Title	A study of translation regulation in cancer by application of ribosome profiling
Author(s)	Tzani, Ioanna
Publication date	2016
Original citation	Tzani, I. 2016. A study of translation regulation in cancer by application of ribosome profiling. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2016, Ionna Tzani. http://creativecommons.org/licenses/by-nc-nd/3.0/
Embargo information	Indefinite
Embargo lift date	10000-01-01
Item downloaded from	http://hdl.handle.net/10468/3646

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

A study of translation regulation in cancer by application of ribosome profiling

A thesis submitted to the National University of Ireland, Cork, in
fulfilment of the requirements for the degree of
Doctor of Philosophy
by
Ioanna Tzani, MSc

**School of Biochemistry and Cell Biology,
University College Cork**

October 2016

**Supervisor: Dr Gary Loughran
Head of Department: Professor David Sheehan**

Declaration

This thesis is my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed:_____
Ioanna Tzani

Acknowledgements

Firstly, I am truly grateful to Dr Gary Loughran, for his tireless day-to-day supervision, guidance and patience. Without his invaluable advice, help and feedback, this thesis would never have been completed.

I am also grateful to Professor Rosemary O'Connor and Dr Kellie Dean, who believed in me and gave me the opportunity to do this PhD. I am deeply indebted to Dr Dean for her support during the most challenging phases and her advice that has had an immense impact on my personal and academic development. I wish to thank the PhD Scholars Programme in Cancer Biology and Health Research Board for the very generous funding that made my research a reality.

I would like to thank Professor John F. Atkins and Dr Pasha V. Baranov for co-supervising my PhD and Dr Ivaylo P. Ivanov for giving me the opportunity to work on his polyamines project. I am grateful to them for sharing their scientific knowledge with me, in their unique mentoring styles, and teaching me how to deal with challenging situations. The social activities and celebrations they inexhaustibly organised, along with past and current members of their labs (Recode and Lapti), made my experience in the lab unforgettable.

I must thank Dr Yuan Chang, Dr Patrick Moore and Dr Masa Shuda for giving me the opportunity to work on their Merkel Cell Carcinoma project, for sharing their expertise and providing the required reagents.

I would also like to thank Dr Tom Moore and his lab members Dr John Williams and Dr Zhenfei Ning for their help with microscopy. I am grateful to Noreen Casey, Pat Allen, Jenny Duane and Tricia Fowler for sharing their expertise and always being helpful.

Last but not least a big thank you goes to Steph, Michael, Sinead, Anjali, Milan, Anita, Joan, Siobhan, Virag and Ciaran for your friendship and for sharing lab protocols, reagents, experiences and making everyday life in the lab more enjoyable.

"Morality, like art, means drawing a line someplace."

attributed to Oscar Wilde

Abbreviations

AdoMet: S-adenosylmethionine/ SAMe

4E-BP: eIF4E- binding protein

AGC: cAMP-dependent, cGMP-dependent and protein kinase C

ALTO: Alternative to Large T Open Reading Frame

AMD1: Adenosylmethionine Decarboxylase 1, AdoMetDC

Ang1: Angiopoietin 1

APC/C: Anaphase-Promoting Complex/ Cyclosome

AZIN: antizyme inhibitor

BGI: Beijing Genomics Institute

BKPyV: BK virus

CHO cells: Chinese Hamster Ovary cells

CDK: Cyclin-dependent Kinase

CK20: cytokeratin 20

dcAdoMet: S-Adenosylmethioninamine / decarboxylated S-adenosylmethionine

dcSAM= S-Adenosylmethioninamine, dcAdoMet

DDR: DNA damage response

DE: Differentially expressed

DFMO: D,L-alpha-difluoromethylornithine

DHPS: Deoxyhypusine synthase

DOHH: Deoxyhypusine hydrolase

DMEM: Dulbecco's modified Eagle's medium

DMSO: Dimethyl sulfoxide

DSN: Duplex-specific nuclease

FBS: Fetal Bovine Serum

EBV: Epstein-Barr Virus

eIF5A: eukaryotic initiation factor 5A

eIF4E: eukaryotic initiation factor 4E

FGF: Fibroblast Growth Factor

FoxO1: Forkhead box protein O1

FoxO3a: Forkhead box protein O3a

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GPCR: G-proteins-coupled receptor

GWIPS: Genome-wide information on protein synthesis

HBV: Hepatitis B virus

HCV: Hepatitis C virus

HGF: Human Growth Factor

HIF1α: Hypoxia-inducible factor 1α

HTLV-I: Human T-lymphotropic virus-I

HPyV: Human Polyomavirus

IKK: IκB kinase

IRS: Insulin Receptor Substrate

JC or JCPyV: John Cunningham

KSHV: Kaposi's Sarcoma Herpesvirus

LCA: Leukocyte Common Antigen

lincRNA: long interspersed non-coding RNAs

LSD: LT- Stabilization Domain

LT: Large T antigen

MAP: Microtubule- Associated Protein

mTOR: mammalian Target of Rapamycin

MTT: Thiazolyl Blue Tetrazolium Bromide

MCV: Merkel Cell Polyomavirus

MCC: Merkel Cell Carcinoma

MMP: Matrix Metalloproteinase

MPyV: Murine Polyomavirus

MTT: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

MUR: MCV T antigen unique region

Mybbp1a: Myb binding protein 1a

MW or MWPyV: Malawi Polyomavirus

MTA: 5'-methylthioadenosine

NCRR: Noncoding Regulatory Region

NLS: Nuclear Localisation Signal

NTC: Non- Template Control

NEMO: NF-kappaB essential modulator

NSE: Neuron-specific Enolase

ori: origin of replication

OAZ: Antizyme

OBD: Origin Binding Domain

ODC: Ornithine Decarboxylase

ORF: Open Reading Frame

PAO: Polyamine oxidase

PBS: Phosphate- buffered saline

PBS-T: PBS- Tween

PKB: Protein Kinase B

RTK: Receptor Tyrosine Kinase

PCR: Polymerase Chain Reaction

PDK1: phosphoinositide-dependent kinase 1

PH: Pleckstrin Homology

PHAS: Phosphorylated Heat and Acid Stable

PHD: Pleckstrin Homology Domain

PI3K: phosphoinositide 3-kinase

PIK3CA: Phosphtidylinositol-4,5-Bisphosphate 3-kinase, Catalytic Subunit Alpha

PI(4,5)P$_2$: phosphoatidylinositol-4,5-bisphosphate

PI(3,4,5)P$_3$/ PIP3: phosphatidylinositol-3,4,5-triphosphate

PLB: Passive Lysis Buffer

PtdIns: phosphatidylinositol

PTEN: Phosphatase and tensin homolog on chromosome 10

PP4C: Protein Phosphatase 4C

PROTOR: protein observed with RICTOR

RACE: Rapid Amplification of cDNA Ends

Rb: Retinoblastoma

RICTOR: Rapamycin-insenitive companion of mTOR

RIPA: Radioimmunoprecipitation

RFC: Replication Factor C

Ribo-seq: Ribosome profiling

RTK: Receptor Tyrosine Kinase

RPF: Ribosome Protected Fragment

SAMe: S-adenosyl-L-methionine

SAT1: spermidine/spermine N(1)-acetyltransferase (SSAT)

SCLC: Small Cell Lung Cancer

SGK: Serum- and glucocorticoid-regulated kinase

SIRT: Sirtuin

SMO: Spermine Oxidase

SMS: spermine synthase

SRM: spermine synthase

SSAT: spermidine/spermine N(1)-acetyltransferase

sT: small T antigen

SV40: simian vacuolating virus 40

TE: Translation Efficiency

TetO$_2$: Tetracycline operators

TetR: Tet Repressor

TOP: 5' Terminal Oligopyrimidine

TSC: Tuberous Sclerosis Complex

TSV or TSPyV: Trichodysplasia Spinulosa- associated polyomavirus

TTF-1:Thyroid Transcription Factor 1

uORF: upstream Open reading Frame

VEGF: Vascular Endothelial Growth Factor

VLP: Virus like particle

Vsp34: Vacuolar protein-sorting defective 34

WHIM: warts, hypogammaglobulinemia infections and myelokathexis syndrome

General Abstract

Cancer is the name given to diverse diseases, whose common characteristic is uncontrolled cell proliferation. In an effort to rationalise the complexity of cancer Hanahan and Weinberg (2000, 2011) proposed the hallmarks that comprise biological capabilities that enable tumour development. Dysregulation of molecular pathways underlie these hallmarks of tumour progression and their study has been central to cancer biology. The PI3K/AKT/mTOR pathway has a central role in translation, cell growth and proliferation and it is frequently dysregulated in various cancers. In this work we aimed to elucidate the role of the PI3K pathway and protein synthesis in cancer.

Our work focused on various levels of the PI3K pathway and protein synthesis, with application of ribosome profiling; a powerful technique, which provides genome-wide information on protein synthesis, by capturing actively translating ribosomes and sequencing of their associated transcripts.

Previous work on bioinformatics and ribosome profiling data suggested the presence of an N-terminal extension in the well-studied tumour suppressor PTEN. Mutational analysis and further investigation of the 5' leader of this gene performed in this study led to the discovery of previously uncharacterized proteoforms with N-terminal extensions. Given the physiological importance of PTEN this discovery is expected to broaden our understanding of the translational regulation of this gene and elucidate its role in molecular pathology.

Polyamines are ubiquitous small basic molecules with known but not well defined roles in a range of physiological processes, including translation. The enzymes involved in this pathway are themselves subject to extensive translational regulation. In an effort to further characterize regulation of polyamines and their effect on translation we applied ribosome profiling in a previously established system available in the lab. This work confirmed our previous knowledge on regulation of enzymes of the polyamine pathway and identified previously uncharacterized translational events.

Translation has a central role in physiology and its aberrations can cause cancer. Merkel Cell Carcinoma is a rare but aggressive tumour with poor prognosis

and increasing incident rates. Approximately 80% of these cancers are induced by the Merkel Cell Polyomavirus (MCV). The small T (sT) viral antigen had been proposed to have a causal role in tumour initiation by affecting cap dependent translation. To study this molecular interaction we aimed to develop a model system that inducibly expresses MCV sT to imitate the initial steps of tumourigenesis and applied ribosome profiling on metastatic cell lines.

1. General introduction

1.1 Biology of cancer

Cancer is a leading cause of death worldwide with 8.2 million deaths and 14.1 million new cases in 2012 (Ferlay et al. 2015; Ryerson et al. 2016). Cancer is not a single disease, but a series of different diseases that are characterized by uncontrolled proliferation. It develops through a process that resembles evolution of species (Greaves & Maley 2012), as it is the result of genetic mutations and epigenetic alterations (Sharma et al. 2010; Berdasco et al. 2010) that are selected, because they promote cellular proliferation and survival (Stratton et al. 2009).

In an effort to rationalise this complex and diverse disease Hanahan and Weinberg (2011) suggested initially six alterations in cellular physiology typically observed in cancer cells. These include the ability of cells to grow without being dependent on growth signals (Feldman & Feldman 2001; Davies & Samuels 2010), lack of response to anti-growth signals (Burkhart & Sage 2008), apoptosis evasion (Adams & Cory 2007; Lowe et al. 2004; Evan & Littlewood 1998), limitless replicative potential acquisition (Blasco 2005; Shay & Wright 2000), promotion of angiogenesis (Hanahan et al. 1996) and tissues invasion and metastasis (Berx & van Roy 2009; Cavallaro & Christofori 2004). Genomic instability (Negrini et al. 2010; Salk et al. 2010) and an inflamed tumour microenvironment (Mantovani 2010; Egeblad et al. 2010; Joyce & Pollard 2009) are considered enabling characteristics for tumour development, while avoidance of immune destruction (Strauss et al. 2010; Yang et al. 2010; Shields et al. 2010; Nelson 2008) and dysregulation of cellular energetics (O Warburg 1956; Otto Warburg 1956) were the most recently proposed hallmarks of cancer (Hanahan & Weinberg 2011).

Some of the molecular pathways that are responsible for causing cancer related aberrations in cell physiology have been characterized and the role of their components is being studied extensively (Downward 2003; Anastas & Moon 2012; Bianco et al. 2006).

1.2 PI3K/ AKT/ mTOR pathway

The PI3K/AKT/mTOR signaling pathway has a central role in cell growth and proliferation, translation, apoptosis regulation and cell metabolism. It integrates a

range of signals that include growth factors such as insulin and insulin-like growth factors, amino acid concentrations, energy signals from AMP-activated kinase and various stress factors such as hypoxia and DNA damage (Zoncu et al. 2011; Ruderman et al. 1990; Rodriguez-Viciana et al. 1994; Parker & Waterfield 1992) (Figure. 1.1). The major nodes of this pathway are frequently dysregulated in various cancers (Yuan & Cantley 2008).

The PI3K pathway is activated by cell surface receptors tyrosine kinases (RTKs), integrins (Guo & Giancotti 2004), B and T cell receptors (Okkenhaug & Vanhaesebroeck 2003), cytokines (Baker et al. 2007) and G-protein-coupled receptors (GPCRs) (Zhang et al. 2007). Several growth factors, including the fibroblast growth factor (FGF), the human growth factor (HGF), the vascular endothelial cell growth factor (VEGF), angiopoietin I (Ang1) and insulin can trigger the PI3K pathway, by activating and autophosphorylating RTKs (Yarden & Ullrich 1988; Blume-Jensen & Hunter 2001). Receptors activate PI3Ks either by directly interacting with them or by interactions mediated by the Insulin Receptor Substrate proteins (IRS), which act as scaffolds and organize signaling complexes (Mardilovich et al. 2009). PI3Ks are subdivided into three classes (I-III) based on their sequence homology and substrate specificity. The different classes of PI3K and the various isoforms that belong to each class have distinct substrate specificity and roles in signal transduction (Engelman et al. 2006).

Class I PI3Ks is the most well studied class of PI3Ks and includes two subclasses (IA and IB), which are coupled to and activated by RTKs and GPCRs respectively (Katso et al. 2001). Class IA PI3Ks are heterodimers between a p85 regulatory and a p110 catalytic subunit. The p85 regulatory subunit is a substrate for a range of phosphoproteins including tyrosine kinases (Vivanco & Sawyers 2002). Class IB PI3Ks are heterodimers between the regulatory subunit p101 and the catalytic subunit p110. Class II PI3Ks includes three isoforms encoded by distinct genes. Members of this class lack regulatory subunits and only have p110-like catalytic subunits (Engelman et al. 2006). They mostly use phosphatidylinositol (PtdIns) lipid as a substrate, but can also use PtdIns-4-phosphate (Vanhaesebroeck et al. 2010). Class III PI3Ks only have one member, the vacuolar protein-sorting defective 34 (Vsp34) (Vanhaesebroeck et al. 2010; Engelman et al. 2006).

