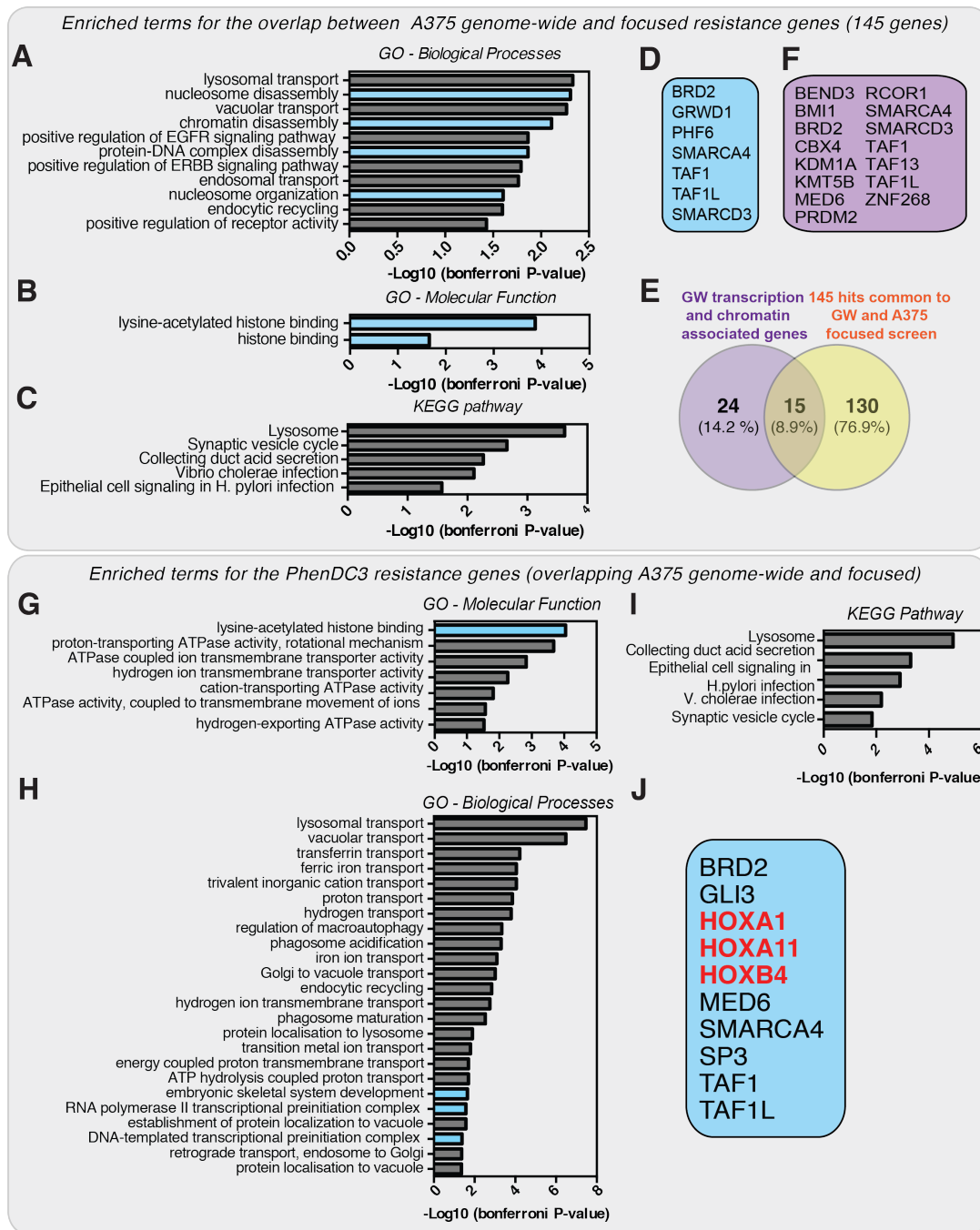
Figure 5.7. Focused screening in A375 cells reflects similar resistance gene enrichment as seen by genome-wide screening

(A-C) Enriched terms ordered by −log10 P-value (most to least significant) for the 145 resistance genes replicated by A375 focused screening; significance determine p < 0.05 for (A) Gene Ontology Molecular Function terms, (B) Gene Ontology Biological Processes and (C) KEGG pathway. In blue are transcription/chromatin related terms. (D) genes that contribute to the enrichment of chromatin and transcription associated terms for the 145 A375 focused screen sensitisers.(E) Venn diagram showing the overlap between the resistance genes from the A375 focused screen and the 39 genes that contributed to chromatin and transcription terms within the genome-wide resistance genes and (F) the 15 genes in common. (G-I) Enriched (G) GO molecular function, (H) GO biological process and (I) KEGG terms for PhenDC3 resistance genes with transcription and chromatin related terms shown in blue which are listed in (J) with HOX genes shown in red.

Next, the HT1080 focused screen was used to identify resistance genes common to both cell lines. Overlapping resistance genes, regardless of ligand (i.e. PDS or PhenDC3), across all screens (Figure 5.8A) gave 40 genes that when knocked down caused consistent resistance to ligand treatment. Several databases were used to categorise these according to their ligand-related (transport, energy and lysosome), nucleic acid related and "other" roles (Figure 5.8B).
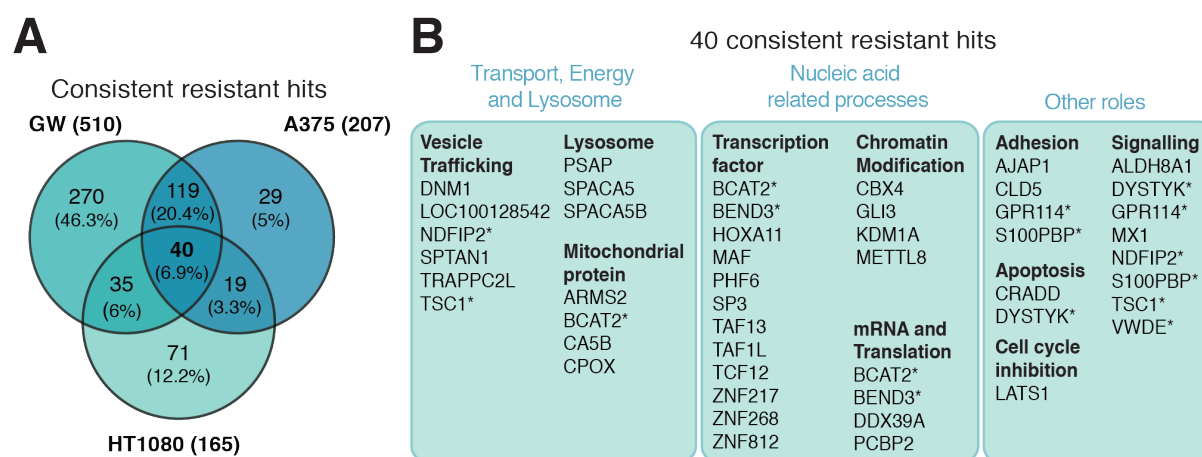
Figure 5.8. 40 high confidence resistance hits identified across all screens
(A) Venn diagram showing the overlap of the resistance genes (FDR ≤ 0.05, 50 % or 3 hairpins $log_2FC > 1$) from the genome-wide screen and the focused screens in A375 and HT1080.
(B) DAVID, STRING (experimental data, co-expression, medium confidence (≥ 0.4)) interaction and UniprotKB data was used to categorise the common resistance hits. These were broadly d into "Transport, Energy and Lysosome", "Nucleic acid related" and "Other" roles

Then, PDS and PhenDC3 hits common across all screens were considered (Figure 5.9A&B). This gave 15 consistent PDS (Figure 5.9A) and four consistent PhenDC3 (Figure 5.9B) hits. Only one gene deficiency, *ZNF217*, was found to be independent of screen and ligand treatment (Figure 5.9C). For the focused screens in isolation, a lack of DHX29 was additionally identified to cause resistance (Figure 5.9D).
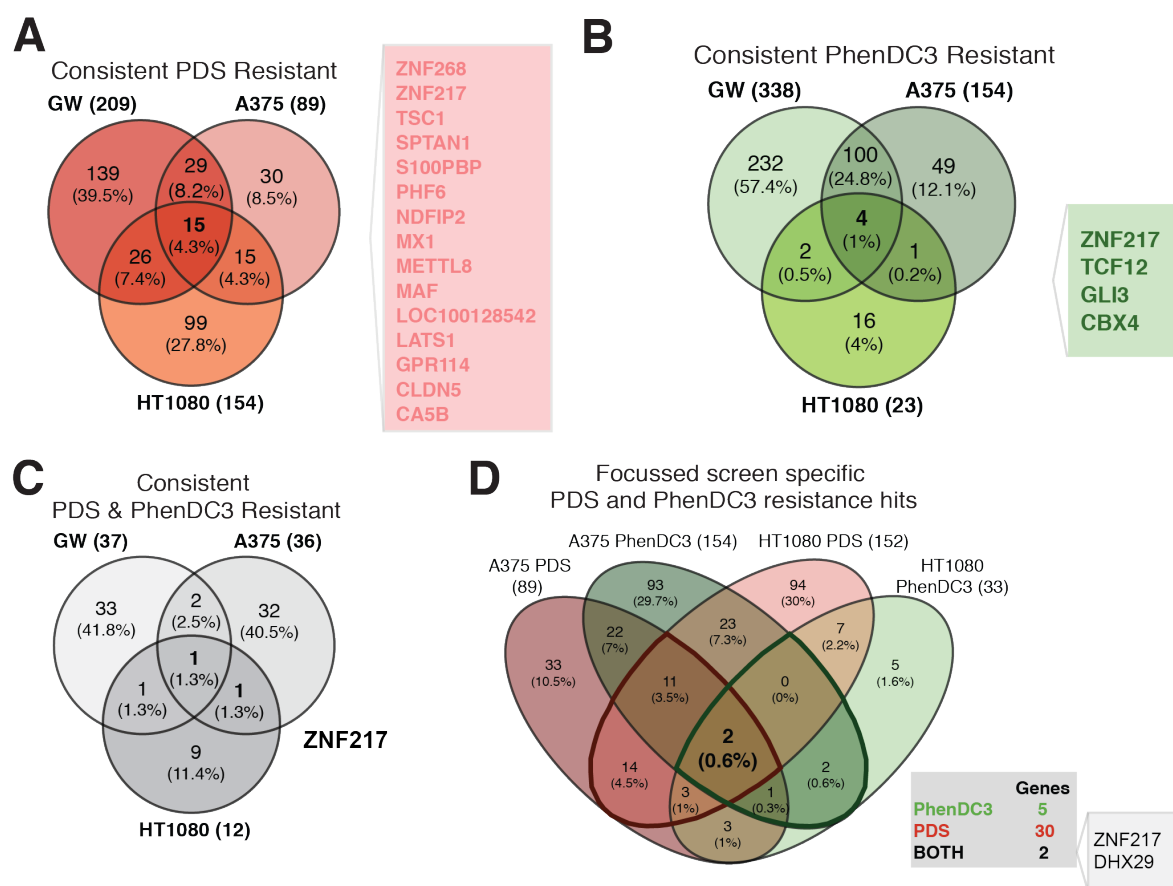
Figure 5.9. Focused screening reveals consistent PDS and PhenDC3 resistance hits
Venn diagrams showing the overlap of resistance genes (FDR <0.05, 50 % or 3 hairpins, log$_2$FC > 1) for (A) PDS for all three screens, (B) PhenDC3 across all three screens, (C) PDS and PhenDC3 across all three screens, (D) PDS and PhenDC3 in HT1080 and A375 focused screen

Finally, the 20 genes that caused a growth advantage following treatment with the top "drug-like" PDS derivative, SA-100-128 (Figure 5.2 K and L) were overlapped with PDS and PhenDC3 resistance genes to investigate commonalities across all screens and revealed six genes, including the transcription factors ZNF268 and HOXA11 (Figure 5.10). Considering only genes uncovered via focused screening, revealed that DHX29 deficiency also provided a growth advantage for A375 cells in the presence of SA-100-128, as was previously seen for PDS and PhenDC3 treatment of A375 and HT1080 cells.
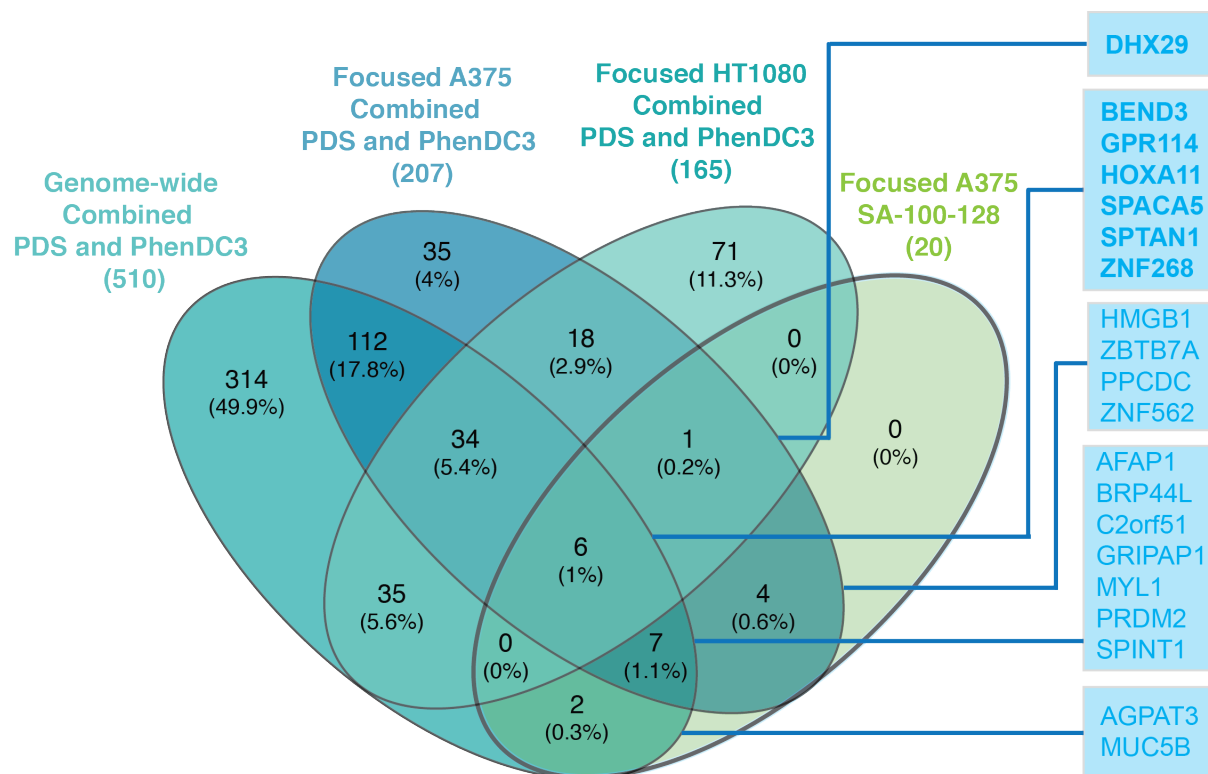
Figure 5.10. SA-100-128 focused screen reveals common resistance genes with PDS and PhenDC3

Venn diagram overlapping the PDS and PhenDC3 resistance genes (FDR ≤ 0.05, 50 % or 3 hairpins, $log_2FC > 1$) from the three screens: A375 PDS and PhenDC3 genome-wide; A375 focused screen PDS and PhenDC3; HT1080 focused screen PDS and PhenDC3, with the upregulated SA-100-128 genes uncovered in the A375 focused screen (FDR ≤ 0.05, 50 % or 3 genes, $log_2FC > 0$)

5.2.4 Validation of resistance genes via a short-term siRNA approach

As for the four key sensitisers (chapter 3: *BRCA1, TOP1, GAR1* and *DDX42*), resistance to PDS and PhenDC3 was independently verified by a short-term siRNA approach. For this, four genes were selected based on their consistent identification in multiple screens with different G4-stabilising ligands and/or enrichment in pathway analyses. The first two genes (see section 5.2.3) chosen were the predicted zinc finger ZNF217 (Lee *et al*, 2016) and RNA-helicase DHX29 (Parsyan *et al*, 2009). The third gene selected was the general transcription factor TAF1 (Wassarman & Sauer, 2001), given the general enrichment of transcription-associated genes and the consistent resistance to PhenDC3 in all shRNA screens upon depletion of this gene. Finally, DDX39A, a DEAD box RNA helicase (Sugiura *et al*, 2007b) was the fourth gene selected due to its functional similarity to DHX29 and to make comparisons to DDX42 which, despite belonging to the same helicase family was validated as a G4-ligand sensitiser (see chapter 3). The significant hairpins and associated median $\log_2$FC values for these four selected genes are summarised below for all shRNA screens (Figure 5.11).

A

A375 GW screen

Gene	Hairpins in t0	Enriched shRNAs (PDS)	Median LogFC (PDS)		Enriched shRNAs (PhenDC3)	Median LogFC (PhenDC3)	
ZNF217	6	3	2.13	✓	3	4.31	✓
DHX29	5	3	0.98	✗	2	2.56	✗
TAF1	15	1	1.69	✗	10	2.73	✓
DDX39A	5	3	1.19	✓	2	2.56	✗

B

A375 Focused screen

Gene	Hairpins in t0	Enriched shRNAs (PDS)	Median LogFC (PDS)		Enriched shRNAs (PhenDC3)	Median LogFC (PhenDC3)		Enriched shRNAs (SA-100-128)	Median LogFC (SA-100-128)	
ZNF217	7	6	1.02	✓	7	1.87	✓	2	0.70	✗
DHX29	6	4	1.21	✓	4	1.06	✓	3	0.65	✓
TAF1	5	0	NA	✗	5	1.62	✓	0	NA	
DDX39A	6	2	1.46	✗	4	1.35	✓	2	0.81	✗

C

HT1080 Focused screen

Gene	Hairpins in t0	Enriched shRNAs (PDS)	Median LogFC (PDS)		Enriched shRNAs (PhenDC3)	Median LogFC (PhenDC3)		Enriched shRNAs (SA-100-128)	Median LogFC (SA-100-128)	
ZNF217	7	4	1.91	✓	3	1.16	✓	0	NA	✗
DHX29	6	4	2.23	✓	4	2.21	✓	1	1.91	✗
TAF1	5	2	2.34	✗	0	NA	✗	0	NA	✗
DDX39A	6	3	1.62	✓	2	1.68	✗	0	NA	✗

Figure 5.11. Four resistance genes identified by shRNA screening chosen for siRNA validation
(A-C) tables summarising the $\log_2 FC$ and significant hairpins (FDR ≤ 0.05) for ZNF217, DHX29, TAF1 and DDX39A for (A) genome-wide A375 shRNA screen, (B) A375 focused shRNA screen and (C) HT1080 focused screen. ✓ = passed, ✗ = did not pass the 50 % or 3 significant hairpin threshold, median $\log_2 FC$ > 1 criteria for a gene knockdown considered to cause PDS and/or PhenDC3 resistance.

A schematic of two possible outcomes expected for siRNA-induced resistance is depicted below (Figure 5.12):

1) siRNA-induced target protein depletion does not cause growth differences under DMSO conditions compared to the control non-targeting siRNA, but causes a growth advantage following G4-ligand treatment

2) Protein knockdown causes growth inhibition compared to the non-targeting siRNA in DMSO treated cells, but this inhibition phenotype is relieved by G4-ligand treatment

Resistance - Scenario (1)
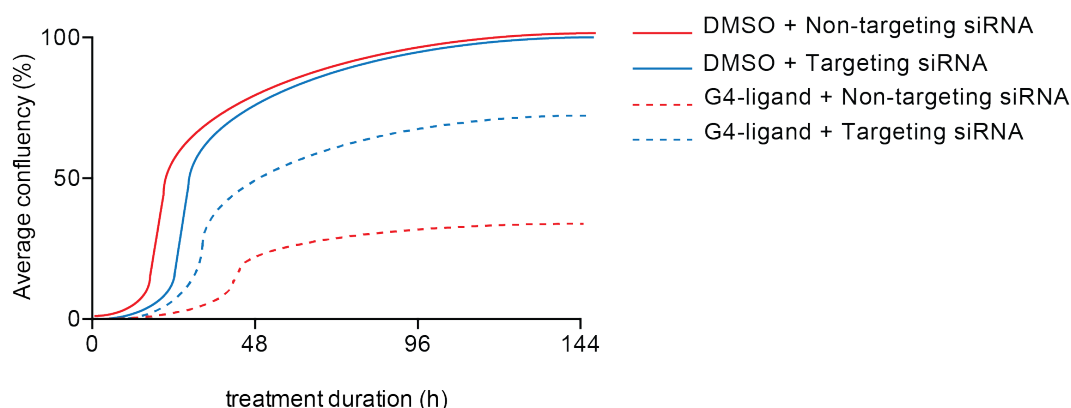
Resistance - Scenario (2)
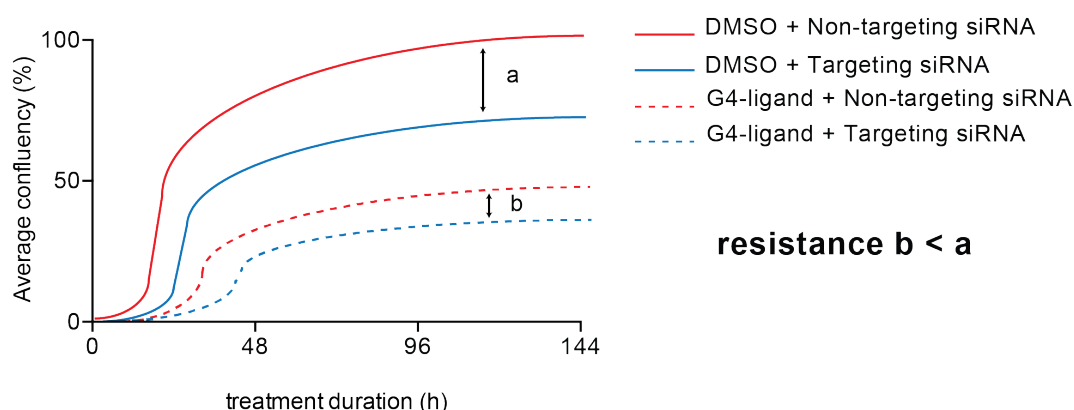
resistance b < a

Figure 5.12. Possible outcomes for a siRNA model for G4-ligand resistance genes
Scenario (1) knocking down the protein does not cause a growth defect compared to the non-targeting in DMSO treatment but causes a growth advantage following G4-ligand treatment. Scenario (2) Knocking down the protein causes a growth defect compared to non-targeting control in the DMSO condition, but this growth defect is reduced in the presence of G4-ligand treatment. Solid lines = DMSO treatment; Dotted lines = G4-ligand treatment; Red = non-targeting siRNA; Blue = Targeting siRNA.

HT1080 and A375 cells were transfected separately with siRNAs targeting *ZNF217, DHX29, DDX39A* and *TAF1* alongside non-targeting siRNA and non-transfected controls. Following 24 h, cells were treated with PDS, PhenDC3 or vehicle control for 144 h (A375: 5 µM or 10 µM PDS, 40 µM or 70 µM PhenDC3; HT1080: 0.25 µM or 0.5 µM PDS, 20 µM or 40 µM PhenDC3) (Figure 5.13 and 5.14, HT1080 and A375 respectively).
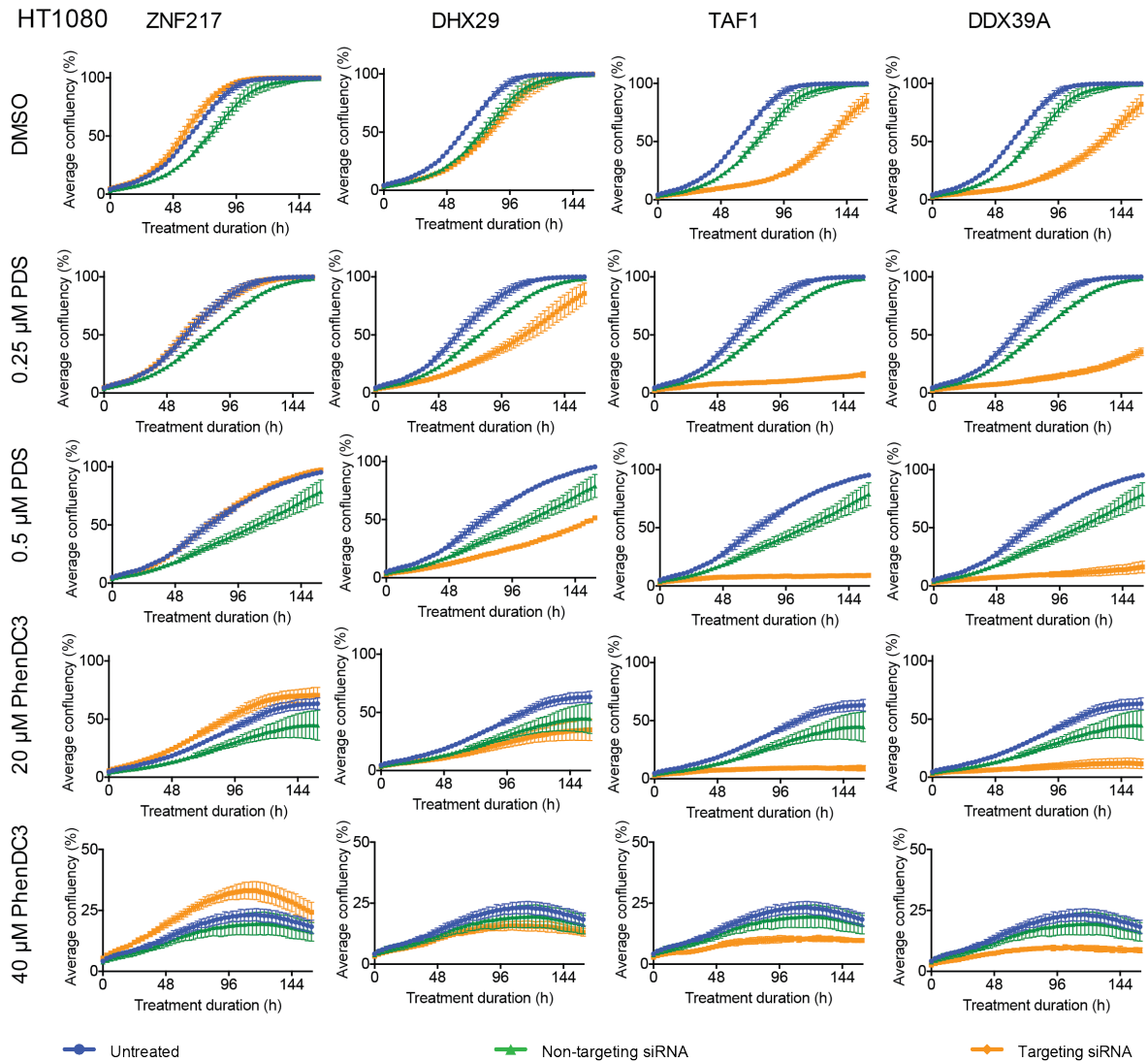
Figure 5.13. Short-term siRNA knockdowns in HT1080 of ZNF217, DHX29, TAF1 and DDX39A ahow altered growth profiles in the presence of PDS and PhenDC3

HT1080 cells were transfected with targeting siRNAs (orange) against ZNF217, DHX29, TAF1 or DDX39A for 24 h before treatment with PDS (0.25 µM and 0.5 µM), PhenDC3 (20 µM and 40 µM) or vehicle control (DMSO). For each knockdown, confluency over 144 h was monitored and plotted against confluency of non-transfected cells (blue) and cells transfected with a non-targeting control siRNA (green). Experiments were performed in triplicate and average confluency accumulation shown (mean ± standard deviation).
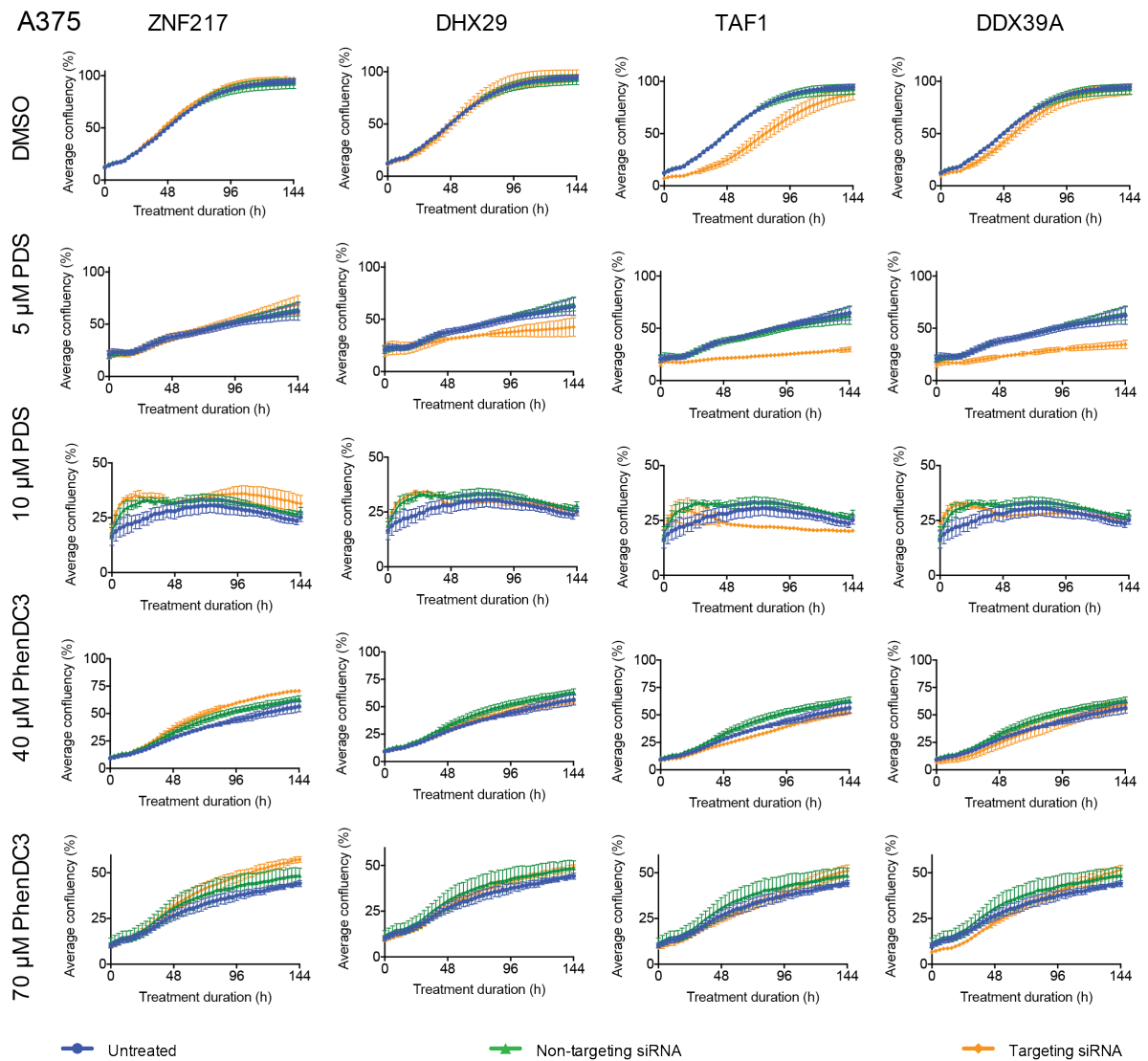
Figure 5.14. Short-term siRNA knockdowns in A375 of ZNF217, DHX29, TAF1 and DDX39A ahow altered growth profiles in the presence of PDS and PhenDC3
A375 cells were transfected with targeting siRNAs (orange) against ZNF217, DHX29, TAF1 or DDX39A for 24 h before treatment with PDS (5 µM and 10 µM), PhenDC3 (40 µM and 70 µM) or vehicle control (DMSO). For each knockdown, confluency over 144 h was monitored and plotted against confluency of non-transfected cells (blue) and cells transfected with a non-targeting control siRNA (green). Experiments were performed in triplicate and average confluency accumulation shown (mean ± standard deviation).