The growth factor activated phosphoinositide-3 kinase (PI3K) phosphorylates and converts the plasma membrane lipid phosphoatidylinositol-4,5-bisphosphate [$PI(4,5)P_2$] to phosphatidylinositol-3,4,5-triphosphate [$PI(3,4,5)P_3$] (Cantley 2002). $PI(3,4,5)P_3$ is a key second messenger (Insall & Weiner 2001), tightly regulated by the antagonistic activity of PTEN, which reduces the cellular levels of PIP_3 with its lipid phosphatase activity (Cully et al. 2006). Many proteins contain a pleckstrin homology (PH) domain, with which they can interact with $PI(3,4)P_2$ and $PI(3,4,5)P_3$. The generation of PIP_3 triggers the translocation of the PH-domain containing phosphoinositide-dependent kinase 1 (PDK1) to the plasma membrane (Mclaughlin et al. 2002).

Interaction of PDK1 with $PI(3,4,5)P_3$, enables phosphorylation and activation of AKT (Lawlor & Alessi 2001), as well as interaction with serum/glucocorticoid-regulated kinases (SGK), protein kinase C (PKC) and p70S6 kinase (S6K) (Biondi et al. 2001). AKT (or Protein kinase B (PKB)) is a serine/threonine kinase expressed as three isoforms (AKT1, AKT2, AKT3 or PKBα, PKBβ and PKBγ), with distinct, yet overlapping roles, encoded by an equal number of genes (Vivanco & Sawyers 2002; Kiely & Kiely 2015). AKT proteins belong to the family of AGC kinases and contain an N-terminal PH domain, a serine/threonine kinase catalytic domain and a C-terminal regulatory domain. AKT activation requires phosphorylation on its activation loop (T308) by PDK1 and on its carboxy-terminal hydrophobic motif (S473) by mTORC2 (Hemmings & Restuccia 2012). Downstream targets of AKT include the pro-apoptotic FOXO proteins (Guertin et al. 2006), BAD, Caspase-9 (Zhou et al. 2000) and GSK3β (Li et al. 2008), which are inactivated upon AKT phosphorylation and TSC1/2 (tuberous sclerorsis 1 and 2), IKKα (Makarov & Romashkova 1999), AS160 (Thong et al. 2007) and Mdm2 (Zhou & Hung 2002).

AKT induced phosphorylation and subsequent inactivation of TSC1/2 activates, through Rheb, the mTORC1 complex (Huang & Manning 2009). Dephosphorylation of AKT at T308 occurs by PP2A (Andjelković et al. 1996) and by PHLPP1/2 at S473 (Brognard et al. 2007).

Activating mutations of proteins of the PI3K-Akt-mTOR pathway are frequently observed in a range of cancers, including head and neck, lung, brain, colorectal, breast and gynecologic tumours (Qiu et al. 2006) and in Merkel Cell

Polyomavirus (MCV) negative Merkel Cell carcinomas (MCC) (Hafner et al. 2012; Nardi et al. 2012). Mutations of the p100α subunit of PI3K (*PIK3CA*), which result in activation of the PI3K-Akt-mTOR pathway have been observed in MCC (Nardi et al. 2012). This pathway has been shown to be activated in MCC through mutations in the pleckstrin homology domain (PHD) of Akt1 (Hafner et al. 2012).
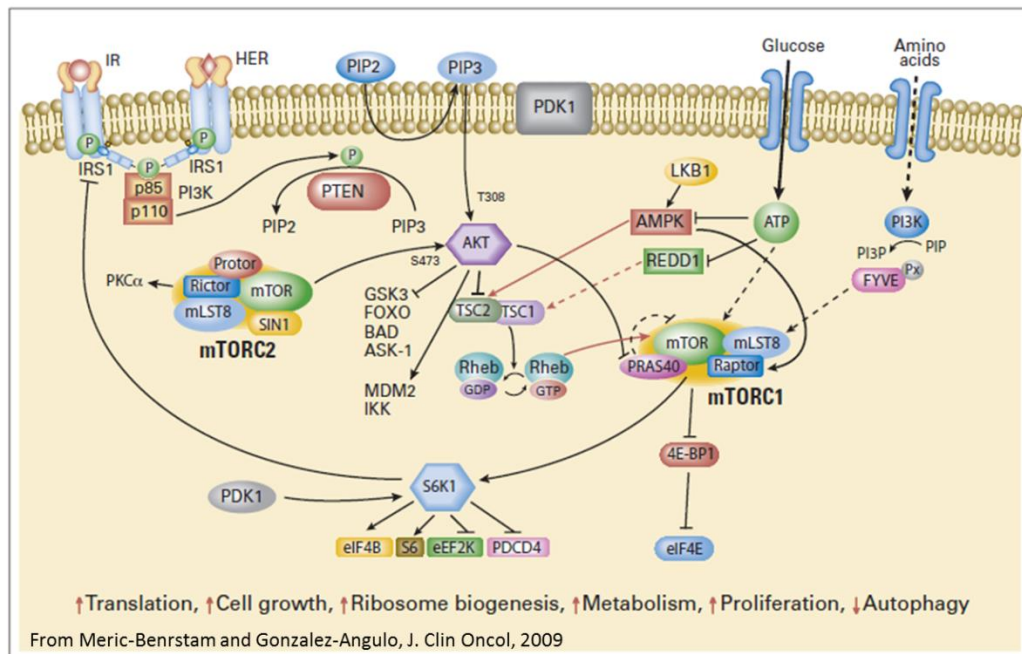
Figure 1.1 The PI3K/Akt/mTOR pathway. The pathway integrates signals from growth factors, insulin, nutrients and energy levels and regulates critical cellular processes, including translation, cell survival and cell cycle progression (Meric-Bernstam & Gonzalez-Angulo 2009).

Mammalian target of rapamycin (mTOR) nucleates two distinct multiprotein complexes mTORC1 and mTORC2 (Zoncu et al. 2011), with different upstream inputs and downstream outputs (Laplante & Sabatini 2012).

mTORC1 consists of five proteins; mammalian Target of Rapamycin (mTOR), which is the catalytic subunit of the complex, the regulatory-associated protein of mTOR (Raptor), the proline-rich AKT substrate 40kDa (PRAS40), the DEP-domain-containing mTOR-interacting protein (Deptor) and the mammalian lethal with Sec13 protein 8 (mLST8, or GβL) (Peterson et al. 2009). mTORC1 is a major regulator of cell growth and metabolism by promoting a range of anabolic processes such as synthesis of proteins and lipids, and inhibiting autophagy. mTORC1 is a downstream target of AKT, and an upstream regulator of S6K1 and eIF-4E-binding protein 1 (4E-

BP1), both of which regulate protein synthesis and cell growth (Richter & Sonenberg 2005). mTORC1 promotes ribosome biogenesis by activating the transcription factor IA (Mayer et al. 2004), and lipid synthesis by positively regulating the activity of the transcription factors sterol regulatory element binding protein 1 (SREBP1) (Porstmann et al. 2008) and peroxisome proliferator-activated receptor-γ (PPARγ) (Kim & Chen 2004). mTORC1 also regulates mitochondrial metabolism and biogenesis and sustains mitochondrial membrane potential, oxygen consumption and cellular ATP levels (Schieke et al. 2006), and inhibits autophagy (Codogno & Meijer 2005).

mTORC2 consists of mTOR, mLST8 (or GβL), RICTOR (rapamycin-insensitive companion of mTOR), PROTOR (protein observed with RICTOR), Deptor and mSIN1 (Sarbassov et al. 2005; Frias et al. 2006; Jacinto et al. 2006; Yang et al. 2006; Pearce et al. 2007). mTORC2 directly binds to and regulates AKT, but it also phosphorylates and regulates PKCα through an poorly characterized mechanism (Guertin & Sabatini 2007). mTORC2 controls the actin cytoskeleton, with a mechanism that is poorly characterized, but involves mTORC2 phosphorylation of PKCα and paxillin (Dos D. Sarbassov et al. 2004; Jacinto et al. 2004). Phosphorylation of AKT at S473 occurs by mTORC2, but only some of the downstream targets of AKT are affected by this phosphorylation event (Guertin et al. 2006; Jacinto et al. 2006). The transcription factors forkhead box protein O1 (FoxO1) and FoxO3a, which are involved in metabolism, cell-cycle arrest and stress resistance, are among the downstream targets of AKT that are affected by mTORC2 (Calnan & Brunet 2008). SGK1 is regulated by mTORC2 and is a regulator of FoxO1 and FoxO3a (García-Martínez & Alessi 2008), suggesting that the effect of mTORC2 on these proteins is mediated through its interaction with SGK1, rather than with AKT (Laplante & Sabatini 2009).

Dysregulation of mTORC1 promotes tumourigenesis in several ways by enabling cap-dependent translation initiation (through eIF4E-binding protein 1 (4E-BP1) inactivation), by inhibiting autophagy, by promoting angiogenesis through hypoxia-inducible factor 1α (HIF1α) upregulation and by activation of the transcription factor SREBP1c, which has oncogenic activity in glioblastomas (Guo et al. 2013), prostate (Huang et al. 2012), endometrial (W. Li et al. 2012), breast (Freed-Pastor et al. 2012) and pancreatic (Sun et al. 2015) cancers. Dysregulation of

mTORC2 exerts its tumourigenic activity through activation of Akt and other proteins of the cAMP-dependent, cGMP-dependent and protein kinase C (AGC) protein kinase family, such as SGK (Zoncu et al. 2011).

Phosphatase and tensin homolog on chromosome 10 (PTEN) is a dual protein and lipid phosphatase (Li et al. 1997; Steck et al. 1997) identified as a negative regulator of PI3K signalling. It exerts its tumour suppressor activity mostly by dephosphorylating PI3K and thus inactivating the PI3K pathway (Stambolic et al. 1998; Wu et al. 1998). PTEN is frequently deleted in melanomas, gliomas, prostate and breast cancer (Hollander et al. 2011) and in MCCs, in which chromosome 10 is frequently missing (Paulson et al. 2009).

The PI3K-Akt-mTOR pathway presents great interest for cancer research because it is one of the most commonly mutated pathways in cancer. Several drugs exist that target various components of this pathway through distinct mechanisms. The main types of drugs that target this pathway are rapamycin, mTOR catalytic inhibitors, AKT inhibitors and dual kinase inibitors targeting both PI3K and mTOR. Rapamycin targets mostly the mTORC1 complex and more specifically ribosomal protein S6K1, but not 4E-BP1 downstream of mTORC1. Ablation of the S6K1-dependent negative feedback loop results in upregulation of PI3K activity following treatment with rapamycin. Depending on the duration of the treatment, rapamycin can block mTORC2 too (Shor et al. 2008; Sarbassov et al. 2006). ATP-competitive inhibitors block both mTORC1 and mTORC2 and present more potent inhibitors of the PI3K pathway, targeting more proteins (Zoncu et al. 2011).

Dual inhibitors of PI3K and mTOR and combinations of drugs with more than one target are more potent in silencing the pathway, but their high toxicity in healthy cells impedes their use in the clinic (Holmes 2011; Schenone et al. 2011).

1.3 Ribosome profiling

Ribosome profiling (Ribo-seq) is a newly developed technique, which is based on deep sequencing of ribosome-protected mRNA fragments (Ingolia et al. 2009) and enables genome-wide information on protein synthesis (GWIPS) (Weiss & Atkins 2011). During translation each ribosome encloses and protects from nuclease degradation, a ~30nt fragment of the translated mRNA (Steitz 1969; Wolin & Walter 1988; Ingolia et al. 2009). In ribosome profiling, ribosomes are immobilized right before cell lysis, either by flash freezing cells or by addition of a translation elongation inhibitor such as cycloheximide (Ingolia et al. 2009). After cell lysis, RNase I (from *E. coli*) is used to remove unprotected RNA and leave ribosome protected fragments (RPFs). Unlike other nucleases (RNase A, RNAse T1 and micrococcal nuclease) Rnase I doesn't show sequence specificity (Meador & Kennell 1990). Total RNA isolation is followed by size selection (~30 nt) of the RPFs, which are then converted into a cDNA library for deep sequencing. Library preparation involves ligation of a preadenylated linker, followed by reverse transcription, circularization and PCR for library amplification (Ingolia et al. 2012) (Figure 1.3).

Following next generation sequencing the exact location of footprints is identified by their alignment to the genome. Sequencing technology (Bentley et al. 2008) enables characterization of the synthesized mRNA molecule, providing information about the amount of protein synthesized and the position of the ribosome on the mRNA molecule (Ingolia 2014). The average density of RPFs on an mRNA, provides an estimate of the protein synthesis rate. mRNA-seq is usually performed in the samples analysed by profiling and used as a control for the abundance of mRNAs. Comparison of the rate of protein synthesis, which can be estimated by ribosome profiling with mRNA abundance, can provide information about translational efficiency (Brar & Weissman 2015).
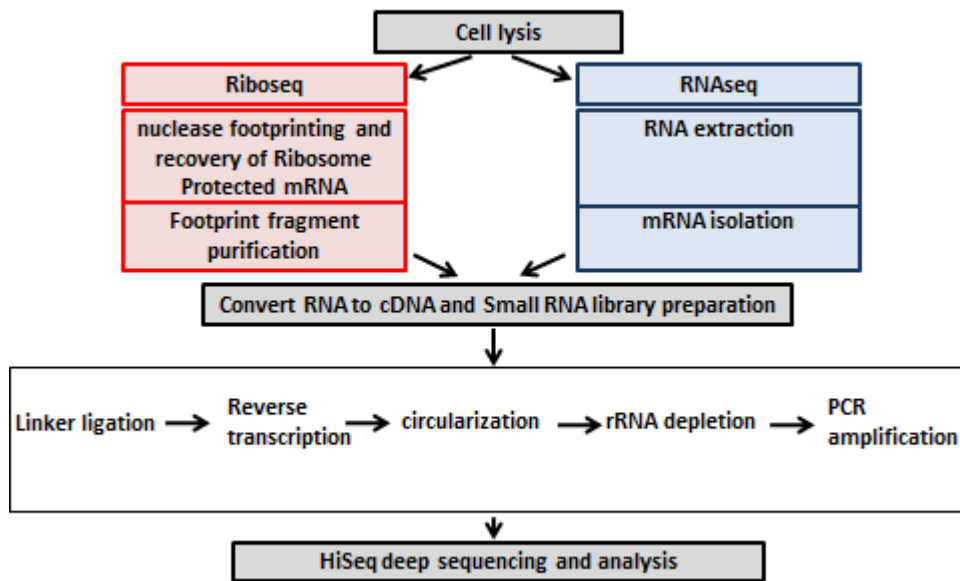
Figure 1.2: Overview of the ribosome profiling protocol. After cell lysis, translating ribosomes are digested with Rnase I. The mRNA fragments, which are protected from the ribosomes are recovered and converted into a cDNA library. Library preparation starts with a ligation of a linker, followed by reverse transcription and subsequent circularisation. The library is amplified by PCR and analysed by deep sequencing. RNAseq is performed in parallel and it is used as an internal control for RNA levels.

1.4 Advantages of using ribosome profiling

Ribosome profiling provides quantitative information about protein synthesis, with unprecedented sensitivity. The stable isotope labelling methods coupled by mass spectrometry, which also enable measurement of protein synthesis and turnover, cannot provide the sensitivity available by ribosome profiling (Vogel & Marcotte 2012).

Ribosome profiling also provides positional information that enables the detection of codon periodicity (Ingolia et al. 2009). Analysis of ribosome footprint positions, facilitated the elucidation of many translational events, such as non-AUG initiation and upstream open reading frame (uORF) translation (Gerashchenko et al. 2012; Ingolia et al. 2011; Brar et al. 2012), ribosomal frameshifting (Michel et al. 2012), pausing (G.-W. Li et al. 2012; Woolstenhulme et al. 2015) and stop codon readthrough (Namy et al. 2004; Dunn et al. 2013).

At the level of gene expression studies, ribosome profiling enables detection of changes in protein synthesis as they occur (Andreev et al. 2015; Ingolia et al. 2009).

Until the advent of ribosome profiling, genome wide gene expression studies were performed by microarray analysis or RNA-seq (Wang et al. 2009), both of which use transcript abundance as a proxy for protein synthesis. Unlike ribosome profiling, which captures both transcript abundance and actively translated mRNAs, these techniques do not take into account translational regulation. The previously used microarray analysis of polysome associated mRNAs (Zong et al. 1999; Johannes et al. 1999), cannot provide information about the exact position of the ribosome and about translation of low-abundance transcripts. More efficiently translated mRNA transcripts have more ribosomes bound to them and can be separated from the least efficiently translated ones, through polysome fractionation by sucrose density gradient centrifugation and fractions' isolation (Arava et al. 2003; Hendrickson et al. 2009). Subsequent microarray analysis of fractions, can give more quantitative results, but requires analysis of many fractions per sample.

Proteomics is useful for monitoring the steady state levels of protein, but it does not provide insight into the dynamics of protein regulation (Vogel & Marcotte 2012).

Ribosome profiling has provided insight into the sequences that are being translated. For example, long interspersed non-coding RNAs (lincRNAs), which were considered nontranslatable transcirpts (Bertone et al. 2004; Carninci et al. 2005; Guttman et al. 2009), have been reported to be translated in profiling experiments (Guttman et al. 2013; Ingolia et al. 2014). Mass spectrometry has confirmed translation of some of the newly identified peptides (Stern-Ginossar et al. 2012; Schwaid et al. 2013; Slavoff et al. 2013). Although translated products can get degraded too quickly to make functional peptides, they are physiologically relevant, because degraded peptides in the ER are the source of antigens presented to circulating cytotoxic T cells and increased ER stress could potentially affect immune responses (Starck et al. 2012).

1.5 Limitations of ribosome profiling

Ribosome profiling has some limitations, which need to be considered in data interpretation.

Ribosomal RNA contamination decreases the amount of informative sequencing data and can increase the cost of the experiment. Removal of rRNA sequences can be achieved by subtractive hybridisation (Ingolia et al. 2012). Alternatively duplex-specific nuclease (DSN) can be used (Zhulidov et al. 2004; Chung et al. 2015). DSN is an enzyme isolated from the Kamchatka crab (*Parlithodes camtschaticus*), which can cleave double-stranded DNA and RNA-DNA hybrid duplexes (Shagin et al. 2002). rRNA removal involves denaturation of cDNA, reannealing and enzymatic degradation of double stranded DNA by DSN (Chung et al. 2015; Zhulidov et al. 2004).

The first step of ribosome profiling is translation inhibition to freeze ribosomes *in situ*. Inhibition has to be fast to avoid artificial accumulation of ribosomes and blurred signal. Although flash freezing of samples can be very efficient at capturing the exact position of ribosomes (Ingolia et al. 2012), it's application can be technically challenging. Alternatives such as cycloheximide treatment, can distort ribosomal distribution especially near translation start sites (Guydosh & Green 2014; Ingolia et al. 2012). This issue can be addressed by avoiding addition of cycloheximide to the cell medium (Gerashchenko et al. 2012).

The library preparation process itself includes many steps, each of which has the potential to distort the data. The current appoach to library preparation involves ligation of preadenylated oligonucleotides at the 3' end of RPFs, followed by cDNA synthesis and first-strand cDNA circularization and subsequent PCR amplification (Levin et al. 2010). This approach reduces sequence dependent biases in library preparation (Jayaprakash et al., 2011; Hafner et al., 2011).

Sequence data alignment can be challenging especially for highly repetitive areas or areas showing high similarity (Brar & Weissman 2015). The protein synthesis rate in profiling is inferred from the average ribosome occupancy along the mRNA transcript. Differences in translation elongation rate among mRNA and ribosomes that do not finish translation can affect the accuracy of the protein synthesis rate (Brar & Weissman 2015).

RPFs are inferred on the basis of their size. As a result of this, it is not always possible to differentiate between RNAs that are translated and contaminating ones (Brar & Weissman 2015). Contaminating RNAs are usually parts of ribonucleoprotein (RNP) complexes, such as RNaseP, telomerase and the vault RNP and thus get protected by RNase digestion and subsequently sediment with ribosomes. A recently developed tool (Ingolia et al. 2014) enables differentiation between RPFs and co-precipitating RNAs, by identifying the stereotypical distribution of ribosome footprints of each experiment.

Visualization of published Ribo-seq data is available online in GWIPS-vis (http://gwips.ucc.ie/) (Michel et al. 2014). GWIPS-vis is based on the Genome Browser of the University of California Santa Cruz (http://genome.ucsc.edu/) (Meyer et al. 2013). Figure (1.4) shows a screenshot of the GWIPS-vis (hg19) RNAseq for gene *PIK3CA*. RNAseq alignments are shown as green histograms and Riboseq alignments are shown as red columns.

Figure 1.3 Screenshot of PIK3CA from GWIPS-vis

Bibliography

Adams, J.M. & Cory, S., 2007. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, 26(9), pp.1324–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17322918 [Accessed July 28, 2016].

Aird, D. et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology*, 12(2), p.R18.

Anastas, J.N. & Moon, R.T., 2012. WNT signalling pathways as therapeutic targets in cancer. *Nature Reviews Cancer*, 13(1), pp.11–26. Available at: http://www.nature.com/doifinder/10.1038/nrc3419 [Accessed July 26, 2016].

Andjelković, M. et al., 1996. Activation and phosphorylation of a pleckstrin homology domain containing protein kinase (RAC-PK/PKB) promoted by serum and protein phosphatase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 93(12), pp.5699–704. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8650155 [Accessed August 21, 2016].

Andreev, D.E. et al., 2015. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome biology*, 16, p.90.

Arava, Y. et al., 2003. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp.3889–3894.

Baker, S.J., Rane, S.G. & Reddy, E.P., 2007. Hematopoietic cytokine receptor signaling. *Oncogene*, 26(47), pp.6724–6737. Available at: http://www.nature.com/doifinder/10.1038/sj.onc.1210757 [Accessed August 18, 2016].

Bentley, D.R. et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–59.

Berdasco, M. et al., 2010. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Developmental cell*, 19(5), pp.698–711. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21074720 [Accessed July 25, 2016].

Bertone, P. et al., 2004. Global Identification of Human Transcribed Sequences with Genome Tiling Arrays. *Science* , 306(5705), pp.2242–2246.

Berx, G. & van Roy, F., 2009. Involvement of members of the cadherin superfamily in cancer. *Cold Spring Harbor perspectives in biology*, 1(6), p.a003129.

Available at: http://www.ncbi.nlm.nih.gov/pubmed/20457567 [Accessed July 27, 2016].

Bianco, R. et al., 2006. Key cancer cell signal transduction pathways as therapeutic targets. *European Journal of Cancer*, 42(3), pp.290–294.

Biondi, R.M. et al., 2001. The PIF-binding pocket in PDK1 is essential for activation of S6K and SGK, but not PKB. *The EMBO journal*, 20(16), pp.4380–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11500365 [Accessed August 20, 2016].

Blasco, M.A., 2005. Telomeres and human disease: ageing, cancer and beyond. *Nature reviews. Genetics*, 6(8), pp.611–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16136653 [Accessed July 29, 2016].

Blume-Jensen, P. & Hunter, T., 2001. Oncogenic kinase signalling. *Nature*, 411(6835), pp.355–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11357143 [Accessed August 22, 2016].

Brar, G.A. et al., 2012. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* , 335(6068), pp.552–557.

Brar, G.A. & Weissman, J.S., 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol*, 16(11), pp.651–664.

Brognard, J. et al., 2007. PHLPP and a Second Isoform, PHLPP2, Differentially Attenuate the Amplitude of Akt Signaling by Regulating Distinct Akt Isoforms. *Molecular Cell*, 25(6), pp.917–931. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276507001177 [Accessed August 21, 2016].

Burkhart, D.L. & Sage, J., 2008. Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nature Reviews Cancer*, 8(9), pp.671–682. Available at: http://www.nature.com/doifinder/10.1038/nrc2399 [Accessed July 28, 2016].

Calnan, D.R. & Brunet, A., 2008. The FoxO code. *Oncogene*, 27(16), pp.2276–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18391970 [Accessed August 22, 2016].

Cantley, L.C., 2002. The Phosphoinositide 3-Kinase Pathway. *Science* , 296(5573), pp.1655–1657.

Carninci, P. et al., 2005. The transcriptional landscape of the mammalian genome. *Science (New York, N.Y.)*, 309(5740), pp.1559–1563.

Cavallaro, U. & Christofori, G., 2004. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nature Reviews Cancer*, 4(2), pp.118–132. Available at: http://www.nature.com/doifinder/10.1038/nrc1276 [Accessed July 27, 2016].

Chung, B.Y. et al., 2015. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA (New York, N.Y.)*, 21(10), pp.1731–1745.

Codogno, P. & Meijer, A.J., 2005. Autophagy and signaling: their role in cell survival and cell death. *Cell death and differentiation*, 12 Suppl 2, pp.1509–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16247498 [Accessed August 22, 2016].

Cully, M. et al., 2006. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nature reviews. Cancer*, 6(3), pp.184–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16453012 [Accessed August 17, 2016].

Dos D. Sarbassov, D.D. et al., 2004. Rictor, a Novel Binding Partner of mTOR, Defines a Rapamycin-Insensitive and Raptor-Independent Pathway that Regulates the Cytoskeleton. *Current Biology*, 14(14), pp.1296–1302. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0960982204004713 [Accessed August 22, 2016].

Davies, M.A. & Samuels, Y., 2010. Analysis of the genome to personalize therapy for melanoma. *Oncogene*, 29(41), pp.5545–5555. Available at: http://www.nature.com/doifinder/10.1038/onc.2010.323 [Accessed July 29, 2016].

Downward, J., 2003. Targeting RAS signalling pathways in cancer therapy. *Nature Reviews Cancer*, 3(1), pp.11–22. Available at: http://www.nature.com/doifinder/10.1038/nrc969 [Accessed July 26, 2016].

Dunn, J.G. et al., 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *eLife*, 2, p.e01179. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3840789&tool=p

mcentrez&rendertype=abstract [Accessed July 17, 2014].

Egeblad, M. et al., 2010. Tumors as organs: complex tissues that interface with the entire organism. *Developmental cell*, 18(6), pp.884–901. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20627072 [Accessed July 28, 2016].

Engelman, J.A., Luo, J. & Cantley, L.C., 2006. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature reviews. Genetics*, 7(8), pp.606–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16847462 [Accessed July 29, 2016].

Evan, G. & Littlewood, T., 1998. A matter of life and cell death. *Science (New York, N.Y.)*, 281(5381), pp.1317–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9721090 [Accessed July 28, 2016].

Feldman, B.J. & Feldman, D., 2001. The development of androgen-independent prostate cancer. *Nature Reviews Cancer*, 1(1), pp.34–45. Available at: http://www.nature.com/doifinder/10.1038/35094009 [Accessed July 29, 2016].

Ferlay, J. et al., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer. Journal international du cancer*, 136(5), pp.E359-86.

Freed-Pastor, W.A. et al., 2012. Mutant p53 Disrupts Mammary Tissue Architecture via the Mevalonate Pathway. *Cell*, 148(1), pp.244–258.

Frias, M.A. et al., 2006. mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. *Current biology : CB*, 16(18), pp.1865–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16919458 [Accessed August 22, 2016].

García-Martínez, J.M. & Alessi, D.R., 2008. mTOR complex 2 (mTORC2) controls hydrophobic motif phosphorylation and activation of serum- and glucocorticoid-induced protein kinase 1 (SGK1). *The Biochemical journal*, 416(3), pp.375–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18925875 [Accessed August 22, 2016].

Gerashchenko, M. V, Lobanov, A. V & Gladyshev, V.N., 2012. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proceedings of the National Academy of Sciences of the United*

States of America, 109(43), pp.17394–17399.

Greaves, M. & Maley, C.C., 2012. Clonal evolution in cancer. *Nature*, 481(7381), pp.306–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22258609 [Accessed July 25, 2016].

Guertin, D.A. et al., 2006. Ablation in Mice of the mTORC Components raptor, rictor, or mLST8 Reveals that mTORC2 Is Required for Signaling to Akt-FOXO and PKCα, but Not S6K1. *Developmental Cell*, 11(6), pp.859–871. Available at: http://linkinghub.elsevier.com/retrieve/pii/S153458070600459X [Accessed August 21, 2016].