For reasons outlined in chapter 3, section 2.1, the non-targeting siRNA transfected cell growth curve (non-targeting control) was used as a reference, for each ligand/DMSO treatment condition. To confirm protein depletion, cell lysates were immunoblotted with appropriate antibodies 48 h post-transfection (Figure 5.15A and Figure 5.16A; HT1080 and A375 respectively). For both HT1080 and A375 (Figure 5.15B and 5.16B), protein depletion was evident for all targeting siRNAs (ranging from 30 – 65 % knockdown for A375, and 38 – 89 % knockdown for HT1080). Next, confluency differences compared to the non-targeting control for treated samples were investigated and analysed as in chapter 3 (Figure 5.15 and 5.16C-J). In summary ZNF217-knockdown provided resistance to PDS and PhenDC3 in HT1080, but only PhenDC3 in A375. For HT1080, this growth advantage only becomes apparent after 96 h. DHX29-deficiency did not replicate the ligand sensitivities shown in the screen, at the doses and treatment durations tested, showing a phenotype more associated with sensitisation. TAF1-deficiency also caused resistance to PhenDC3, and to a lower extent PDS, in both cell lines for the short-term siRNA experiment. Lastly, DDX39A knockdown caused both PDS and PhenDC3 resistance for HT1080, but not A375. Moreover, the phenotype seen for ZNF217-deficient cells reflects an increased growth rate in the presence of G4-ligands (scenario 1) whereas for TAF1 and DDX39A, the knockdown seems to rescue the G4-ligand growth inhibition phenotype (scenario 2).
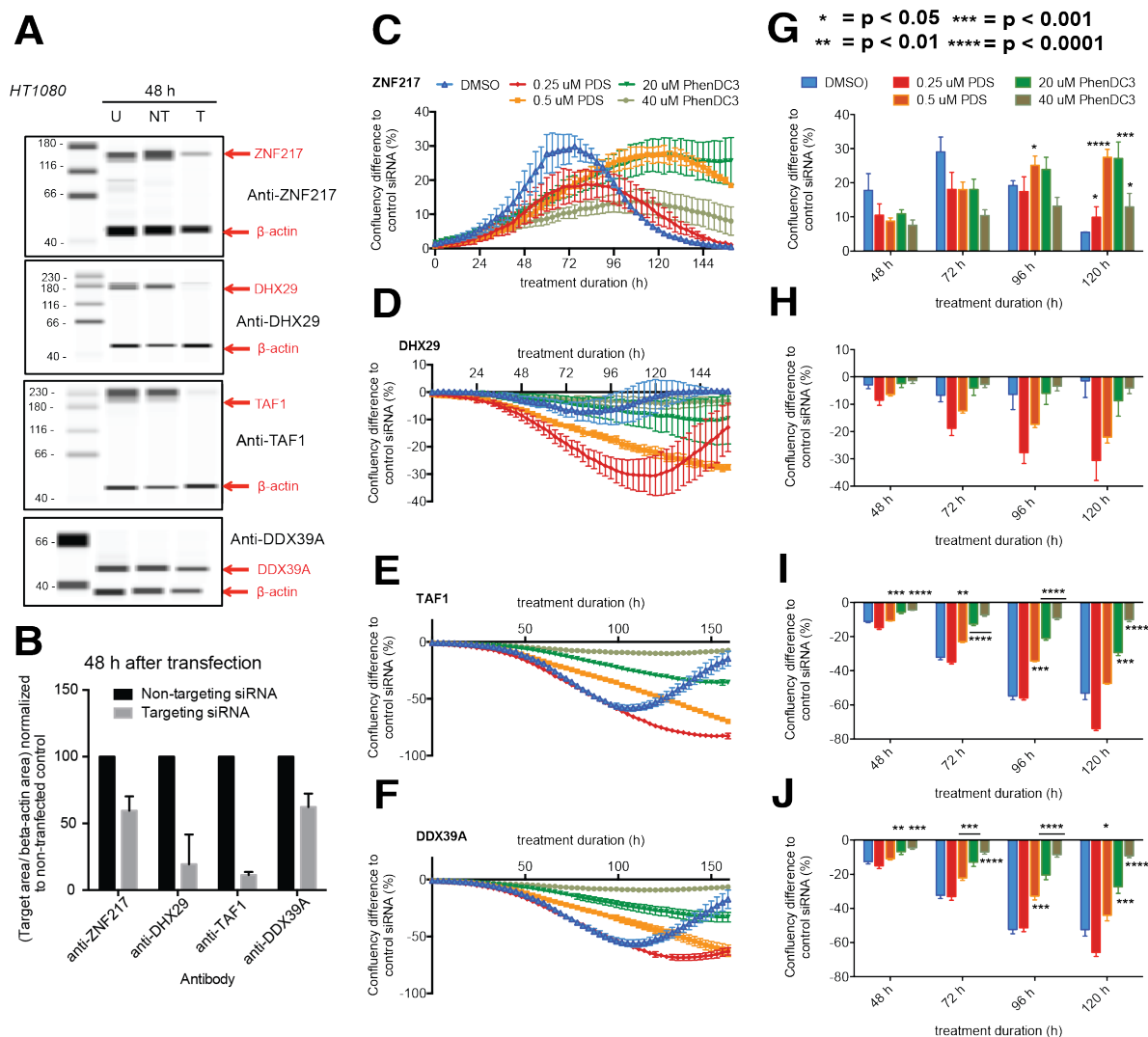
Figure 5.15. Short-term siRNA knockdowns to validate resistance genes identified by shRNA screening in HT1080

(A) Representative western blots showing lysates immunoblotted for ZNF217, DHX29, TAF1, DDX39A and beta-actin loading control for non-transfected cells (U) and cells transfected with targeting (T) and non-targeting control (NT) siRNA, 48 h after transfection

(B) ZNF217, DHX29, TAF1, DDX39A proteins levels in targeting lysates were normalized internally to the beta-actin loading controls and expressed as a percentage of the levels in the non-targeting control for two independent blots (mean ± standard deviation).

(C-J) HT1080 cells were transfected with the targeting siRNAs for 24 h before PDS, PhenDC3 or DMSO treatment. Confluency difference (mean ± standard deviation) compared to NT siRNA were plotted across two biological replicates for A) ZNF217, B) DHX29, C) TAF1, D) DDX39A. Confluency differences at 48, 72, 96 and 120 h were plotted for comparison for E) ZNF217, F) DHX29, G) TAF1, H) DDX39A. Significant confluency differences for G4-ligand versus DMSO treatment were determined using an unpaired parametric t-test, assuming equal standard deviation.
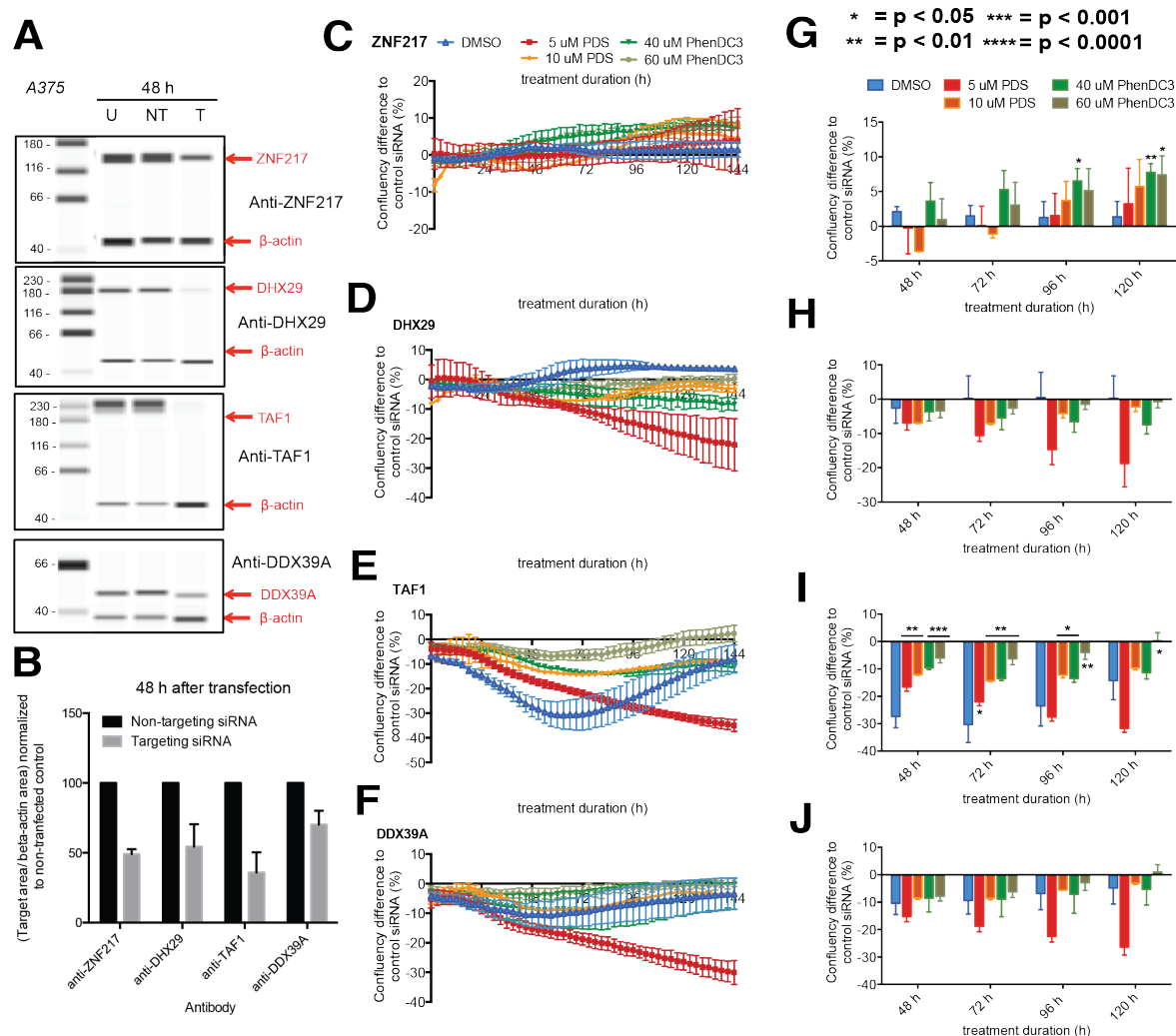
Figure 5.16. Short-term siRNA knockdowns to validate resistance genes identified by shRNA screening in A375

(A) Representative western blots showing lysates immunoblotted for ZNF217, DHX29, TAF1, DDX39A and beta-actin loading control for non-transfected cells (U) and cells transfected with targeting (T) and non-targeting control (NT) siRNA, 48 h after transfection

(B) ZNF217, DHX29, TAF1, DDX39A proteins levels in targeting lysates were normalized internally to the beta-actin loading controls and expressed as a percentage of the levels in the non-targeting control for two independent blots (mean ± standard deviation).

(C-J) A375 cells were transfected with the targeting siRNAs for 24 h before PDS, PhenDC3 or DMSO treatment. Confluency difference (mean ± standard deviation) compared to NT siRNA were plotted across two biological replicates for A) ZNF217, B) DHX29, C) TAF1, D) DDX39A. Confluency differences at 48, 72, 96 and 120 h were plotted for comparison for E) ZNF217, F) DHX29, G) TAF1, H) DDX39A. Significant confluency differences for G4-ligand versus DMSO treatment were determined using an unpaired parametric t-test, assuming equal standard deviation.

5.3 Discussion

5.3.1 General observations

The studies described in this chapter aimed to investigate gene depletions that cause resistance to G4-stabilising ligand treatment. Ligand resistance in cells may arise due to decreased uptake, greater extrusion and/or decreased target accessibility (reviewed in Gottesman, 2002). In this case, the target is the G4 structure. Other common mechanisms of resistance to chemotherapeutic drugs include mutation of the drug target, for example active site mutation, or upregulation of compensating pathways, to circumvent the pathway and/or protein that has been pharmacologically inhibited (Zheng, 2017). I propose that the resistance genes uncovered, and discussed in this chapter, fall into two main categories: 1) lysosome and energy related genes and 2) transcription and chromatin modification related genes. The latter may reflect a reduction in binding sites for PDS and PhenDC3, as transcriptional and chromatin compaction alteration is hypothesised to reduce the number of, and/or accessibility to, DNA-G4s and RNA-G4s. Conversely therefore, increased transcription and chromatin relaxation may facilitate ligand G4-binding, and increase sensitivity to these molecules.

In support of the transcription/chromatin-related hypothesis, the strand separation and torsional stress associated with transcription encourages G4 formation (reviewed in Kim, 2017) and treatment of SV40-transformed MRC-5 cells with the pol II transcriptional inhibitor DRB, prevents PDS-induced DNA damage (Rodriguez *et al*, 2012; Bensaude, 2011). Supporting a correlation between chromatin relaxation and a potential increase in G4-ligand binding

sites, treatment of HaCaT cells with the HDAC inhibitor entinostat expanded open chromatin sites, correlating with an increase of 4,000 G4 peaks, as visualised by G4 ChIP-seq (Hänsel-Hertsch *et al*, 2016).

Extending this hypothesis, both transcription and translation are increased in many cancers to support increased proliferation and biosynthesis (Martín-Martín *et al*, 2017; Bhat *et al*, 2015; Bywater *et al*, 2013). This could contribute to the increased G4s in several cancers compared normal tissues (Biffi *et al*, 2014b). One way of chemotherapeutically exploiting cancer-associated transcriptional upregulation in solid tumour xenograft models is through RNA polymerase (RNAP) inhibition by TAS-106 (Friday *et al*, 2012). Based on the results of this chapter, an alternative approach could be the use of G4-ligands to target the potential increase in G4-structures (this will be discussed in more detail in section 5.3.2).

The global lysosome term enrichment, particularly evident in the A375 genome-wide screen, suggests that lysosomal deficiencies provide a growth advantage to G4-stabilising ligand treatment. I hypothesise that these reflect off-target ligand effects and that, in lysosome-proficient cells, these organelles are involved in the cellular response to PDS and PhenDC3 treatment. In particular, they contribute to sensitivity unrelated to stabilisation of G4-structures, as indicated by the fact that depletion of a single lysosome component (e.g. individual components of the lysosomal ATPase) is sufficient to induce G4-ligand resistance.

This could be a result of 'lysosomal trapping' a phenomenon reported for other bulky, cationic lipophilic drugs (Kazmi *et al*, 2013). Given that G4-ligand targets are nuclear (DNA/RNA) and cytoplasmic (RNA), lysosomal accumulation would prevent PDS and PhenDC3 from reaching the desired binding sites. This may indicate a requirement to reduce lysosomal trapping in the future development of clinical G4-stabilising ligands. However, the best *in vitro* G4 stabilisers are inherently planar, aromatic and amine-rich, features known to exacerbate lysosome trapping for other molecules (Kazmi *et al*, 2013). Therefore weaker G4-stabilising ligands with reduced lysosome accumulation may be more clinically appropriate. To test this hypothesis, one could exploit the focused G4 shRNA pool (developed in chapter 2), to screen future clinical G4-ligands and investigate whether such molecules engage similar synthetic lethalities and chromatin/transcription-related resistance genes to PDS and PhenDC3 while reducing the emergence of lysosome related hits.

Deficiencies in several solute carrier components (SLC) proteins that perform facilitative and secondary transporter roles within the cell and golgi membranes (He *et al*, 2009) were also found to cause a G4 ligand resistance phenotype. Examples include three of nine SLC12 isoforms (A7, A10 and A9) that function as cation cotransporters responsible for of K+ and Cl- uptake. One interpretation of this resistance phenotype is that by reducing the intracellular concentration of K+, a cation that stabilises endogenous G4 structures, through the depletion of SLC, the ability for intracellular G4 to form might be reduced. As SLC12 transporter deficiencies have been linked to

peripheral neuropathy (Howard *et al*, 2002), it would be interesting to investigate whether reduced G4-structure formation contributes to this pathology. Other SLC component deficiencies identified as resistance genes may represent off-target effects and thus provide insights into improving the specificity of G4-stabilising ligands. For example SLC17A1 and SLC18A3 are vesicular glutamate and amine transporters respectively (Eiden, Schafer, Weihe, & Schatz, 2004; Reimer & Edwards, 2004) and SLC47A1 is involved in multidrug and toxin extrusion (Otsuka *et al*, 2005).

5.3.2 Hypothesised resistance mechanisms illustrated by specific gene-deficiency examples

Resistance to G4-ligand treatment induced by the deficiency in several core transcription components suggests that reduced transcription may decrease available G4-binding sites (section 5.3.1). Thus, the converse scenario, where transcription is increased may encourage G4-permissiveness and thus G4-stabilising ligand target accessibility. Among these resistance hits are gene-deficiencies in TATA-binding proteins associated factors (TAF): 1,2,4,6,12 and 13, required for pol II transcription (Furukawa & Tanese, 2000). These TAFs are overexpressed in and contribute to the malignancy of several cancers: TAF2 and TAF4 in high-grade serous ovarian cancers (Ribeiro *et al*, 2014); TAF12 in colorectal cancer (Voulgari *et al*, 2008) and acute myeloid leukaemia (Xu *et al*, 2018); TAF1 in lung and breast cancers (Wada *et al*, 1992), HPV-dependent cervical (Centeno, Ramirez-Salazar, Garcia-Villa, Gariglio, & Garrido, 2008) and uterine serous cancers (Zhao *et al*, 2013). Inhibiting TAFs has been explored for treatment of such malignancies. An

alternative may be G4-ligand usage to target a hypothesised transcription-associated G4 increase.

Within the genome-wide resistance investigations, specialised transcription factors were also identified, including the tumour suppressors *TP53* and *RB1*. TP53 protein can cause cellular apoptosis and senescence following DNA damage and/or oncogene overexpression (Schmitt *et al*, 2002) and is mutated in >50 % of human cancers (Ko & Prives, 1996). Given the enrichment of DNA damage genes in the PDS and PhenDC3 sensitiser list (Chapter 2), and the well-established hypothesis that ligand stabilised G4s can induce a DDR (see sections 1.4.2.2 and 1.6), it is perhaps unsurprising, but never before shown, that *TP53* depletion would provide a growth advantage in the presence of a G4-stabilising ligand.

Similarly, Rb1 is mutated/inactivated in the majority of cancers including paediatric retinoblastoma (Chinnam & Goodrich, 2011) and prevents cell proliferation at the S-phase mitotic checkpoint following genotoxic stress, including DNA-damage (Sherr, 2001). Here, Rb1 depletion provided G4-ligand resistance and similarly, imparts resistance to other DNA-damaging chemotherapies, such as cisplatin in breast, lung and prostate cancer cells (Sharma *et al*, 2007), suggesting comparable mechanisms whereby the Rb1 response to these molecules causes lethality.

Further specific transcription factor examples include three homeobox genes (HOXA1, HOXA11 and HOXB4). Deficiencies in these genes showed

consistent PhenDC3 resistance in the genome-wide and focused screens. HOXA11 deficiencies also caused PDS resistance in HT1080 cells and a growth advantage following SA-100-128 treatment of A375 cells. This phenotype may reflect a reduction of transcription causing decreased DNA-G4 ligand binding sites. Non-transcriptional HOX functions include promoting replication and translational regulation (Rezsohazy, 2014; Miotto & Graba, 2010). As mRNA G4s represent an abundant target and G4-ligand synthetic lethality is linked to replication-associated DNA damage, reducing both via HOX gene suppression may also explain the observed resistance.

Chromatin remodelling deficiencies also caused G4-ligand resistance, which may be linked with the observation that G4-structures are predominantly found in nucleosome-depleted chromatin (Hänsel-Hertsch *et al*, 2016). Consequently, in cells lacking genome decompaction machinery, G4s may be less abundant and/or inaccessible to PDS and PhenDC3. This is highlighted by BRD2- and BRD9- (Bromodomain 2 and 9) deficiencies causing G4-ligand resistance, proteins that bind transcriptionally active (i.e. G4-permissiveness) euchromatin, increasing chromatin decompaction by recruiting histone acetylases and remodellers (reviewed in Belkina & Denis, 2012). Depletion of these proteins removes this chromatin decompaction role whereas overexpression may increase open chromatin (reviewed in Josling, Selvarajah, Petter, & Duffy, 2012). For example, several leukaemias exhibit increased BRD2 activity and two BRD inhibitors JQ1 and I-BET, show promise in mouse xenograft BRD-driven malignancies (reviewed in Belkina & Denis, 2012). Similarly BRD9 depletion prevents mouse and human AML

proliferation (Hohmann *et al*, 2016). If the mechanism behind BRD-deficiency induced G4-ligand resistance stems from reduced G4 numbers and/or accessibility in closed chromatin, then BRD overexpression in these malignancies may increase G4-ligand sensitivity. Therefore targeting G4s rather than BRD proteins may provide therapeutic alternatives in these cancers.

5.3.3. Four frequent "resistance" genes uncovered via genetic screening for further exploration

Based on the outcome of the PDS and PhenDC3 screens, four resistance genes were chosen for siRNA validation: *ZNF217, TAF1, DHX29* and *DDX39A*. As discussed in chapter 2, siRNA is a related but independent approach to shRNA-induced knockdown. Both *ZN7217-* and *TAF1*-depletion reflected the resistance seen for the screens. *DDX39A*-deficiencies reflected the resistance phenotype shown by HT1080, but not A375. Intriguingly, *DHX29*-deficiency seemed to cause PDS and PhenDC3 sensitivity in both cell lines. The possible mechanisms of the G4-ligand phenotypes following depletion of the four proteins are discussed below, alongside putative areas for their therapeutic exploitation.

While TAF1 represents a core transcription factor, ZNF217 is a promoter-specific Kruppel-like transcription factor (Lee *et al*, 2016). TAF1 validation by siRNA supports the hypothesis that general transcription decrease can cause G4-ligand resistance, possibly by decreasing the G4 target formation and accessibility. Transcription deficiencies may reduce the number of 1) DNA

G4-structures at single-stranded transcription bubbles and/or 2) RNA-G4s, due to less cytoplasmic transcripts. Similarly to TAFs (section 5.3.2), ZNF217 is overexpressed in several cancers, correlating with poor prognosis and metastasis (Cohen *et al*, 2015; Quinlan *et al*, 2007; Plevova *et al*, 2010). ZNF217 activates oncogenes and pluripotency genes (Littlepage *et al*, 2012; Krig *et al*, 2007) but represses tumour suppressors (Quinlan *et al*, 2007; Banck *et al*, 2009; Thillainadesan *et al*, 2008). Bioinformatic analysis of promoters of these gene classes show enrichments and depletions of PQS respectively (Eddy & Maizels, 2006; Huppert & Balasubramanian, 2007). This may explain the G4-ligand resistance observed with ZNF217-depletion, as reducing oncogene expression and increasing tumour suppressor transcription would cause a net decrease in transcriptionally active, accessible promoter G4s. For ZNF217-overexpression, the opposite may occur perhaps suggesting such malignancies would be acutely sensitive to treatment with G4-stabilising ligands, though this requires further investigation.

DDX39A, an ATP-dependent DEAD box RNA helicase that unwinds dsRNA *in vitro* and increases HeLa cell proliferation when overexpressed (Sugiura *et al*, 2007a) was chosen due to homology with DDX42. Whilst DDX42 deficiencies resulted in G4-ligand sensitivity (chapter 3), DDX39A-depletion caused a growth advantage in the presence of PDS and PhenDC3. As other known G4-interactors were identified as sensitisers, including DHX36 and XRN1, this converse resistance phenotype suggests that DDX39A is unlikely to directly bind and/or unwind G4 structures. DDX39A also interacts with the telomeric G4-binder TRF2, with depletion causing telomere shortening and DDR

induction (Biffi, Tannahill, & Balasubramanian, 2012; Pedroso, Hayward, & Fletcher, 2009; Yoo & Chung, 2011). However, as G4-ligands can cause cell lethality by telomere deprotection, this is unlikely to be the G4-ligand resistance mechanism.

One explanation could arise from the interaction of DDX39A with CIP29, an mRNA export factor that regulates gene expression after DDR induction in *X. laevis* (Holden *et al*, 2017). This interaction increases DDX39A RNA-helicase activity *in vitro* (Sugiura *et al*, 2007b). Perhaps DDX39A-deficiency reduces detection of G4-induced DNA damage, providing a PDS and PhenDC3 resistance mechanism analogously to that described above for RB1- and TP53-deficiency. From a therapeutic perspective, DDX39A is upregulated in pancreatic, mesothelioma, bladder and gastrointestinal tumours (Kuramitsu *et al*, 2013a, 2013b; Kato *et al*, 2012; Kikuta *et al*, 2012), which may respond to G4-stabilising ligand treatment, based on the checkpoint mechanism hypothesised here. However, it remains to be explored if DDX39A overexpression causes sensitivity, a converse phenotype to what we have revealed by depletion.

Of the four proteins, only DHX29 did not replicate any resistance phenotypes seen via shRNA screening in the siRNA approach. This is surprising, as shRNA-induced DHX29 knockdown also caused a growth advantage following SA-100-128 treatment of A375 cells. This suggests a different mechanism for short-term (causing PDS and PhenDC3 sensitisation) versus long-term treatment. DHX29 is a processive mRNA helicase, necessary for

translation initiation for transcripts with structured 5'UTRs *in vitro* and *in vivo* (Parsyan *et al*, 2009; Pisareva *et al*, 2008; Pisareva & Pisarev, 2016). DHX29 deficiency impedes tumour growth in mouse xenografts (Parsyan *et al*, 2009), congruous with an increased translation requirement. Short-term sensitivity may be consistent with DHX29-deficiencies preventing 5' UTR G4s resolution, similar to that seen for DHX36 (Chen *et al*, 2015). Thus further RNA-G4 stabilisation caused by PDS or PhenDC3 treatment could cause cell lethality in the 6-day siRNA experiment. However, during long-term G4 ligand treatment, a translation and proliferation decrease may reduce DNA and RNA accessibility to PDS and PhenDC3, circumventing the initial sensitivity seen at earlier timepoints.

Overall, resistance gene analysis complements the synthetic lethality interactions uncovered in chapter 2, revealing transcription, both core and promoter specific, and chromatin remodeling as key determinants of G4 formation and accessibility by G4-stabilising ligands. Analogously to DDR-deficiencies representing synthetic lethal interactions with G4-stabilisation, here depletion of genes sensing such damage causes resistance. I therefore suggest three possible mechanisms for responding to stabilised G4s: i) helicases that resolve G4s (e.g. DHX36, RTEL1, DDX42); ii) genes that repair damage caused by targeting G4s (e.g. BRCA1, BRCA2, PALB2); iii) removing genes encoding senescent/lethality inducing checkpoints that detect G4-induced damage (e.g. TP53, RB1, DDX39A). Each of these could be exploited in a cancer-associated context as discussed above. Additionally, a novel candidate for an RNA 5' UTR G4 helicase was uncovered, DHX29,

which is reported to be cytoplasmic, rather than nuclear as for DDX42 and thus may act on a different set of RNA-G4s. Resistance analysis further indicated lysosome deficiencies as resistance mechanisms to PDS and PhenDC3; I propose that these molecules undergo lysosome trapping, highlighting an area that needs to be optimised in the generation of "drug-like" G4-stabilising ligands for clinical application.

Chapter 6

Overall conclusions and future objectives

6.1 General aim and objectives

In this thesis I have outlined a genetic screening approach and preliminary validation experiments to identify genes, that when knocked down via RNAi result in a difference in the growth (either sensitisation or resistance) of cells treated with G4-stabilising ligands (compared to DMSO). The overall aim was to understand the genes and pathways that interact with or mediate G4 formation and function (both RNA- and DNA-G4s) and thus explore the biological factors affecting G4-ligand sensitivity. As previously discussed in Section 1.9, the outcome of these investigations provide insight into the following three interlinked areas:

1) G4 Biology – Understanding the cellular response to stabilised G4s and, by inference the processes normally involving and regulating G4s

2) Putative clinical exploitation – Identifying possible genotypes acutely sensitive to G4-stabilising ligand treatment and increase our understanding of these ligands

3) Technical – Create and optimise a set of tools and resources for future testing of different cell lines, novel G4-stabilising ligands and for the community to establish new directions in G4-biology.

In addition to generally exploring proteins and pathways interacting with G4s and identifying G4-ligand sensitive genotypes, I set out to address several

areas within the field. Firstly, whether G4-structures perform functional roles or are merely problematic structures that require resolution. Secondly, recent evidence has suggested that RNA-G4s are globally unfolded in the cell (Guo & Bartel, 2016) despite multiple prior studies suggesting regulatory RNA-G4s roles (reviewed in Rhodes & Lipps, 2015), a dichotomy requiring resolution. More generally, available G4-stabilising ligands still require further exploration, including methods and/or features to improve their G4-specifity and medicinal chemistry properties. Additionally, a side-by-side comparison of the similarities and/or differences following application of G4-ligands to cells was lacking. A systematic comparison of the synthetic lethalities and resistance mechanisms for PDS and PhenDC3 was discussed in chapter 2 and 3, and revealed differences between the two molecules, going some way to address this area. As such this will not be further discussed. In this chapter, I will summarise how work within this thesis has partly explored these areas and also provided the foundations for future investigations.

6.2 Endogenous pathways involving G4-structures

Insights into the endogenous roles and regulation of RNA and DNA-G4s were provided by gene depletions causing either sensitisation or resistance to G4-stabilisation by small molecule ligands. The biological concepts identified within this thesis are summarised in Figure 6.1 and will be discussed below, providing a few illustrative examples for each (gene deficiencies imparting sensitivity in green; imparting resistance in purple). Further examples for each can be found within individual chapters. Generally, biological insights into G4s

are discussed as either complications associated with unresolved G4s or the normal, positive functions for these dynamic structures.