Guertin, D.A. & Sabatini, D.M., 2007. Defining the role of mTOR in cancer. *Cancer cell*, 12(1), pp.9–22.

Guo, D., Bell, E.H. & Chakravarti, A., 2013. Lipid metabolism emerges as a promising target for malignant glioma therapy. *CNS Oncology*, 2(3), pp.289–299. Available at: http://www.futuremedicine.com/doi/abs/10.2217/cns.13.20 [Accessed June 1, 2016].

Guo, W. & Giancotti, F.G., 2004. Integrin signalling during tumour progression. *Nature Reviews Molecular Cell Biology*, 5(10), pp.816–826. Available at: http://www.nature.com/doifinder/10.1038/nrm1490 [Accessed August 18, 2016].

Guttman, M. et al., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235), pp.223–227.

Guttman, M. et al., 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, 154(1), pp.240–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3756563&tool=pmcentrez&rendertype=abstract [Accessed May 27, 2014].

Guydosh, N.R. & Green, R., 2014. Dom34 rescues ribosomes in 3' untranslated regions. *Cell*, 156(5), pp.950–962.

Hafner, C. et al., 2012. Activation of the PI3K/AKT Pathway in Merkel Cell Carcinoma K. Smalley, ed. *PLoS ONE*, 7(2), p.e31255. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3281946/.

Hafner, M. et al., 2011. RNA-ligase-dependent biases in miRNA representation in

deep-sequenced small RNA cDNA libraries. *RNA*, 17(9), pp.1697–1712.

Hanahan, D. et al., 1996. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, 86(3), pp.353–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8756718 [Accessed July 27, 2016].

Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, 144(5), pp.646–674.

Hemmings, B.A. & Restuccia, D.F., 2012. PI3K-PKB/Akt Pathway. *Cold Spring Harbor Perspectives in Biology*, 4(9), pp.a011189–a011189. Available at: http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a011189 [Accessed August 21, 2016].

Hendrickson, D.G. et al., 2009. Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA P. D. Zamore, ed. *PLoS Biology*, 7(11), p.e1000238.

Hollander, M.C., Blumenthal, G.M. & Dennis, P.A., 2011. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nature reviews. Cancer*, 11(4), pp.289–301.

Holmes, D., 2011. PI3K pathway inhibitors approach junction. *Nat Rev Drug Discov*, 10(8), pp.563–564. Available at: http://dx.doi.org/10.1038/nrd3527.

Huang, J. & Manning, B.D., 2009. A complex interplay between Akt, TSC2 and the two mTOR complexes. *Biochemical Society transactions*, 37(Pt 1), pp.217–222.

Huang, W.-C. et al., 2012. Activation of Androgen Receptor, Lipogenesis, and Oxidative Stress Converged by SREBP-1 Is Responsible for Regulating Growth and Progression of Prostate Cancer Cells. *Molecular Cancer Research*, 10(1), pp.133–142. Available at: http://mcr.aacrjournals.org/cgi/doi/10.1158/1541-7786.MCR-11-0206 [Accessed June 1, 2016].

Ingolia, N.T. et al., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. , 324(April), pp.218–223.

Ingolia, N.T., 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet*, 15(3), pp.205–213.

Ingolia, N.T. et al., 2014. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell reports*, 8(5), pp.1365–1379.

Ingolia, N.T. et al., 2012. The ribosome profiling strategy for monitoring translation

in vivo by deep sequencing of ribosome-protected mRNA fragments.

Ingolia, N.T., Lareau, L.F. & Weissman, J.S., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), pp.789–802.

Insall, R.H. & Weiner, O.D., 2001. PIP3, PIP2, and cell movement--similar messages, different meanings? *Developmental cell*, 1(6), pp.743–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11740936 [Accessed August 17, 2016].

Jacinto, E. et al., 2004. Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. *Nature cell biology*, 6(11), pp.1122–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15467718 [Accessed August 22, 2016].

Jacinto, E. et al., 2006. SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity. *Cell*, 127(1), pp.125–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16962653 [Accessed August 22, 2016].

Jayaprakash, A.D. et al., 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic acids research*, 39(21), p.e141.

Johannes, G. et al., 1999. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23), pp.13118–13123.

Joyce, J.A. & Pollard, J.W., 2009. Microenvironmental regulation of metastasis. *Nature reviews. Cancer*, 9(4), pp.239–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19279573 [Accessed July 28, 2016].

Katso, R. et al., 2001. Cellular function of phosphoinositide 3-kinases: implications for development, homeostasis, and cancer. *Annual review of cell and developmental biology*, 17, pp.615–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11687500 [Accessed August 18, 2016].

Kiely, M. & Kiely, P.A., 2015. PP2A: The Wolf in Sheep's Clothing? *Cancers*, 7(2), pp.648–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25867001 [Accessed August 17, 2016].

Kim, J.E. & Chen, J., 2004. regulation of peroxisome proliferator-activated receptor-gamma activity by mammalian target of rapamycin and amino acids in adipogenesis. *Diabetes*, 53(11), pp.2748–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15504954 [Accessed August 22, 2016].

Laplante, M. & Sabatini, D.M., 2009. mTOR signaling at a glance. *J Cell Sci*, 122(Pt 20), pp.3589–3594. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19812304 [Accessed August 22, 2016].

Laplante, M. & Sabatini, D.M., 2012. mTOR signaling in growth control and disease. *Cell*, 149(2), pp.274–293. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331679/.

Lawlor, M.A. & Alessi, D.R., 2001. PKB/Akt. *Journal of Cell Science*, 114(16), pp.2903–2910.

Levin, J.Z. et al., 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, 7(9), pp.709–715. Available at: http://www.nature.com/doifinder/10.1038/nmeth.1491 [Accessed June 2, 2016].

Li, G.-W., Oh, E. & Weissman, J.S., 2012. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, 484(7395), pp.538–541.

Li, J. et al., 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science (New York, N.Y.)*, 275(5308), pp.1943–1947.

Li, W. et al., 2012. Repression of endometrial tumor growth by targeting SREBP1 and lipogenesis. *Cell Cycle*, 11(12), pp.2348–2358. Available at: http://www.tandfonline.com/doi/abs/10.4161/cc.20811 [Accessed June 1, 2016].

Li, Y. et al., 2008. Regulation of Akt/FOXO3a/GSK-3beta/AR signaling network by isoflavone in prostate cancer cells. *The Journal of biological chemistry*, 283(41), pp.27707–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18687691 [Accessed August 21, 2016].

Lowe, S.W., Cepero, E. & Evan, G., 2004. Intrinsic tumour suppression. *Nature*,

432(7015), pp.307–315. Available at: http://www.nature.com/doifinder/10.1038/nature03098 [Accessed July 28, 2016].

Makarov, S.S. & Romashkova, J.A., 1999. NF-|[kappa]|B is a target of AKT in anti-apoptotic PDGF signalling. *Nature*, 401(6748), pp.86–90. Available at: http://www.nature.com/doifinder/10.1038/43474 [Accessed August 21, 2016].

Mantovani, A., 2010. Molecular pathways linking inflammation and cancer. *Current molecular medicine*, 10(4), pp.369–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20455855 [Accessed July 28, 2016].

Mardilovich, K. et al., 2009. Expression and function of the insulin receptor substrate proteins in cancer. *Cell Communication and Signaling*, 7(1), p.14. Available at: http://biosignaling.biomedcentral.com/articles/10.1186/1478-811X-7-14 [Accessed August 16, 2016].

Mayer, C. et al., 2004. mTOR-dependent activation of the transcription factor TIF-IA links rRNA synthesis to nutrient availability. *Genes & development*, 18(4), pp.423–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15004009 [Accessed August 22, 2016].

Mclaughlin, S. et al., 2002. PIP 2 AND PROTEINS: Interactions, Organization, and Information Flow. *Annu. Rev. Biophys. Biomol. Struct*, 31, pp.151–75.

Meador, J. & Kennell, D., 1990. Cloning and sequencing the gene encoding Escherichia coli ribonuclease I: exact physical mapping using the genome library. *Gene*, 95(1), pp.1–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2253883 [Accessed June 1, 2016].

Meric-Bernstam, F. & Gonzalez-Angulo, A.M., 2009. Targeting the mTOR signaling network for cancer therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(13), pp.2278–2287.

Meyer, L.R. et al., 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic acids research*, 41(Database issue), pp.D64-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23155063 [Accessed June 3, 2016].

Michel, A.M. et al., 2014. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*, 42. Available at: http://dx.doi.org/10.1093/nar/gkt1035.

Michel, A.M. et al., 2012. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Research* , 22(11), pp.2219–2229. Available at: http://genome.cshlp.org/content/22/11/2219.abstract.

Namy, O. et al., 2004. Reprogrammed genetic decoding in cellular gene expression. *Molecular cell*, 13(2), pp.157–168.

Nardi, V. et al., 2012. Activation of PI3K signaling in Merkel cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18(5), pp.1227–1236.

Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D., 2010. Genomic instability--an evolving hallmark of cancer. *Nature reviews. Molecular cell biology*, 11(3), pp.220–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20177397 [Accessed July 27, 2016].

Nelson, B.H., 2008. The impact of T-cell immunity on ovarian cancer outcomes. *Immunological Reviews*, 222(1), pp.101–116. Available at: http://doi.wiley.com/10.1111/j.1600-065X.2008.00614.x [Accessed July 28, 2016].

Okkenhaug, K. & Vanhaesebroeck, B., 2003. PI3K in lymphocyte development, differentiation and activation. *Nature Reviews Immunology*, 3(4), pp.317–330. Available at: http://www.nature.com/doifinder/10.1038/nri1056 [Accessed August 18, 2016].

Parker, P.J. & Waterfield, M.D., 1992. Phosphatidylinositol 3-kinase: a novel effector. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research*, 3(10), pp.747–752.

Paulson, K.G. et al., 2009. Array-CGH reveals recurrent genomic changes in Merkel cell carcinoma including amplification of L-Myc. *The Journal of investigative dermatology*, 129(6), pp.1547–1555.

Pearce, L.R. et al., 2007. Identification of Protor as a novel Rictor-binding component of mTOR complex-2. *The Biochemical journal*, 405(3), pp.513–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17461779 [Accessed August 22, 2016].

Peterson, T.R. et al., 2009. DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. *Cell*, 137(5), pp.873–

86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19446321 [Accessed August 22, 2016].

Porstmann, T. et al., 2008. SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell metabolism*, 8(3), pp.224–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18762023 [Accessed August 22, 2016].

Qiu, W. et al., 2006. PIK3CA Mutations in Head and Neck Squamous Cell Carcinoma. *Clinical Cancer Research* , 12(5), pp.1441–1446. Available at: http://clincancerres.aacrjournals.org/content/12/5/1441.abstract.

Richter, J.D. & Sonenberg, N., 2005. Regulation of cap-dependent translation by eIF4E inhibitory proteins. , pp.477–480.

Rodriguez-Viciana, P. et al., 1994. Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature*, 370(6490), pp.527–532.

Ruderman, N.B. et al., 1990. Activation of phosphatidylinositol 3-kinase by insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 87(4), pp.1411–1415.

Ryerson, A.B. et al., 2016. Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer. *Cancer*, 122(9), pp.1312–1337.

Salk, J.J., Fox, E.J. & Loeb, L.A., 2010. Mutational heterogeneity in human cancers: origin and consequences. *Annual review of pathology*, 5, pp.51–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19743960 [Accessed July 27, 2016].

Sarbassov, D.D. et al., 2006. Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB. *Molecular cell*, 22(2), pp.159–168.

Sarbassov, D.D., Ali, S.M. & Sabatini, D.M., 2005. Growing roles for the mTOR pathway. *Current opinion in cell biology*, 17(6), pp.596–603. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16226444 [Accessed August 22, 2016].

Schenone, S. et al., 2011. ATP-competitive inhibitors of mTOR: an update. *Current medicinal chemistry*, 18(20), pp.2995–3014.

Schieke, S.M. et al., 2006. The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity. *The Journal of biological chemistry*, 281(37), pp.27643–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16847060 [Accessed August 22, 2016].

Schwaid, A.G. et al., 2013. Chemoproteomic Discovery of Cysteine-Containing Human Short Open Reading Frames. *Journal of the American Chemical Society*, 135(45), pp.16750–16753.

Shagin, D.A. et al., 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. *Genome research*, 12(12), pp.1935–1942.

Sharma, S., Kelly, T.K. & Jones, P.A., 2010. Epigenetics in cancer. *Carcinogenesis*, 31(1), pp.27–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19752007 [Accessed July 25, 2016].

Shay, J.W. & Wright, W.E., 2000. Hayflick, his limit, and cellular ageing. *Nature reviews. Molecular cell biology*, 1(1), pp.72–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11413492 [Accessed July 29, 2016].

Shields, J.D. et al., 2010. Induction of lymphoidlike stroma and immune escape by tumors that express the chemokine CCL21. *Science (New York, N.Y.)*, 328(5979), pp.749–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20339029 [Accessed July 28, 2016].

Shor, B. et al., 2008. A new pharmacologic action of CCI-779 involves FKBP12-independent inhibition of mTOR kinase activity and profound repression of global protein synthesis. *Cancer research*, 68(8), pp.2934–2943.

Slavoff, S.A. et al., 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nature chemical biology*, 9(1), pp.59–64.

Stambolic, V. et al., 1998. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*, 95(1), pp.29–39.

Starck, S.R. et al., 2012. Leucine-tRNA Initiates at CUG Start Codons for Protein Synthesis and Presentation by MHC Class I. *Science*, 336(6089), pp.1719–1723.

Steck, P.A. et al., 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature genetics*, 15(4), pp.356–362.

Steitz, J.A., 1969. Polypeptide Chain Initiation: Nucleotide Sequences of the Threee Ribosomal Binding Sites in Bacteriophage R17 RNA. *Nature*, 224(5223), pp.957–964.

Stern-Ginossar, N. et al., 2012. Decoding Human Cytomegalovirus. *Science* ,

338(6110), pp.1088–1093.

Stratton, M.R., Campbell, P.J. & Futreal, P.A., 2009. The cancer genome. *Nature*, 458(7239), pp.719–724.

Strauss, D.C. et al., 2010. Transmission of donor melanoma by organ transplantation. *The Lancet. Oncology*, 11(8), pp.790–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20451456 [Accessed July 28, 2016].

Sun, Y. et al., 2015. SREBP1 regulates tumorigenesis and prognosis of pancreatic cancer through targeting lipid metabolism. *Tumor Biology*, 36(6), pp.4133–4141. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25589463 [Accessed June 1, 2016].

Thong, F.S.L., Bilan, P.J. & Klip, A., 2007. The Rab GTPase-activating protein AS160 integrates Akt, protein kinase C, and AMP-activated protein kinase signals regulating GLUT4 traffic. *Diabetes*, 56(2), pp.414–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17259386 [Accessed August 21, 2016].

Vanhaesebroeck, B. et al., 2010. The emerging mechanisms of isoform-specific PI3K signalling. *Nature Reviews Molecular Cell Biology*, 11(5), pp.329–341. Available at: http://www.nature.com/doifinder/10.1038/nrm2882 [Accessed August 20, 2016].

Vivanco, I. & Sawyers, C.L., 2002. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nature reviews. Cancer*, 2(7), pp.489–501. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12094235 [Accessed August 17, 2016].

Vogel, C. & Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics*, 13(4), pp.227–232.

Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63.

Warburg, O., 1956. Injuring of Respiration the Origin of Cancer Cells. *Science*, 123(3191), pp.309–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13298683 [Accessed July 27, 2016].

Warburg, O., 1956. On Respiratory Impairment in Cancer Cells. *Science*, 124(3215), pp.269–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13351639 [Accessed July 27, 2016].

Weiss, R.B. & Atkins, J.F., 2011. Translation Goes Global. *Science* , 334(6062), pp.1509–1510.

Wolin, S.L. & Walter, P., 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO Journal*, 7(11), pp.3559–3569.

Woolstenhulme, C.J. et al., 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell reports*, 11(1), pp.13–21.

Wu, X. et al., 1998. The PTEN/MMAC1 tumor suppressor phosphatase functions as a negative regulator of the phosphoinositide 3-kinase/Akt pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 95(26), pp.15587–15591.

Yang, L. et al., 2010. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends in immunology*, 31(6), pp.220–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20538542 [Accessed July 28, 2016].

Yang, Q. et al., 2006. Identification of Sin1 as an essential TORC2 component required for complex formation and kinase activity. *Genes & development*, 20(20), pp.2820–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17043309 [Accessed August 22, 2016].

Yarden, Y. & Ullrich, A., 1988. Growth factor receptor tyrosine kinases. *Annual review of biochemistry*, 57, pp.443–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3052279 [Accessed August 22, 2016].

Yuan, T.L. & Cantley, L.C., 2008. PI3K pathway alterations in cancer: variations on a theme. *Oncogene*, 27(41), pp.5497–5510. Available at: http://www.nature.com/doifinder/10.1038/onc.2008.245 [Accessed July 29, 2016].

Zhang, H. et al., 2007. PDGFRs are critical for PI3K/Akt activation and negatively regulated by mTOR. *The Journal of clinical investigation*, 117(3), pp.730–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17290308 [Accessed August 17, 2016].

Zhou, B.P. & Hung, M.-C., 2002. Novel targets of Akt, p21(Cipl/WAF1), and MDM2. *Seminars in oncology*, 29(3 Suppl 11), pp.62–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12138399 [Accessed August 21, 2016].

Zhou, H. et al., 2000. Akt Regulates Cell Survival and Apoptosis at a Postmitochondrial Level. *The Journal of Cell Biology*, 151(3), pp.483–494. Available at: http://www.jcb.org/cgi/content/full/151/3/483 [Accessed August 21, 2016].

Zhulidov, P.A. et al., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic acids research*, 32(3), p.e37.

Zoncu, R., Efeyan, A. & Sabatini, D.M., 2011. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol*, 12(1), pp.21–35.

Zong, Q. et al., 1999. Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proceedings of the National Academy of Sciences of the United States of America*, 96(19), pp.10632–10636.

2   Systematic analysis of the PTEN 5' leader identifies a major AUU initiated proteoform

Gary Loughran generated Figure 2.3.1b, 2.3.3b, 2.3.4a, 2.3.5a, 2.3.5c and
Supplementary Figures 5, 7, 8, 10, 11
Ivaylo P. Ivanov generated Figure2.3.1a and Supplementary Figure 1
Pavel V. Baranov generated Supplementary Figure 2
Ruslan I. Dmitriev and Ioanna Tzani generated Supplementary Figure 4

Ioanna Tzani generated all the other figures of this chapter

This chapter has been published in Open Biology
http://rsob.royalsocietypublishing.org/content/6/5/150203

Abstract

Abundant evidence for translation within the 5' leaders of many human genes is rapidly emerging, especially, because of the advent of ribosome profiling. In most cases, it is believed that the act of translation rather than the encoded peptide is important. However, the wealth of available sequencing data in recent years allows phylogenetic detection of sequences within 5' leaders that have emerged under coding constraint and therefore allow for the prediction of functional 5' leader translation. Using this approach, we previously predicted a CUG-initiated, 173 amino acid N-terminal extension to the human tumour suppressor PTEN. Here, a systematic experimental analysis of translation events in the *PTEN* 5' leader identifies at least two additional non-AUG-initiated PTEN proteoforms that are expressed in most human cell-lines tested. The most abundant extended PTEN proteoform initiates at a conserved AUU codon and extends the canonical AUG-initiated PTEN by 146 amino acids. All N-terminally extended PTEN proteoforms tested retain the ability to downregulate the PI3K pathway. We also provide evidence for the translation of two conserved AUG-initiated upstream open reading frames within the *PTEN* 5' leader that control the ratio of PTEN proteoforms.

2.1   Introduction

The process of translation can be described in four steps: initiation, which is usually tightly regulated; elongation; termination and ribosome recycling (Sonenberg & Hinnebusch 2009). In eukaryotes, the scanning model for translation initiation postulates that the small ribosomal subunit, in complex with initiation factors and Met-tRNA$_i$, binds first to the 5' cap then scans 3' until a suitable initiation codon is found (Kozak 1980). Base-pairing interactions between the anticodon loop of the Met-tRNA$_i$ bound to the ribosome and an AUG codon in the mRNA cause the ribosome to stop scanning and set the reading frame for protein synthesis (Cigan et al. 1988). Typically, the ribosome initiates protein synthesis at the AUG codon closest to the 5' end of the mRNA, though the efficiency of initiation is dependent on the nucleotide sequence surrounding the initiator codon with the optimal sequence known as the Kozak context (Kozak 1999). The Kozak context— comprising 6 nt before and 1 nt immediately following a potential initiation codon—has significant influence on the recognition of an initiation site, through partially understood mechanisms requiring the activities of eIF1 (Mitchell & Lorsch 2008; Richard J Jackson et al. 2010; Llácer et al. 2015; Hussain et al. 2014) and eIF5 (Nanda et al. 2009; Valásek et al. 2001; Hinnebusch 2014). Using multiple sequence alignments, the consensus context in mammals was identified as GCC<u>R</u>CC**AUG**<u>G</u> (Kozak 1987) with the identity of the underlined nucleotides in the -3 and +4 positions (relative to the 'A' of the AUG) being the most important. However, a recent high-throughput analysis of all possible initiation contexts revealed RYMRMVAUGGC as the optimal context in human and mouse cells and additionally revealed synergistic effects of neighbouring nucleotides (Noderer et al. 2014).

Initiation can occur at most codons that differ from AUG by a single nucleotide (non-cognate or non-AUG). Seven out of the nine possible single-nucleotide substitutions at the AUG start codon of dihydrofolate reductase were functional as translation start sites in mammalian cells (Peabody 1989). In all of the cases in which it was examined, the N-terminal residue of these proteins was methionine (Peabody 1989), suggesting that translation initiation relied on mis-pairing between the anticodon of Met-tRNA$_i$ and the non-AUG start codon in the

mRNA. However, a report exists of CUG initiation by elongator Leu-tRNA functioning as initiator tRNA (Starck et al. 2012).

Initiation is the only step where an incoming aminoacyl tRNA is bound directly in the ribosomal P-site (Ramakrishnan 2002; Simonetti et al. 2008). Unlike the A-site, where mRNA : tRNA interactions are strictly monitored by the decoding centre (Ogle et al. 2001), the P-site can tolerate mismatches in the codon : anticodon duplex (Potapov et al. 1995; Baranov et al. 2004; Svidritskiy & Korostelev 2015; Herr et al. 2004). This allows the incorporation of Met-tRNA$_i$ at a wider range of codons compared with elongator Met-tRNA whose incorporation is strictly limited to AUG codons. The most favourable context for non-AUG initiation is believed to be identical to that for AUG starts (Chen et al. 2008; Kozak 1989; Portis et al. 1994). In addition, a strong RNA secondary structure starting approximately 15 nt downstream of the non-AUG codon may significantly increase initiation efficiency (Kozak 1990). Another important factor for non-AUG initiation is that it is located upstream of the most 5' AUG codon (Audrey M Michel et al. 2014).

When initiation codons occur in the 5' leaders of transcripts they give rise to either N-terminal extensions to the main ORF or else upstream open reading frames (uORFs). It has been estimated that AUG-initiated uORFs are present in approximately half of the human protein coding genes (Calvo et al. 2009). Furthermore, ribosome profiling provided evidence for the presence of translating ribosomes on more than 200 non-AUG-initiated uORFs in yeast (Ingolia et al. 2009) and much more widespread non-AUG initiation in mammals (Ingolia et al. 2011; Lee et al. 2012; Fritsch et al. 2012). In general, the translation of uORFs has an inhibitory effect on translation of the main protein coding ORF, because ribosomes terminating a uORF are often unable to reinitiate owing to the loss of necessary initiation factors. However, the 40S subunits of ribosomes translating short ORFs (less than 35 codons) may retain some initiation factors after termination (Kozak 1999; Rajkowitsch et al. 2004)—although efficient re-initiation is precluded until all necessary initiation factors have been reloaded onto the 40S subunit. In most instances, there is a requirement for a ternary complex of eIF2, GTP and Met-tRNAi which is regulated by the phosphorylation status of eIF2α (Hinnebusch 2005).

An earlier finding that sequences in the 5' leaders are highly conserved and that the level of conservation globally increases towards the leader/main ORF boundaries (Shabalina et al. 2004) suggests that this conservation could be due, in part, to the 3' ends of a portion of 5' leaders encoding N-terminal extensions to the annotated AUG-initiated proteins. In total, more than 60 instances of non-AUG-initiated N-terminal extensions have been predicted or verified experimentally in mammals (Tikole & Sankararamakrishnan 2006; Ivanov et al. 2011; Damme et al. 2014). In most cases, the non-AUG initiation provides an alternative longer proteoform in addition to a proteoform resulting from initiation at a standard AUG codon downstream via a process termed 'leaky scanning'. Where alternative proteoforms are produced as a result of leaky scanning, the longer isoform frequently contains a signal for subcellular localization that is absent in the shorter form (Arnaud et al. 1999; Tee & Jaffe 2001; Packham et al. 1997; Hann et al. 1988; Coldwell et al. 2004).

Previously, we performed a systematic analysis of the 5' leaders of human GenBank RefSeq mRNAs to investigate the extent of non-AUG initiation in humans (Ivanov et al. 2011). This involved analysis of codon substitution rates in pairwise alignments of human and mice orthologous sequences to identify regions of 5' leaders evolving under the constraints of protein coding evolution. When a region within a 5' leader evolves under such constraints, it is very likely that the encoded protein can improve an organism's fitness and is thus functional. This approach predicts a CUG-initiated, 173 amino acid, N-terminal extension within the 5' leader of *phosphatase and tensin homologue on chromosome ten (PTEN). PTEN* is a powerful tumour suppressor gene that encodes a dual-specificity phosphatase (Li et al. 1997; Steck et al. 1997) frequently mutated in human cancers (Stambolic et al. 2000) and autism spectrum disorders (Varga et al. 2009). Its best characterized function is its ability to negatively regulate cell survival by dephosphorylating phosphatidylinositol 3,4,5 triphosphate (PIP3) and thus inhibiting phosphoinositide 3-kinase (PI3K) signalling (Maehama & Dixon 1998).

Independently, two other groups subsequently identified the same PTEN N-terminal extension (Hopkins et al. 2013; Liang et al. 2014). Here we extended our

analysis of the *PTEN* 5' leader and identify non-AUG-initiated translation that leads to the synthesis of at least two additional N-terminally extended proteoforms.

2.2 Material and methods

2.2.1 Plasmids

The *PTEN* 5' leader was amplified by PCR from HEK-293T genomic DNA using appropriate primers (Integrated DNA Technologies) that incorporated a 5' *Hind*III restriction site and a 30 *Bam*HI restriction site. *PTEN* 5' leaders were mutated by two-step PCR with appropriately designed primers. Amplicons were cloned *Hind*III/*Bam*HI into the dual luciferase plasmid p2-Luc (Grentzmann et al. 1998) such that the PTEN 5' leader replaced the *Renilla*-encoding sequences and were fused directly to the firefly-encoding sequences. The coding sequence of PTEN was obtained as a gblock (Integrated DNA Technologies) with incorporated 5' *Hind*III and a 3' *Xba*I restriction site and cloned into phRL-CMV (Promega). The coding sequence of PTEN was also subcloned downstream of the *PTEN* 5' leader to replace the firefly encoding sequence of the *PTEN* 5' leader-FLuc constructs made previously.

For *PTEN* 5' leader GFP fusions, EGFP was digested from pEGFP-N3 (Clontech) with *Bam*HI and *Xba*I restriction enzymes and cloned *Bam*HI/*Xba*I using standard cloning techniques into the *PTEN* 5' leader firefly-encoding plasmids described above, such that the EGFP sequence replaced the firefly-encoding sequences and were fused directly to the *PTEN* 5' leader. The EGFP AUG to AAA mutation was made by two-step PCR with appropriately designed primers. For signal peptide –*Gaussia* luciferase fusions, amplicons generated by two-step PCR were cloned *Bam*HI/*Xba*I into pCMV-GLuc (NEB). All clones were verified by sequencing. eIF1 and eIF5 overexpression constructs were described previously (Ivanov et al. 2010; Loughran et al. 2012).