1) DNA replication

RECQL4
ATRX
DNA polymerase
DNA damage

RTEL1 WEE1 TP53
ORC1 CHEK1 TP53I11
PCNA CCND1 RB1
POLR1A ANAPC4 PHB
POLA2 ANAPC11
POLD2 ANKRD11 TOP1
POLQ USP37
RFC2 CUL1
RFC4 UBA3
AHCYL1
CCAR1

BRCA1
BRCA2 RNF20 DDX39A?
CHD8 USP1 TP53
POLQ SAE1 TP53I11
POLD2 UBA2 RB1
RPA1 AHCYL1
BAP SRCAP
PALB2

7) Ubiquitin non-proteolytic modification

UBA3
UBA2
SAE1
UBE2H
RBX1

proteosome

PSMC2
CUL1

8) Ubiquitin mediated proteolysis

2) Transcription

BRCA1 GAR1?
ADAR1 ATRX
SMARCA4
SMARCB1
SMARCE1
SRCAP?
CHD8

3) Chromatin remodelling/ histone modification

Histones

MLL4
WHSC1
ANKRD11
HCF-1
RNF20

Histone tails

RNA polymerase
TP53 GTFs
TP53I11 TAFs
RB1 MED6
PHB MED8
HOXA1 MED23
HOXA11 TAF1
HOXB4
ZNF217

HDAC4
HDAC11
PRDM2
PRDM4
BRD2
BRD9

DDX42
SF3B3
SRSF10
HNRNPM

pre-mRNA

5) Splicing

DDX3X DHX9
DDX5 DHX15
DDX6 DHX30
DDX10 DHX38
DDX46

6) Translation and mRNA decay

RPS11,13,29
RPL7,9,24,29
RPL736,37A,39
eIF3H
GAR1
DHX36
DHX29

Ribosome
5'
mature mRNA
3'

3'
mRNA decay
5'
XRN1
XRN2

4) Telomere

RPA3
RTEL1
ATRX DNA2
GAR1
FUS

TERRA

9) Mitochondrial roles

Mitochondria

MRPS 5,9,15
MRPL 2,3,11,12,13
MRPL 24,28,30,35 TFAM

NUCLEUS
CYTOPLASM

11) Ion Channels

Vesicle

SLC17A1
SLC18A3

K+
SLC12A7
SLC12A9
SLC12A10

SLC47A

Cl-

10) Lysosome

VPS52
VPS53
VPS54

ATPase6:
V1C1
VOD1
VOD2
V1E1
V1G1

Vesicle
Lysosome

CELL MEMBRANE

Gene deficiency causes G4-ligand sensitisation
Gene deficiency causes G4-ligand resistance

Figure 6.1. Possible locations and roles of RNA- and DNA-G4s in the cell as revealed by shRNA screening

225

6.2.1 DNA replication

DNA G4s are thought to perform regulatory roles in DNA replication, alongside dedicated helicase machinery that unwinds these structures (Rhodes & Lipps, 2015). In the absence of helicase regulation, DNA polymerase obstruction via DNA-G4 can induce DNA damage (see section 1.4.2.3). Consistent with these proposed functions, and regulation of DNA-G4s, deficiencies in genes associated with DNA replication and DNA damage (both detection and repair) were identified as causing sensitisation and also resistance to stabilisation of these structures via G4-ligands (DNA damage will be discussed in section 6.2.3). Among these were DNA-G4 helicases including *RECQL4*, *ATRX* and *RTEL1*, with the latter known to be required for both telomeric replication (Vannier *et al*, 2012) and genomic DNA replication as part of the replisome complex (Vannier *et al*, 2013). DNA topoisomerase 1 (TOP1), which is required to relieve superhelical tension ahead of the replication fork (Wang, 2002), was also identified as a synthetic lethality to G4-ligand treatment. This supports an additional key method of regulating DNA-G4s, other than helicases (as was discussed in section 3.2.1).

The synthetic lethality screening methodology also highlighted that small reductions in DNA replication proficiency can induce G4-stabilisation synthetic lethality. This is supported by the sensitivity seen following deficiency in several independent polymerase subunits including *POLA2* (a DNA polymerase alpha-primase required for replication initiation*,* Fioani, Lucchini, & Plevani, 1997)*; POLD2,* (a DNA polymerase for lagging strand synthesis*;* Tumini, Barroso, -Calero, & Aguilera, 2016) and *POLQ* (DNA polymerase

theta which recruits ORC2 and ORC4 to allow correct timing of DNA replication; Fernandez-Vidal et al., 2014). This indicates that stabilised G4s can compromise the activity of several polymerases and that their absence leads to cell death.

Synthetic lethalities were also seen with other replication components, *ORC1, PCNA, RFC2* and *RFC4* (Cullmann *et al*, 1995)*.* Replication factor C components (*RFC*) and PCNA are required to coordinate leading and lagging strand synthesis (Ellison & Stillman, 1998; Tsurimoto & Stillman, 1991) and their identification as sensitisers suggests a strong relationship between G4 stabilisation and DNA replication. Also identified were replication-associated proteins that have not previously been implicated with G4s, including ligases responsible for non-proteolytic ubiquitin modifications. For example the NEDD8-activating enzyme E1 ligase catalytic subunit, UBA3, which modifies Cullins to ensure controlled DNA replication (Read *et al*, 2000). UBA3 deficiencies are associated with uncontrolled S-phase DNA replication and DNA damage (Xu *et al*, 2014a).

6.2.2 Cell cycle

Linked to DNA replication, synthetic lethalities also extend to other phases of the cell cycle including mitotic entry (*WEE1, CCND1*), metaphasic chromosome alignment (*USP37*; Yeh et al., 2015) and anaphase (*ANAPC4, ANAPC11, CUL1*). For the latter, G4s were previously linked to sister chromatid alignment in yeast during meiosis (Anuradha & Muniyappa, 2004). Here, the identification of cell cycle components as sensitivities in mitotic cell

lines could implicate a role for G4-structures in normal chromosome alignment and correct cell cycle progression.

ATRX-deficiencies also cause abnormalities in spindle alignment and mitosis in mice (De La Fuente *et al*, 2004; Baumann *et al*, 2010; Ritchie *et al*, 2008), in addition to causing DNA replication hindrance and induction of DNA damage (section 6.2.3 Juhász, Elbakry, Mathes, & Löbrich, 2018). The DNA-damage phenotypes and G4-ligand sensitivity associated with ATRX-deficiency are reminiscent of those portrayed by G4-helicase deficiencies. Based on these phenotypes, ATRX conceivably directly interacts with and aids in the resolution of DNA-G4 structures in cells.

Synthetic lethality associated with cell cycle deficiencies is complemented by the identification of the mitotic checkpoint *RB1* (Giacinti & Giordano, 2006) as a resistance gene. *RB1* functions to inhibit problematic cell cycle progression, such as that putatively caused by aberrant G4-stabilisation (e.g. DNA replication inhibition). This checkpoint control is removed following shRNA-induced deficiency, thus cells are able to continue dividing rather than stalling. Similarly, deficiencies in *TP53* and its target gene *TP53I11,* which contributes to the apoptotic induction/proliferation inhibition of the former (Liang *et al*, 2004) were also associated with resistance. Knock down of a further replication checkpoint, *PHB,* similarly provided G4-ligand resistance, a potent tumour suppressor that interacts with both p53 and Rb1 as a negative cell proliferation regulator (Wang *et al*, 1999). Additionally, DDX39A deficiency was identified as a resistance mechanism, a gene recently linked to the DDR

in *X.laevis* (see section 5.3.3; Holden, Taylor, & Lindsay, 2017) in a manner similar to p53 and Rb1. If also true of human cells, this may explain the DDX39A-deficient resistance to PDS and PhenDC3 treatment.

6.2.3 DNA damage

Homologous recombination (HR) DNA damage repair deficiencies, have previously been exploited for BRCA1 and BRCA2 as synthetic lethality strategies with several G4-stabilising ligands (McLuckie et al., 2013; Tauchi et al., 2003; H. Xu et al., 2017). Here DDR synthetic lethalities were extended to other HR pathway components, including *SRCAP* (see section 6.2.5), *PALB2* (Nepomuceno *et al*, 2017), *BAP1* (Carbone *et al*, 2013), *POLD2* (Tumini *et al*, 2016) and *RPA3* (Safa *et al*, 2014)*.* Additionally non-HR synthetic lethalities were identified including the Fanconi anaemia pathway (*USP1*; Nijman et al., 2005) and microhomology-mediated end joining (*POLQ* and POLA2*;* Dantzer, Nasheuer, Vonesch, de Murcia, & Ménissier-de Murcia, 1998; Wood & Doublié, 2016). Emphasising this, *TP53, RB1, TP53I11* and *PHB* in addition to presenting cell cycle checkpoints, are important players in the induction of cell cycle arrest or apoptosis in response to DNA damage, thus a depletion in any one of these components may allow continued cell proliferation in the presence of stabilised G4-induced DNA damage.

6.2.4 Transcriptional roles for G4s

Similarly to DNA polymerase deficiencies, RNA polymerase depletions also resulted in G4-ligand sensitivity, as revealed by the identification of POLR1A, a component of polymerase 1 (Pol1) as a sensitiser. Pol1 is responsible for

ribosomal RNA transcription (Boisvert *et al*, 2007) and previous studies show that Pol1 inhibition combined with treatment with the G4-stabilising ligand CX-4561, induces p53-mediated apoptosis (Quin *et al*, 2016). This supports the hypothesis that stabilised G4s can act as regulatory obstacles to transcription.

Contrasting the link between Pol1 transcription and synthetic lethality with G4-ligands, a number of gene deficiencies in components of the core transcription machinery lead to resistance (rather than sensitivities), compared to the DMSO control. These genes were outlined extensively in chapter 5 and included general transcription factors (GTFs), transcription associated factors (TAFs), and components of mediator (MED6). The formation of single stranded DNA in the transcription bubble is thought to favour G4 structures, by removing competition for DNA duplex formation (reviewed in Kim, 2017). If loss of a transcription machinery component leads to a lower transcriptional output, then this suggests there are fewer DNA-G4s and RNA-G4s, the latter due to decreased mRNA production. This would suggest that in the absence of any gene deficiency, G4s exist, and that they may perform a transcriptional regulatory role, providing experimental support for a mechanistic hypothesis.

Despite this, many integral general transcription machinery deficiencies were not identified as resistance genes. This could be technical (reflecting the strength of the shRNA knockdown, an experimental artefact and/or protein levels within the cell) or that only specific transcription components are specifically important regarding DNA-G4s. An example is the RNA polymerase 2 Mediator complex, a transcriptional coactivator at enhancers

(Soutourina, 2017). The Mediator complex is composed of 25 subunits, yet only MED6, 8, 23 deficiencies were identified as resistance genes. MED6 and MED8 interact in the mammalian complex, and are on the surface on the structure, while the position and function of Med23 is less understood (Soutourina, 2017). Isolated deletion of yeast homologues of Med6 and Med8 are lethal in yeast, although not explored in human cells (Soutourina, Wydau, Ambroise, Boschiero, & Werner, 2011), suggestive of an important transcriptional role. Thus the components most important for transcription, such as specific mediator components, may be the most susceptible to G4 interference following their depletion.

In addition to the general transcription machinery, deficiencies in specific transcription factors, including TP53, RB1 and ZNF217, also caused resistance to G4-stabilising ligands (see section 5.3). Conversely, deficiencies in some specific transcription factors caused PDS and PhenDC3 sensitivity. For example depletion of *ADAR1*, a known DNA G4-binding protein associated with transcriptional repression at genes with promoter G4s (Kang *et al*, 2014) caused a sensitisation to G4-ligands. Perhaps shRNA-induced *ADAR1* depletion alleviated the transcriptional repression and increased exposed G4s for ligand binding at genes containing promoter G4s. In addition to a role in DDR, the key sensitiser BRCA1 has also been linked to transcriptional roles (Mullan *et al*, 2006), perhaps exacerbating the synthetic lethality phenotype. Overall, the multitude of transcriptionally linked proteins that were identified via my screen supports that regulation of transcription can

be modulated by dynamically forming G4s, and thus perturbed via ligand-induced G4 persistence.

Extending a G4-transcriptional role to the mitochondria, depletion of the mitochondrial transcription factor TFAM (Ngo *et al*, 2014), shown to bind both RNA and DNA G4 within the mitochondrial genome and transcriptome *in vitro* (Lyonnais *et al*, 2017), was associated with G4-ligand sensitivity. However, this sensitivity seems to contradict the hypothesis outlined here: TFAM deficiencies would be predicted to reduce mitochondrial transcription, and thus a putative reduction in ligand accessible G4-structures. Distinct to this transcriptional function, TFAM also provides a chromatin-like role by multimerising to coat, condense and protect the mtDNA (Ngo *et al*, 2014). Studies *in vitro* support that mitochondrial G4 structures can drive this multimerisation (Lyonnais *et al*, 2017). This latter role for G4s is arguably more compatible with the synthetic lethality phenotype: in the absence of TFAM, mitochondrial DNA G4s are less compact and more accessible to G4-stabilising ligands. Regardless, my screen has provided experimental evidence in human cells, that the *in vitro* observation of TFAM binding G4s has a functional role, which can be manipulated by G4-ligand treatment. A more general link between altering genomic DNA compaction, via nucleosome remodelling, and how this can alter G4 accessibility is next described.

6.2.5 G4s link to chromatin structure

BG4 ChIP experiments have shown that G4s preferentially form in open chromatin (Hänsel-Hertsch *et al*, 2016). It is hypothesised that chromatin decompaction increases G4 accessibility and/or formation and ligand sensitivity whilst nucleosomal compaction would limit G4 ligand binding, causing resistance. The ability of chromatin remodellers to alter chromatin compaction, and in this way influence G4 accessibility, is reflected by their identification as both sensitisers and resistance genes. Examples of deficiencies causing sensitisation include ATP-dependent SWI/SNF DNA translocases: ATRX, SMARCA4, SMARCB1 and SMARCE1 (discussed in section 2.5.3). Another example is CHD8, an ATP-dependent DNA helicase that causes transcriptional repression due to chromatin remodelling (Ronan *et al*, 2013) and may promote G4-ligand sensitivity via chromatin accessibility. Deficiencies in the SRCAP chromatin remodeller, associated with transcriptionally active chromatin (Ruhl *et al*, 2006) also caused G4-ligand synthetic lethality. As SRCAP depletion is associated with a reduction in open chromatin, this is incompatible with the simple G4-DNA ligand accessibility hypothesis presented above. However, SRCAP has recently been identified as a HR DDR protein, and SRCAP depletion increased HeLa cell sensitivity to DNA damaging agents including ionising radiation and mitomycin C (Dong *et al*, 2014). Thus, for SRCAP, synthetic lethality may arise from DNA damage deficiencies rather than nucleosomal compaction alteration.

In addition to direct nucleosome remodellers, deficiencies in histone tail modifiers and the proteins that recruit them were also identified as sensitisers.

These include H3K4 methyltransferase MLL4 (Froimchuk *et al*, 2017), H3K36 methyltransferase WHSC1 (Bennett *et al*, 2017), HDAC recruiter ANKRD11 (Neilsen *et al*, 2008) and HCF-1, shown to interact with HDAC1, HDAC2 and MLL methyltransferases (Tyagi *et al*, 2007). A set of different modifier deficiencies were also identified as resistance genes: histone deacetylases, HDAC4 and HDAC11, which act as transcriptional repressors (Bottomley *et al*, 2008); BRD2 and BRD9 (Belkina & Denis, 2012), which cause chromatin decompaction via histone lysine acetylase recruitment and the H3 methyltransferases PRDM2 and PRDM4 (Bogani *et al*, 2013) acting as both transcriptional repressors and activators. Many 'resistance' chromatin remodellers, perform inhibitory, closed chromatin roles. Therefore, their depletion would lead to chromatin opening, increasing G4 accessibility and G4-ligand sensitivity. The resistance phenotype observed however, indicates that the mechanism that I propose is too simplistic and requires further exploration. To evaluate this mechanism further, a possible next step would be to pharmacologically inhibit each of these individual chromatin remodellers and directly investigate how this alters G4 formation via G4 ChIP, such as in Hänsel-Hertsch et al., 2016. More importantly, there is a need to identify where in the genome (i.e. intro or exon) these G4s are formed, and how this differs and can be altered via specific inhibition of the individual proteins. This more detailed understanding may help expand the mechanistic association between chromatin remodelling and G4-ligand sensitivity.

6.2.6 Telomeric G4 roles

The telomere is enriched in G4-structures (see section 1.3.3) and many deleterious phenotypes are seen following cell treatment with various G4-stabilising ligands and have been attributed to telomere damage (Neidle, 2010). Telomere G4s are hypothesised to perform a protective role via protein interactions at the chromosome end. Several synthetic lethalities were associated with telomere proteins, including, but not limited to deficiencies in FUS, which regulates telomere length via ternary complex formation with TERRA RNA- and telomeric DNA-G4s (Takahama *et al*, 2013); ATRX and RTEL1, which are required to unwind and allow replication of telomeric DNA G4 (Wong *et al*, 2010; Vannier *et al*, 2012); RPA3, deletion of which results in telomere shortening (Kobayashi et al., 2010); and DNA2 which binds to and cleaves inter and intramolecular G4 at the telomere (Masuda-Sasa *et al*, 2008) to maintain telomere integrity. This perhaps supports the concept that hypothesised G4 interactions within these proteins have a biological role.

Telomere binding proteins not identified as synthetic lethalities, include shelterin components, and helicases such as WRN and BLM that reportedly unwind telomeric G4s (Aggarwal *et al*, 2011; Sfeir *et al*, 2009; Palm & de Lange, 2008). This may point to a hierarchy of biological importance of telomeric G4-interacting proteins, a comparison only achievable via a systematic and unbiased investigation, such as that provided within this thesis. For DNA2, nucleolytic processing of telomeric G4s is hypothesised to allow quicker and more efficient removal of G4 structures than helicase unwinding (Lin *et al*, 2013) on a timescale more compatible with accurate

replication, perhaps explaining why nuclease but not helicase deficiency is associated with synthetic lethality, in the context of the telomere.

6.2.7 RNA G4 roles: splicing, translation and mRNA decay

A role for G4-structures in pre-mRNA processing is supported by the emergence of several key splicing factors as sensitisers, including *SRSF10*, *HNRNPM* and *SF3B3*. SF3B3 interacts with U2 snRNP at the splicing branchpoint and is hypothesised to bind G4-structures (Mori *et al*, 2013). It is notable that DDX42, identified here as a potential novel RNA G4-helicase (chapter 4), was shown to interact with SF3b (product of *SF3B3*) in early mass spectrometric studies (Will *et al*, 2002). SF3b is currently being explored as an anti-tumour pharmacologic target with pladienolide derivatives (Kotake *et al*, 2007). G4-stabilising ligands might therefore be candidates for further exploration from the same pharmacological perspective.

RNA-G4s in mRNA UTRs are proposed to perform regulatory translational roles (Beaudoin & Perreault, 2010; Bugaut & Balasubramanian, 2012). Supporting the hypothesis that aberrant G4 regulation can lead to ribosome translocation impediment (Bugaut & Balasubramanian, 2012), multiple small and large cytoplasmic ribosome subunit components, (*RPS* and *RPL* respectively) and the initiation factor, eIF3H were identified as synthetic lethal interactions. Mitochondrial ribosome deficiencies were also identified as sensitivities (*MRPS* and *MRPL*), suggesting that G4 forming sequences exist within mitochondrial transcripts and may also perform regulatory roles. The key sensitiser GAR1, in conjunction with other H/ACA snoRNPs is also

essential in ribosomal biogenesis (reviewed in Watkins & Bohnsack, 2012). Also identified were several predicted helicases, including members of the DDX and DHX families (see section 2.4.5), whose deficiencies imparted sensitivity to PDS and/or PhenDC3. This included the known RNA-G4 unwinder DHX36 (Vaughn *et al*, 2005). Currently, other than DHX36, these helicases have not been characterised with respect to G4-biology and it remains an open question as to their role. Overall, my observations suggest that gene deficiencies that compromise ribosome function, i.e. rRNA biogenesis, ribosomal protein synthesis or mRNA helicase deficiency, are synthetic lethal with G4-ligand treatment, supporting a role for G4s in translation regulation.

Extending the possible roles of G4-interactions in RNA biology, XRN1 and XRN2 5' and 3' exoribonuclease deficiencies were sensitive to PDS and PhenDC3 treatment. Previous studies have shown that XRN1 preferentially binds and degrades mRNA containing G4 structures (Bashkirov *et al*, 1997), suggesting that G4s provide a docking role to allow efficient turnover and translational control of an mRNA subset. The phenotypic data presented here suggest that XRN2 may function similarly to XRN1.

6.2.8 Cellular ionic composition

Depletion of several ion SLC channels caused G4-ligand resistance, many of which I hypothesise reflect off-target ligand effects. As discussed in chapter 5, loss of several components of the SLC12 chloride/cation transporters caused a growth advantage to both PDS and PhenDC3. These channels are

responsible for the cellular uptake of potassium and sodium cations, and alterations in the ionic balance may alter the stability of endogenous G4 structures. This may highlight an additional novel regulatory mechanism, in addition to dedicated helicase machinery and DDR, by which the cell can regulate G4 formation.

6.2.9 Ubiquitination as a novel player involved in G4-biology

Another novel G4-link identified by the screen is the synthetic lethality seen with deficiencies of several components of the ubiquitin pathway. Several of these have been discussed above, in the context of the DNA damage response, DNA replication and mitosis, and involve components in both proteolytic and non-proteolytic modification of substrates. For the former, this includes deficiencies in the proteasome itself (*PSMC2*). Thus while the ubiquitin pathway itself has not previously been linked to G4-biology, many of the processes post-translationally regulated by ubiquitination have been studied from a G4-perspective.

6.2.10 Summary insights into the role and regulation of G4s

The results of the shRNA screens have provided systematic and unbiased experimental evidence for roles and subcellular locations of G4s. Many of these roles and loci have been hypothesised and shown *in vitro,* but never demonstrated in human cells. These include several RNA related proteins, covering a wide range of RNA functions from splicing to mRNA decay. This may provide a counter argument to the proposal that RNA-G4s are globally unfolded endogenously (Guo & Bartel, 2016). More generally, the breadth of

pathways that are engaged via stabilisation of G4s with low ligand concentrations strongly support that G4s are involved in wide-ranging DNA and RNA roles.

Here the functions of G4s within the cell have been probed by perturbing the system with small molecules. Not all of these G4s may exist endogenously and small molecule treatment may promote the formation of physiologically unstable G4-structures by favouring the equilibrium of the folded state. Therefore, although the genes and pathways discussed above provide insights into positive functions for G4s, to further understand the specific biological roles of G4s will require a combination of experiments to genetically remove a specific G4, or to modulate a protein binding to G4s and characterise the resultant phenotypes. The genes and areas identified here provide likely candidates for future investigation via such approaches.

The screen presented here uses G4-ligand treatment as a surrogate for G4-stabilisation. However, it is unknown whether the biophysical data on these molecules reflects the situation within cells. In addition to G4-ligands, another set of tools developed to probe G4 function are G4-specific antibodies, such as BG4 (see section 1.4). Unlike small molecules, BG4 binds endogenously forming G4s, that have been trapped by cell fixation, rather than altering their dynamics in living cells. Further exploration of the targets identified here could come from integrating the results of the screen with BG4 ChIP-seq (Hänsel-Hertsch *et al*, 2016) to identify the locations of G4s in HT1080 and A375 cells. Additionally, BG4 ChIP-seq would provide insights into chromatin landscape

differences between the two cell lines, which may partially contribute to the differences in gene sensitivities and resistance genes between the two cell lines.

It is of note that the synthetic lethalities and resistance genes identified here are based on a single gene knockdown. From an evolutionary standpoint, it is perhaps surprising that a significant G4-ligand response is observed following perturbation of single gene and/or G4-role. One explanation for this could be that the G4-stabilisation by small molecules is greater or more widespread than normally seen within a cell, and that simultaneously disrupting several pathways makes it easier to see a defect of a single gene, because the system is 'primed' to fail. Such an unphysiological onslaught may allow effective chemotherapy, particularly seeing as cancer cells in general seem to show a higher level of G4-structures compared to normal cells (Biffi *et al*, 2014b). General chemotherapeutic possibilities will be next discussed, although more detailed and specific examples can be found elsewhere within the thesis.

6.3 Towards a chemotherapeutic use of G4 ligands

In addition to probing G4 functions, my project was also aimed to identify disease-associated genotypes, particularly in cancer, that may be susceptible to G4-stabilising ligands. This aim has three interdisciplinary parts:

Part 1) Identify disease-associated genetic backgrounds.

Extend our knowledge of the genetic contribution to diseases such as cancer. This is not addressed within this thesis and refers more widely to the continued understanding of genetic diseases and/or the development of the personalised medicine field.

Part 2) Identify genetic backgrounds that are sensitive to G4-stabilising ligands. Some of these genotypes identified may already be known to be disease related, while other sensitivities have not been linked with disease but may emerge in the future.

Part 3) Support the clinical medicinal development of G4-stabilising ligands.
Moving a research grade molecule towards a more drug-like entity will allow the eventual clinical exploitation of any acutely susceptible disease-associated genotypes identified.

Throughout this thesis, the tools and resources that have been developed may contribute to this aim and will be discussed below, now considering the same genes and pathways highlighted in Figure 6.1 from a therapeutic perspective.

6.3.1 Non-G4 associated effects of G4-targeting ligands

A central assumption to investigations in this thesis is that the sensitivities or resistance genes uncovered following PDS and/or PhenDC3 treatment arise

from G4-stabilisation. However, suggestive of non-G4 associated interactions, deficiencies in vesicle (*VPS52, VPS53, VPS54*) and lysosome components (*ATPase V1C1, V0D1, V1E1, V1G1*) (Figure 6.1) were uncovered as causing ligand resistance. These V-ATPases consist of V0 transmembrane and V1 cytosolic domains and are important for lysosome acidification (Holliday, 2014). Conversely, VPS52, 53, 54, comprising all three of the Golgi-associated retrograde protein (GARP) complex subunits are required for recycling of the lysosomal hydrolases. The observation that impairing lysosome formation and/or function provides G4-ligand resistance suggests that normally these ligands accumulate in the lysosome and also, that lysosomal stress may contribute to non-G4 associated ligand induced death. In support of this, work in our lab suggests PDS can gather in the lysosome (personal communication, Dr. Marco di Antonio, University of Cambridge).

For other polar lipophilic drugs, including the anti-depressants thioridazine and perazine, acidification in the lysosome prevented diffusion back into the cytosol (Daniel *et al*, 2001). Such molecules are referred to as lysosomotropic and their excessive lysosomal accumulation can cause dysfunction, stress and ultimately apoptosis (Lu *et al*, 2017). For G4-ligands, lysosomal accumulation prevents the molecules from reaching their G4 targets in the nucleus and/or cytoplasm. As lysosome function is found in both normal and cancer cells, minimising such unwanted side effects in the development of future ligands, will increase the therapeutic window between normal and cancer cells, and increase the proportion of G4-ligands that reach their intended G4-target. Furthermore, the identification of vesicular and drug

export ion channels deficiencies (e.g. *SLC17*, *SLC18* and *SLC47*) as resistance mechanisms, may support that the G4-stabilising molecules enter the cell via active uptake rather than simple diffusion across the cell membrane, something that also needs to be optimised for future drug development. Alternatively, this may suggest that ion balance of the cell is important for the formation of G4s and that this can be disrupted by SLC deficiency. Evaluating the impact of SLC depletion on the G4 formation, for example via BG4 ChIP, would provide more insight into this.

To date, G4-ligands have been tailored for *in vitro* rather than clinical work. The features that make G4-stabilising ligands potent and selective stabilisers of G4-structures *in vitro* often make them poor clinical drugs, with respect to molecular weight, cell permeability and potential for lysosome trapping (Kazmi et al., 2013; Lipinski, Lombardo, Dominy, & Feeney, 2001; see chapters 3 and 5). Therefore despite the considerable chemotherapeutic potential for G4-stabilising ligands (discussed throughout this thesis), there is an unmet need to develop their chemistry for application to the clinic. A different method to ameliorate lysosomal stress would be co-treatment of G4-ligands with a lysosomal inhibitor such as chloroquine, to increase the 'active' G4-ligand concentration in cells and encourage engagement with their G4 targets.

As G4-stabilising molecules with more desirable medicinal and pharmacokinetic properties are needed, my G4-specific focused screening approach can be exploited to benchmark the synthetic lethalities of new molecules against PDS and PhenDC3. This was exemplified with SA-100-

128, a PDS derivative with improved medicinal chemistry properties. As an additional tool, a five-day experiment has been designed with siRNA knockdown of the top synthetic lethalities *BRCA1, TOP1, DDX42* and *GAR1* in HT1080 cells, that enables the rapid screening of a panel of molecules. These four knockdowns were sensitive to PDS, PhenDC3 and CX-5461, and represent positive controls for comparisons with future developed ligands. As proof-of-principle, this screening capacity was demonstrated with 12 candidate G4-stabilising ligands with improved medicinal properties. Overall, I propose that a clinically effective G4-stabilising ligand would show synthetic lethalities similar to that seen for PDS and PhenDC3, with reduced engagement of lysosome/vesicle associated genes.

6.3.2 Tailoring G4-stabilising ligands towards specific therapeutic niches

Much of the work within the G4-field, and also within this thesis, focuses on pan-G4 stabilising ligands. The use of pan G4-ligands for the genetic screening investigation was imperative to the unbiased aim of systematically investigating the response to global G4-stabilisation. Targeting multiple processes simultaneously via such pan G4-stabilisation may be chemotherapeutically advantageous in a) killing cells and b) preventing the subsequent emergence of resistance cells. For this to work optimally, however, there needs to be selectivity in targeting of cancerous over normal cells. As several cancers exhibit increased G4s compared to controls (Biffi *et al*, 2014b), this may be one method to gauge their sensitivity.

Alternatively, cancers with certain genotypes may be acutely sensitive to G4-ligands. Examples, where I hypothesise G4-ligands may be of particular benefit are summarised in Figure 6.2 and have been discussed and referenced in greater detail in earlier chapters. For example, given the efficacy of CX-5461 in BRCA-deficient tumours (Xu *et al*, 2017), myeloid leukaemia and uveal melanoma, which are characterised by DDR deficiencies in PALB2 and BAP1 respectively, may show similar therapeutic success. Breast cancer associated genotypes beyond BRCA-deficiencies may also be exploitable, such as in KEAP1-deficiency and/or TAF1 overexpression. Similarly, the TAF2 and TAF4 status of HGSOC may impact the effectiveness of using G4-ligands in combination with CDK12 inhibitors (see Figure 6.3). Finally, multiple myeloma offers several distinct combinatorial opportunities with G4-ligands, including pharmacological inhibition of WHSC1, XRN2 and components of the ubiquitin pathway.

Cancer	Gene/Pathway	Gene Deficiency identified as Sensitiser/Resistance?	Gene Status in Cancer?
Myeloid leukaemia	PALB2	sensitiser	Deficient
Uveal melanoma	BAP1	sensitiser	Deficient
Multiple Myeloma	WHSC1	sensitiser	Overexpressed
	XRN2	sensitiser	Overexpressed
	Ubiquitin	sensitiser	Overexpressed
Breast cancer	KEAP1	sensitiser	Deficient
	BRCA1	sensitiser	Deficient
	BRCA2	sensitiser	Deficient
	TAF1	resistant	Overexpressed
HGSOC	CDK12	sensitiser	Overexpressed
	TAF2	resistant	Overexpressed
	TAF4	resistant	Overexpressed

Figure 6.2. Examples of cancers that may be targeted by G4-ligand based therapies

Within both the genome-wide and focused screens, several synthetic lethalities were recurrently enriched including DNA replication, DNA damage, translation and splicing. Within the literature and clinic to date, cancer

chemotherapies often target DNA: the nucleoside analogue gemcitabine is a chain terminator inhibiting DNA polymerase and disrupting dNTP balance within the cell also via ribonucleotide reductase (de Sousa Cavalcante & Monteiro, 2014); methotrexate inhibits thymine production and thus DNA replication (Rajagopalan *et al*, 2002); cisplatin inhibits mitosis and induces a DNA damage response via DNA crosslinking (Dasari & Tchounwou, 2014) and olaparib prevents activation of an appropriate DNA damage response via PARP inhibition (Rottenberg *et al*, 2008). These current therapies, while somewhat effective, are not applicable to all patients, can have toxic and undesirable side effects and cancers can develop resistance. Thus, there is scope for G4-ligands to augment the current panel of DNA-targeting chemotherapies. Additionally, as highlighted in our screen, many non-DNA associated synthetic lethalities were uncovered, for example genes with a hypothesised RNA function, perhaps supporting a niche chemotherapeutic target for G4 stabilising ligands. Specific examples of this have been discussed in chapter 2, such as deficiencies in: splicing factors including oncogenes SRSF10 and HNRNPM, mRNA exoribonucleases XRN1 and XRN2 and, the snoRNPs GAR1, DKC1 and NHP2.