2.2.2 Cell culture

HEK-293T, MDA-MB-231, MCF-7, HeLa, HUH-7, U2OS and A172 cells were maintained in DMEM supplemented with 10% FBS, 1 mM L-glutamine and antibiotics. PC3 cells were maintained in RPMI supplemented with 10% FBS, 1 mM L-glutamine and antibiotics.

2.2.3 Luciferase assay

HEK-293T cells were transfected with Lipofectamine 2000 reagent (Invitrogen), using the 1 day protocol (reverse transfection) in which suspended cells are added directly to the DNA complexes in full-area 96-well plates. For each transfection, the following were added to each well: 100 ng of each firefly luciferase-expressing plasmid, 10 ng of each *Renilla* luciferase-expressing plasmid plus 0.4 µl Lipofectamine 2000 (Invitrogen) in 48.4 µl Opti-Mem (Gibco). The transfecting DNA complexes in each well were incubated with $4x10^4$ cells suspended in 50 µl DMEM (RPMI for PC3 cells) plus 10% FBS. Transfected cells were incubated at 37°C in 5% CO_2 for 24 h. On the next day, cells were washed once with 1 x PBS and then lysed in 25 µl of 1 x passive lysis buffer (PLB; Promega) and firefly and *Renilla* luciferase activities were determined using the Dual Luciferase Stop & Glow Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems).

Firefly luciferase activity was calculated relative to the activity of the co-transfected control plasmid expressing *Renilla* luciferase (pSV40-Renilla). All data points were averaged, and the standard deviation calculated. Data represent the mean and standard deviation of at least three independent experiments each done in quadruplicate.

For secretion luciferase assays (see figure 2.3.5c), only *Gaussia* luciferase activities were assayed and the percentage activity in both the cell lysate (intracellular) and culture media (extracellular) calculated. For sodium arsenite treatment, HEK-293T cells were transfected as above, and sodium arsenite (5 mM) was added 6 h post-transfection for either 2 or 4 h.

2.2.4 Immunoblotting

Cells were transfected in six-well plates using Lipofectamine 2000 reagent, again using the 1 day protocol described above, with 1 µg of each indicated plasmid. Where FLuc- and RLuc-expressing plasmids were cotransfected (see figure 2.3.6b and supplementary material, figure S10b) a ratio of 10 : 1 was used. The transfecting DNA complexes in each well were incubated with $0.8 x 10^6$ HEK-293T cells suspended in 3 ml DMEM plus 10% FBS and incubated overnight at 37°C in 5%

CO$_2$. Transfected cells were lysed in 100 µl 1 x PLB and 10 µl removed for dual luciferase assay.

For PC3 transfections, 1.2 x 10^5 cells were plated in triplicate wells (12-well plates) 1 day prior to forward transfection with Lipofectamine 2000 reagent (4 µl) and 500 ng of each indicated plasmid in 500 µl of Opti-Mem. Cells were replenished with fresh RPMI media 6 h post-transfection, then after 18 h, cells were washed and treated with serum free RPMI for a further 24 h. Transfected cells were lysed in radioimmunoprecipitation assay (RIPA) buffer plus protease inhibitors (Sigma) and NaF (20 mM).

Proteins were resolved by 4–12% gradient Bis/Tris–SDS/ PAGE (Bolt^{TM}: Thermo Fisher Scientific) under constant voltage (165 V) for 90 min and transferred to nitrocellulose membranes (Protran), which were incubated at 48°C overnight with primary antibodies. Immunoreactive bands were detected on membranes after incubation with appropriate fluorescently labelled secondary antibody using a LI-COR Odyssey® Infrared Imaging Scanner. Densitometry analysis was performed using I$_{MAGE}$J software (NIH) and GraphPad P$_{RISM}$ used for statistical analysis.

2.2.5   Immunoprecipitation and GFP-trap®

Cells were lysed in RIPA buffer plus protease inhibitors (Sigma), then lysates were incubated with 25 µl of protein G agarose beads (Pierce) plus anti-PTEN (138G6) overnight at 48°C with gentle rocking. The beads were washed (three times) with ice-cold RIPA buffer and then immunoprecipitated proteins removed from the beads by boiling for 5 min. in 20 µl of 2 x SDS–PAGE sample buffer for electrophoresis and immunoblotting.

GFP-trap®_A beads (ChromoTek) were equilibrated according to the manufacturer's protocol. For collection of the extracellular fractions, culture medium was centrifuged at 200g for 5 min at 4°C to remove debris. For intracellular fractions, cells were lysed in RIPA buffer as above. 10 µl of equilibrated beads were added to each fraction and incubated rotating at 48°C for 1 h. The beads were washed (three times) with ice-cold dilution buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then immunoprecipitated proteins removed from the

beads by boiling for 10 min in 40 µl of 2 x SDS–PAGE sample buffer for electrophoresis and immunoblotting.

2.2.6 **RT-qPCR**

HEK-293T cells were transfected in triplicate wells (six-well plate) as above with either construct 1 (wild-type *PTEN* leader fused to firefly luciferase) or construct 5 (*PTEN* leader with all uAUGs mutated to AGG). 24 h post-transfection cells were removed (trypsin), divided into two aliquots of 20% and 80% and then collected by centrifugation. 20% of cells were resuspended in 20 µl 1 PLB for dual luciferase assay. RNA was isolated using Trizol reagent (Invitrogen) from the remaining 80% of cells, and 500 ng of DNAse-treated (RQ1: Promega) RNA was reverse transcribed using oligo-dT and random hexamers according to the manufacturer's instructions (Superscritpt III: Invitrogen). Reactions minus reverse transcriptase were included to control for contaminating genomic or plasmid DNA. SYBR green (Qiagen) qPCR was performed on an Applied Biosystems 7300 real-time PCR system with firefly luciferase primers (sense TGGAGAGCAACTGCATAAGG and antisense ATTCCGCGTACGTGATGTT) and a set of intron-spanning control primers for GAPDH (sense AGCCT CCCGCTTCGCTCTCT and antisense CCAGGCGCCCAA TACGACCA). Relative RNA quantitation was analyzed using the Livak method ($2^{-\Delta\Delta Ct}$) and used to normalize relative luciferase activities to relative RNA levels.

2.2.7 **Antibodies**

An affinity-purified rabbit polyclonal antibody (anti-PTEN-L) directed to a predicted antigen (PRHQQLLPSLSSFFFSHR LPD) within all four extended PTEN proteoforms was prepared by GenScript. The following commercially available antibodies were also used. Mouse anti-PTEN (6H2.1; Millipore), rabbit anti-PTEN (138G6: Cell Signalling), rabbit anti-GFP (A6455: Novex), goat anti-firefly luciferase (G7451: Promega), rabbit anti-S473-phospho-AKT (D9E: Cell Signalling), mouse anti-pan-AKT (40D4: Cell Signalling), mouse anti-*Renilla* luciferase (1D5.2 Millipore), rabbit anti-eIF5 (ab85913: Abcam) and mouse anti-β-actin (AC-15: Sigma). Anti-eIF1 was a generous gift from Ariel Stanhill (Technion-Israel Institute of Technology).

2.2.8 Fluorescence microscopy

Live cell imaging was performed as described before (Dmitriev et al. 2013) using an inverted Axiovert 200 fluorescence microscope (Zeiss), equipped with 100/1.4 Plan Apochromat oil-immersion objective (Zeiss), pulsed excitation module (470 nm, 590 nm LEDs), bandpass filters 510–560 nm and gated CCD camera (LaVision, Biotec). Briefly, HeLa cells were seeded onto eight well chambers pre-coated with a mixture of collagen IV and poly-D-lysine (Ibidi), allowed to attach (24 h) and forward transfected for 24 h with plasmid DNAs encoding PTEN leader –GFP fusions as indicated. Images were processed using IMSPECTOR software (LaVision, Biotec) and combined in Adobe I$_{LLUSTRATOR}$ CS2.

2.3   Results

Previous searches for evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences predicted a CUG-initiated, 173 amino acid extension to the tumour suppressor PTEN (Ivanov et al. 2011). Further phylogenetic analysis of the *PTEN* 5' leader with additional sequence data reveals deep nucleotide conservation in mammals (figure 2.3.1a and supplementary material, figure S1). Two independent groups (Hopkins et al. 2013; Liang et al. 2014) have recently provided experimental evidence for a human PTEN N-terminal extension that reportedly initiates at the same CUG predicted by Ivanov *et al*. (2011). Preliminary results in our laboratory indicated the existence of multiple N-terminally extended PTEN proteoforms. Here we set out to systematically investigate these multiple proteoforms and also to determine the effect, if any, of two conserved AUG initiated uORFs on translation of these PTEN proteoforms (boxed in figure 2.3.1a).

We noted that the 5' end of both GenBank RefSeq PTEN mRNA isoforms (NM_000314.6, NM_001304718.1), which have identical first exons, do not correspond to the transcription start site predicted by the Fantom Projects' CAGE analysis (Kawaji et al. 2011) which finds that the transcription start site is a further 187 nt 3' of the Genbank RefSeq annotated *PTEN* mRNA 5' end (supplementary material, figure S2). A +187 transcription start site is also in agreement with mRNAseq data obtained as controls to multiple ribosome profiling experiments, available in GWIPS-viz (A M Michel et al. 2014) (see supplementary material, figure S2) as well as with the majority of publicly available human *PTEN* expressed sequence tags. Therefore, in this study, all test constructs with the *PTEN* 5' leader start at +187 relative to the 5' end of GenBank RefSeq *PTEN* mRNAs.

We first transfected HEK-293T cells with plasmid DNA expressing the human *PTEN* 5' leader fused to sequences encoding firefly luciferase (FLuc; figure 2.3.1b). Immunoblots generated from transfected cell lysates and probed with FLuc antibodies detected four slower migrating FLuc proteoforms when FLuc is preceded by the *PTEN* 5' leader (lane 2, top panel, figure 3.2.1c). These four proteoforms are absent from cells transfected with FLuc expressing constructs lacking the *PTEN* 5' leader (lane 1, top panel, figure 2.3.1c) indicating that the multiple proteoforms are

not post-translationally modified variants of FLuc. Furthermore, the same pattern of proteoforms was also detected when these same lysates were probed with a custom antibody (anti-PTEN-L) directed against a peptide predicted from sequences immediately 5' of, and in-frame with, the *PTEN* main ORF (lane 2, middle panel, figure 2.3.1c).

These multiple proteoforms are most likely N-terminally extended variants generated by initiation at in-frame non-AUG codons within the *PTEN* 5' leader. Alternatively, some of these proteoforms could be proteolytically cleaved variants of the previously reported (Hopkins et al. 2013; Liang et al. 2014) CUG-initiated 173 amino acid N-terminally extended PTEN proteoform. To investigate this latter possibility, we transfected HEK-293T cells with *PTEN* 5' leader-FLuc expressing constructs in which this CUG was changed to either a non-initiating AGG triplet or to a UAA stop codon, that not only prevents initiation but also terminates translation from upstream initiation sites. Both anti-FLuc and anti-PTEN-L immunoblots from these transfected lysates indicate that mutation of the CUG leads to the disappearance of only the most slowly migrating proteoform, thus ruling out the possibility that some of the other proteoforms are cleavage products (lanes 3 and 4, top and middle panels, figure 2.3.1c). These observations suggest that the *PTEN* 5' leader has the potential to generate at least four N-terminally extended proteoforms (that are within the detection limits of these experiments) and that the previously reported proteoform initiated at CUG, while likely the longest proteoform may not be the most abundant.

Because of the recent reporting of an N-terminally extended PTEN proteoform (Ivanov et al. 2011; Hopkins et al. 2013; Liang et al. 2014), a unified nomenclature for PTEN proteoforms was proposed (Pulido et al. 2014) where newly identified PTEN proteins are named alphabetically as PTEN-L, PTEN-M, PTEN-N, etc. We have adopted this proposed nomenclature and henceforth refer to these four PTEN isoforms as PTEN-L, PTEN-M, PTEN-N and PTEN-O, where PTEN-L is the presumed longest variant and is initiated at the previously reported CUG codon (Ivanov et al. 2011; Hopkins et al. 2013; Liang et al. 2014), whereas proteoform PTEN-M appears to be the most abundant (table 2.1 and figure 2.3.1c).

Table 2.1 Details of PTEN N-terminally extended proteoforms. Nucleotide distances from aAUG (annotated AUG of canonical PTEN) are indicated where A of the aAUG is +1.

| Column1	nt from aAUG	N-term ext AA	total nt	total AA	MW (kDa)
PTEN-L	-519	173	1728	576	64.9
PTEN-M	-438	146	1647	549	62.5
PTEN-N	-393	131	1602	534	61.0
PTEN-O	-216	72	1425	475	55.0
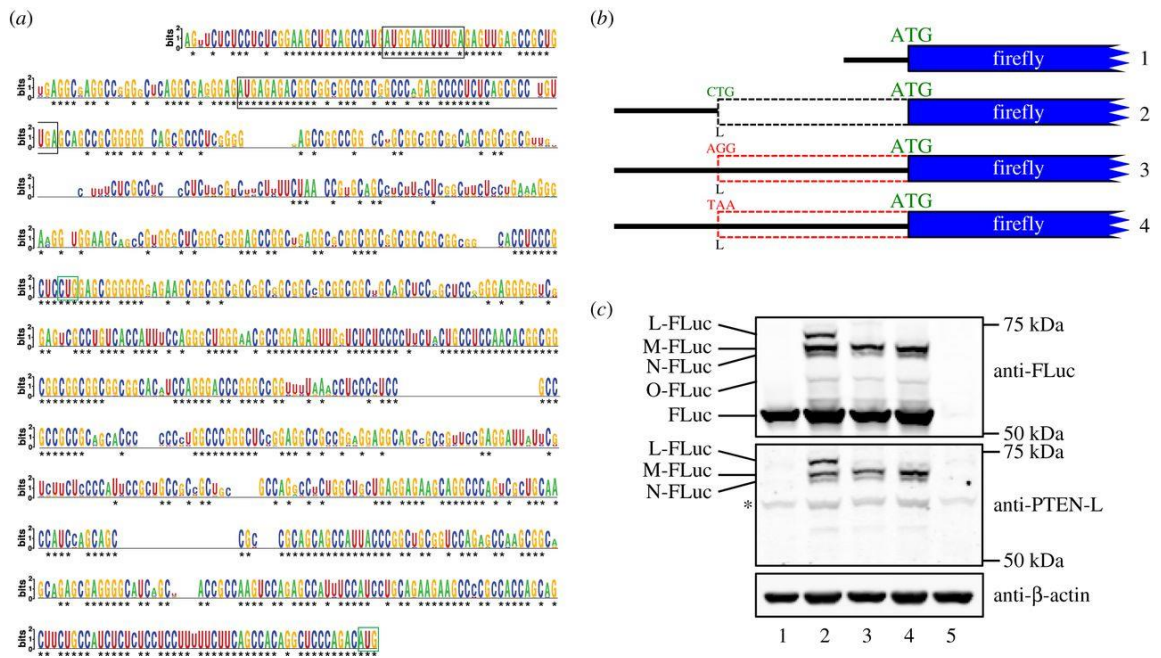
Figure 2.3.1

(a) Sequence logo representation (produced with WebLogo (Crooks et al. 2004)) of a multiple sequence alignment of *PTEN* 5' leaders from 52 mammals. The alignment was generated with C$_{LUSTAL}$X (Heringa 1999) and corrected manually. Asterisks indicate nucleotides conserved in all 52 species. Open black boxes represent two conserved uORFs. Open green boxes indicate the main ORF AUG and previously predicted CUG initiation codons. (b) Illustration of transfected firefly luciferase constructs 1–4 used for immunoblotting. (c) Immunoblot of cell lysates prepared from HEK-293T cells transfected with firefly luciferase expressing constructs as indicated and probed with antibodies against firefly luciferase (anti-FLuc: top panel), the PTEN N-terminal extension (anti-PTEN-L: middle panel) and β-actin (bottom panel). The four proteoforms with extended PTEN N-termini are named as L, M, N and O. This nomenclature was recently proposed for novel PTEN proteoforms by Pulido et al. (2014). Asterisk indicates a non-specific protein that co-migrates with the O-proteoform, thus precluding its detection with anti-PTEN-L. Lane 5 contains cell lysates from mock-transfected cells.

It is conceivable that PTEN-M, PTEN-N and PTEN-O are not non-AUG-initiated N-terminally extended proteoforms. Instead, PTEN-M and PTEN-N could be post-translationally modified variants of PTEN-O, whereas PTEN-N and PTEN-O could be cleavage products of PTEN-M. To address these possibilities, and to determine whether non-AUG initiation could explain the presence of these proteoforms, we made *PTEN* 5' leader-FLuc constructs in which potential non-AUG initiation codons were systematically mutated in turn to either AGG or UAA. Potential in-frame near-cognate initiation codons from the relevant region within the *PTEN* 5' leader are highlighted in figure 2.3.2a and those in a favourable Kozak context (purine at -3 or G at +4) are underlined. Mutation of the most 5' in-frame AUU codon to AGG completely abolishes synthesis of the most prevalent proteoform M-FLuc (lanes 4 and 5, figure 2.3.2b). As expected, mutation of this AUU to a termination codon causes premature termination of L-FLuc (lane 5, figure 2.3.2b). It is difficult to be certain about the nature of the initiation codon for N-FLuc, because a cross-reacting endogenous protein migrates at the same position; however, it seems likely that the second-most 5' in-frame CUG is responsible as there is a clear decrease in N-FLuc intensity when this CUG and all 3' non-AUGs are mutated individually to UAA (see lanes 9, 11, 13 and 15, figure 2.3.2b). Mutation of the most 3' in-frame CUG completely abrogates expression of O-FLuc (lanes 14 and 15, figure 2.3.2b). In summary, the three minor proteoforms (L, N and O) are all initiated at CUG, whereas the major proteoform (M) is initiated at AUU.

We next tested for the existence of endogenous human PTEN N-terminally extended proteoforms by immunoprecipitating PTEN from several different cell lines using commercially available antibodies directed against antigens within the annotated PTEN (CDS). The predicted molecular weight of PTEN is 47.2 kDa, but there are many reports indicating that the apparent molecular weight of PTEN is approximately 55 kDa. In agreement with this, immunoblots of PTEN immunocomplexes reveal the presence of an approximately 55 kDa protein in all cell lines tested (figure 2.3.3a). In addition to the canonical AUG-initiated PTEN, three slower migrating proteins are observed in all
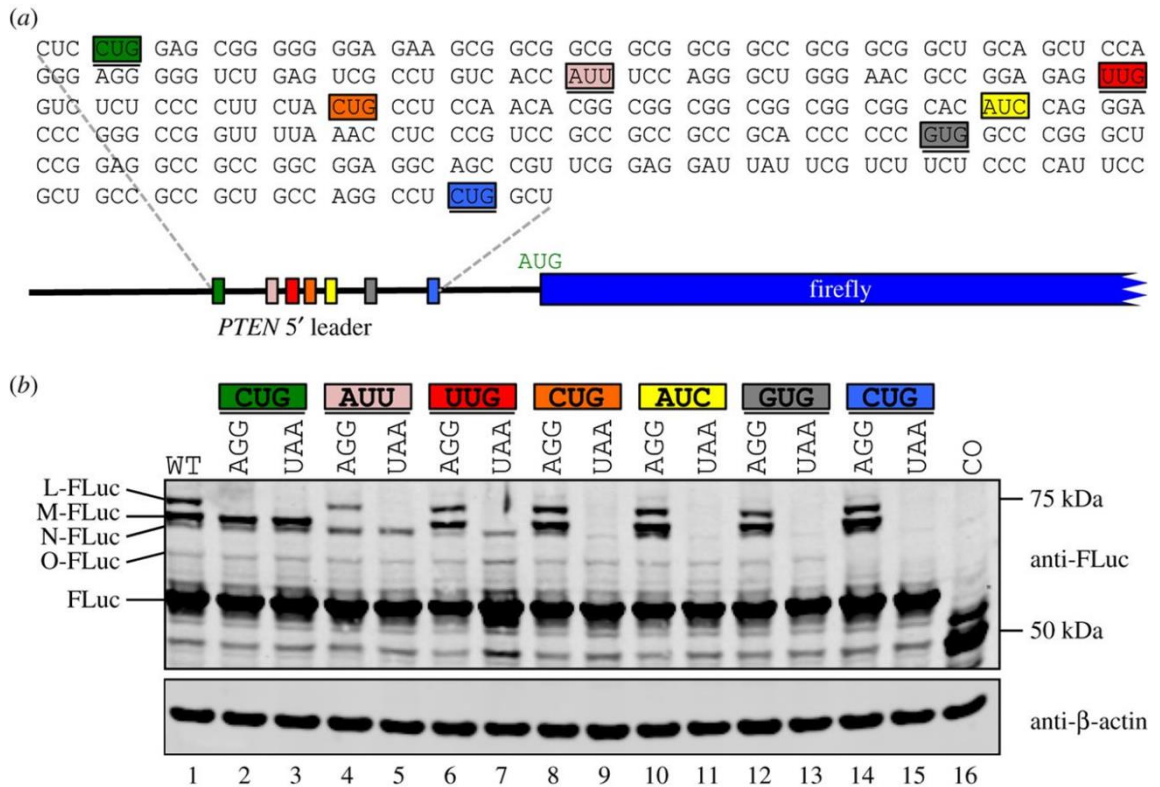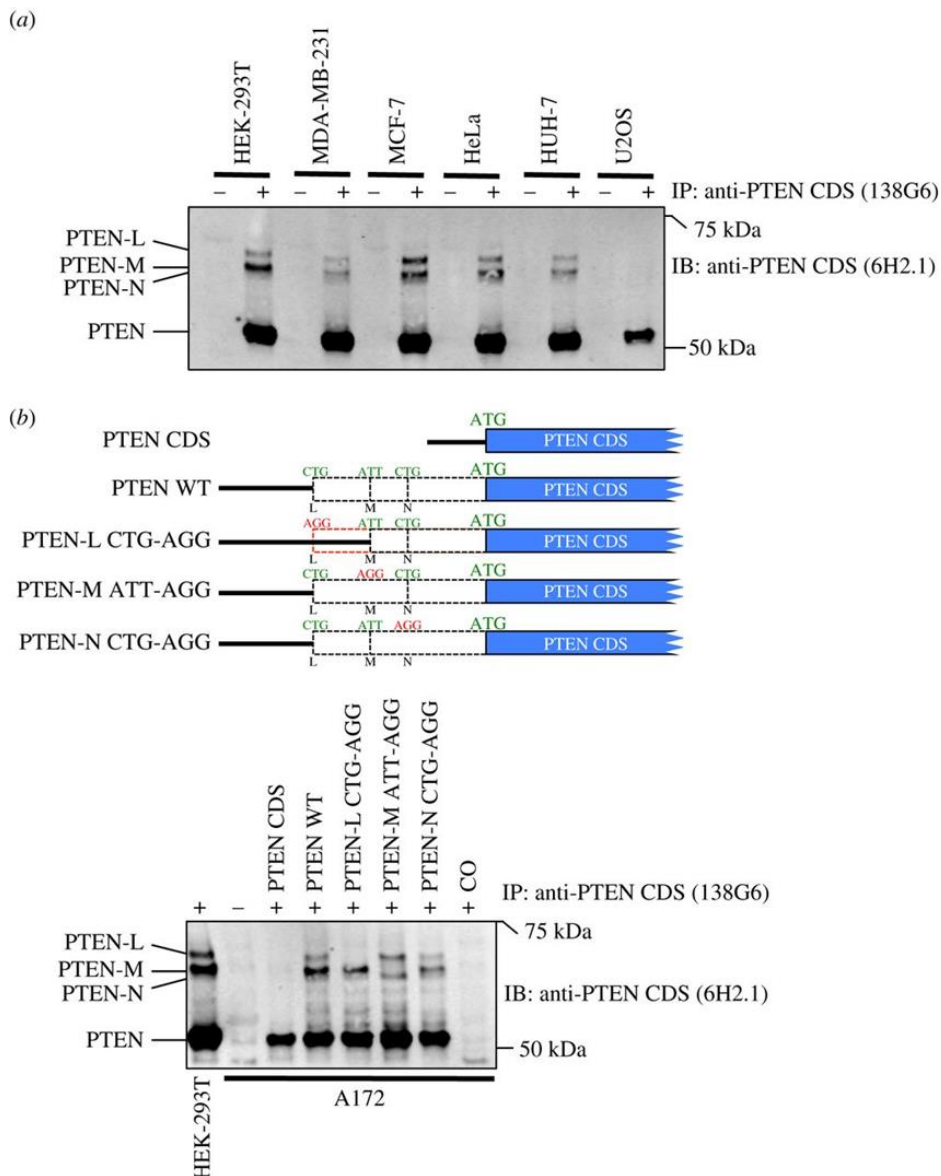
Figure 2.3.2
(a) Illustration of human *PTEN* 5' leader fused to firefly luciferase and sequence of 5' leader region in which potential non-AUG codons are highlighted in different colours with those in favourable Kozak context (-3 purine and/or +4 G) underlined. The previously reported CUG initiation site (Ivanov et al. 2011; Hopkins et al. 2013; Liang et al. 2014) is highlighted in green and is the most 5' potential non-AUG initiation site. The *PTEN* CDS reading frame is indicated with spaces between the codons. (b) Anti-FLuc and anti-β-actin immunoblots of cell lysates prepared from HEK-293T cells transfected with firefly luciferase encoding sequences fused to the wild-type or non-AUG codon mutated (indicated) PTEN 5' leader. The four FLuc proteoforms with extended PTEN N-termini are indicated as L-, M-, N- and O-FLuc. In the control (CO, lane 16), the main ORF (firefly) AUG is mutated to UAA.

cell lines tested other than U2OS. The molecular weights of these three proteins correlate well with those predicted for PTEN proteoforms PTEN-L, PTEN-M and PTEN-N. To determine whether the migration of these putative endogenously expressed PTEN N-terminally extended proteoforms correlate with exogenously expressed proteoforms PTEN-L, PTEN-M and PTEN-N (we could not detect any endogenous proteoform that could correspond in molecular weight to PTEN-O in these experiments), we transfected the human *PTEN* CDS fused to either wild-type or mutated *PTEN* 5' leaders into A172 cells which do not express endogenous PTEN. Immunoprecipitates from transfected A172 cells were compared with

immunoprecipitates from HEK-293T cells expressing endogenous *PTEN* and show that exogenously expressed proteoforms PTEN-L, PTEN-M and PTEN-N co-migrate with endogenous PTEN proteins from HEK-293T cells (compare the first and fourth lanes in figure 2.3.3b). Furthermore, in agreement with the FLuc reporter constructs (figure 2.3.2b), mutation of the first and third in-frame CUGs prevents initiation of the two minor PTEN proteoforms PTEN-L and PTEN-N, whereas mutation of the first in-frame AUU abolishes expression of the major PTEN-M proteoform (figure 2.3.3b). Together, these results indicate that three slower migrating PTEN proteoforms apparent in immunoprecipitates from several cell lines correlate with non-AUG-initiated proteoforms PTEN-L, PTEN-M and PTEN-N.

Several studies report that PTEN-L is an active phosphatase and retains the ability to downregulate the PI3K pathway (Hopkins et al. 2013; Liang et al. 2014; Sean B. Johnston & Raines 2015; Wang et al. 2015; Masson et al. 2016). To determine whether the PTEN proteoforms described here are active phosphatases, we measured the phosphorylation status of the major PI3K substrate, AKT, in PC3 cells (no endogenous PTEN expression) transfected with either wild-type or N-terminally extended PTEN proteoforms. Exogenous expression of PTEN reduced the levels of AKT phosphorylation almost twofold, and similar levels of reduction were observed for all four N-terminally extended proteoforms (figure 2.3.4 and supplementary material, figure S3). Therefore, similar to previous observations for PTEN-L, the phosphatase activities of PTEN-M, PTEN-N and PTEN-O are not overtly affected by their N-terminal extensions.

N-terminal extensions often harbour signals for subcellular targeting or secretion. However, live cell imaging of AUG-initiated PTEN N-terminal extensions fused to GFP reveal diffuse cytoplasmic localization for all four PTEN–GFP fusions indistinguishable from the localization of GFP alone (supplementary material, figure

Figure 2.3.3

(a) Anti-PTEN CDS (6H2.1) immunoblot of anti-PTEN CDS (138G6) immunoprecipitates prepared from several cell lines as indicated showing detection of endogenous canonical AUG-initiated PTEN as well as the three non-AUG-initiated PTEN proteoforms PTEN-L, PTEN-M and PTEN-N. (b) Anti-PTEN CDS (6H2.1) immunoblot of anti-PTEN CDS (138G6) immunoprecipitates prepared from A172 cells (which lack endogenous PTEN) transfected with constructs expressing either wild-type PTEN or mutants of PTEN proteoforms PTEN-L, PTEN-M and PTEN-N (top panel illustration) showing that the non-AUG initiation codons of PTEN-L, PTEN-M and PTEN-N are CUG, AUU and CUG, respectively. The left-most lane shows control immunoprecipitates from HEK-293T cells indicating endogenous canonical AUG-initiated PTEN and N-terminally extended proteoforms PTEN-L, PTEN-M and PTEN-N. CO indicates control immunoprecipitates from mock-transfected cells.