6.3.3 Identifying and validating genotypes susceptible to G4-ligand treatment

My shRNA screening approach identifies synthetic lethal gene deficiencies, many of which are cancer-related and may offer future therapeutic possibilities with G4-stabilising ligands either as single agent (for synthetic lethal interactions with tumour suppressor gene deficiencies) or combinatorial

therapies (where the sensitivity is based on deficiency of an oncogene). Also, some 'resistance' genotypes might offer scope for chemotherapeutic exploitation. These have been discussed extensively elsewhere in the thesis and now I discuss future experiments aimed at validating such chemotherapeutic possibilities.

6.3.3.1 Paired tumour and normal tissue samples

The cell lines (A375 and HT1080) used within this thesis were chosen for technical rather than biological reasons. The next step would be to validate top hits in cancer cell lines with appropriate mutations. However, a caveat with any cancer cell line study is identifying the appropriate control. Several candidate cases from the literature are outlined below, chosen due to an appropriate control for reference. For each paired tumour and normal tissue samples, it would be informative to compare G4-ligand sensitivity and G4 quantification (via BG4 IF or G4-ChIP).

1) **GAR1 –** liquid biopsies from patients with chronic lymphocytic leukaemia (CLL) from a randomised clinical group alongside normal biopsies (Dos Santos *et al*, 2017). CLL samples showed significantly upregulated GAR1 mRNA levels compared to controls

2) **XRN2 –** Various spontaneous lung tumour samples in mice versus normal margin lung tissue are reported to have a positive correlation between the incidence rate and XRN2 transcript levels (Lu *et al*, 2010)

3) **SMARCE1/BAF57 –** xenograft mouse samples of MDA-MD-231 lung metastasis versus metastatic tumours derived from cells with a SMARCE1 knockdown (Sethuraman *et al*, 2016).

4) **SMARCA4 –** Gene is highly expressed in high-grade serous ovarian cancer but depleted in small cell carcinoma of the ovary. For each, paired tumour and normal margin tissue is available (Jelinic *et al*, 2014).

6.3.3.2 Comparison with other chemotherapeutic molecules

Other ways to explore and validate the chemotherapeutic options for G4-stabilising ligands would be via the use of other molecules that target the same pathway and/or deficiency. In chapters 3 and 5, I discussed and referenced several drugs (both clinical and experimental, referred to as 'available drug treatments', Figure 6.3) that were used to target cancers with genotypes that emerged from to the shRNA screen results. I also hypothesised areas where G4-ligands could be applied (summarised Figure 6.3). These drugs are split into two categories: 'single agent' refers to a genotype susceptible to a particular drug treatment, which is also predicted to be sensitive to G4-ligands. For these scenarios, the possible next steps would be to compare the efficacy of these drugs to G4-ligands (PDS, PhenDC3 and CX-5461). I also present 'combinatorial' options, where G4-ligands may increase the potency of pharmacological inhibition of a drug target. This remains to be validated via drug combination synergy screening. One route for further exploration would be via the use of relevant mouse models.

6.3.3.3 Validation in mouse models

Once G4-ligands with improved medicinal chemistry properties have been developed, the next stage before clinical use in human trials would be their investigation in appropriate mouse models. Of particular interest would be investigating the response of tumours that occur in genetically engineered mouse models (GEMMs). GEMMs are cited as better models than their cancer cell inoculation models as the tumours develop in a completely immune proficient environment (Kersten *et al*, 2017). The success of G4-ligands in GEMM trials relies on clinically tolerable version of G4-ligands that would be amenable to administration, such as via tail vein injection, which as discussed above, are currently lacking.

Several mouse models that would be useful to study the sensitisers/resistance genes uncovered in the screen (Figure 6.3) have been developed as part of the Wellcome Trust Sanger Institute Sanger Mouse Genetics Project (Sanger MGP; www.sanger.ac.uk/science/collaboration/mouse-resource-portal). In addition to these GEMMs, there also exists a *Trp53* GEMM in which both alleles are replaced with tamoxifen inducible *Trp53* alleles (Martins, Brown-Swigart, & Evan, 2006). Deficiency of p53 was identified as a putative resistance mechanism to G4-ligand treatment. This mouse model would allow investigation into whether p53 deficient tumours are also resistant to G4-ligand treatment, and by using tamoxifen to restore p53 in established tumours, investigate whether we can restore G4-ligand sensitivity. The use of such GEMM models has been integral in refining and bringing other drugs to

clinical trial. For example, the BRCA1-deficient KBIP mouse model, which closely mimicks the human cancer response to PARP inhibtor and platinum drugs (Rottenberg *et al*, 2008, 2007) was imperative in establishing the clinical use of olaparib for ovarian, breast and colorectal cancer treatment (Lee *et al*, 2014). Such models could similarly refine the clinical advancement of current and future G4-stabilising ligands.

Gene Deficiency	Sensitiser/ Resistance	oncogene/ tumour suppressor	Combinatorial/ single agent	Available Drug treatments	Available mouse models
CDK12	Sensitiser	Oncogene	Combinatorial	Dinaclib (SCH77965)	Cdk12em1(IMPC)Tcp
WEE1	Sensitiser	Oncogene/ Tumour suppressor	Combinatorial/ single agent	AZD1775	Wee1tm1b(EUCOMM)Wtsi
KEAP1	Sensitiser	Oncogene/ Tumour suppressor	Combinatorial/ single agent	CDDO-Me CPUY192018	Keap1tm1a(EUCOMM)Wtsi
SMAD4	Sensitiser	Tumour suppressor	Single agent	**GSKi:** NCT01632306 NCT01214603 NCT01287520	Smad4tm1a(EUCOMM)Wtsi
CCDC6	Sensitiser	Tumour suppressor	Single agent	olaparib	Ccdc6em1(IMPC)Wtsi
WHSC1	Sensitiser	Oncogene	Combinatorial	Suramin N-alkylsinefungin DA3003-1 PF-03882845 Chaetocin TC-LPA5-4 ABT-199	n/a
PSMC2	Sensitiser	Oncogene	Combinatorial	**proteosome inhibitor:** bortezomib CEP187710 carfizomib	Psmc2tm1b(EUCOMM)Hmgu
USP1	Sensitiser	Oncogene/ Tumour suppressor	Combinatorial/ single agent	pimozide GW7947	Usp1tm1b(KOMP)Wtsi
UBA3	Sensitiser	Oncogene	Combinatorial	MLN4924	Uba3tm1a(EUCOMM)Wtsi
SRSF10	Sensitiser	Oncogene	Combinatorial	E7107 1C8	Srsf10tm1a(KOMP)Wtsi
BRCA1/2	Sensitiser	Tumour suppressor		olaparib CX-5461	Brca2tm1a(EUCOMM)Hmgu Brca1tm1a(KOMP)Wtsi
TOP1	Sensitiser	Tumour suppressor		etoposide camptothecin	Top1tm1a(KOMP)Wtsi
GAR1	Sensitiser	Oncogene			Gar1tm1a(EUCOMM)Wtsi
DDX42	Sensitiser	-			Ddx42tm1b(EUCOMM)Wtsi
GTFs, TAFs	Resistance	Oncogene	Single agent	TAS-106	
BRD2/BRD9	Resistance	Oncogene	Single agent	JQ1 I-BET	Brd2em1(IMPC)Wtsi Brd9em1(IMPC)Tcp

Figure 6.3. Chemotherapeutic options for G4-stabilising ligands and possible mouse models for validation

6.4 Adapting the screening approach for other investigations

In this thesis I have described a successful and optimised screening technique for investigating genotypes sensitive or resistant to G4-stabilising ligands. To identify the strongest sensitisers and the foremost resistance mechanisms, low ligand doses and readout of cell death were used. This methodology can be adapted to answer several other questions. One obvious step would be to extend the screen to more ligands with several cell lines, and look for universal commonalities, completely independent of cell type and G4-ligand. This would indicate a core set of synthetic lethalities arising from G4-stabilisation. I have gone someway to investigate this, performing the focused screen in two cell lines of discrete lineages: HT1080 (mesenchymal) and A375 (neural crest).

Also, it would be prudent to perform the screen with control molecules, including isoforms of G4-ligands that do not stabilise G4-structures such as iPDS, a non-G4 stabilising isomer of PDS (Dr. Marco Di Antonio, personal communication), in order to distinguish between cellular responses to treatment with a large, polar molecule from phenotypes derived from stabilisation of G4-structures. Another interesting comparison would be to perform the screen with TMPyP4 and TMPyP2, cationic porphyrin isomers that both stabilise G4s *in vitro*, but do so via different binding mechanisms (Frank Xiaoguang Han, Richard T. Wheelhouse, & Hurley, 1999). It is currently unknown if differences in biophysical binding between molecules such as the 'TMPyP' molecules would engage independent synthetic lethalities or resistance mechanisms in cells. Furthermore, to investigate gene

deficiencies that cause resistance to G4-stabilising ligands, the screen could be performed with higher ligand concentrations to create a greater selection pressure and look for an outgrowth of resistant cells, as is more characteristic of positive selection screens (Miles *et al*, 2016).

Throughout these investigations with the pan-G4 ligands PDS and PhenDC3, I have separated biological G4 roles and/or sensitive genotypes into either RNA or DNA G4-stabilisation responses (Figure 6.1). However, many RNA processes are inevitably also DNA related (at the level of transcription), preventing precise differentiation between the two. One way to discern between the two is to perform the screen using molecules that solely bind either DNA- or RNA-G4. As a starting point, within the focused screen results, PDS and PhenDC3 enriched synthetic lethality terms seemed to show a DNA and RNA bias respectively, perhaps suggesting that the molecules have an endogenous preferentially affinity for DNA-G4 and RNA-G4 respectively. Within the literature however, both molecules are reported to cause transcription and translational changes via binding to RNA and DNA G4s (see introduction for details). For a RNA-G4 screen, cPDS, a molecule reported to increase cytoplasmic not nuclear BG4 foci, could be performed. However, this limits the investigation to cytoplasmic RNA-G4, thus omitting investigations into the cellular response to pre-mRNA G4 stabilisation (i.e. those involved in splicing). Thus there is an experimental need to develop RNA and DNA specific G4-ligands, which can then be explored via the tools provided here.

Both cell lines used for our investigations (A375 and HT1080) are p53-proficient. This was important for the initial setup, as p53 is an important regulator of the BRCA1 and BRCA2 damage response, which represented positive controls in terms of synthetic lethality with G4-stabilising ligands. Indeed, p53 proficiency was integral in expanding the knowledge of the DDR to stabilised G4s. However, many cancers are p53 deficient or mutated (Ko & Prives, 1996), and p53 deficiency has been uncovered as a resistance mechanism. Thus, it would be interesting to perform the screen in a p53-deficient cell line, and investigate how this alters the synthetic lethalities and resistance mechanisms uncovered.

6.5 Concluding remarks

Here, I present a set of tools and resources towards answering key outstanding areas in the G4 field. The first was expanding our knowledge of the genes and processes involved in the cellular response to stabilised G4s. The second was to identify disease related genotypes in which G4-stabilising ligands, both current and future, could be successfully applied to in the clinic. This body of work provides a significant step forward in answering these questions and has identified a new *in vitro* G4-binder and putative helicase, DDX42. In addition, I created a set of resources to test future G4-binding molecules, as they are developed to become more clinically appropriate. I provide a proof-of-principle on this here with the G4-ligand currently in clinical trials (CX-5461) and also with a panel of pharmacologically improved PDS derivatives. This identified a top candidate SA-100-128, which shows some similarities in sensitivity to other G4-ligands e.g. *BRCA2* and *ATRX*

deficiencies. I envisage that these tools can be employed to identify new clinical pan-G4 ligands for use in cancer, by pinpointing genotypes that are truly sensitised to stabilisation of G4s. Ultimately my work will refine our knowledge of the key players in G4-biology, and the application and exploitation of these structures in disease.

Chapter 7

Materials and Methods

7.1 General Methods

Unless otherwise stated oligonucleotides were obtained from Sigma Aldrich as lyophilised powders and dissolved in TE buffer (pH 7.4) and stored at -20 ^{0}C as 100 µM stock. All DNA and RNA (plasmid and oligonucleotides) stocks were stored at -20 ^{0}C. The UV spectrum of all nucleic acid samples quantified was recorded via Nanodrop One^c (Thermo Fisher Scientific).

7.2 Ligand Synthesis

Pyridostatin (PDS): 4-(2-aminoethoxy)-N2,N6-bis(4-(2-aminoethoxy)quinolin-2- yl)pyridine-2,6-dicarboxamide. Synthesis was described by Rodriguez *et al*. (Rodriguez *et al*, 2008). PhenDC3: 3,3'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))bis(1- methylquinolin-1-ium). Synthesis was described by De Cian *et al* (De Cian *et al*, 2007b). Synthesis was performed by Dr. Marco di Antonio. All PDS derivatives with improved medicinal chemistry properties (Chapter 3, section 3.2.4-3.2.5) were synthesised in-house by Dr. S. Adhikari and Mr. J Patterson.

7.3 Cell Lines

HT1080, A375 (ATCC CRL-1619, CRL-121) and Plat-A (Cell Biolabs Inc, RV-102) cell lines were cultured in DMEM medium (ThermoFisher Scientific, cat #41966029) supplemented with 10% (v/v) heat inactivated FBS

(ThermoFisher Scientific, cat #10500064). All cell lines were grown at 37 °C in a 5% CO_2 humidified atmosphere and regularly checked to be mycoplasma-free by RNA-capture ELISA. Cell line genotypes were certified by STR profiling (by CRUK Cambridge Insitute Biorepository).

7.4 Quantification of live cell numbers

Live cell numbers (e.g. for plating cells for CellTitre-Glo assays, the screens and Incucyte experiments) were determined using the Muse Cell Analyser (Merck), 'Count & Viability' assay according to manufacturer's instructions. Cells were diluted either 1:10 or 1:20 in 'Muse Count & Viability kit' solution (Merck, cat # MCH60013), to give a viable cell concentration of 1-2 x 10^6 cells/mL, with 'Events to Acquire' parameter set at 1000 events.

7.5 Determination of G4 ligand concentration for shRNA screens

PDS, PhenDC3 and SA-100-128 were used as 50 mM stocks, dissolved in DMSO (Thermofisher Scientific, cat # 20688). GI_{20} values were calculated by treating A375 and HT1080 cells with serial dilutions of PDS, PhenDC3 and SA-100-128 (maximum concentration 100 µM, 300 µM and 25 µM respectively) for 96 h and determining cell death via a CellTitre-Glo One Solution assay (Promega, cat # G8461) according to manufacturer's protocol. Each serial dilution was replicated 4 times for 2 cell-seeding densities (1000/1500 cells per well). For both cell densities, curves were plotted averaging the 4 replicates in Prism (GraphPad v6) using a Non-Linear regression model, "dose-response – inhibition" equation [log(inhibitor) vs. normalized response – variable slope] and GI_{20} values calculated. The GI_{20}

concentrations used represent an average for three separate assays per cell line and yielded the following concentrations used for the screens - A375: 10 µM PhenDC3, 1.5 µM PDS and 1 µM SA-100-128; HT1080: 1 µM PhenDC3, 0.5 µM PDS and 1.7 µM SA-100-128.

7.6 Composition of and recombinant DNA production from shRNA libraries

The genome-wide screen uses the Transomic LMN shRNA library against the human protein coding genome (a kind gift from Professor Greg Hannon, University of Cambridge, vector shown in Figure 2.3), consisting of 113,002 total shRNAs, split between 12 pools for ease-of-handling (approximately 10,000 shRNAs per pool) with an average number of 6 optimised hairpins per gene. The G-quadruplex-focused screen consists of a custom shRNA pool (Purchased from Transomic) with the same LMN vector (8,018 shRNAs); this includes 1247 genes (7,436 shRNAs) uncovered in the genome-wide screen (751 sensitizers and 496 upregulated genes), 116 additional genes identified from the literature as potentially G4-associated (439 shRNAs) and shRNAs targeting 37 olfactory receptors as non-targeting controls (143 shRNAs). The 496 upregulated genes (FDR ≤ 0.05, 50 % or 3 hairpins; $log_2FC \geq 1$) were included to mimic the genome-wide screen on a smaller scale by maintaining the population ratio of sensitisation and resistance. These resistance genes were secondly used to understand the mechanisms that can cause resistance to G4-stabilising ligands. In this custom pool, unlike the genome-wide library, the number of shRNAs was capped at seven per gene. The backbone of both libraries contains Neo[R] and ZsGreen markers to allow monitoring of infected

cell lines by Geneticin (Gibco, cat # 10131035) selection and fluorescence (MacsQUANT) respectively. Both libraries were provided as glycerol stocks. Bacterial density was determined by calculating the colony forming units (CFU) from dilutions of the original glycerol stock after plating on agar (O/N, 37 °C, 100 µg/ml ampicillin). Then, glycerol stocks were thawed completely and sufficient volume was taken (based on CFU) to ensure a minimum of 1000-fold hairpin representation, and inoculated in Gigaprep bacterial cultures (LB media + 100 µg/ml ampicillin). Plasmid DNA was isolated from these cultures using a ZR Gigaprep kit D4057 (Zymo research) according to the manufacturer's protocol and DNA was quantified by Nanodrop One[c] (Thermo Fisher Scientific).

7.7 shRNA stable cell line creation

For the genome-wide screen, each pool was treated independently, necessitating the creation of 12 different polymorphic A375 cell lines each containing an average of 10,000 shRNAs, per replica (3 replicas, 36 polymorphic cell lines). For the focused screen, virus was created from all shRNAs simultaneously to create a single polymorphic cell line for both HT1080 and A375, (3 replicas per cell line). Virus was produced using the Platinum-A packaging cell line (4-6 15 cm plates per pool) and calcium phosphate transfection. 24 h after plating Platinum-A cells (70-80 % confluency), media was replaced with DMEM medium supplemented with 1 % (v/v) PenStrep (Thermo Fisher Scientific, cat # 150763) and 10 % (v/v) heat inactivated FBS, shRNA library plasmid (75 µg) was then mixed with pCMV-VSV-G plasmid (7.5 µg, Addgene cat # 8454), Pasha/DGCR8 siRNA (2.7 µM,

Qiagen cat # 1027423), to increase viral titre and 0.75 M $CaCl_2$ in a total volume of 1.5 mL per 15 cm dish and bubbled with 1.5 mL 2 x HBS (50 mM HEPES, 10 mM KCl, 12 mM Dextrose, 280 mM NaCl, 1.5 mM Na_2PO_4 at pH 7.00) buffer and added to the Platinum-A cells in a dropwise fashion. Immediately before adding the DNA-Pasha-transfection mixture to the Platinum-A cells, chloroquine diphosphate (lysosomal inhibitor, Acros Organics cat # 455200250) was added to the plates at a final concentration 2.5 µM. 14-16 h after transfection, fresh media was added with 1:1000 1M sodium butyrate (Merck, cat # 303410) for enhanced mammalian expression of the shRNA LMN plasmid. Virus was then harvested 48 h after transfection and filter sterilised (0.45 µM) and stored at 4 °C for a maximum of 7 days. Viral titre was determined by performing mock infections with the experimental cell line and quantifying fluorescent cells, via flow cytometry (MacsQUANT, Miltenyi Biotec Ltd.) 48 h after infection. For both the genome-wide and focused-screen, 3.6 x 10^6 target cells were infected at a viral volume predicted to cause 30 % infection (MOI 0.3) to minimise multiple shRNA integrations per cell. This provides approximately 10 x 10^6 shRNA expressing cells (1000-fold shRNA representation). Virus was diluted in serum free media plus polybrene (8 µg/mL). Infections were carried out in triplicate, and treated as independent replicates hereafter. 48 h after infection cells were selected in 800 µg/ml (HT1080) and 1000 µg/ml (A375) geneticin for 7-9 days. These concentrations were determined via 7 day geneticin toxicity curves prior to transfection setup. 'Complete' selection was determined via flow cytometry (MacsQUANT, Miltenyi Biotec Ltd.) as a minimum of 95 % fluorescent cells.

7.8 Cell culture process for genome-wide and focused pool of shRNAs

Following complete selection (see section 7.7), a reference time point was harvested (t0) and cells were split into 3 x 15 cm plates per replica: PDS, PhenDC3 and DMSO vehicle control, each containing 8-10 x 10^6 cells to maintain 1000-fold hairpin representation. Every 72 h, cells were trypsinised, counted to determine the number of population doublings, and 10 x 10^6 (A375 genome-wide and focused) or 8 x 10^6 (HT1080) cells per replica replated in fresh drug/DMSO and media (17 mL media per plate). At all times, sufficient cell numbers were used so that a minimum of 1000 or 800 cells per shRNA was maintained (A375 and HT1080 respectively). The volume of DMSO used in the 'vehicle' condition is equal to the volume for 10 μM PhenDC3. The remaining drug treatments were also supplemented with DMSO to match this volume and keep the DMSO concentration constant between treatments, cell lines and screens. A final time point (tF) was harvested after 15 population doublings. For each pool of the genome-wide screen, 12 samples were generated (t0, DMSO tF, PDS tF, PhenDC3 tF; 3 replicas each). Therefore, 144 samples of 10 x 10^6 cells were generated to cover the entire screen. For the focused screen, for each cell line, 12 samples were generated (t0, DMSO tF, PDS tF, PhenDC3 tF, 3 replicas each). An additional focused screen was also performed with SA-100-128, using the same conditions, experiment duration and parameters as above, again with the volume of DMSO used in the 'vehicle' equal to the volume for 10 μM PhenDC3, to facilitate interscreen comparisons. Cells were treated with 1 μM (A375) and 1.7 μM (HT1080) SA-

100-128 and this screen generated 9 samples per cell line (t0, DMSO tF, SA-100-128 tF, 3 replicas each).

7.9 Barcode recovery, adapter ligation and sequencing

All PCR and sequencing oligonucleotides were supplied by Sigma Aldrich and sequences are summarised in the table below. Cell pellets (10×10^6 cells) were resuspended in PBS and genomic DNA (gDNA) extracted using QIAmp DNA Blood Maxi Kit (Qiagen, cat # 51194) according to the manufacturer's spin protocol, eluted in a final volume of 1200 µl and quantified by a Qubit DNA HS Assay Kit (Thermo Fisher Scientific, cat # Q32851). The shRNA inserts were PCR-amplified from all gDNA in each sample, in multiple 50 µl reactions each using 1.5 µg gDNA, with KOD Hot Start DNA Polymerase and the following reagents (Merck, cat # 710864): 5 µl 10 x buffer, 5 µl 2 mM each dNTPs, 4 µl $MgSO_4$ (25 mM), 1.5 µl polymerase, 4 µl DMSO. Forward (Mir-F) and reverse (PGKpro-R) primers flanking the loop and antisense sequence of the hairpin region were used at a final concentration of 300 nM. PCR was performed under the following conditions: 98 °C for 5 min, then for 25 cycles of 98 °C for 35 s, 58 °C 35 s, 72 °C for 35 s, followed by 72 °C for 5 min. 1.2 mL of pooled PCR reaction were cleaned-up using QIAquick PCR purification kit (Qiagen, cat # 28104) according to manufacturer's protocol. 2 µg of this purified PCR product were PCR amplified in a second PCR step, using forward (P5-Seq-P-Mir-Loop) and reverse (P7-Index-n-TruSeq-PGKpro-R) primers containing the P5 and P7 flowcell adapters respectively. The PCR was performed in 8 x 50 µL reactions each containing 500 ng template DNA. The reverse primer contains TruSeq adapter small RNA Indexes (Illumina) to

allow multiplexing of the samples and a 6 nucleotide barcode, denoted 'nnnnnn' below. PCR reagents were used as for the first PCR, with the exception of the primers, which were used at a final concentration 1.5 µM. The second PCR was performed under the following conditions: 98 °C for 5 min, then for 25 cycles of 94 °C for 35 s, 52 °C 35 s, 72 °C for 35 s, followed by 72 °C for 5 min. All PCR reactions were performed using the BioRad T^{100} Thermocycler. 200 µl of pooled secondary PCR product was cleaned up as previously and the desired product (~340 bp) was size selected using a BluePippin instrument with 2 % Internal Standard Marker Kit (DF marker 100-600 bp; Sage Science, BDF2010), according to manufacturer's protocol using a broad range elution (300-400 bp). Individual samples were quantified with a KAPA library quantification kit (KAPA Biosystems, cat # 0796-6014-0001) using a BioRad CFX96 Real Time PCR instrument with no ROX according to manufacturer's protocols. Libraries were diluted to 4 nM in RNAse free water. The following were combined to create 4 nM pooled libraries: for the genome-wide screen samples 24 libraries (2 pools); for the PDS and PhenDC3 focused screen samples, 24 libraries (both cell lines), and for the SA-100-128 focused screen, 18 libraries (both cell lines), with each sample within the 4 nM pool, having a unique TruSeq adapter. The genome-wide screen samples were sequenced in 6 batches; all focused screen samples were sequenced simultaneously. DNA-Seq libraries were prepared from these samples using the NextSeq Illumina Platform v2 High Output Kit 75 cycles, followed by 36 base pair single-read sequencing performed on an Illumina NextSeq instrument, using a custom sequencing primer.

Table 7.1. PCR and sequencing primers for shRNA recovery and quantification

Oligo name	Description	Sequence 5'-3'
Mir5-F	Primary PCR Forward Primer	5'-CAGAATCGTTGCCTGCACATCTTGGAAAC- 3'
PGKpro-R	Primary PCR Reverse Primer	5' -CTGCTAAAGCGCATGCTCCAGACTGC- 3'
P5-Seq-P-Mir-Loop	Secondary PCR forward Primer	5'-AATGATACGGCGACCACCGAGATCTACACT AGCCTGCGCACGTAGTGAAGCCACAGATGTA-3'
P7-Index-n-Truseq-PGKpro-R	Secondary PCR barcoded reverse primer	5'-CAAGCAGAAGACGGCATACGAGAT **nnnnnn**GTGACTGGAGTTCAGACGTGTGCTCTTCC GATCTCTGCTAAAGCGCATGCTCCAGACTGC – 3'
SeqPrimer MirLoop	Custom sequencing primer	5'-AGCCTGCGCACGTAGTGAAGCCACAGATGTA-3

7.10 siRNA preliminary optimisation experiments

ON-TARGETplus Set of 4 - Human siRNAs (Dharmacon/GE healthcare) were ordered for each protein, in addition to Non-targeting siRNA 1,2,3 and 4 (Dharmacon/GE healthcare), each as individual stocks (5 nmol lyophilised powder) and diluted to 200 nM stocks in nuclease free water (according to manufacturer's recommendations (Ambion). Preliminary transfection/knockdown experiments (as described below) for both HT1080 and A375 were performed. For each protein the level 48 h after transfection with the 4 targeting and the 4 non-targeting siRNAs was quantified via immunoblotting (as described below). Based on this optimisation, the targeting siRNA that caused the greatest knockdown compared to the non-transfected control for each protein was selected for further experimental use. The non-targeting control that was least toxic and gave a protein level most similar to the non-transfected control was also selected (Non-targeting siRNA number 2). These optimised siRNAs are summarised in the section below.

7.11 siRNA – transfection, experimental outline and immunoblotting

Cells were transfected with either targeting or non-targeting control siRNAs using lipofectamine RNAiMAX (Thermo Fisher Scientific, cat # 13778150) and OptiMEM Reduced Serum medium (Thermo Fisher Scientific) according to the manufacturer's protocol (Reagent protocol 2013) alongside a non-transfected control. 24 h after transfection, cells were trypsinised, counted and replated in media supplemented with ligand (PDS, PhenDC3, cPDS, CX-5461, SA-100-128, and 11 PDS derivatives with improved medicinal chemistry properties – concentrations can be found within sections 3.2.1-3.2.4 and 5.2.4) or DMSO vehicle control (minimum 2 replicates per condition) in either a 48-well plate format (seeding density - 8,000 cells per well A375; 4,000 cells per well HT1080) or a 96-well plate format (seeding density 1,500 cells per well HT1080). For all experiments, cell growth was monitored for 144 h using IncuCyteZOOM live cell analysis (Sartorius), which approximates cell confluency as a percentage of the well area covered. Scans were performed every 3 h; 9 and 3 scans per well (48- and 96-well plate respectively). To monitor protein levels, cells transfected simultaneously with the same siRNA-reagent mixture were harvested 48 h and 144 h after transfection, by cell scraping and lysed on ice (30 min) with RIPA lysis buffer with protease inhibitor + EDTA (Thermo Fisher Scientific, cat # 8990). Lysates were quantified via Direct Detect Spectrometry (Merck) and analysed by capillary electrophoresis via the Protein Simple WES platform, which separates and blots with antibody of interest, according to manufacturer's protocol with antibodies summarised in the table below. Lysates from non-transfected and

siRNA-treated (targeting and non-targeting) samples were probed with antibodies against the target plus mouse or rabbit anti-beta actin antibody (as appropriate) by multiplexing, with the exception of DDX39A, where beta-actin was run in a separate lane. For non-targeting and targeting lysates, the area of the desired band was normalized to beta-actin and then normalised to the protein level in the non-targeting sample, for three (48 h after transfection lysates) or two independent WES runs (144 h after transfection). Protein depletion is expressed as an average of these normalised values. Growth differences for targeting siRNAs were taken from the average growth curve of the non-targeting control, and then expressed as individual time points (as shown in the text). Significant differences between targeting and the non-targeting siRNA for ligand treatment versus DMSO ($p < 0.05$) were determined using a one-tailed parametric t-test, assuming equal standard deviation.

Table 7.2. optimised siRNAs used for validation of G4-ligand resistance and sensitivity

Dharmacon siRNA	Catalogue number	Sequence 5'-3'
Non-targeting 2	D-001810-02-05	UGGUUUACAUGUUGUGUGA
BRCA1 (A375)	J-003461-08	CAACAUGCCCAUAGAUCAA
BRCA1 (HT1080)	J-003461-12	GAAGGAGCUUUCAUCAUUUC
TOP1 (both cell lines)	J-005278-08	CGAAGAAGGUAGUAGAGGUC
DDX42 (both cell lines)	J-012393-11	GGAGAUCGACUAACGGCAA
GAR1 (both cell lines)	J-013386-06	UCCAGAACGUGUAGUCUUA
ZNF217 (A375)	J-004987-12	GUGCAGGCCUCUCGCAAGA
ZNF217 (HT1080)	J-004987-11	UGAUAAAAGUCAAGUGCGA
DHX29 (both cell lines)	J-013759-09	CUGCAGAUCAUUACGGAAC
TAF1 (both cell lines)	J-005041-10	GGACAAGACAGGGUUACUA
DDX39A (both cell lines)	J-004920-12	UGGAGGUGUUUGUGGACGA

Table 7.3. Optimised antibody conditions for WES Simple Western Platform detection

Protein	Antibody	Protein Concentration (mg/ml)	Antibody dilution	Loading control (LC)	LC antibody dilution
BRCA1	Cell Signalling Technology, cat # 4970-CST	1.0	1:50	**Rabbit beta-actin** 4970-CST	1:100
TOP1	Abcam, cat # AB109374	0.5	1:500	**Rabbit beta-actin** 4970-CST	1:500
DDX42	NovusBio cat #NBP2-31742	0.5	1:250	**mouse beta-actin** Merck cat # A5441	1:250
GAR1	Abcam cat #AB80975	0.8	1:100	**Rabbit beta-actin** 4970-CST	1:500
ZNF217	Abcam cat #AB48133	0.5	1:250	**Rabbit beta-actin** 4970-CST	1:500
DHX29	Cell Signalling Technology, cat # 4159-CST	0.8	1:50	**Rabbit beta-actin** 4970-CST	1:500
TAF1	Cell Signalling Technology, cat #12781-CST	0.8	1:100	**Rabbit beta-actin** 4970-CST	1:250
DDX39A	Abcam cat #AB50697	0.8	1:250	**mouse beta-actin** Merck cat # A5441	1:250

7.12 Sequencing, read processing, alignment and counting of shRNAs

Bioinformatic analyses were performed by Sergio Martinez Cuesta, with input and discussion from Darcie Mulhearn, Katie Zyner and Nicolas Erard. Reads were trimmed to 22 nucleotides, base qualities were evaluated with FastQC v0.11.3 (Andrews, 2010) and bases were filtered from the 3' end with a Phred quality threshold of 33 using the FASTX-Toolkit v0.0.14 (Gordon, 2010). Trimmed reads were aligned to the 113,002 reference shRNA sequences provided by transOMIC technologies (Knott et al., 2014) using Bowtie 2 v2.2.6 with default parameters (Langmead and Salzberg, 2012), which resulted in

overall alignment rates of 90-95 % with an average of 98 % of reference sequences detected. The generated SAM files were processed to obtain shRNA counts using Unix tools (https://opengroup.org/unix) and Python scripts (v2.7.10, https://www.python.org), and library purity and potential contaminations were investigated with stacked bar plots and multidimensional scaling (MDS) using the R programming language v3.2.1 (https://cran.r-project.org). The code and scripts developed during the development of the project are available in the Balasubramanian group's GitHub website (https://github.com/sblab-bioinformatics/GWscreen_G4sensitivity).

7.13 Filtering, normalisation, differential representation analysis and defining sensitisation

To discard shRNAs bearing low counts, each library was filtered based on a counts-per-million threshold of 0.5 for all initial time points (t0), e.g. in a library of 10×10^6 reads, with a threshold of at least 5 counts for each shRNA for all initial time points (t0). Normalisation factors were calculated to scale the raw library sizes using the weighted trimmed mean of M-values (TMM) approach (Robinson and Oshlack, 2010). To compare groups of replicates (time points and chemical treatments) for each pool, differential representation analysis of shRNA counts was performed using edgeR (Robinson *et al*, 2010). Common and shRNA-specific dispersions were estimated to allow the fitting of a negative binomial generalised linear model (glm) to the treatment counts. Contrasts between the initial time point and the treatments were defined (PDS-t0, PhenDC3-t0, and DMSO-t0) and likelihood ratio tests were carried out accordingly (Dai et al., 2014). Fold changes (FC) were then computed for

every shRNA, and false discovery rates (FDR) were estimated using the Benjamini-Hochberg method. A gene was defined as significantly differentially represented for a given treatment if at least 50 % or a minimum of 3 shRNAs were significant (FDR ≤ 0.05); sensitisation was additionally determined by applying a median $\log_2 FC \leq -1$ and resistance was determined by applying a median $\log_2 FC \geq -1$. For the SA-100-128 cell line, as the number of hits was too low this additional sensitisation/resistance threshold was not included and "hits" were considered if they were significantly depleted or upregulated (FDR ≤ 0.05; 50 % or 3 hairpins median $\log_2 FC < 0$ or median $\log_2 FC > 0$, respectively).

7.14 Exploring genes associated to G-quadruplexes in databases and biomedical literature

Three different approaches were developed to uncover genes linked to G-quadruplexes. 18 high confidence G4-related genes were obtained by scanning for genes where the corresponding UniprotKB (The UniProt Consortium, 2017) entry is annotated with the term 'quadruplex' or genes annotated with at least one of the following 11 GO terms with any evidence assertion method (Ashburner et al., 2000, Table 7.4). Furthermore, 55 confirmed human G4 interacting proteins as defined by the G4IPB database (Mishra *et al*, 2016) (http://bsbe.iiti.ac.in/bsbe/ipdb/index.php) were also used to determine predefined G4-interacting proteins from our genome-wide shRNA screen. To obtain this list we removed from the G4IPBD, gene entries where the only G4-relationship was a predicted G4-forming sequence in the mRNA or DNA (i.e. not a direct protein interaction) or where the protein wasn't

human. To expand these lists the text-mining tool PolySearch2 (Liu *et al*,

2015b) was used to explore associations between G-quadruplex terms and

gene names in the biomedical literature.

Table 7.4. 11 G4-associated gene ontology terms for database searching

GO id	Name	Type	Link
GO:0051880	G-quadruplex DNA binding	Molecular function	https://www.ebi.ac.uk/QuickGO/term/GO:0051880
GO:0002151	G-quadruplex RNA binding	Molecular function	https://www.ebi.ac.uk/QuickGO/term/GO:0002151
GO:0061849	Telomeric G-quadruplex DNA binding	Molecular function	https://www.ebi.ac.uk/QuickGO/term/GO:0061849
GO:0071919	G-quadruplex DNA formation	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:0071919
GO:0044806	G-quadruplex DNA unwinding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:0044806
GO:1905493	regulation of G-quadruplex DNA binding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905493
GO:1905494	negative regulation of G-quadruplex DNA binding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905494
GO:1905495	positive regulation of G-quadruplex DNA binding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905495
GO:1905465	regulation of G-quadruplex DNA unwinding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905465
GO:1905466	negative regulation of G-quadruplex DNA unwinding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905466
GO:1905467	positive regulation of G-quadruplex DNA unwinding	Biological process	https://www.ebi.ac.uk/QuickGO/term/GO:1905467

This algorithm assumes that the greater the co-occurence frequency of terms within sentences or database records, the stronger the association. The word span between co-occurring terms in the text also influences the association score. The set of G-quadruplex entities and synonyms was defined using the corresponding MeSH term id D054856 (https://www.ncbi.nlm.nih.gov/mesh/68054856) and the thesaurus of gene names obtained from the PolySearch2 website (http://polysearch.cs.ualberta.ca/). A total of 5477 pieces of text were identified in PubMed and PubMed Central where any of the G-quadruplex terms co-occur with more than 500 human gene names. The PolySearch2 relevancy score measures the strength of association between two text groups, in this case human protein-coding gene names and G4-terms and synonyms (as defined in MeSH; see methods). The higher the score, the more likely terms from the two groups co-occur within the same abstract. In addition to co-occurrence frequency, the score also accounts for the distance between terms from the two groups, using both parameters to determine G4-association. Overall, this generated 526 G4-associated genes of which 54 (10%) were uncovered as G4-sensitisers (https://github.com/sblab-bioinformatics/GWscreen_G4sensitivity), which were manually edited to 16 genes as discussed in chapter 2.

7.15 KEGG Pathway, Gene Ontology and Protein domain enrichment analysis

ClueGO v2.3.3 (Bindea *et al*, 2013, 2009), a plugin for Cytoscape (Shannon *et al*, 2003) (v3.5.1) was used to determine and visualise networks of enriched

KEGG pathways and Gene Ontology terms (Biological Process and Molecular Function) for the lists of significantly depleted genes upon G4 ligand treatment. Specifically, a right-sided (Enrichment) test based on the hyper-geometric distribution was performed on the corresponding Entrez gene IDs for each gene list and the Bonferroni adjustment (p < 0.05) was applied to correct the p-value for multiple hypothesis testing. Only experimental evidence codes (EXP,IDA,IPI,IMP,IGI,IEP) were used. The Kappa-statistics score threshold was set to 0.4 and GO term fusion was used to diminish redundancy of terms shared by similar proteins. Other parameters include: GO level intervals (3-8 genes) and Group Merge (50%). Protein domains were investigated using DAVID (v6.8) to integrate GENE3D crystallographic data and PFAM sequence information and enrichment was considered significant if the EASE score p < 0.05 (Finn *et al*, 2016; Yeats *et al*, 2006).

7.16 COSMIC analysis

Cancer mutation data (CosmicMutantExport.tsv) from COSMIC database v82 (Forbes *et al*, 2015) was used to investigate the association between G4 sensitisers and cancer genes. ~ 150,000 were mutations available in COSMIC for 702 (93%) sensitiser genes, with some predicted to be pathogenic by the fathmm algorithm embedded within the resource. The Cancer Gene Census (http://cancer.sanger.ac.uk/census) was used to investigate whether G4 sensitisers are often enriched in genes containing mutations causally implicated in cancer. Fisher's exact tests as implemented the R programming language were used to calculate the significance of the

old enrichments for the three subsets of PDS only, PhenDC3 only and PDS+PhenDC3 sensitisers with cancer genes as defined by COSMIC.

7.17 DDX42 characterisation - nuclear versus cytoplasmic localisation

HT1080, HEK and HeLa cells were harvested from a 70 % confluent 15 cm plate by cell scraping in PBS on ice and pelleted by centrifugation (500 g, 5 min, 4 ^{o}C). Pellet was resuspended in 3 volumes of low salt buffer (20 mM HEPES pH7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.2 mM EDTA, 1 mM DTT) plus protease inhibitor (cOmplete mini, Roche cat#11836153001), lysed on ice (15 min) and added 0.5 % Igepal. Samples were vortexed (1 min) and centrifuged (900 g, 15 min, 4 ^{o}C). Supernatant was collected for cytoplasmic extract. Residual pellets were washed in low salt buffer, and supernatant discarded. Nuclei pellets were resuspended in high salt buffer (20 mM HEPES pH7.4, 500 mM NaCl, 3 mM MgCl$_2$, 0.5 % Igepal, 0.2 mM EDTA, 1 mM DTT) plus protease inihibitors and lysed on ice with intermittent vortexing (30 min). Lysates were syringed to promote lysis and shear gDNA and nuclei lysis confirmed by trypan blue staining according to manufacturer's protocols (Thermofisher Scientific cat#15250061), followed by centrifugation (13,000 g, 10 min, 4 ^{o}C). The supernatant was then collected as nuclear extract. Cytoplasmic and nuclear lysates were quantified by Direct Detect and analysed via Wes at a concentration 0.5 mg/mL (both as described above) for DDX42 expression. Samples were also immunoblotted with antibodies targeting nuclear laminB1 (CST 1258; 1:250) and cytoplasmic GAPDH to confirm subcellular fractionation efficiency (CST 5172, 1:50).

7.18 DDX42 characterisation – annealing of G4s for ELISA treatment

Biotinylated oligonucleotides for G-quadruplex and non-G-quadruplex forming sequences (IDT technologies; see Table 7.5) were annealed in 10mM TrisHCl pH 7.4, 100 mM KCl by heating at 95 ^{0}C, 10 min followed by slow cooling to RT O/N at a controlled rate of 0.2 ^{o}C/min. Annealed oligonucleotides were stored at 4 ^{0}C for maximum 1 month. Biotinylated oligonucleotides were also annealed in 10 mM TrisHCl pH 7.4, 100 mM LiCl, which is less permissive for G4 folding to act as negative controls in the ELISA assay.

Table 7.5. Biotinylated oligonucleotides for ELISA investigation

Oligo	RNA/ DNA	Sequence
nRAS G4	RNA	5' [Btn] GGG A GGGG C GGG UCU GGG 3'
nRAS mut	RNA	5' [Btn] GAG A GAGG C GAG UCU GAG 3'
Bcl-2 intraG4	RNA	5' [Btn] UUA GGGGG CCGU GGGG U GGG AGCU GGGG 3'
Bcl-2 interG4	RNA	5' [Btn] UUA GGGGG CCGU UUUU U GGG AGCU GGGG 3'
Bcl-2 mutant	RNA	5' [Btn] UUA GAGAG CCGU UUUU U GAG AGCU GAGG 3'
TERRA-comp (ssRNA)	RNA	5' [Btn] CCC TAA CCC TAA CCC TAA CCC TAA CCC TAA 3'
Stem loop	RNA	5' [Btn] ACA GGG CUC CGC GAU GGC GGA GAA 3'
Myc G4	DNA	5' [Btn] TGA GGG T GGG TA GGG T GGG TAA 3'
Myc mut	DNA	5' [Btn] TGA GAG T GAG TA GAG T GAG TAA 3'

7.19 DDX42 characterisation – Enzyme Linked Immunosorbent Assay

Recombinant human DDX42 with an N-terminal GST tag was used (NovusBio H0001325-P01). UltraPure distilled water (Invitrogen 10977-035) was used for all buffers. Streptavidin Coated High Binding Capacity 96-well plates (ThermoScientific prod #15501) were hydrated with PBS (30 min) and coated with 50 nM biotinylated oligonucleotides (1 h, shaking 450 rpm). Wells were

washed three times with ELISA buffer (50 mM K_2HPO_4 pH 7.4 and 100 mM KCl/100 mM LiCl); 1 min shaking, 450 rpm. Wells were blocked in 3 % (w/v) BSA in ELISA buffer for 1 h, at RT and then incubated with serial dilutions of hDDX42 up either 300 nM or 200 nM for 1 h (as outlined in the text). Wells were washed 3 times with 0.1 % TWEEN-20 in ELISA buffer and incubated for 1 h with anti-GST HRP-conjugated antibody (AbCam AB3416) diluted 1:10000 in blocking buffer. Wells were again washed 3 times with ELISA-Tween, and the bound anti-GST HRP detected with TMB substrate (Sigma T4444) for 2 min. Reactions were stopped with 2 M HCl. Absorbance at 450 nm was measured with PheraSTAR plate reader (BMG labtech). Binding curves with standard error of the mean (SEM) were fitted using GraphPad Prism software, using a non-linear regression fit, one site, specific binding model with saturation kinetics. The following equation was used: $y=(Bmax*x)/(K_d + x)$, where x = concentration of DDX42 (nM) and Bmax is the maximum specific binding (i.e. saturation).

7.20 Circular dichroism spectroscopy analysis of oligonucleotides

200 μL of 10 μM oligonucleotide were prepared in assay buffer and annealed as described above. Circular dichroism (CD) experiments were then performed by Dr. Marco di Antonio. CD spectra were recorded on an Applied Photo-physics Chirascan CD spectropolarimeter using a 1 mm path length quartz cuvette. CD measurements were performed at 298 K over a range of 200-320 nm using a response time of 0.5 s, 1 nm pitch and 0.5 nm bandwidth. The recorded spectra represent a smoothed average of three scans, zero-

corrected at 320 nm (Molar ellipticity θ is quoted in 10^5 deg cm^2 $dmol^{-1}$). The absorbance of the buffer was subtracted from the recorded spectra.

7.21   siRNA   transfection   and   cell   fixation   for   BG4 Immunofluorescence in DDX42-depleted cells

HT1080 cells were transfected with non-targeting and anti-DDX42 siRNA using the optimised siRNAs (GE healthcare/Dharmacon) described above (section 7.9), in a 6-well plate format; 2 biological replicas performed for each sample with each replica transfected independently. 24 h after transfection, cells were trypsinised, counted and 60,000 cells replated per well in a 6-well plate containing autoclaved 18 mm coverslips (VWR, cat#631-0153) in 3 mL media supplemented with 0.5 µM PDS or the equivalent volume of DMSO, and left overnight. For each biological replica and sample sufficient coverslips were prepared for two timepoints: 72 h and 120 h after transfection. For each timepoint NT and DDX42 siRNA transfected cells were treated with PDS or DMSO.   For each biological replica, two further 72 h NT + DMSO wells/coverslips were prepared for antibody controls (see below). Additionally, an extra well of DDX42 and NT transfected cells were plated for immunoblot analysis, to confirm protein knockdown after 72 h (performed and analysed as in section 7.10). Following 72 h and 120 h, media was removed from the wells, cells were washed with PBS (2 x 2 mL), and fixed with 4 % PFA in PBS (2 mL per well, 10 min, RT). After washing with PBS (2 x 2mL), cells were permeabilised with 0.1 % Triton-PBS (10 min), and washed again with PBS (2 x 2mL). At this stage, 72 h timepoints were kept at 4 ^{o}C and immunofluorescence   staining   as   described   below   was   performed

simultaneously with the 120 h timepoints. Biological replicas were stained on separate days.

7.22 DDX42 characterisation - BG4 Immunofluorescence staining and microscopy quantification analysis

BG4 was produced and quantified in-house according to Hansel-Hertsch 2018 *Nature Protocols* (in press). Briefly, 8 µM BG4 was used for the IF experiment. Coverslips were incubated in on parafilm with 200 µL blocking buffer (2 % Marvel in PBS; 1 h, 37 ^{o}C). Next, cells were incubated with 1/800 BG4 in blocking buffer (50 µL per silde; 1 h, 37 ^{0}C). One 72 h NT + DMSO sample was incubated in blocking buffer without BG4 (50 µL; 1 h, 37 ^{0}C; no BG4 control). Cells were then washed 3 x PBS + 0.1 % Tween (PBS-T; 5 min per wash, shaking), before incubating with rabbit anti-FLAG antibody (Cell Signalling Technology; cat# CST-2368S 1:800 in blocking buffer; 50 µL per slide; 1 h, 37 ^{o}C). Also included was a '72 h NT + DMSO' slide untreated with anti-FLAG (50 µL blocking buffer; 1 h, 37 ^{o}C). Following a further 3 x PBS-T washes (5 min; shaking), all slides were incubated with goat anti-rabbit Alexa-fluor594 (45 min, 37 ^{o}C; LifeTech A11037), washed 1 x PBS-T (5 min) and incubated with DAPI (0.4 µg/mL in PBS-T, 10 min, shaking) and again washed with PBS-T (5 min). Then coverslips were quickly washed with distilled water (3 x) and air-dried (30 min). Coverslips were mounted onto slides using ProLong Diamond Antifade Mountant (Thermofisher Scientific, B36961), dried O/N at RT and stored at 4°C until imaged. Confocal images were acquired using a Leica TCS SP5 microscope with 63x oil objective (Leica Microsystems). The 405 nm diode laser was used to excite the DAPI

channel and tertiary Alexa594 fluorophore was excited via the 543 laser line. Fluorescence was performed with sequential acquisition (2 seqeuntial scans, switch between frames) detected via hybrid detectors (Leica HyD Photon Counter), with the following wavelengths collected: 410-485 nm and 640-800 nm for DAPI and Alexa594 respectively. Pixel size was set to 2048 X 2048 with zoom factor 2, with a frame average of 3 to reduce background signal. Pinhole, laser power, and gain settings were consistent between samples within the same biological replica, and set to prevent oversaturation and ensure no 594 nm signal in the no antibody controls. 10 frames were acquired per condition representing a total 70-100 nuclei per sample. Images were processed via CellProfiler v3.0.0 (Carpenter *et al*, 2006) using the DAPI stain to denote nuclei localisation. The following parameters were used: nuclei "primary objects" diameter restricted to 150-500 pixel units and objects outside the image border or diameter range discarded and "speckle" counter settings, with BG4 foci restricted to 2-35 pixels. For each sample, the average BG4 foci per nuclei were plotted.

7.23 DDX42 characterisation – DDX42 level quantification in response to G4-ligand treatment

HT1080 cells were seeded in a 6 well plate format at 20 % confluency. 24 h after seeding, one well was transfected with anti-DDX42 siRNA. The remaining wells were treated for 96 h with either DMSO vehicle control or the following ligands and concentrations: PDS (0.25 and 0.5 µM), PhenDC3 (20 and 40 µM), CX-5461 (10 and 50 nM) and SA-100-128 (0.5 µM). Samples were then lysed and immunoblotted for DDX42 as outlined in section in 7.10.

For each biological replica, DDX42 level was normalised to the internal beta-actin control and expressed as a percentage of that the DMSO treated control. This was averaged for two biological replicas and significant difference was determined by a parametric t-test assuming equal standard deviation.

References

Abed DA, Goldstein M, Albanyan H, Jin H & Hu L (2015) Discovery of direct inhibitors of Keap1–Nrf2 protein–protein interaction as potential therapeutic and preventive agents. *Acta Pharm. Sin. B* **5:** 285–299

Agarwala P, Pandey S, Mapa K & Maiti S (2013) The G-quadruplex augments translation in the 5' untranslated region of transforming growth factor β2. *Biochemistry* **52:** 1528–1538

Aggarwal M, Sommers JA, Shoemaker RH & Brosh RM (2011) Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc. Natl. Acad. Sci. U. S. A.* **108:** 1525–1530

Aguilera A & García-Muse T (2013) Causes of genome instability. *Annu. Rev. Genet.* **47:** 1–32

Ajmal M, Khan MI, Neveling K, Khan YM, Azam M, Waheed NK, Hamel CP, Ben-Yosef T, De Baere E, Koenekoop RK, Collin RWJ, Qamar R & Cremers FPM (2014) A missense mutation in the splicing factor gene *DHX38* is associated with early-onset retinitis pigmentosa with macular coloboma. *J. Med. Genet.* **51:** 444–448

Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR & Cheadle JP (2002) Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat. Genet.* **30:** 227–232

Ambrus A, Chen D, Dai J, Jones RA & Yang D (2005) Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **44:** 2048–2058

Ameres SL & Zamore PD (2013) Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14:** 475–488

Anderson SF, Schlegel BP, Nakajima T, Wolpin ES & Parvin JD (1998) BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nat. Genet.* **19:** 254–256

Ando H, Kawaai K & Mikoshiba K (2014) IRBIT: A regulator of ion channels and ion transporters. *Biochim. Biophys. Acta - Mol. Cell Res.* **1843:** 2195–2204

Ando H, Mizutani A, Kiefer H, Tsuzurugi D, Michikawa T & Mikoshiba K (2006) IRBIT suppresses IP3 receptor activity by competing with IP3 for the common binding site on the IP3 receptor. *Mol. Cell* **22:** 795–806

Ando H, Mizutani A, Matsu-ura T & Mikoshiba K (2003) IRBIT, a Novel Inositol 1,4,5-Trisphosphate (IP$_3$) Receptor-binding Protein, Is Released from the IP$_3$ Receptor upon IP$_3$ Binding to the Receptor. *J. Biol. Chem.* **278:** 10602–10612

Anuradha S & Muniyappa K (2004) Meiosis-specific yeast Hop1 protein promotes synapsis of double-stranded DNA helices via the formation of guanine quartets. *Nucleic Acids Res.* **32:** 2378–2385

Aoude LG, Pritchard AL, Robles-Espinoza CD, Wadt K, Harland M, Choi J, Gartside M, Quesada V, Johansson P, Palmer JM, Ramsay AJ, Zhang X, Jones K, Symmons J, Holland EA, Schmid H, Bonazzi V, Woods S, Dutton-Regester K, Stark MS, et al (2015) Nonsense mutations in the shelterin complex genes ACD and TERF2IP in familial melanoma. *J. Natl. Cancer Inst.* **107:** 1-7

Arimondo PB, Riou JF, Mergny JL, Tazi J, Sun JS, Garestier T & Hélène C (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.* **28:** 4832–4838

Arnaoutov A & Dasso M (2014) IRBIT is a novel regulator of ribonucleotide reductase in higher eukaryotes. *Science.* **345:** 1512–1515

Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S & Kurreck J (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* **14:** 1290–1296

Auyeung VC, Ulitsky I, McGeary SE & Bartel DP (2013) Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* **152:** 844–858

Azorsa DO, Gonzales IM, Basu GD, Choudhary A, Arora S, Bisanz KM, Kiefer JA, Henderson MC, Trent JM, Von Hoff DD & Mousses S (2009) Synthetic lethal RNAi screening identifies sensitizing targets for gemcitabine therapy in pancreatic cancer. *J. Transl. Med.* **7:** 43

Azzalin CM & Lingner J (2014) Telomere functions grounding on TERRA firma. *Trends Cell Biol.* **25:** 29-36

Backert S, Gelos M, Kobalz U, Hanski ML, Böhm C, Mann B, Lövin N, Gratchev A, Mansmann U, Moyer MP, Riecken EO & Hanski C (1999) Differential gene expression in colon carcinoma cells and tissues detected with a cDNA array. *Int. J. cancer* **82:** 868–874

Bailey SM & Murnane JP (2006) Telomeres, chromosome instability and cancer. *Nucleic Acids Res.* **34:** 2408–2417

Bainbridge MN, Armstrong GN, Gramatges MM, Bertuch AA, Jhangiani SN, Doddapaneni H, Lewis L, Tombrello J, Tsavachidis S, Liu Y, Jalali A, Plon SE, Lau CC, Parsons DW, Claus EB, Barnholtz-Sloan J, Il'yasova D, Schildkraut J, Ali-Osman F, Sadetzki S, et al (2015) Germline mutations in shelterin complex genes are associated with familial glioma. *J. Natl. Cancer Inst.* **107:** 384-387

Bajrami I, Frankum JR, Konde A, Miller RE, Rehman FL, Brough R, Campbell J, Sims D, Rafiq R, Hooper S, Chen L, Kozarewa I, Assiotis I, Fenwick K, Natrajan R, Lord CJ & Ashworth A (2014) Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res.* **74:** 287–297

Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M & Sasisekharan V (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20:** 4061–4067

Balasubramanian S, Hurley LH & Neidle S (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10:** 261–275

Balasubramanian S & Neidle S (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13:** 345–353

Balkwill GD, Derecka K, Garner TP, Hodgman C, Flint APF & Searle MS (2009) Repression of translation of human estrogen receptor alpha by G-quadruplex formation. *Biochemistry* **48:** 11487–11495

Banck MS, Li S, Nishio H, Wang C, Beutler AS & Walsh MJ (2009) The ZNF217 oncogene is a candidate organizer of repressive histone modifiers. *Epigenetics* **4:** 100–106

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

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA & Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709–1712

Bartz SR, Zhang Z, Burchard J, Imakura M, Martin M, Palmieri A, Needham R, Guo J, Gordon M, Chung N, Warrener P, Jackson AL, Carleton M, Oatley M, Locco L, Santini F, Smith T, Kunapuli P, Ferrer M, Strulovici B, et al (2006) Small Interfering RNA Screens Reveal Enhanced Cisplatin Cytotoxicity in Tumor Cells Having both BRCA Network and TP53 Disruptions. *Mol. Cell. Biol.* **26:** 9377–9386

Bashkirov VI, Scherthan H, Solinger JA, Buerstedde JM & Heyer WD (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell Biol.* **136:** 761–773

Bates GJ, Nicol SM, Wilson BJ, Jacobs A-MF, Bourdon J-C, Wardrop J, Gregory DJ,

Lane DP, Perkins ND & Fuller-Pace F V (2005) The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J.* **24:** 543–553

Baumann C, Viveiros MM & De La Fuente R (2010) Loss of Maternal ATRX Results in Centromere Instability and Aneuploidy in the Mammalian Oocyte and Pre-Implantation Embryo. *PLoS Genet.* **6:** e1001137

Beà S, Valdés-Mas R, Navarro A, Salaverria I, Martín-Garcia D, Jares P, Giné E, Pinyol M, Royo C, Nadeu F, Conde L, Juan M, Clot G, Vizán P, Di Croce L, Puente DA, López-Guerra M, Moros A, Roue G, Aymerich M, et al (2013) Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 18250–18255

Beaudoin J-D & Perreault J-P (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **38:** 7022–7036

Beaudoin J-D & Perreault J-P (2013) Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **41:** 5898–5911

Beaume N, Pathak R, Yadav VK, Kota S, Misra HS, Gautam HK & Chowdhury S (2013) Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans involves G4 DNA-mediated regulation. *Nucleic Acids Res.* **41:** 76–89

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

Belkina AC & Denis G V (2012) BET domain co-regulators in obesity, inflammation and cancer. *Nat. Rev. Cancer* **12:** 465–477

Bennett RL, Swaroop A, Troche C & Licht JD (2017) The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb. Perspect. Med.* **7:** a026708

Bensaude O (2011) Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2:** 103–108

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL & Bernards R (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428:** 431–437

Berridge MJ, Lipp P & Bootman MD (2000) The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **1:** 11–21

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

Bharti SK, Sommers JA, George F, Kuper J, Hamon F, Shin-ya K, Teulade-Fichou M-P, Kisker C & Brosh RM (2013) Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *J. Biol. Chem.* **288:** 28217–28229

Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J & Topisirovic I (2015) Targeting the translation machinery in cancer. *Nat. Rev. Drug Discov.* **14:** 261–278

Biffi G, Di Antonio M, Tannahill D & Balasubramanian S (2014a) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6:** 75–80

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

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

182–186

Biffi G, Tannahill D, Miller J, Howat WJ & Balasubramanian S (2014b) Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One* **9:** e102711

Bindea G, Galon J & Mlecnik B (2013) CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* **29:** 661–663

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z & Galon J (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25:** 1091–1093

Blackburn EH (1991) Structure and function of telomeres. *Nature* **350:** 569–573

Bochman ML, Paeschke K & Zakian VA (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13:** 770–780

Bochman ML, Sabouri N & Zakian VA (2010) Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst).* **9:** 237–249

Boettcher M & McManus MT (2015) Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Mol. Cell* **58:** 575–585

Bogani D, Morgan MAJ, Nelson AC, Costello I, McGouran JF, Kessler BM, Robertson EJ & Bikoff EK (2013) The PR/SET domain zinc finger protein Prdm4 regulates gene expression in embryonic stem cells but plays a nonessential role in the developing mouse embryo. *Mol. Cell. Biol.* **33:** 3936–3950

Boisvert F-M, van Koningsbruggen S, Navascués J & Lamond AI (2007) The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **8:** 574–585

Bommi-Reddy A, Almeciga I, Sawyer J, Geisen C, Li W, Harlow E, Kaelin WG & Grueneberg DA (2008) Kinase requirements in human cells: III. Altered kinase requirements in VHL-/- cancer cells detected in a pilot synthetic lethal screen. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 16484–16489

Booy EP, Howard R, Marushchak O, Ariyo EO, Meier M, Novakowski SK, Deo SR, Dzananovic E, Stetefeld J & McKenna SA (2014) The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.* **42:** 3346–3361

Bose P, Hermetz KE, Conneely KN & Rudd MK (2014) Tandem repeats and G-rich sequences are enriched at human CNV breakpoints. *PLoS One* **9:** e101607

Bottomley MJ, Lo Surdo P, Di Giovine P, Cirillo A, Scarpelli R, Ferrigno F, Jones P, Neddermann P, De Francesco R, Steinkühler C, Gallinari P & Carfí A (2008) Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **283:** 26694–26704

Boudreau RL, Monteys AM & Davidson BL (2008) Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. *RNA* **14:** 1834–1844

Bourdoncle A, Estévez Torres A, Gosse C, Lacroix L, Vekhoff P, Le Saux T, Jullien L & Mergny J-L (2006) Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc.* **128:** 11094–11105

Brázda V, Hároníková L, Liao JCC & Fojta M (2014) DNA and RNA quadruplex-binding proteins. *Int. J. Mol. Sci.* **15:** 17493–17517

Brosh RM (2013) DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer* **13:** 542–558

Brosh RM & Cantor SB (2014) Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia. *Front. Genet.* **5:** 372-386

Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, Darnell RB & Warren ST (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107:** 477–487

Brummelkamp TR, Bernards R & Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296:** 550–553

Bryan TM & Baumann P (2011) G-quadruplexes: from guanine gels to chemotherapeutics. *Mol. Biotechnol.* **49:** 198–208

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, Kyle S, Meuth M, Curtin NJ & Helleday T (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434:** 913–917

Budhathoki JB, Ray S, Urban V, Janscak P, Yodh JG & Balci H (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res.* **42:** 11528–11545

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

Bugaut A & Balasubramanian S (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40:** 4727–4741

Bukrinsky MI, Sharova N, Dempsey MP, Stanwick TL, Bukrinskaya AG, Haggerty S & Stevenson M (1992) Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. U. S. A.* **89:** 6580–6584

Burge S, Parkinson GN, Hazel P, Todd AK & Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34:** 5402–5415

Bywater MJ, Pearson RB, McArthur GA & Hannan RD (2013) Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat. Rev. Cancer* **13:** 299–314

Cahoon LA & Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science* **325:** 764–767

Cammas A & Millevoi S (2016) RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* **45:** 1584-1595

Capra JA, Paeschke K, Singh M & Zakian VA (2010) G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae. *PLoS Comput. Biol.* **6:** e1000861

Carbone M, Yang H, Pass HI, Krausz T, Testa JR & Gaudino G (2013) BAP1 and cancer. *Nat. Rev. Cancer* **13:** 153–159

Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang I, Friman O, Guertin DA, Chang J, Lindquist RA, Moffat J, Golland P & Sabatini DM (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7:** R100

Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136:** 642–655

Castanotto D, Sakurai K, Lingeman R, Li H, Shively L, Aagaard L, Soifer H, Gatignol A, Riggs A & Rossi JJ (2007) Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic Acids Res.* **35:** 5154–5164

Castillo Bosch P, Segura-Bayona S, Koole W, van Heteren JT, Dewar JM, Tijsterman M & Knipscheer P (2014) FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **33:** 2521–2533

Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E & Méchali M (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle* **11:** 658–667

Centeno F, Ramirez-Salazar E, Garcia-Villa E, Gariglio P & Garrido E (2008) TAF1 Interacts with and Modulates Human Papillomavirus 16 E2-Dependent Transcriptional Regulation. *Intervirology* **51:** 137–143

Chakraborty P & Grosse F (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst).* **10:** 654–665

Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP & Balasubramanian S (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33:** 877-881

Chan DA & Giaccia AJ (2011) Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat. Rev. Drug Discov.* **10:** 351–364

Chang S, Khoo C & DePinho RA (2001) Modeling chromosomal instability and epithelial carcinogenesis in the telomerase-deficient mouse. *Semin. Cancer Biol.* **11:** 227–239

Chang T-S, Wei K-L, Lu C-K, Chen Y-H, Cheng Y-T, Tung S-Y, Wu C-S & Chiang M-K (2017) Inhibition of CCAR1, a Coactivator of β-Catenin, Suppresses the Proliferation and Migration of Gastric Cancer Cells. *Int. J. Mol. Sci.* **18:** 2189-2198

Chao C-H, Chen C-M, Cheng P-L, Shih J-W, Tsou A-P & Lee Y-HW (2006) DDX3, a DEAD box RNA helicase with tumor growth-suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor. *Cancer Res.* **66:** 6579–6588

Charley PA, Wilusz CJ & Wilusz J (2018) Identification of phlebovirus and arenavirus RNA sequences that stall and repress the exoribonuclease XRN1. *J. Biol. Chem.* **293:** 285–295

Chen D, Frezza M, Schmitt S, Kanwar J & P. Dou Q (2011) Bortezomib as the First Proteasome Inhibitor Anticancer Drug: Current Status and Future Perspectives. *Curr. Cancer Drug Targets* **11:** 239–253

Chen J, Dexheimer TS, Ai Y, Liang Q, Villamil MA, Inglese J, Maloney DJ, Jadhav A, Simeonov A & Zhuang Z (2011b) Selective and Cell-Active Inhibitors of the USP1/ UAF1 Deubiquitinase Complex Reverse Cisplatin Resistance in Non-small Cell Lung Cancer Cells. *Chem. Biol.* **18:** 1390–1400

Chen MC, Murat P, Abecassis K, Ferré-D'Amaré AR & Balasubramanian S (2015) Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* **43:** 2223–2231

Chen X-C, Chen S-B, Dai J, Yuan J-H, Ou T-M, Huang Z-S & Tan J-H (2018) Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells. *Angew. Chemie Int. Ed.* **57:** 4702–4706

Chen ZJ & Sun LJ (2009) Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Mol. Cell* **33:** 275–286

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

Cheung PK, Horhant D, Bandy LE, Zamiri M, Rabea SM, Karagiosov SK, Matloobi M, McArthur S, Harrigan PR, Chabot B & Grierson DS (2016) A Parallel Synthesis Approach to the Identification of Novel Diheteroarylamide-Based Compounds Blocking HIV Replication: Potential Inhibitors of HIV-1 Pre-mRNA Alternative Splicing. *J. Med. Chem.* **59:** 1869–1879

Chiarella S, De Cola A, Scaglione GL, Carletti E, Graziano V, Barcaroli D, Lo Sterzo C, Di Matteo A, Di Ilio C, Falini B, Arcovito A, De Laurenzi V & Federici L (2013) Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Res.* **41:** 3228–3239

Chinnam M & Goodrich DW (2011) RB1, development, and cancer. *Curr. Top. Dev. Biol.* **94:** 129–169

Chiu Y-L & Rana TM (2002) RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol. Cell* **10:** 549–561

Choi H, Lee H, Kim SR, Gho YS & Lee SK (2013) Epstein-Barr virus-encoded microRNA BART15-3p promotes cell apoptosis partially by targeting BRUCE. *J. Virol.* **87:** 8135–8144

Chong CR & Sullivan DJ (2007) New uses for old drugs. *Nature* **448:** 645–646

Christiansen J, Kofod M & Nielsen FC (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Res.* **22:** 5709–5716

Chung K-H, Hart CC, Al-Bassam S, Avery A, Taylor J, Patel PD, Vojtek AB & Turner

DL (2006) Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. *Nucleic Acids Res.* **34:** e53

Churikov D, Wei C & Price CM (2006) Vertebrate POT1 restricts G-overhang length and prevents activation of a telomeric DNA damage checkpoint but is dispensable for overhang protection. *Mol. Cell. Biol.* **26:** 6971–6982

De Cian A, Cristofari G, Reichenbach P, De Lemos E, Monchaud D, Teulade-Fichou M-P, Shin-ya K, Lacroix L, Lingner J & Mergny J-L (2007a) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl. Acad. Sci.* **104:** 17347–17352

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

Clynes D, Higgs DR & Gibbons RJ (2013) The chromatin remodeller ATRX: a repeat offender in human disease. *Trends Biochem. Sci.* **38:** 461–466

Cogoi S, Paramasivam M, Filichev V, Géci I, Pedersen EB & Xodo LE (2009) Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells. *J. Med. Chem.* **52:** 564–568

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

Cogoi S, Zorzet S, Rapozzi V, Géci I, Pedersen EB & Xodo LE (2013) MAZ-binding G4-decoy with locked nucleic acid and twisted intercalating nucleic acid modifications suppresses KRAS in pancreatic cancer cells and delays tumor growth in mice. *Nucleic Acids Res.* **41:** 4049–4064

Cohen PA, Donini CF, Nguyen NT, Lincet H & Vendrell JA (2015) The dark side of ZNF217, a key regulator of tumorigenesis with powerful biomarker value. *Oncotarget* **6:** 41566–41581

Cordin O, Banroques J, Tanner NK & Linder P (2006) The DEAD-box protein family of RNA helicases. *Gene* **367:** 17–37

Coussens NP, Kales SC, Henderson MJ, Lee OW, Horiuchi KY, Wang Y, Chen Q, Kuznetsova E, Wu J, Cheff DM, Cheng KC-C, Shinn P, Brimacombe KR, Shen M, Simeonov A, Ma H, Jadhav A & Hall MD (2017) Small Molecule Inhibitors of the Human Histone Lysine Methyltransferase NSD2 / WHSC1 / MMSET Identified from a Quantitative High-Throughput Screen with Nucleosome Substrate. *bioRxiv***: 208439

Crabbe L (2004) Defective Telomere Lagging Strand Synthesis in Cells Lacking WRN Helicase Activity. *Science (80-.).* **306:** 1951–1953

Creacy SD, Routh ED, Iwamoto F, Nagamine Y, Akman SA & Vaughn JP (2008) G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.* **283:** 34626–34634

Crozat A, Åman P, Mandahl N & Ron D (1993) Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature* **363:** 640–644

Cruz JA & Westhof E (2009) The dynamic landscapes of RNA architecture. *Cell* **136:** 604–609

Cullmann G, Fien K, Kobayashi R & Stillman B (1995) Characterization of the five replication factor C genes of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **15:** 4661–4671

Dacheux E, Vincent A, Nazaret N, Combet C, Wierinckx A, Mazoyer S, Diaz J-J, Lachuer J & Venezia ND (2013) BRCA1-Dependent Translational Regulation in Breast Cancer Cells. *PLoS One* **8:** e67313

Daly AK (2017) Pharmacogenetics: a general review on progress to date. *Br. Med. Bull.* **124:** 1–15

Daniel WA, Wójcikowski J & Pałucha A (2001) Intracellular distribution of psychotropic drugs in the grey and white matter of the brain: the role of lysosomal trapping. *Br. J. Pharmacol.* **134:** 807–814

Dantzer F, Nasheuer HP, Vonesch JL, de Murcia G & Ménissier-de Murcia J (1998) Functional association of poly(ADP-ribose) polymerase with DNA polymerase alpha-primase complex: a link between DNA strand break detection and DNA replication. *Nucleic Acids Res.* **26:** 1891–1898

Danussi C, Bose P, Parthasarathy PT, Silberman PC, Van Arnam JS, Vitucci M, Tang OY, Heguy A, Wang Y, Chan TA, Riggins GJ, Sulman EP, Lang F, Creighton CJ, Deneen B, Miller CR, Picketts DJ, Kannan K & Huse JT (2018) Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling. *Nat. Commun.* **9:** 1057; 1-15

Darnell JC, Jensen KB, Jin P, Brown V, Warren ST & Darnell RB (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107:** 489–499

Dasari S & Tchounwou PB (2014) Cisplatin in cancer therapy: molecular mechanisms of action. *Eur. J. Pharmacol.* **740:** 364–378

Davis JT (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.* **43:** 668–698

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

Demagny H & De Robertis EM (2016) Point mutations in the tumor suppressor Smad4/DPC4 enhance its phosphorylation by GSK3 and reversibly inactivate TGF-β signaling. *Mol. Cell. Oncol.* **3:** e1025181

Denli AM, Tops BBJ, Plasterk RHA, Ketting RF & Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* **432:** 231–235

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22:** 1775–1789

Di Antonio M, Biffi G, Mariani A, Raiber E-A, Rodriguez R & Balasubramanian S (2012) Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chemie Int. Ed.* **51:** 11073–11078

DiMasi JA, Hansen RW & Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22:** 151–185

DiMasi JA, Hansen RW, Grabowski HG & Lasagna L (1995) Research and development costs for new drugs by therapeutic category. A study of the US pharmaceutical industry. *Pharmacoeconomics* **7:** 152–169

Dizin E, Gressier C, Magnard C, Ray H, Décimo D, Ohlmann T & Dalla Venezia N (2006) BRCA1 interacts with poly(A)-binding protein: implication of BRCA1 in translation regulation. *J. Biol. Chem.* **281:** 24236–24246

Dong S, Han J, Chen H, Liu T, Huen MSY, Yang Y, Guo C & Huang J (2014) The Human SRCAP Chromatin Remodeling Complex Promotes DNA-End Resection. *Curr. Biol.* **24:** 2097–2110

Donlic A & Hargrove AE (2018) Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip. Rev. RNA* **9:** e1477

Doudna JA & Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346:** 1077-1084

Downey M, Houlsworth R, Maringele L, Rollie A, Brehme M, Galicia S, Guillard S, Partington M, Zubko MK, Krogan NJ, Emili A, Greenblatt JF, Harrington L, Lydall D & Durocher D (2006) A genome-wide screen identifies the evolutionarily conserved KEOPS complex as a telomere regulator. *Cell* **124:** 1155–1168

Dragon F, Pogačić V & Filipowicz W (2000) In vitro assembly of human H/ACA small

nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol. Cell. Biol.* **20:** 3037–3048

Drosopoulos WC, Kosiyatrakul ST & Schildkraut CL (2015) BLM helicase facilitates telomere replication during leading strand synthesis of telomeres. *J. Cell Biol.* **210:** 191–208

Drygin D, Lin A, Bliesath J, Ho CB, O'Brien SE, Proffitt C, Omori M, Haddach M, Schwaebe MK, Siddiqui-Jain A, Streiner N, Quin JE, Sanij E, Bywater MJ, Hannan RD, Ryckman D, Anderes K & Rice WG (2011) Targeting RNA Polymerase I with an Oral Small Molecule CX-5461 Inhibits Ribosomal RNA Synthesis and Solid Tumor Growth. *Cancer Res.* **71:** 1418–1430

Dvinge H, Kim E, Abdel-Wahab O & Bradley RK (2016) RNA splicing factors as oncoproteins and tumour suppressors. **16:** 413-430

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

Eddy J & Maizels N (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **36:** 1321–1333

Edelmann MJ, Nicholson B & Kessler BM (2011) Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases. *Expert Rev. Mol. Med.* **13:** e35

Edenberg ER, Downey M & Toczyski D (2014) Polymerase stalling during replication, transcription and translation. *Curr. Biol.* **24:** 445-452

Eiden LE, Schafer MK-H, Weihe E & Schatz B (2004) The vesicular amine transporter family (SLC18): amine/proton antiporters required for vesicular accumulation and regulated exocytotic secretion of monoamines and acetylcholine. *Pflugers Arch. Eur. J. Physiol.* **447:** 636–640

El-Brolosy MA & Stainier DYR (2017) Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.* **13:** e1006780

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K & Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494–498

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5:** 435–445

Ellison V & Stillman B (1998) Reconstitution of recombinant human replication factor C (RFC) and identification of an RFC subcomplex possessing DNA-dependent ATPase activity. *J. Biol. Chem.* **273:** 5979–5987

Enokizono Y, Matsugami A, Uesugi S, Fukuda H, Tsuchiya N, Sugimura T, Nagao M, Nakagama H & Katahira M (2003) Destruction of quadruplex by proteins, and its biological implications in replication and telomere maintenance. *Nucleic Acids Res. Suppl.*: 231–232

Fabbro M & Henderson BR (2003) Regulation of tumor suppressors by nuclear-cytoplasmic shuttling. *Exp. Cell Res.* **282:** 59–69

Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L, La Starza R, Diverio D, Colombo E, Santucci A, Bigerna B, Pacini R, Pucciarini A, Liso A, Vignetti M, Fazi P, Meani N, Pettirossi V, Saglio G, Mandelli F, et al (2005) Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype. *N. Engl. J. Med.* **352:** 254–266

Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NMB, Jackson SP, Smith GCM & Ashworth A (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434:** 917–921

Fatica A & Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15:** 7–21

Fellmann C, Hoffmann T, Sridhar V, Hopfgartner B, Muhar M, Roth M, Lai DY, Barbosa IAM, Kwon JS, Guan Y, Sinha N & Zuber J (2013) An optimized

microRNA backbone for effective single-copy RNAi. *Cell Rep.* **5:** 1704–13

Fellmann C, Zuber J, McJunkin K, Chang K, Malone CD, Dickins RA, Xu Q, Hengartner MO, Elledge SJ, Hannon GJ & Lowe SW (2011) Functional identification of optimized RNAi triggers using a massively parallel sensor assay. *Mol. Cell* **41:** 733–746

Fernandez-Vidal A, Guitton-Sert L, Cadoret J-C, Drac M, Schwob E, Baldacci G, Cazaux C & Hoffmann J-S (2014) A role for DNA polymerase θ in the timing of DNA replication. *Nat. Commun.* **5:** 4285; 1-10

Fernando H, Rodriguez R & Balasubramanian S (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* **47:** 9365–9371

Feuerhahn S, Chen L-Y, Luke B & Porro A (2015) No DDRama at chromosome ends: TRF2 takes centre stage. *Trends Biochem. Sci.* **40:** 275-285

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J & Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44:** 279–285

Fioani M, Lucchini G & Plevani P (1997) The DNA polymerase α-primase complex couples DNA replication, cell-cycle progression and DNA-damage response. *Trends Biochem. Sci.* **22:** 424–427

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE & Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806–811

Fisette J-F, Montagna DR, Mihailescu M-R & Wolfe MS (2012) A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. *J. Neurochem.* **121:** 763–773

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S & Bentley DL (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell* **60:** 256–267

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U & Campbell PJ (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43:** 805-811

Frank Xiaoguang Han, Richard T. Wheelhouse and & Hurley LH (1999) Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *Journal of the American Chemical Society* **121**: 3561-3570

Franklin RE & Gosling RG (1953) Molecular Configuration in Sodium Thymonucleate. *Nature* **171:** 740–741

Friday B, Lassere Y, Meyers CA, Mita A, Abbruzzese JL & Thomas MB (2012) A phase I study to determine the safety and pharmacokinetics of intravenous administration of TAS-106 once per week for three consecutive weeks every 28 days in patients with solid tumors. *Anticancer Res.* **32:** 1689–1696

Friedemann J, Grosse F & Zhang S (2005) Nuclear DNA Helicase II (RNA Helicase A) Interacts with Werner Syndrome Helicase and Stimulates Its Exonuclease Activity. *J. Biol. Chem.* **280:** 31303–31313

Froimchuk E, Jang Y & Ge K (2017) Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627:** 337–342

Fukuchi K, Martin GM & Monnat RJ (1989) Mutator phenotype of Werner syndrome is characterized by extensive deletions. *Proc. Natl. Acad. Sci. U. S. A.* **86:** 5893–5897

Furukawa T & Tanese N (2000) Assembly of Partial TFIID Complexes in Mammalian Cells Reveals Distinct Activities Associated with Individual TATA Box-binding Protein-associated Factors. *J. Biol. Chem.* **275:** 29847–29856

Gagou ME, Ganesh A, Phear G, Robinson D, Petermann E, Cox A & Meuth M

(2014) Human PIF1 helicase supports DNA replication and cell growth under oncogenic-stress. *Oncotarget* **5:** 11381–11398

Gaj T, Gersbach CA & Barbas CF (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31:** 397–405

Ganot P, Bortolin ML & Kiss T (1997a) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89:** 799–809

Ganot P, Caizergues-Ferrer M & Kiss T (1997b) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11:** 941–956

Gao C, Pang L, Ren C & Ma T (2012) Decreased expression of Nedd4L correlates with poor prognosis in gastric cancer patient. *Med. Oncol.* **29:** 1733–1738

García-Santisteban I, Peters GJ, Giovannetti E & Rodríguez J (2013) USP1 deubiquitinase: cellular functions, regulatory mechanisms and emerging potential as target in cancer therapy. *Mol. Cancer* **12:** 1-12

Gebhart E, Bauer R, Raub U, Schinzel M, Ruprecht KW & Jonas JB (1988) Spontaneous and induced chromosomal instability in Werner syndrome. *Hum. Genet.* **80:** 135–139

Gerhards NM & Rottenberg S (2018) New tools for old drugs: Functional genetic screens to optimize current chemotherapy. *Drug Resist. Updat.* **36:** 30–46

Giacinti C & Giordano A (2006) RB and cell cycle progression. *Oncogene* **25:** 5220–5227

Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D & Lapeyre B (1992) GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J.* **11:** 673–682

Giri B, Smaldino PJ, Thys RG, Creacy SD, Routh ED, Hantgan RR, Lattmann S, Nagamine Y, Akman SA & Vaughn JP (2011) G4 Resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res.* **39:** 7161–7178

Gisselsson D, Jonson T, Petersén A, Strömbeck B, Dal Cin P, Höglund M, Mitelman F, Mertens F & Mandahl N (2001) Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc. Natl. Acad. Sci. U. S. A.* **98:** 12683–12688

Gomez D, Guédin A, Mergny J-L, Salles B, Riou J-F, Teulade-Fichou M-P & Calsou P (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.* **38:** 7187–7198

Gomez D, Lemarteleur T, Lacroix L, Mailliet P, Mergny J-L & Riou J-F (2004a) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* **32:** 371–379

Gomez D, Paterski R, Lemarteleur T, Shin-Ya K, Mergny J-L & Riou J-F (2004b) Interaction of telomestatin with the telomeric single-strand overhang. *J. Biol. Chem.* **279:** 41487–41494

Gomez D, Wenner T, Brassart B, Douarre C, O'Donohue M-F, El Khoury V, Shin-Ya K, Morjani H, Trentesaux C & Riou J-F (2006) Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J. Biol. Chem.* **281:** 38721–38729

González V, Guo K, Hurley L & Sun D (2009) Identification and Characterization of Nucleolin as a c- *myc* G-quadruplex-binding Protein. *J. Biol. Chem.* **284:** 23622–23635

González V & Hurley LH (2010) The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **50:** 111–129

Gottesman MM (2002) Mechanisms of Cancer Drug Resistance. *Annu. Rev. Med.* **53:** 615–627

Gowan SM, Heald R, Stevens MF & Kelland LR (2001) Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with

G-quadruplexes. *Mol. Pharmacol.* **60:** 981–988

Gray LT, Vallur AC, Eddy J & Maizels N (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10:** 313–318

Griffith M, Griffith OL, Coffman AC, Weible J V, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, Kumar RD, Bose R, Ding L, Walker JR, Larson DE, Dooling DJ, Smith SM, et al (2013) DGIdb: mining the druggable genome. *Nat. Methods* **10:** 1209–1210

Grimm D, Pandey K & Kay MA (2005) Adeno-associated virus vectors for short hairpin RNA expression. *Methods Enzymol.* **392:** 381–405

Guédin A, Gros J, Alberti P & Mergny J-L (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38:** 7858–7868

Guo JU & Bartel DP (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353:** 1382-1390

Guo Y, Kartawinata M, Li J, Pickett HA, Teo J, Kilo T, Barbaro PM, Keating B, Chen Y, Tian L, Al-Odaib A, Reddel RR, Christodoulou J, Xu X, Hakonarson H & Bryan TM (2014) Inherited bone marrow failure associated with germline mutation of ACD, the gene encoding telomere protein TPP1. *Blood* **124:** 2767–2774

Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH & Kern SE (1996) DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271:** 350–353

Haider SM, Neidle S & Parkinson GN (2011) A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **93:** 1239–1251

Halder K & Chowdhury S (2007) Quadruplex-coupled kinetics distinguishes ligand binding between G4 DNA motifs. *Biochemistry* **46:** 14762–14770

Halder K, Halder R & Chowdhury S (2009) Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. *Mol. Biosyst.* **5:** 1703–1712

Halder R, Riou J-F, Teulade-Fichou M-P, Frickey T & Hartig JS (2012) Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res. Notes* **5:** 1-11

Han H, Bennett RJ & Hurley LH (2000) Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N'-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry* **39:** 9311–9316

Han H & Hurley LH (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **21:** 136–142

Han K, Jeng EE, Hess GT, Morgens DW, Li A & Bassik MC (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35:** 463–474

Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144:** 646–674

Hänsel-Hertsch R, Di Antonio M & Balasubramanian S (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18:** 279–284

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

Harkness RW & Mittermaier AK (2017) G-quadruplex dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **1865:** 1544–1554

Harrigan JA, Jacq X, Martin NM & Jackson SP (2017) Deubiquitylating enzymes and drug discovery: emerging opportunities. *Nat. Rev. Drug Discov.* **17:** 57-78

Hartwell L (1992) Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell* **71:** 543–546

Hartwell LH, Szankasi P, Roberts CJ, Murray AW & Friend SH (1997) Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278:** 1064–1068

Hayashi N & Murakami S (2002) STM1, a gene which encodes a guanine quadruplex binding protein, interacts with CDC13 in Saccharomyces cerevisiae. *Mol. Genet. Genomics* **267:** 806–813

Hazel P, Huppert J, Balasubramanian S & Neidle S (2004) Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **126:** 16405–16415

He L, Vasiliou K & Nebert DW (2009) Analysis and update of the human solute carrier (SLC) gene superfamily. *Hum. Genomics* **3:** 195–206

Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, Brosh RM, Sen D & Lansdorp PM (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42:** 860–869

Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang L-S & Johnson FB (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.* **36:** 144–156

Hickson ID (2003) RecQ helicases: caretakers of the genome. *Nat. Rev. Cancer* **3:** 169–178

Hirashima K & Seimiya H (2015) Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* **43:** 2022–2032

Hohmann AF, Martin LJ, Minder JL, Roe J-S, Shi J, Steurer S, Bader G, McConnell D, Pearson M, Gerstberger T, Gottschamel T, Thompson D, Suzuki Y, Koegl M & Vakoc CR (2016) Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat. Chem. Biol.* **12:** 672–679

Holden J, Taylor EM & Lindsay HD (2017) Cip29 is phosphorylated following activation of the DNA damage response in Xenopus egg extracts. *PLoS One* **12:** e0181131

Holliday LS (2014) Vacuolar H $^+$ -ATPase: An Essential Multitasking Enzyme in Physiology and Pathophysiology. *New J. Sci.* **2014:** 1–21

Hoogsteen K (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16:** 907–916

Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Nakatsuka A, Kunichika T, Obuse C & Waga S (2013) Human Origin Recognition Complex Binds Preferentially to G-quadruplex-preferable RNA and Single-stranded DNA. *J. Biol. Chem.* **288:** 30161–30171

Howard HC, Mount DB, Rochefort D, Byun N, Dupré N, Lu J, Fan X, Song L, Rivière J-B, Prévost C, Horst J, Simonati A, Lemcke B, Welch R, England R, Zhan FQ, Mercado A, Siesser WB, George AL, McDonald MP, et al (2002) The K–Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat. Genet.* **32:** 384–392

Huang TT, Nijman SMB, Mirchandani KD, Galardy PJ, Cohn MA, Haas W, Gygi SP, Ploegh HL, Bernards R & D'Andrea AD (2006) Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nat. Cell Biol.* **8:** 341–347

Huang W-C, Tseng T-Y, Chen Y-T, Chang C-C, Wang Z-F, Wang C-L, Hsu T-N, Li P-T, Chen C-T, Lin J-J, Lou P-J & Chang T-C (2015) Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.* **43:** 10102–13

Huang W, Smaldino PJ, Zhang Q, Miller LD, Cao P, Stadelman K, Wan M, Giri B, Lei M, Nagamine Y, Vaughn JP, Akman SA & Sui G (2012) Yin Yang 1 contains G-quadruplex structures in its promoter and 5′-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res.* **40:** 1033–1049

Huang X & Dixit VM (2016) Drugging the undruggables: exploring the ubiquitin system for drug development. *Cell Res.* **26:** 484–498

Huber MD, Duquette ML, Shiels JC & Maizels N (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J. Mol. Biol.* **358:** 1071–1080

Huber MD, Lee DC & Maizels N (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.* **30:** 3954–3961

Hud N V, Schultze P, Sklenář V & Feigon J (1999) Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J. Mol. Biol.* **285:** 233–243

Hudson JS, Ding L, Le V, Lewis E & Graves D (2014) Recognition and binding of human telomeric G-quadruplex DNA by unfolding protein 1. *Biochemistry* **53:** 3347–3356

Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, Meloon B, Engel S, Rosenberg A, Cohen D, Labow M, Reinhardt M, Natt F & Hall J (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.* **23:** 995–1001

Huppert JL & Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33:** 2908–2916

Huppert JL & Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35:** 406–413

Hwang H, Buncher N, Opresko PL & Myong S (2012) POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure* **20:** 1872–1880

Jackson RJ, Hellen CUT & Pestova T V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11:** 113–127

Jayaraj GG, Pandey S, Scaria V & Maiti S (2012) Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol.* **9:** 81–86

Jelinic P, Mueller JJ, Olvera N, Dao F, Scott SN, Shah R, Gao J, Schultz N, Gonen M, Soslow RA, Berger MF & Levine DA (2014) Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat. Genet.* **46:** 424–426

Jeong W, Kim HS, Kim YB, Kim MA, Lim W, Kim J, Jang H-J, Suh DH, Kim K, Chung HH, Bazer FW, Song YS, Han JY & Song G (2012) Paradoxical expression of *AHCYL1* affecting ovarian carcinogenesis between chickens and women. *Exp. Biol. Med.* **237:** 758–767

Johnson JE, Cao K, Ryvkin P, Wang L-S & Johnson FB (2010) Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* **38:** 1114–1122

Johnson JE, Smith JS, Kozak ML & Johnson FB (2008) In vivo veritas: Using yeast to probe the biological functions of G-quadruplexes. *Biochimie* **90:** 1250–1263

Josling GA, Selvarajah SA, Petter M & Duffy MF (2012) The role of bromodomain proteins in regulating gene expression. *Genes (Basel).* **3:** 320–343

Juhász S, Elbakry A, Mathes A & Löbrich M (2018) ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. *Mol. Cell*. **71:** 11-24

Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J & Crabtree GR (2013) Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45:** 592–601

Kafri R, Springer M & Pilpel Y (2009) Genetic redundancy: new tricks for old genes. *Cell* **136:** 389–92

Kamath-Loeb AS, Loeb LA, Johansson E, Burgers PM & Fry M (2001) 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.* **276:** 16439–16446

Kamileri I, Karakasilioti I & Garinis GA (2012) Nucleotide excision repair: new tricks with old bricks. *Trends Genet.* **28:** 566–573

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael

JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ & Ding L (2013) Mutational landscape and significance across 12 major cancer types. *Nature* **502:** 333–339

Kang H-J, Le TVT, Kim K, Hur J, Kim KK & Park H-J (2014) Novel interaction of the Z-DNA binding domain of human ADAR1 with the oncogenic c-Myc promoter G-quadruplex. *J. Mol. Biol.* **426:** 2594–2604

Kang H-J & Park H-J (2009) Novel molecular mechanism for actinomycin D activity as an oncogenic promoter G-quadruplex binder. *Biochemistry* **48:** 7392–7398

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

Kato M, Wei M, Yamano S, Kakehashi A, Tamada S, Nakatani T & Wanibuchi H (2012) DDX39 acts as a suppressor of invasion for bladder cancer. *Cancer Sci.* **103:** 1363–1369

Kavanaugh GM, Wise-Draper TM, Morreale RJ, Morrison MA, Gole B, Schwemberger S, Tichy ED, Lu L, Babcock GF, Wells JM, Drissi R, Bissler JJ, Stambrook PJ, Andreassen PR, Wiesmüller L & Wells SI (2011) The human DEK oncogene regulates DNA damage response signaling and repair. *Nucleic Acids Res.* **39:** 7465–7476

Kazmi F, Hensley T, Pope C, Funk RS, Loewen GJ, Buckley DB & Parkinson A (2013) Lysosomal sequestration (trapping) of lipophilic amine (cationic amphiphilic) drugs in immortalized human hepatocytes (Fa2N-4 cells). *Drug Metab. Dispos.* **41:** 897–905

Keniry MA Quadruplex structures in nucleic acids. *Biopolymers* **56:** 123–46

Kersten K, de Visser KE, van Miltenburg MH & Jonkers J (2017) Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol. Med.* **9:** 137–153

Khodadoust MS, Verhaegen M, Kappes F, Riveiro-Falkenbach E, Cigudosa JC, Kim DSL, Chinnaiyan AM, Markovitz DM & Soengas MS (2009) Melanoma Proliferation and Chemoresistance Controlled by the DEK Oncogene. *Cancer Res.* **69:** 6405–6413

Khvorova A, Reynolds A & Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115:** 209–216

Kikin O, Zappala Z, D'Antonio L & Bagga PS (2008) GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs. *Nucleic Acids Res.* **36:** 141-148

Kikuma T, Ohtsu M, Utsugi T, Koga S, Okuhara K, Eki T, Fujimori F & Murakami Y (2004) Dbp9p, a Member of the DEAD Box Protein Family, Exhibits DNA Helicase Activity. *J. Biol. Chem.* **279:** 20692–20698

Kikuta K, Kubota D, Saito T, Orita H, Yoshida A, Tsuda H, Suehara Y, Katai H, Shimada Y, Toyama Y, Sato K, Yao T, Kaneko K, Beppu Y, Murakami Y, Kawai A & Kondo T (2012) Clinical proteomics identified ATP-dependent RNA helicase DDX39 as a novel biomarker to predict poor prognosis of patients with gastrointestinal stromal tumor. *J. Proteomics* **75:** 1089–1098

Kim M-Y, Gleason-Guzman M, Izbicka E, Nishioka D & Hurley LH (2003) The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Res.* **63:** 3247–3256

Kim M-Y, Vankayalapati H, Shin-ya K, Wierzba K & Hurley LH (2002) Telomestatin, a Potent Telomerase Inhibitor That Interacts Quite Specifically with the Human Telomeric Intramolecular G-Quadruplex. *J. Am. Chem. Soc.* **124:** 2098–2099

Kim N (2017) The interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **25:** 1-19

Kinzler KW & Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell*

87: 159–170

Kishikawa S, Murata T, Kimura H, Shiota K & Yokoyama KK (2002) Regulation of transcription of the Dnmt1 gene by Sp1 and Sp3 zinc finger proteins. *Eur. J. Biochem.* **269:** 2961–2970

Kiss T (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20:** 3617–3622

Knott SR V, Maceli A, Erard N, Chang K, Marran K, Zhou X, Gordon A, Demerdash O El, Wagenblast E, Kim S, Fellmann C & Hannon GJ (2014) A computational algorithm to predict shRNA potency. *Mol. Cell* **56:** 796–807

Ko LJ & Prives C (1996) p53: puzzle and paradigm. *Genes Dev.* **10:** 1054–1072

Kobayashi Y, Sato K, Kibe T, Seimiya H, Nakamura A, Yukawa M, Tsuchiya E & Ueno M (2010) Expression of Mutant RPA in Human Cancer Cells Causes Telomere Shortening. *Biosci. Biotechnol. Biochem.* **74:** 382–385

Kondo K, Mashima T, Oyoshi T, Yagi R, Kurokawa R, Kobayashi N, Nagata T & Katahira M (2018) Plastic roles of phenylalanine and tyrosine residues of TLS/FUS in complex formation with the G-quadruplexes of telomeric DNA and TERRA. *Sci. Rep.* **8:** 1-11

König SLB, Evans AC & Huppert JL (2010) Seven essential questions on G-quadruplexes. *Biomol. Concepts* **1:** 197–213

Kotake Y, Sagane K, Owa T, Mimori-Kiyosue Y, Shimizu H, Uesugi M, Ishihama Y, Iwata M & Mizui Y (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.* **3:** 570–575

Kreig A, Calvert J, Sanoica J, Cullum E, Tipanna R & Myong S (2015) G-quadruplex formation in double strand DNA probed by NMM and CV fluorescence. *Nucleic Acids Res.* **43:** 7961–7970

Krig SR, Jin VX, Bieda MC, O'Geen H, Yaswen P, Green R & Farnham PJ (2007) Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. *J. Biol. Chem.* **282:** 9703–9712

Krüger AC, Raarup MK, Nielsen MM, Kristensen M, Besenbacher F, Kjems J & Birkedal V (2010) Interaction of hnRNP A1 with telomere DNA G-quadruplex structures studied at the single molecule level. *Eur. Biophys. J.* **39:** 1343–1350

Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E & Tijsterman M (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. *Curr. Biol.* **18:** 900–905

Kumari S, Bugaut A, Huppert JL & Balasubramanian S (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3:** 218–221

Kuramitsu Y, Suenaga S, Wang Y, Tokuda K, Kitagawa T, Tanaka T, Akada J, Maehara S-I, Maehara Y & Nakamura K (2013a) Up-regulation of DDX39 in human pancreatic cancer cells with acquired gemcitabine resistance compared to gemcitabine-sensitive parental cells. *Anticancer Res.* **33:** 3133–3136

Kuramitsu Y, Tominaga W, Baron B, Tokuda K, Wang Y, Kitagawa T & Nakamura K (2013b) Up-regulation of DDX39 in human malignant pleural mesothelioma cell lines compared to normal pleural mesothelial cells. *Anticancer Res.* **33:** 2557–2560

Kuscu C, Arslan S, Singh R, Thorpe J & Adli M (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32:** 677–683

Kwok CK & Balasubramanian S (2015) Targeted Detection of G-Quadruplexes in Cellular RNAs. *Angew. Chemie Int. Ed.* **54:** 6751–6754

Kwok CK, Marsico G, Sahakyan AB, Chambers VS & Balasubramanian S (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13:** 841–844

De La Fuente R, Baumann C & Viveiros MM (2011) Role of ATRX in chromatin

structure and function: implications for chromosome instability and human disease. *Reproduction* **142:** 221–234

De La Fuente R, Viveiros MM, Wigglesworth K & Eppig JJ (2004) ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev. Biol.* **272:** 1–14

Lagah S, Tan I-L, Radhakrishnan P, Hirst RA, Ward JH, O'Callaghan C, Smith SJ, Stevens MFG, Grundy RG & Rahman R (2014) RHPS4 G-quadruplex ligand induces anti-proliferative effects in brain tumor cells. *PLoS One* **9:** e86187

Lam EYN, Beraldi D, Tannahill D & Balasubramanian S (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **4:** 1-8

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, et al (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921

de Lange T (2009) How telomeres solve the end-protection problem. *Science* **326:** 948–952

Law MJ, Lower KM, Voon HPJ, Hughes JR, Garrick D, Viprakasit V, Mitson M, De Gobbi M, Marra M, Morris A, Abbott A, Wilder SP, Taylor S, Santos GM, Cross J, Ayyub H, Jones S, Ragoussis J, Rhodes D, Dunham I, et al (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143:** 367–378

Le DD, Di Antonio M, Chan LKM & Balasubramanian S (2015) G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun. (Camb).* **51:** 8048–50

Lee D-F, Walsh MJ & Aguiló F (2016) ZNF217/ZFP217 Meets Chromatin and RNA. *Trends Biochem. Sci.* **41:** 986–988

Lee J -m., Ledermann JA & Kohn EC (2014) PARP Inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies. *Ann. Oncol.* **25:** 32–40

Lee SC-W & Abdel-Wahab O (2016) Therapeutic targeting of splicing in cancer. *Nat. Med.* **22:** 976–986

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S & Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415–419

Lemarteleur T, Gomez D, Paterski R, Mandine E, Mailliet P & Riou J-F (2004) Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.* **323:** 802–808

Lengauer C, Kinzler KW & Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* **396:** 643–649

León-Ortiz AM, Svendsen J & Boulton SJ (2014) Metabolism of DNA secondary structures at the eukaryotic replication fork. *DNA Repair (Amst).* **19:** 152–162

Leonard AC & Méchali M (2013) DNA replication origins. *Cold Spring Harb. Perspect. Biol.* **5:** a010116

Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP & Joenje H (2005) The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat. Genet.* **37:** 934–935

Liang X-Q, Cao E-H, Zhang Y & Qin J-F (2004) A P53 target gene, PIG11, contributes to chemosensitivity of cells to arsenic trioxide. *FEBS Lett.* **569:** 94–98

Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN & Phan AT (2009) Structure of the Human Telomere in K $^+$ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **131:** 4301–4309

Lim SP, Wong NC, Suetani RJ, Ho K, Ng JL, Neilsen PM, Gill PG, Kumar R & Callen DF (2012) Specific-site methylation of tumour suppressor ANKRD11 in breast cancer. *Eur. J. Cancer* **48:** 3300–3309

Lin C-W, Cheng C-W, Yang T-C, Li S-W, Cheng M-H, Wan L, Lin Y-J, Lai C-H, Lin W-Y & Kao M-C (2008) Interferon antagonist function of Japanese encephalitis virus NS4A and its interaction with DEAD-box RNA helicase DDX42. *Virus Res.* **137:** 49–55

Lin L, Piao J, Ma Y, Jin T, Quan C, Kong J, Li Y & Lin Z (2014) Mechanisms Underlying Cancer Growth and Apoptosis by DEK Overexpression in Colorectal Cancer. *PLoS One* **9:** e111260

Lin W, Sampathi S, Dai H, Liu C, Zhou M, Hu J, Huang Q, Campbell J, Shin-Ya K, Zheng L, Chai W & Shen B (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *EMBO J.* **32:** 1425–1439

Linder P & Jankowsky E (2011) From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12:** 505–516

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46:** 3–26

Littlepage LE, Adler AS, Kouros-Mehr H, Huang G, Chou J, Krig SR, Griffith OL, Korkola JE, Qu K, Lawson DA, Xue Q, Sternlicht MD, Dijkgraaf GJP, Yaswen P, Rugo HS, Sweeney CA, Collins CC, Gray JW, Chang HY & Werb Z (2012) The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression. *Cancer Discov.* **2:** 638–651

Liu J, Shaik S, Dai X, Wu Q, Zhou X, Wang Z & Wei W (2015a) Targeting the ubiquitin pathway for cancer treatment. *Biochim. Biophys. Acta* **1855:** 50–60

Liu Y, Beyer A & Aebersold R (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell.* **165:** 535-550

Liu Y, Liang Y & Wishart D (2015b) PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res.* **43:** 535–542

Liu YP, Haasnoot J, ter Brake O, Berkhout B & Konstantinova P (2008) Inhibition of HIV-1 by multiple siRNAs expressed from a single microRNA polycistron. *Nucleic Acids Res.* **36:** 2811–2824

Liu Z & Gilbert W (1994) The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure. *Cell* **77:** 1083–1092

Loeb LA (1991) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51:** 3075–3079

London TBC, Barber LJ, Mosedale G, Kelly GP, Balasubramanian S, Hickson ID, Boulton SJ & Hiom K (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283:** 36132–36139

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

Lu M-C, Ji J-A, Jiang Y-L, Chen Z-Y, Yuan Z-W, You Q-D & Jiang Z-Y (2016) An inhibitor of the Keap1-Nrf2 protein-protein interaction protects NCM460 colonic cells and alleviates experimental colitis. *Sci. Rep.* **6:** 1-13

Lu S, Sung T, Lin N, Abraham RT & Jessen BA (2017) Lysosomal adaptation: How cells respond to lysosomotropic compounds. *PLoS One* **12:** e0173771

Lu Y, Liu P, James M, Vikis HG, Liu H, Wen W, Franklin A & You M (2010) Genetic variants cis-regulating Xrn2 expression contribute to the risk of spontaneous lung tumor. *Oncogene* **29:** 1041–1049

Lyonnais S, Tarrés-Solé A, Rubio-Cosials A, Cuppari A, Brito R, Jaumot J, Gargallo

R, Vilaseca M, Silva C, Granzhan A, Teulade-Fichou M-P, Eritja R & Solà M (2017) The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. *Sci. Rep.* **7:** 1-15

MacKeigan JP, Murphy LO & Blenis J (2005) Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nat. Cell Biol.* **7:** 591–600

Madireddy A, Purushothaman P, Loosbroock CP, Robertson ES, Schildkraut CL & Verma SC (2016) G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. *Nucleic Acids Res.* **44:** 3675–3694

Maicher A, Lockhart A & Luke B (2014) Breaking new ground: digging into TERRA function. *Biochim. Biophys. Acta* **1839:** 387–394

Makarov VL, Hirose Y & Langmore JP (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88:** 657–666

Mali P, Esvelt KM & Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10:** 957–963

Manjunath N, Wu H, Subramanya S & Shankar P (2009) Lentiviral delivery of short hairpin RNAs. *Adv. Drug Deliv. Rev.* **61:** 732–745

Marcel V, Tran PLT, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou M-P, Hall J, Mergny J-L, Hainaut P & Van Dyck E (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32:** 271–278

Martadinata H, Heddi B, Lim KW & Phan AT (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50:** 6455–6461

Martens-de Kemp SR, Brink A, van der Meulen IH, de Menezes RX, Te Beest DE, Leemans CR, van Beusechem VW, Braakhuis BJM & Brakenhoff RH (2017) The FA/BRCA Pathway Identified as the Major Predictor of Cisplatin Response in Head and Neck Cancer by Functional Genomics. *Mol. Cancer Ther.* **16:** 540–550

Martín-Martín N, Carracedo A & Torrano V (2017) Metabolism and Transcription in Cancer: Merging Two Classic Tales. *Front. cell Dev. Biol.* **5:** 1-8

Martin SA, Hewish M, Sims D, Lord CJ & Ashworth A (2011) Parallel high-throughput RNA interference screens identify PINK1 as a potential therapeutic target for the treatment of DNA mismatch repair-deficient cancers. *Cancer Res.* **71:** 1836–1848

Martin SA, McCabe N, Mullarkey M, Cummins R, Burgess DJ, Nakabeppu Y, Oka S, Kay E, Lord CJ & Ashworth A (2010) DNA polymerases as potential therapeutic targets for cancers deficient in the DNA mismatch repair proteins MSH2 or MLH1. *Cancer Cell* **17:** 235–248

Marzec P, Armenise C, Pérot G, Roumelioti F-M, Basyuk E, Gagos S, Chibon F & Déjardin J (2015) Nuclear-Receptor-Mediated Telomere Insertion Leads to Genome Instability in ALT Cancers. *Cell* **160:** 913–927

Maser RS & DePinho RA (2002) Keeping telomerase in its place. *Nat. Med.* **8:** 934–6

Masuda-Sasa T, Polaczek P, Peng XP, Chen L & Campbell JL (2008) Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition. *J. Biol. Chem.* **283:** 24359–24373

Matheson CJ, Backos DS & Reigan P (2016) Targeting WEE1 Kinase in Cancer. *Trends Pharmacol. Sci.* **37:** 872–881

Mattern MR, Wu J & Nicholson B (2012) Ubiquitin-based anticancer therapy: Carpet bombing with proteasome inhibitors vs surgical strikes with E1, E2, E3, or DUB inhibitors. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823:** 2014–2021

McBride JL, Boudreau RL, Harper SQ, Staber PD, Monteys AM, Martins I, Gilmore

BL, Burstein H, Peluso RW, Polisky B, Carter BJ & Davidson BL (2008) Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 5868–5873

McCubrey JA, Steelman LS, Bertrand FE, Davis NM, Sokolosky M, Abrams SL, Montalto G, D'Assoro AB, Libra M, Nicoletti F, Maestro R, Basecke J, Rakus D, Gizak A, Demidenko ZN, Cocco L, Martelli AM & Cervello M (2014) GSK-3 as potential target for therapeutic intervention in cancer. *Oncotarget* **5:** 2881–2911

McGeary RP, Bennett AJ, Tran QB, Cosgrove KL & Ross BP (2008) Suramin: clinical uses and structure-activity relationships. *Mini Rev. Med. Chem.* **8:** 1384–1394

McLuckie KIE, Di Antonio1. McLuckie, K. I. E. et al. G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. J. Am. Chem. Soc. 135, 9640–3 (2013). M, Zecchini H, Xian J, Caldas C, Krippendorff B-F, Tannahill D, Lowe C & Balasubramanian S (2013) G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.* **135:** 9640–9643

McMahon M, Contreras A & Ruggero D (2015) Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip. Rev. RNA* **6:** 173–189

McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11:** 786–799

Meissner M, Lopato S, Gotzmann J, Sauermann G & Barta A (2003) Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. *Exp. Cell Res.* **283:** 184–195

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

Michelini F, Pitchiaya S, Vitelli V, Sharma S, Gioia U, Pessina F, Cabrini M, Wang Y, Capozzo I, Iannelli F, Matti V, Francia S, Shivashankar G V., Walter NG & d'Adda di Fagagna F (2017) Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat. Cell Biol.* **19:** 1400–1411

Miles LA, Garippa RJ & Poirier JT (2016) Design, execution, and analysis of pooled *in vitro* CRISPR/Cas9 screens. *FEBS J.* **283:** 3170–3180

Miotto B & Graba Y (2010) Control of DNA replication: A new facet of Hox proteins? *BioEssays* **32:** 800–807

Mirihana Arachchilage G, Dassanayake AC & Basu S (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* **22:** 262–272

Mishra SK, Tawani A, Mishra A & Kumar A (2016) G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **6:** 1-9

Mitani K & Kubo S (2002) Adenovirus as an integrating vector. *Curr. Gene Ther.* **2:** 135–144

Mitchell JR, Cheng J & Collins K (1999a) A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19:** 567–576

Mitchell JR, Wood E & Collins K (1999b) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402:** 551–555

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, Carpenter AE, Foo SY, Stewart SA, Stockwell BR, Hacohen N, Hahn WC, Lander ES, Sabatini DM & Root DE (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124:** 1283–1298

Mohr SE, Smith JA, Shamu CE, Neumüller RA & Perrimon N (2014) RNAi screening comes of age: improved techniques and complementary approaches. *Nat. Rev. Mol. Cell Biol.* **15:** 591–600

Mor-Vaknin N, Saha A, Legendre M, Carmona-Rivera C, Amin MA, Rabquer BJ,

Gonzales-Hernandez MJ, Jorns J, Mohan S, Yalavarthi S, Pai DA, Angevine K, Almburg SJ, Knight JS, Adams BS, Koch AE, Fox DA, Engelke DR, Kaplan MJ & Markovitz DM (2017) DEK-targeting DNA aptamers as therapeutics for inflammatory arthritis. *Nat. Commun.* **8:** 1-13

Mori K, Lammich S, Mackenzie IRA, Forné I, Zilow S, Kretzschmar H, Edbauer D, Janssens J, Kleinberger G, Cruts M, Herms J, Neumann M, Van Broeckhoven C, Arzberger T & Haass C (2013) hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol.* **125:** 413–423

Morra F, Luise C, Visconti R, Staibano S, Merolla F, Ilardi G, Guggino G, Paladino S, Sarnataro D, Franco R, Monaco R, Zitomarino F, Pacelli R, Monaco G, Rocco G, Cerrato A, Linardopoulos S, Muller MT & Celetti A (2015) New therapeutic perspectives in CCDC6 deficient lung cancer cells. *Int. J. Cancer* **136:** 2146–2157

Morris MJ & Basu S (2009) An Unusually Stable G-Quadruplex within the 5′-UTR of the MT3 Matrix Metalloproteinase mRNA Represses Translation in Eukaryotic Cells. *Biochemistry* **48:** 5313–5319

Morris MJ, Negishi Y, Pazsint C, Schonhoft JD & Basu S (2010) An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc.* **132:** 17831–17839

Moyal L, Lerenthal Y, Gana-Weisz M, Mass G, So S, Wang S-Y, Eppink B, Chung YM, Shalev G, Shema E, Shkedy D, Smorodinsky NI, van Vliet N, Kuster B, Mann M, Ciechanover A, Dahm-Daphi J, Kanaar R, Hu MC-T, Chen DJ, et al (2011) Requirement of ATM-Dependent Monoubiquitylation of Histone H2B for Timely Repair of DNA Double-Strand Breaks. *Mol. Cell* **41:** 529–542

Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, Lovrecz GO, Beck JL & Bryan TM (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6:** 1-12

Mukundan VT & Phan AT (2013) Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **135:** 5017–5028

Mullan PB, Quinn JE & Harkin DP (2006) The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene* **25:** 5854–5863

Müller S, Sanders DA, Di Antonio M, Matsis S, Riou J-F, Rodriguez R & Balasubramanian S (2012) Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.* **10:** 6537–6546

Murat P & Balasubramanian S (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.* **25:** 22–29

Murat P, Gormally M V, Sanders D, Di Antonio M & Balasubramanian S (2013) Light-mediated in cell downregulation of G-quadruplex-containing genes using a photo-caged ligand. *Chem. Commun. (Camb).* **49:** 8453–8455

Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, Balasubramanian S, Khanna R & Tellam J (2014) G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **10:** 358–364

Nambiar M, Goldsmith G, Moorthy BT, Lieber MR, Joshi M V, Choudhary B, Hosur R V & Raghavan SC (2011) Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Res.* **39:** 936–948

Neidle S (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277:** 1118–1125

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

Neilsen PM, Cheney KM, Li C-W, Chen JD, Cawrse JE, Schulz RB, Powell JA, Kumar R & Callen DF (2008) Identification of ANKRD11 as a p53 coactivator. *J.*

Cell Sci. **121:** 3541–3552

Nepomuceno T, De Gregoriis G, de Oliveira FMB, Suarez-Kurtz G, Monteiro A & Carvalho M (2017) The Role of PALB2 in the DNA Damage Response and Cancer Predisposition. *Int. J. Mol. Sci.* **18:** 1-20

Ngo HB, Lovely GA, Phillips R & Chan DC (2014) Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat. Commun.* **5:** 1-12

Nguyen GH, Tang W, Robles AI, Beyer RP, Gray LT, Welsh JA, Schetter AJ, Kumamoto K, Wang XW, Hickson ID, Maizels N, Monnat RJ & Harris CC (2014) Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci. U. S. A.* **111:** 9905–9910

Ni J, Tien AL & Fournier MJ (1997) Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89:** 565–573

Nickoloff JA, Jones D, Lee S-H, Williamson EA & Hromas R (2017) Drugging the Cancers Addicted to DNA Repair. *J. Natl. Cancer Inst.* **109:** 1-13

Nijman SMB (2011) Synthetic lethality: General principles, utility and detection using genetic screens in human cells. *FEBS Lett.* **585:** 1–6

Nijman SMB, Huang TT, Dirac AMG, Brummelkamp TR, Kerkhoven RM, D'Andrea AD & Bernards R (2005) The Deubiquitinating Enzyme USP1 Regulates the Fanconi Anemia Pathway. *Mol. Cell* **17:** 331–339

Noll JE, Jeffery J, Al-Ejeh F, Kumar R, Khanna KK, Callen DF & Neilsen PM (2012) Mutant p53 drives multinucleation and invasion through a process that is suppressed by ANKRD11. *Oncogene* **31:** 2836–2848

O'Sullivan RJ & Karlseder J (2010) Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11:** 171–81

Oganesian L, Moon IK, Bryan TM & Jarstfer MB (2006) Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* **25:** 1148–1159

Ogloblina AM, Bannikova VA, Khristich AN, Oretskaya TS, Yakubovskaya MG & Dolinnaya NG (2015) Parallel G-quadruplexes formed by guanine-rich microsatellite repeats inhibit human topoisomerase I. *Biochem.* **80:** 1026–1038

Ohnmacht SA, Marchetti C, Gunaratnam M, Besser RJ, Haider SM, Di Vita G, Lowe HL, Mellinas-Gomez M, Diocou S, Robson M, Šponer J, Islam B, Barbara Pedley R, Hartley JA & Neidle S (2015) A G-quadruplex-binding compound showing anti-tumour activity in an in vivo model for pancreatic cancer. *Sci. Rep.* **5:** 1-11

Opresko PL, von Kobbe C, Laine J-P, Harrigan J, Hickson ID & Bohr VA (2002) Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. *J. Biol. Chem.* **277:** 41110–41119

Otsuka M, Matsumoto T, Morimoto R, Arioka S, Omote H & Moriyama Y (2005) A human transporter protein that mediates the final excretion step for toxic organic cations. *Proc. Natl. Acad. Sci.* **102:** 17923–17928

Ou T-M, Lu Y-J, Zhang C, Huang Z-S, Wang X-D, Tan J-H, Chen Y, Ma D-L, Wong K-Y, Tang JC-O, Chan AS-C & Gu L-Q (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J. Med. Chem.* **50:** 1465–1474

Paddison PJ, Caudy AA, Bernstein E, Hannon GJ & Conklin DS (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16:** 948–958

Paeschke K, Bochman ML, Garcia PD, Cejka P, Friedman KL, Kowalczykowski SC & Zakian VA (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497:** 458–462

Paeschke K, Capra JA & Zakian VA (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* **145:** 678–691

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

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

Pai C-C & Kearsey SE (2017) A Critical Balance: dNTPs and the Maintenance of Genome Stability. *Genes (Basel).* **8:** 1-14

Palm W & de Lange T (2008) How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42:** 301–334

Pan J, Deng Q, Jiang C, Wang X, Niu T, Li H, Chen T, Jin J, Pan W, Cai X, Yang X, Lu M, Xiao J & Wang P (2015) USP37 directly deubiquitinates and stabilizes c-Myc in lung cancer. *Oncogene* **34:** 3957–3967

Parry D, Guzi T, Shanahan F, Davis N, Prabhavalkar D, Wiswell D, Seghezzi W, Paruch K, Dwyer MP, Doll R, Nomeir A, Windsor W, Fischmann T, Wang Y, Oft M, Chen T, Kirschmeier P & Lees EM (2010) Dinaciclib (SCH 727965), a Novel and Potent Cyclin-Dependent Kinase Inhibitor. *Mol. Cancer Ther.* **9:** 2344–2353

Parsyan A, Shahbazian D, Martineau Y, Petroulakis E, Alain T, Larsson O, Mathonnet G, Tettweiler G, Hellen CU, Pestova T V, Svitkin Y V & Sonenberg N (2009) The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 22217–22222

Patel SS & Donmez I (2006) Mechanisms of helicases. *J. Biol. Chem.* **281:** 18265–18268

Paulsson K, Harrison CJ, Andersen MK, Chilton L, Nordgren A, Moorman A V & Johansson B (2013) Distinct patterns of gained chromosomes in high hyperdiploid acute lymphoblastic leukemia with t(1;19)(q23;p13), t(9;22)(q34;q22) or MLL rearrangements. *Leukemia* **27:** 974–977

Pearson CE, Nichol Edamura K & Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* **6:** 729–742

Pedroso IM, Hayward W & Fletcher TM (2009) The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures. *Nucleic Acids Res.* **37:** 1541–1554

Pelkmans L (2005) Secrets of caveolae- and lipid raft-mediated endocytosis revealed by mammalian viruses. *Biochim. Biophys. Acta - Mol. Cell Res.* **1746:** 295–304

Perwitasari O, Bakre A, Tompkins SM & Tripp RA (2013) siRNA Genome Screening Approaches to Therapeutic Drug Repositioning. *Pharmaceuticals (Basel).* **6:** 124–160

Phan AT (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277:** 1107–1117

Phan AT, Modi YS and & Patel DJ (2004) Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. **126:** 8710-8716

Piazza A, Adrian M, Samazan F, Heddi B, Hamon F, Serero A, Lopes J, Teulade-Fichou M-P, Phan AT & Nicolas A (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.* **34:** 1718–1734

Piazza A, Boulé J-B, Lopes J, Mingo K, Largy E, Teulade-Fichou M-P & Nicolas A (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.* **38:** 4337–4348

Piazza A, Serero A, Boulé J-B, Legoix-Né P, Lopes J & Nicolas A (2012) Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in Saccharomyces cerevisiae depends on G-quadruplexes or Cdc13. *PLoS Genet.* **8:** e1003033

Piekna-Przybylska D, Sullivan MA, Sharma G & Bambara RA (2014) U3 region in the HIV-1 genome adopts a G-quadruplex structure in its RNA and DNA sequence.

Biochemistry **53:** 2581–2593

Pikor L, Thu K, Vucic E & Lam W (2013) The detection and implication of genome instability in cancer. *Cancer Metastasis Rev.* **32:** 341–352

Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CL, Miles HT & Becker ED (1978) Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. *J. Am. Chem. Soc.* **100:** 3625–3627

Pisareva VP & Pisarev A V. (2016) DHX29 and eIF3 cooperate in ribosomal scanning on structured mRNAs during translation initiation. *RNA* **22:** 1859–1870

Pisareva VP, Pisarev A V, Komar AA, Hellen CUT & Pestova T V (2008) Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell* **135:** 1237–1250

Plevova P, Cerna D, Balcar A, Foretova L, Zapletalova J, Silhanova E, Curik R & Dvorackova J (2010) CCND1 and ZNF217 gene amplification is equally frequent in BRCA1 and BRCA2 associated and non-BRCA breast cancer. *Neoplasma* **57:** 325–332

Pogacić V, Dragon F & Filipowicz W (2000) Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell. Biol.* **20:** 9028–9040

Porro A, Feuerhahn S, Reichenbach P & Lingner J (2010) Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol. Cell. Biol.* **30:** 4808–4817

Postberg J, Tsytlonok M, Sparvoli D, Rhodes D & Lipps HJ (2012) A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene* **497:** 147–54

Premsrirut PK, Dow LE, Kim SY, Camiolo M, Malone CD, Miething C, Scuoppo C, Zuber J, Dickins RA, Kogan SC, Shroyer KR, Sordella R, Hannon GJ & Lowe SW (2011) A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* **145:** 145–158

Pui C-H & Evans WE (1998) Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **339:** 605–615

Puxeddu E, Zhao G, Stringer JR, Medvedovic M, Moretti S & Fagin JA (2005) Characterization of novel non-clonal intrachromosomal rearrangements between the H4 and PTEN genes (H4/PTEN) in human thyroid cell lines and papillary thyroid cancer specimens. *Mutat. Res.* **570:** 17–32

Quante T, Otto B, Brázdová M, Kejnovská I, Deppert W & Tolstonog G V (2012) Mutant p53 is a transcriptional co-factor that binds to G-rich regulatory regions of active genes and generates transcriptional plasticity. *Cell Cycle* **11:** 3290–3303

Quin J, Chan KT, Devlin JR, Cameron DP, Diesch J, Cullinane C, Ahern J, Khot A, Hein N, George AJ, Hannan KM, Poortinga G, Sheppard KE, Khanna KK, Johnstone RW, Drygin D, McArthur GA, Pearson RB, Sanij E & Hannan RD (2016) Inhibition of RNA polymerase I transcription initiation by CX-5461 activates non-canonical ATM/ATR signaling. *Oncotarget* **7:** 49800–49818

Quinlan KGR, Verger A, Yaswen P & Crossley M (2007) Amplification of zinc finger gene 217 (ZNF217) and cancer: when good fingers go bad. *Biochim. Biophys. Acta* **1775:** 333–340

Qureshi MH, Ray S, Sewell AL, Basu S & Balci H (2012) Replication Protein A Unfolds G-Quadruplex Structures with Varying Degrees of Efficiency. *J. Phys. Chem. B* **116:** 5588–5594

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

Rajagopalan PTR, Zhang Z, McCourt L, Dwyer M, Benkovic SJ & Hammes GG (2002) Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *Proc. Natl. Acad. Sci.* **99:** 13481–13486

Ramsay AJ, Quesada V, Foronda M, Conde L, Martínez-Trillos A, Villamor N, Rodríguez D, Kwarciak A, Garabaya C, Gallardo M, López-Guerra M, López-Guillermo A, Puente XS, Blasco MA, Campo E & López-Otín C (2013) POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45:** 526–530

Rao RC & Dou Y (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer* **15:** 334–346

Read MA, Brownell JE, Gladysheva TB, Hottelet M, Parent LA, Coggins MB, Pierce JW, Podust VN, Luo RS, Chau V & Palombella VJ (2000) Nedd8 modification of cul-1 activates SCF(beta(TrCP))-dependent ubiquitination of IkappaBalpha. *Mol. Cell. Biol.* **20:** 2326–2333

Reilly PT, Wysocka J & Herr W (2002) Inactivation of the retinoblastoma protein family can bypass the HCF-1 defect in tsBN67 cell proliferation and cytokinesis. *Mol. Cell. Biol.* **22:** 6767–6778

Reimer RJ & Edwards RH (2004) Organic anion transport is the primary function of the SLC17/type I phosphate transporter family. *Pflugers Arch. Eur. J. Physiol.* **447:** 629–635

Relling M V. & Evans WE (2015) Pharmacogenomics in the clinic. *Nature* **526:** 343–350

Ren J & Chaires JB (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry* **38:** 16067–16075

Rezler EM, Seenisamy J, Bashyam S, Kim M-Y, White E, Wilson WD & Hurley LH (2005) Telomestatin and Diseleno Sapphyrin Bind Selectively to Two Different Forms of the Human Telomeric G-Quadruplex Structure. *J. Am. Chem. Soc.* **127:** 9439–9447

Rezsohazy R (2014) Non-transcriptional interactions of Hox proteins: Inventory, facts, and future directions. *Dev. Dyn.* **243:** 117–131

Rhodes D & Lipps HJ (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43:** 8627–8637

Ribeiro JR, Lovasco LA, Vanderhyden BC & Freiman RN (2014) Targeting TBP-Associated Factors in Ovarian Cancer. *Front. Oncol.* **4:** 1-14

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

Richer AL, Cala JM, O'Brien K, Carson VM, Inge LJ & Whitsett TG (2017) WEE1 Kinase Inhibitor AZD1775 Has Preclinical Efficacy in LKB1-Deficient Non–Small Cell Lung Cancer. *Cancer Res.* **77:** 4663–4672

Richmond TJ & Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* **423:** 145–150

Risitano A & Fox KR (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **32:** 2598–2606

Ritchie K, Seah C, Moulin J, Isaac C, Dick F & Bérubé NG (2008) Loss of ATRX leads to chromosome cohesion and congression defects. *J. Cell Biol.* **180:** 315–324

Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140

Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, Tiffen JC, Petljak M, Palmer JM, Symmons J, Johansson P, Stark MS, Gartside MG, Snowden H, Montgomery GW, Martin NG, Liu JZ, Choi J, et al (2014) POT1 loss-of-function variants predispose to familial melanoma. *Nat. Genet.* **46:** 478–481

Rodriguez R, Miller KM, Forment J V, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S & Jackson SP (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human

genes. *Nat. Chem. Biol.* **8:** 301–310

Rodriguez R, Müller S, Yeoman JA, Trentesaux C, Riou J-F & Balasubramanian S (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130:** 15758–15759

Ronan JL, Wu W & Crabtree GR (2013) From neural development to cognition: unexpected roles for chromatin. *Nat. Rev. Genet.* **14:** 347–359

Rondón-Lagos M, Verdun Di Cantogno L, Marchiò C, Rangel N, Payan-Gomez C, Gugliotta P, Botta C, Bussolati G, Ramírez-Clavijo SR, Pasini B & Sapino A (2014) Differences and homologies of chromosomal alterations within and between breast cancer cell lines: a clustering analysis. *Mol. Cytogenet.* **7:** 1-14

Rottenberg S, Jaspers JE, Kersbergen A, van der Burg E, Nygren AOH, Zander SAL, Derksen PWB, de Bruin M, Zevenhoven J, Lau A, Boulter R, Cranston A, O'Connor MJ, Martin NMB, Borst P & Jonkers J (2008) High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 17079–17084

Rottenberg S, Nygren AOH, Pajic M, van Leeuwen FWB, van der Heijden I, van de Wetering K, Liu X, de Visser KE, Gilhuijs KG, van Tellingen O, Schouten JP, Jonkers J & Borst P (2007) Selective induction of chemotherapy resistance of mammary tumors in a conditional mouse model for hereditary breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 12117–12122

Rouleau M, Patel A, Hendzel MJ, Kaufmann SH & Poirier GG (2010) PARP inhibition: PARP1 and beyond. *Nat. Rev. Cancer* **10:** 293–301

Rudolph KL, Millard M, Bosenberg MW & DePinho RA (2001) Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nat. Genet.* **28:** 155–159

Ruhl DD, Jin J, Cai Y, Swanson S, Florens L, Washburn MP, Conaway RC, Conaway JW & Chrivia JC (2006) Purification of a Human SRCAP Complex That Remodels Chromatin by Incorporating the Histone Variant H2A.Z into Nucleosomes [†]. *Biochemistry* **45:** 5671–5677

Van Rysselberghe P (1935) The Fundamentals of Chemical Thermodynamics. *Chem. Rev.* **16:** 37–51

Saccà B, Lacroix L & Mergny J-L (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **33:** 1182–1192

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

Sahakyan AB, Chambers VS, Marsico G, Santner T, Di Antonio M & Balasubramanian S (2017) Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.* **7:** 1-11

Saharia A, Guittat L, Crocker S, Lim A, Steffen M, Kulkarni S & Stewart SA (2008) Flap endonuclease 1 contributes to telomere stability. *Curr. Biol.* **18:** 496–500

Salk D, Au K, Hoehn H & Martin GM (1981) Cytogenetics of Werner's syndrome cultured skin fibroblasts: variegated translocation mosaicism. *Cytogenet. Cell Genet.* **30:** 92–107

Salvati E, Rizzo A, Iachettini S, Zizza P, Cingolani C, D'Angelo C, Porru M, Mondello C, Aiello A, Farsetti A, Gilson E, Leonetti C & Biroccio A (2015) A basal level of DNA damage and telomere deprotection increases the sensitivity of cancer cells to G-quadruplex interactive compounds. *Nucleic Acids Res.* **43:** 1759–1769

Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S, Chakravarty D, Daian F, Gao Q, Bailey MH, Liang W-W, Foltz SM, Shmulevich I, Ding L, Heins Z, Ochoa A, et al (2018) Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173:** 321–337

Sanders CM (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.* **430:** 119–128

Dos Santos PC, Panero J, Stanganelli C, Palau Nagore V, Stella F, Bezares R & Slavutsky I (2017) Dysregulation of H/ACA ribonucleoprotein components in chronic lymphocytic leukemia. *PLoS One* **12:** e0179883

Sarek G, Vannier J-B, Panier S, Petrini JHJ & Boulton SJ (2015) TRF2 Recruits RTEL1 to Telomeres in S Phase to Promote T-Loop Unwinding. *Mol. Cell* **57:** 622–635

Sarkies P, Murat P, Phillips LG, Patel KJ, Balasubramanian S & Sale JE (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **40:** 1485–1498

Sarkies P, Reams C, Simpson LJ & Sale JE (2010) Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* **40:** 703–713

Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C & Moine H (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20:** 4803–4813

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

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

Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, Zhao Z, Smogorzewska A, Sowa ME, Ang XL, Westbrook TF, Liang AC, Chang K, Hackett JA, Harper JW, Hannon GJ & Elledge SJ (2008) Cancer proliferation gene discovery through functional genomics. *Science* **319:** 620–624

Schmitt CA, Fridman JS, Yang M, Baranov E, Hoffman RM & Lowe SW (2002) Dissecting p53 tumor suppressor functions in vivo. *Cancer Cell* **1:** 289–298

Schoeftner S & Blasco MA (2009) A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J.* **28:** 2323–2336

Schubert T, Pusch MC, Diermeier S, Benes V, Kremmer E, Imhof A & Längst G (2012) Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol. Cell* **48:** 434–444

Schutte M, Hruban RH, Hedrick L, Cho KR, Nadasdy GM, Weinstein CL, Bova GS, Isaacs WB, Cairns P, Nawroz H, Sidransky D, Casero RA, Meltzer PS, Hahn SA & Kern SE (1996) DPC4 gene in various tumor types. *Cancer Res.* **56:** 2527–2530

Schwab RA, Nieminuszczy J, Shin-ya K & Niedzwiedz W (2013) FANCJ couples replication past natural fork barriers with maintenance of chromatin structure. *J. Cell Biol.* **201:** 33–48

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N & Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208

Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR & Rahman N (2006) Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat. Genet.* **38:** 1239–1241

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

Senft D, Qi J & Ronai ZA (2018) Ubiquitin ligases in oncogenic transformation and cancer therapy. *Nat. Rev. Cancer* **18:** 69–88

Sessler JL, Sathiosatham M, Doerr K, Lynch V & Abboud KA (2000) A G-Quartet Formed in the Absence of a Templating Metal Cation: A New 8-(N,N-dimethylaniline)guanosine Derivative. *Angew. Chemie* **112:** 1356–1359

Sethuraman A, Brown M, Seagroves TN, Wu Z-H, Pfeffer LM & Fan M (2016) SMARCE1 regulates metastatic potential of breast cancer cells through the HIF1A/PTK2 pathway. *Breast Cancer Res.* **18:** 1-15

Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL & de Lange T (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138:** 90–103

Shahid R, Bugaut A & Balasubramanian S (2010) The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49:** 8300–8306

Shain AH & Pollack JR (2013) The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One* **8:** e55119

Shalaby T, von Bueren AO, Hürlimann M-L, Fiaschetti G, Castelletti D, Masayuki T, Nagasawa K, Arcaro A, Jelesarov I, Shin-ya K & Grotzer M (2010) Disabling c-Myc in childhood medulloblastoma and atypical teratoid/rhabdoid tumor cells by the potent G-quadruplex interactive agent S2T1-6OTD. *Mol. Cancer Ther.* **9:** 167–179

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG & Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343:** 84–87

Shammas MA, Shmookler Reis RJ, Li C, Koley H, Hurley LH, Anderson KC & Munshi NC (2004) Telomerase inhibition and cell growth arrest after telomestatin treatment in multiple myeloma. *Clin. Cancer Res.* **10:** 770–776

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13:** 2498–2504

Sharma A, Comstock CES, Knudsen ES, Cao KH, Hess-Wilson JK, Morey LM, Barrera J & Knudsen KE (2007) Retinoblastoma Tumor Suppressor Status Is a Critical Determinant of Therapeutic Response in Prostate Cancer Cells. *Cancer Res.* **67:** 6192–6203

Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N, Raver-Shapira N, Minsky N, Pirngruber J, Tarcic G, Hublarova P, Moyal L, Gana-Weisz M, Shiloh Y, Yarden Y, Johnsen SA, Vojtesek B, Berger SL & Oren M (2008) The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev.* **22:** 2664–2676

Sherr CJ (2001) The INK4a/ARF network in tumour suppression. *Nat. Rev. Mol. Cell Biol.* **2:** 731–737

Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, Ghiorzo P, Bressac-de Paillerets B, Nagore E, Avril MF, Caporaso NE, McMaster ML, Cullen M, Wang Z, Zhang X, Bruno W, Pastorino L, Queirolo P, Banuls-Roca J, Garcia-Casado Z, et al (2014) Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. *Nat. Genet.* **46:** 482–486

Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, Hayakawa Y & Seto H (2001) Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* **123:** 1262–1263

Shivalingam A, Izquierdo MA, Marois A Le, Vyšniauskas A, Suhling K, Kuimova MK & Vilar R (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6:** 1-10

Shkreta L, Blanchette M, Toutant J, Wilhelm E, Bell B, Story BA, Balachandran A, Cochrane A, Cheung PK, Harrigan PR, Grierson DS & Chabot B (2017) Modulation of the splicing regulatory function of SRSF10 by a novel compound that impairs HIV-1 replication. *Nucleic Acids Res.* **45:** 4051–4067

Siddiqui-Jain A, Grand CL, Bearss DJ & Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to

repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* **99:** 11593–11598

Silva JM, Marran K, Parker JS, Silva J, Golding M, Schlabach MR, Elledge SJ, Hannon GJ & Chang K (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* **319:** 617–620

Simonsson T (2001) G-quadruplex DNA structures-variations on a theme. *Biol. Chem.* **382:** 621–628

Simonsson T, Pecinka P & Kubista M (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26:** 1167–1172

Simpson KJ, Selfors LM, Bui J, Reynolds A, Leake D, Khvorova A & Brugge JS (2008) Identification of genes that regulate epithelial cell migration using an siRNA screening approach. *Nat. Cell Biol.* **10:** 1027–1038

Singh S, Narang AS & Mahato RI (2011) Subcellular fate and off-target effects of siRNA, shRNA, and miRNA. *Pharm. Res.* **28:** 2996–3015

Sliva K & Schnierle BS (2010) Selective gene silencing by viral delivery of short hairpin RNA. *Virol. J.* **7:** 1-11

Soldatenkov VA, Vetcher AA, Duka T & Ladame S (2008) First evidence of a functional interaction between DNA quadruplexes and poly(ADP-ribose) polymerase-1. *ACS Chem. Biol.* **3:** 214–219

Song B, Bian Q, Zhang Y-J, Shao C-H, Li G, Liu A-A, Jing W, Liu R, Zhou Y-Q, Jin G & Hu X-G (2015) Downregulation of ASPP2 in pancreatic cancer cells contributes to increased resistance to gemcitabine through autophagy activation. *Mol. Cancer* **14:** 1-12

Soucy TA, Smith PG, Milhollen MA, Berger AJ, Gavin JM, Adhikari S, Brownell JE, Burke KE, Cardin DP, Critchley S, Cullis CA, Doucette A, Garnsey JJ, Gaulin JL, Gershman RE, Lublinsky AR, McDonald A, Mizutani H, Narayanan U, Olhava EJ, et al (2009) An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **458:** 732–736

de Sousa Cavalcante L & Monteiro G (2014) Gemcitabine: Metabolism and molecular mechanisms of action, sensitivity and chemoresistance in pancreatic cancer. *Eur. J. Pharmacol.* **741:** 8–16

Soutourina J (2017) Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19:** 262–274

Soutourina J, Wydau S, Ambroise Y, Boschiero C & Werner M (2011) Direct Interaction of RNA Polymerase II and Mediator Required for Transcription in Vivo. *Science (80-.).* **331:** 1451–1454

Spillare EA, Wang XW, von Kobbe C, Bohr VA, Hickson ID & Harris CC (2006) Redundancy of DNA helicases in p53-mediated apoptosis. *Oncogene* **25:** 2119–2123

Subramanian M, Rage F, Tabet R, Flatter E, Mandel J-L & Moine H (2011) G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* **12:** 697–704

Sugiura T, Nagano Y & Noguchi Y (2007a) DDX39, upregulated in lung squamous cell cancer, displays RNA helicase activities and promotes cancer cell growth. *Cancer Biol. Ther.* **6:** 957–964

Sugiura T, Sakurai K & Nagano Y (2007b) Intracellular characterization of DDX39, a novel growth-associated RNA helicase. *Exp. Cell Res.* **313:** 782–790

Suhasini AN & Brosh RM (2013) DNA helicases associated with genetic instability, cancer, and aging. *Adv. Exp. Med. Biol.* **767:** 123–144

Suk K, Kim S, Kim YH, Oh SH, Lee MK, Kim KW, Kim HD & Seo YS (2000) Identification of a novel human member of the DEAD box protein family. *Biochim. Biophys. Acta* **1501:** 63–69

Sun D, Guo K & Shin Y-J (2011) Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Res.* **39:** 1256–1265

Sun H, Karow JK, Hickson ID & Maizels N (1998) The Bloom's syndrome helicase

unwinds G4 DNA. *J. Biol. Chem.* **273:** 27587–27592

Sung YH, Baek I-J, Kim DH, Jeon J, Lee J, Lee K, Jeong D, Kim J-S & Lee H-W (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* **31:** 23–24

Tahara H, Shin-Ya K, Seimiya H, Yamada H, Tsuruo T & Ide T (2006) G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene* **25:** 1955–1966

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

Tang C-F & Shafer RH (2006) Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes. *J. Am. Chem. Soc.* **128:** 5966–5973

Tanner NK & Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell* **8:** 251–262

Tarailo-Graovac M, Wong T, Qin Z, Flibotte S, Taylor J, Moerman DG, Rose AM & Chen N (2015) Spectrum of variations in dog-1/FANCJ and mdf-1/MAD1 defective Caenorhabditis elegans strains after long-term propagation. *BMC Genomics* **16:** 1-10

Tascilar M, Skinner HG, Rosty C, Sohn T, Wilentz RE, Offerhaus GJ, Adsay V, Abrams RA, Cameron JL, Kern SE, Yeo CJ, Hruban RH & Goggins M (2001) The SMAD4 protein and prognosis of pancreatic ductal adenocarcinoma. *Clin. Cancer Res.* **7:** 4115–4121

Tauchi T, Shin-Ya K, Sashida G, Sumi M, Nakajima A, Shimamoto T, Ohyashiki JH & Ohyashiki K (2003) Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* **22:** 5338–5347

Tauchi T, Shin-ya K, Sashida G, Sumi M, Okabe S, Ohyashiki JH & Ohyashiki K (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene* **25:** 5719–5725

Thakur RK, Kumar P, Halder K, Verma A, Kar A, Parent J-L, Basundra R, Kumar A & Chowdhury S (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **37:** 172–183

Thandapani P, Song J, Gandin V, Cai Y, Rouleau SG, Garant J-M, Boisvert F-M, Yu Z, Perreault J-P, Topisirovic I & Richard S (2015) Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *Elife* **4:** 1-30

Thibodeau S, Bren G & Schaid D (1993) Microsatellite instability in cancer of the proximal colon. *Science (80-.).* **260:** 816–819

Thillainadesan G, Isovic M, Loney E, Andrews J, Tini M & Torchia J (2008) Genome analysis identifies the p15ink4b tumor suppressor as a direct target of the ZNF217/CoREST complex. *Mol. Cell. Biol.* **28:** 6066–6077

Tisi D, Chiarparin E, Tamanini E, Pathuri P, Coyle JE, Hold A, Holding FP, Amin N, Martin ACL, Rich SJ, Berdini V, Yon J, Acklam P, Burke R, Drouin L, Harmer JE, Jeganathan F, van Montfort RLM, Newbatt Y, Tortorici M, et al (2016) Structure of the Epigenetic Oncogene MMSET and Inhibition by *N*-Alkyl Sinefungin Derivatives. *ACS Chem. Biol.* **11:** 3093–3105

Tlučková K, Marušič M, Tóthová P, Bauer L, Šket P, Plavec J & Viglasky V (2013) Human Papillomavirus G-Quadruplexes. *Biochemistry* **52:** 7207–7216

Todd AG, Lin H, Ebert AD, Liu Y & Androphy EJ (2013) COPI transport complexes bind to specific RNAs in neuronal cells. *Hum. Mol. Genet.* **22:** 729–736

Todd AK, Johnston M & Neidle S (2005) Highly prevalent putative quadruplex

sequence motifs in human DNA. *Nucleic Acids Res.* **33:** 2901–2907

Toyoshima M, Howie HL, Imakura M, Walsh RM, Annis JE, Chang AN, Frazier J, Chau BN, Loboda A, Linsley PS, Cleary MA, Park JR & Grandori C (2012) Functional genomics identifies therapeutic targets for MYC-driven cancer. *Proc. Natl. Acad. Sci.* **109:** 9545–9550

Tsurimoto T & Stillman B (1991) Replication factors required for SV40 DNA replication in vitro. II. Switching of DNA polymerase alpha and delta during initiation of leading and lagging strand synthesis. *J. Biol. Chem.* **266:** 1961–1968

Tumini E, Barroso S, -Calero CP & Aguilera A (2016) Roles of human POLD1 and POLD3 in genome stability. *Sci. Rep.* **6:** 1-13

Turcotte S, Chan DA, Sutphin PD, Hay MP, Denny WA & Giaccia AJ (2008) A Molecule Targeting VHL-Deficient Renal Cell Carcinoma that Induces Autophagy. *Cancer Cell* **14:** 90–102

Tyagi S, Chabes AL, Wysocka J & Herr W (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol. Cell* **27:** 107–119

Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, Mupo A, Grinkevich V, Li M, Mazan M, Gozdecka M, Ohnishi S, Cooper J, Patel M, McKerrell T, Chen B, Domingues AF, Gallipoli P, Teichmann S, Ponstingl H, et al (2016) A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* **17:** 1193–1205

Uhlmann-Schiffler H, Jalal C & Stahl H (2006) Ddx42p--a human DEAD box protein with RNA chaperone activities. *Nucleic Acids Res.* **34:** 10–22

Uhlmann-Schiffler H, Kiermayer S & Stahl H (2009) The DEAD box protein Ddx42p modulates the function of ASPP2, a stimulator of apoptosis. *Oncogene* **28:** 2065–2073

Vallur AC & Maizels N (2008) Activities of human exonuclease 1 that promote cleavage of transcribed immunoglobulin switch regions. *Proc. Natl. Acad. Sci. U. S. A.* **105:** 16508–16512

Vannier J-B, Pavicic-Kaltenbrunner V, Petalcorin MIR, Ding H & Boulton SJ (2012) RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* **149:** 795–806

Vannier J-B, Sandhu S, Petalcorin MIR, Wu X, Nabi Z, Ding H & Boulton SJ (2013) RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. *Science* **342:** 239–242

Vannier J-B, Sarek G & Boulton SJ (2014) RTEL1: functions of a disease-associated helicase. *Trends Cell Biol.* **24:** 416–425

Vaughn JP, Creacy SD, Routh ED, Joyner-Butt C, Jenkins GS, Pauli S, Nagamine Y & Akman SA (2005) The DEXH Protein Product of the DHX36 Gene Is the Major Source of Tetramolecular Quadruplex G4-DNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.* **280:** 38117–38120

Vert J-P, Foveau N, Lajaunie C & Vandenbrouck Y (2006) An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7:** 520

Voulgari A, Voskou S, Tora L, Davidson I, Sasazuki T, Shirasawa S & Pintzas A (2008) TATA Box-Binding Protein-Associated Factor 12 Is Important for RAS-Induced Transformation Properties of Colorectal Cancer Cells. *Mol. Cancer Res.* **6:** 1071–1083

Wada C, Kasai K, Kameya T & Ohtani H (1992) A general transcription initiation factor, human transcription factor IID, overexpressed in human lung and breast carcinoma and rapidly induced with serum stimulation. *Cancer Res.* **52:** 307–313

Walia R & Chaconas G (2013) Suggested role for G4 DNA in recombinational switching at the antigenic variation locus of the Lyme disease spirochete. *PLoS One* **8:** e57792

Wang C, Zhao L & Lu S Role of TERRA in the Regulation of Telomere Length. *Int. J. Biol. Sci.* **11:** 316–323

Wang JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3:** 430–440

Wang S, Nath N, Fusaro G & Chellappan S (1999) Rb and prohibitin target distinct regions of E2F1 for repression and respond to different upstream signals. *Mol. Cell. Biol.* **19:** 7447–7460

Wang Y-Y, Yang Y-X, Zhe H, He Z-X & Zhou S-F (2014) Bardoxolone methyl (CDDO-Me) as a therapeutic agent: an update on its pharmacokinetic and pharmacodynamic properties. *Drug Des. Devel. Ther.* **8:** 2075–2088

Wanzek K, Schwindt E, Capra JA & Paeschke K (2017) Mms1 binds to G-rich regions in Saccharomyces cerevisiae and influences replication and genome stability. *Nucleic Acids Res.* **45:** 7796–7806

Wassarman DA & Sauer F (2001) TAF(II)250: a transcription toolbox. *J. Cell Sci.* **114:** 2895–2902

Watkins NJ & Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip. Rev. RNA* **3:** 397–414

Watson JD & Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171:** 737–738

Watson LA, Solomon LA, Li JR, Jiang Y, Edwards M, Shin-ya K, Beier F & Bérubé NG (2013) Atrx deficiency induces telomere dysfunction, endocrine defects, and reduced life span. *J. Clin. Invest.* **123:** 2049–2063

Wei W & Lin H-K (2012) The key role of ubiquitination and sumoylation in signaling and cancer: a research topic. *Front. Oncol.* **2:** 1-2

White MF (2009) Structure, function and evolution of the XPD family of iron-sulfur-containing 5'-->3' DNA helicases. *Biochem. Soc. Trans.* **37:** 547–551

Wijdeven RH, Pang B, van der Zanden SY, Qiao X, Blomen V, Hoogstraat M, Lips EH, Janssen L, Wessels L, Brummelkamp TR & Neefjes J (2015) Genome-Wide Identification and Characterization of Novel Factors Conferring Resistance to Topoisomerase II Poisons in Cancer. *Cancer Res.* **75:** 4176–4187

Wilkins MHF, Stokes AR & Wilson HR (1953) Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **171:** 738–740

Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M & Lührmann R (2002) Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J.* **21:** 4978–4988

Williamson JR, Raghuraman MKK & Cech TR (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59:** 871–880

Wilson RC & Doudna JA (2013) Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* **42:** 217–239

Wittig R, Nessling M, Will RD, Mollenhauer J, Salowsky R, Münstermann E, Schick M, Helmbach H, Gschwendt B, Korn B, Kioschis P, Lichter P, Schadendorf D & Poustka A (2002) Candidate genes for cross-resistance against DNA-damaging drugs. **62:** 6698–6705

Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, Schatz JH, Rodrigo CM, Zhao C, Rondou P, de Stanchina E, Teruya-Feldstein J, Kelliher MA, Speleman F, Porco JA, Pelletier J, et al (2014) RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513:** 65–70

Wong HM & Huppert JL (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* **5:** 1713–1719

Wong LH, McGhie JD, Sim M, Anderson MA, Ahn S, Hannan RD, George AJ, Morgan KA, Mann JR & Choo KHA (2010) ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res.* **20:** 351–360

Wood RD & Doublié S (2016) DNA polymerase θ (POLQ), double-strand break repair, and cancer. *DNA Repair (Amst).* **44:** 22–32

Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, Jaenisch R, Zhang F & Sharp PA (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32:** 670–676

Wu Y & Brosh RM (2010) G-quadruplex nucleic acids and human disease. *FEBS J.* **277:** 3470–3488

Wu Y & Brosh RM (2012) DNA helicase and helicase-nuclease enzymes with a conserved iron-sulfur cluster. *Nucleic Acids Res.* **40:** 4247–4260

Wu Y, Shin-ya K & Brosh RM (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.* **28:** 4116–4128

Wysocka J, Reilly PT & Herr W (2001) Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol. Cell. Biol.* **21:** 3820–3829

Xu GW, Toth JI, da Silva SR, Paiva S-L, Lukkarila JL, Hurren R, Maclean N, Sukhai MA, Bhattacharjee RN, Goard CA, Gunning PT, Dhe-Paganon S, Petroski MD & Schimmer AD (2014a) Mutations in UBA3 Confer Resistance to the NEDD8-Activating Enzyme Inhibitor MLN4924 in Human Leukemic Cells. *PLoS One* **9:** e93530

Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, Santos N Dos, Silvester J, Wei V, Garcia J, Kabeer F, Lai D, Soriano P, Banáth J, Chiu DS, Yap D, Le DD, Ye FB, Zhang A, Thu K, et al (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8:** 1-18

Xu Y, Gao XD, Lee J-H, Huang H, Tan H, Ahn J, Reinke LM, Peter ME, Feng Y, Gius D, Siziopikou KP, Peng J, Xiao X & Cheng C (2014b) Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev.* **28:** 1191–1203

Xu Y, Ishizuka T, Yang J, Ito K, Katada H, Komiyama M & Hayashi T (2012) Oligonucleotide models of telomeric DNA and RNA form a Hybrid G-quadruplex structure as a potential component of telomeres. *J. Biol. Chem.* **287:** 41787–41796

Xu Y, Milazzo JP, Somerville TDD, Tarumoto Y, Huang Y-H, Ostrander EL, Wilkinson JE, Challen GA & Vakoc CR (2018) A TFIID-SAGA Perturbation that Targets MYB and Suppresses Acute Myeloid Leukemia. *Cancer Cell* **33:** 13–28

Xu Y, Suzuki Y, Ito K & Komiyama M (2010) Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **107:** 14579–14584

Yadav P, Harcy V, Argueso JL, Dominska M, Jinks-Robertson S & Kim N (2014) Topoisomerase I Plays a Critical Role in Suppressing Genome Instability at a Highly Transcribed G-Quadruplex-Forming Sequence. *PLoS Genet.* **10:** e1004839

Yang D & Hurley LH (2006) Structure of the Biologically Relevant G-Quadruplex in The c-MYC Promoter. *Nucleosides, Nucleotides and Nucleic Acids* **25:** 951–968

Yang WS & Stockwell BR (2008) Synthetic Lethal Screening Identifies Compounds Activating Iron-Dependent, Nonapoptotic Cell Death in Oncogenic-RAS-Harboring Cancer Cells. *Chem. Biol.* **15:** 234–245

Yeats C, Maibaum M, Marsden R, Dibley M, Lee D, Addou S & Orengo CA (2006) Gene3D: modelling protein structure, function and evolution. *Nucleic Acids Res.* **34:** 281–284

Yeh C, Coyaud É, Bashkurov M, van der Lelij P, Cheung SWT, Peters JM, Raught B & Pelletier L (2015) The Deubiquitinase USP37 Regulates Chromosome Cohesion and Mitotic Progression. *Curr. Biol.* **25:** 2290–2299

Yoo HH & Chung IK (2011) Requirement of DDX39 DEAD box RNA helicase for

genome integrity and telomere protection. *Aging Cell* **10:** 557–571

Yoshida T, Tanaka S, Mogi A, Shitara Y & Kuwano H (2004) The clinical significance of Cyclin B1 and Wee1 expression in non-small-cell lung cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **15:** 252–256

Yu C-H, Teulade-Fichou M-P & Olsthoorn RCL (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.* **42:** 1887–92

Zahler AM, Williamson JR, Cech TR & Prescott DM (1991) Inhibition of telomerase by G-quartet DMA structures. *Nature* **350:** 718–720

Zalfa F, Panasiti V, Carotti S, Zingariello M, Perrone G, Sancillo L, Pacini L, Luciani F, Roberti V, D'Amico S, Coppola R, Abate SO, Rana RA, De Luca A, Fiers M, Melocchi V, Bianchi F, Farace MG, Achsel T, Marine J-C, et al (2017) The fragile X mental retardation protein regulates tumor invasiveness-related pathways in melanoma cells. *Cell Death Dis.* **8:** e3169

Zaug AJ, Podell ER & Cech TR (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci.* **102:** 10864–10869

Zeng Y, Wagner EJ & Cullen BR (2002) Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9:** 1327–1333

Zhang AYQ, Bugaut A & Balasubramanian S (2011) A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **50:** 7251–7258

Zhang H, Lu Y, Chen E, Li X, Lv B, Vikis HG & Liu P (2017) XRN2 promotes EMT and metastasis through regulating maturation of miR-10a. *Oncogene* **36:** 3925–3933

Zhang J, Zheng K, Xiao S, Hao Y & Tan Z (2014) Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.* **136:** 1381–1390

Zhang K, Dion N, Fuchs B, Damron T, Gitelis S, Irwin R, O'Connor M, Schwartz H, Scully SP, Rock MG, Bolander ME & Sarkar G (2002) The human homolog of yeast SEP1 is a novel candidate tumor suppressor gene in osteogenic sarcoma. *Gene* **298:** 121–127

Zhang Z, Dai J, Veliath E, Jones RA & Yang D (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.* **38:** 1009–1021

Zhao F, Gong X, Liu A, Lv X, Hu B & Zhang H (2018) Downregulation of Nedd4L predicts poor prognosis, promotes tumor growth and inhibits MAPK/ERK signal pathway in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* **495:** 1136–1143

Zhao J, Bacolla A, Wang G & Vasquez KM (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **67:** 43–62

Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, Guzzo F, English DP, Varughese J, Gasparrini S, Bortolomai I, Buza N, Hui P, Abu-Khalaf M, Ravaggi A, Bignotti E, Bandiera E, Romani C, Todeschini P, Tassi R, et al (2013) Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 2916–2921

Zheng H-C (2017) The molecular mechanisms of chemoresistance in cancers. *Oncotarget* **8:** 59950–59964

Zheng K, Wu R, He Y, Xiao S, Zhang J, Liu J, Hao Y & Tan Z (2014) A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.* **42:** 10832–10844

Zheng S, Vuong BQ, Vaidyanathan B, Lin J-Y, Huang F-T & Chaudhuri J (2015) Non-coding RNA Generated following Lariat Debranching Mediates Targeting of

AID to DNA. *Cell* **161:** 762–773

Zhou X, Li X, Cheng Y, Wu W, Xie Z, Xi Q, Han J, Wu G, Fang J & Feng Y (2014) BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat. Commun.* **5:** 1-11

Zimmer J, Tacconi EMC, Folio C, Badie S, Porru M, Klare K, Tumiati M, Markkanen E, Halder S, Ryan A, Jackson SP, Ramadan K, Kuznetsov SG, Biroccio A, Sale JE & Tarsounas M (2016) Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Mol. Cell* **61:** 449–460