S4). Hopkins *et al*. (2013) suggest that the PTEN-L proteoform harbours an N-terminal signal peptide secretion signal and provide evidence that PTEN-L is secreted and can re-enter cells via a cell re-entry motif similar to the HIV TAT protein. The predicted amino acid sequences of PTEN-M and PTEN-N lack the putative secretion signal yet still retain the putative cell re-entry motif (supplementary material, figure S5) reported by Hopkins *et al*. To rule out the possibility that the N-terminal extension of PTEN-M might possess a cryptic signal peptide sequence, we overexpressed, in HEK-293T cells, PTEN-L (as a control) and PTEN-M N-terminal extensions fused to GFP (same constructs as described in supplementary material, figure S4). A fusion of GFP with the signal peptide sequence from *Gaussia* luciferase (GLuc) (Verhaegent & Christopoulos 2002) was used as a positive control (figure 2.3.5a). Extracellular and intracellular GFP fusion proteins were concentrated by immunoprecipitation using GFP-trap® (immobilized camelid anti-GFP antibody) followed by immunoblotting with anti-GFP. Even though we immunoblotted 50% of the total extracellular fraction and only 5% of the intracellular fraction, we did not detect PTEN-L-GFP or PTEN-M-GFP in the extracellular fraction (see lanes 1 and 3, middle panel, figure 2.3.5b). Furthermore, when HEK- 293T or U2OS cells were transfected with constructs expressing either the PTEN-L or PTEN-M N-terminal extensions, or the putative PTEN-L signal peptide alone, fused to GLuc, we failed to detect GLuc activity in the cell media at levels above background (figure 2.3.5c). Yet GLuc-fused downstream of the PTEN-L N-terminal extension in which the putative PTEN-L signal peptide was replaced with the signal peptide from either GLuc or interleukin-2 efficiently targeted GLuc from cells (figure 2.3.5c).

We next sought to ascertain whether the non-AUG initiation of the PTEN proteoforms is regulated. The selection of poor context initiation codons (including non-AUG start codons) is modulated by intracellular levels of initiation factors eIF1 (increases stringency) and eIF5 (decreases stringency) (Mitchell & Lorsch 2008; Richard J. Jackson et al. 2010; Llácer et al. 2015; Hussain et al. 2014; Nanda et al. 2009; Valásek et al. 2001; Hinnebusch 2014; Ivanov et al. 2010; Loughran et al. 2012). To determine whether initiation of the N-terminally extended PTEN proteoforms are
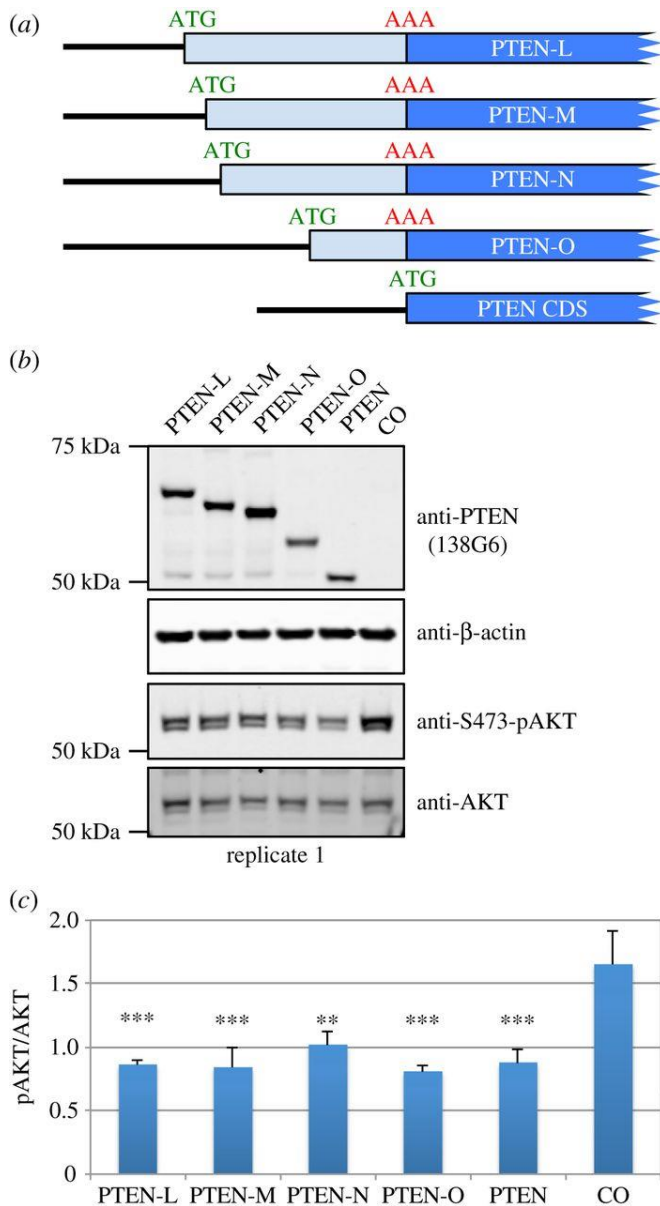
Figure 2.3.4

(a) Illustration shows PTEN expressing constructs transfected into PC3 cells. (b) Immunoblots of cell lysates prepared from PTEN-null PC3 cells transfected with PTEN expressing constructs as indicated for 48 h (serum starved for last 24 h) and probed with antibodies against PTEN (138G6), β-actin, phospho-AKT (S473) and pan-AKT. Additional replicates (replicates 2 and 3) are shown in the supplementary material, figure S3. (c) Mean and standard deviations of relative protein intensities determined by densitometry from three biological replicates. Phospho-AKT intensities were calculated relative to pan-AKT intensities. Relative pAKT levels in lysates from cells transfected with each N-terminally extended PTEN proteoform were compared with the control sample. **$p < 0.01$, ***$p < 0.001$ by one-way ANOVA followed by Tukey's test.
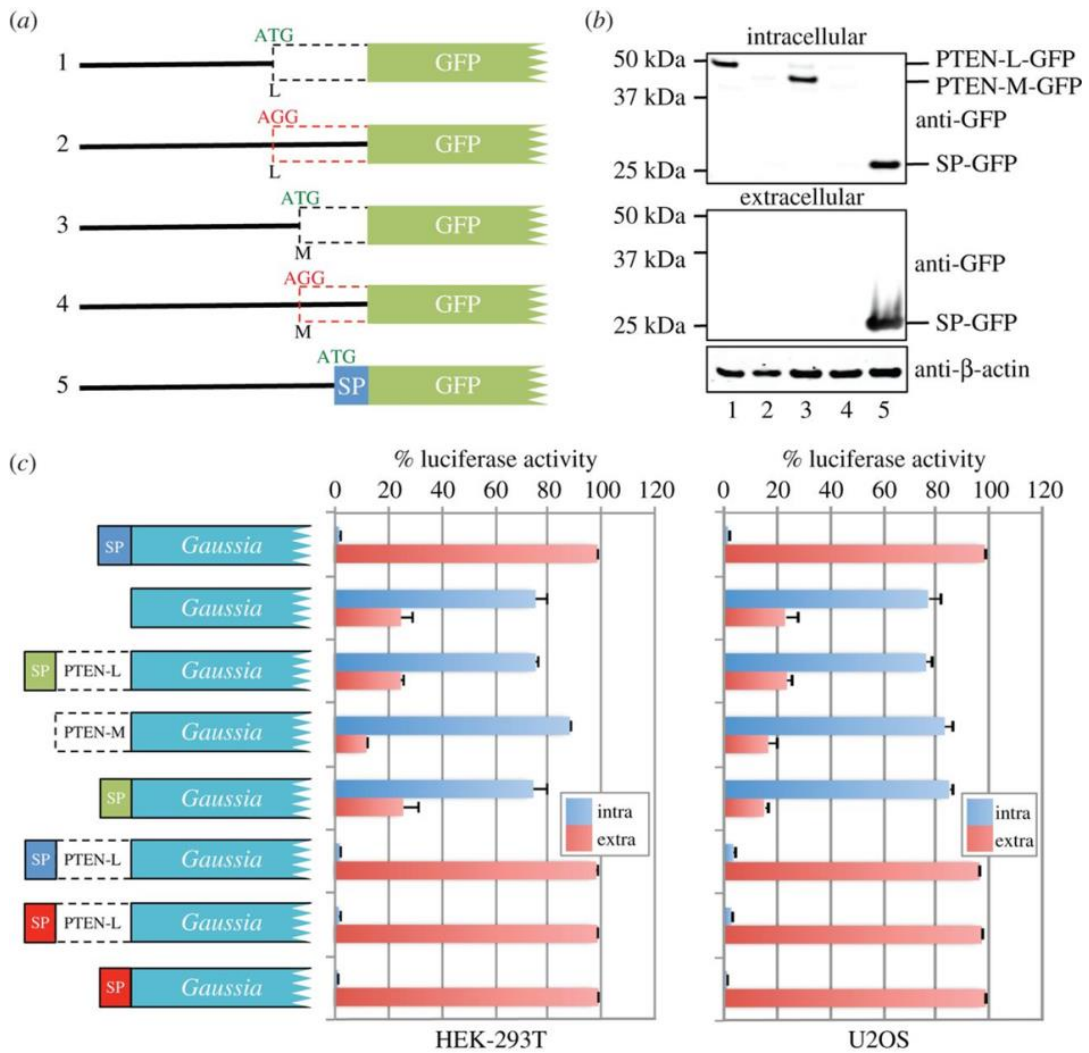
Figure 2.3.5

(a) Illustration of mutant PTEN-L and PTEN-M 5' leader–GFP fusion constructs transfected into HEK-293T cells. SP is the secretion peptide from *Gaussia* luciferase. (b) Anti-GFP and anti-b-actin immunoblots from GFP-trap immunoprecipitates prepared from either HEK-293T cell lysates (intracellular: top panel) or culture media (extracellular: middle panel) transfected with the GFP fusion constructs shown in (a). (c) *Gaussia* luciferase assays show the percentage luciferase activity in the cell lysate (blue: intracellular) and culture media (red: extracellular) when either HEK-293T or U2OS cells (as indicated) were transfected with the constructs indicated for 22 h. The blue signal peptide (SP) is from *Gaussia* luciferase (MGVKVLFALICIAVAEAK), the putative PTEN-L SP (MERGGEAAAAAAAAAAAPGRG) is in green and the interleukin-2 (IL2) SP (MYRMQLLSCIALSLALVTNSA) is in red.

regulated by altered eIF1 or eIF5 levels, we overexpressed each initiation factor separately in HEK-293T cells and then immunoprecipitated endogenous PTEN proteoforms (supplementary material, figure S6). Even though eIF1 and eIF5 levels are robustly expressed, we note no discernible change in the ratio of PTEN proteoforms compared with cells transfected with an empty vector (supplementary material, figure S6).

As shown in figure 2.3.1a, there are two conserved AUG-initiated uORFs (uORF1 and uORF2) close to the *PTEN* 5' cap (figure 2.3.6a and supplementary material, figure S7). The most 5' uORF (uORF1) is only four codons long and starts with tandem AUG codons, both in good Kozak context. uORF2 is much longer (45 codons) and also starts with a good context AUG. Ribosomal profiling data compiled in GWIPS-viz (A M Michel et al. 2014) for *PTEN* show a large number of ribosome protected fragments aligning to uORF1 in comparison with uORF2 (supplementary material, figure S7). In general, translation of uORFs has an inhibitory effect on main ORF translation although this relationship between uORFs and main ORFs is not so simple where multiple uORFs exist. Usually, the translation of short uORFs is less inhibitory than the translation of long uORFs, because the ribosomes' ability to reinitiate after translation of ORFs more than 35 codons is normally greatly reduced (see Introduction). According to the scanning model of eukaryotic translation initiation, we would predict that the majority of scanning 43S complexes would initiate uORF1; however, because it is only four codons long, a high level of re-initiation is expected. Because re-initiation is precluded until all necessary initiation factors have reloaded onto the 40S, translation of uORF1 may favour re-initiation at ORFs more 3' than uORF2. We set out to determine the possible role of these two uORFs on regulation of the different PTEN proteoforms by transfecting HEK-293T cells with *PTEN* 5' -leader-FLuc reporter constructs in which the uORFs were mutated and testing by dual luciferase assay (figure 2.3.6a) and immunoblotting (figure 2.3.6b).

Even though we expect a high level of uORF1 translation, mutation of the uORF1 tandem AUG codons to non-initiating AGG codons has only a minor inhibitory effect (less than 10%) on FLuc activity (compare construct 1 (wild-type PTEN 5' leader) with construct 2 in figure 2.3.6a). Mutating uORF1 is inhibitory

rather than stimulatory presumably because more ribosomes now have access to uORF2 which we predict should be inhibitory. However, the fact that the inhibition is only minor suggests that either uORF2 may not be as inhibitory as we expect or that normally ribosomes translating uORF1 can efficiently reinitiate uORF2. Another explanation could be that in the wild-type context many ribosomes do not initiate uORF1. However, mutation of the uORF1 stop codon to a sense codon (extending the length of uORF1 to 161 codons) severely diminishes FLuc activity (construct 3, figure 2.3.6a) which, in agreement with published ribosome profiling data (supplementary material, figure S7), confirms that ribosomes do initiate at uORF1. A mutation that disables uORF2 translation, either alone or in combination with mutations that prevent uORF1 initiation increased FLuc activity more than twofold (compare construct 1 (wild-type) with constructs 4 and 5 in figure 2.3.6a). This indicates that uORF2 is inhibitory and thus suggests that the level of re-initiation on uORF2 is high. Similar results were observed in several other cell lines tested, including breast carcinoma (MCF-7), prostate carcinoma (PC3) and cervical carcinoma (HeLa; supplementary material, figure S8). In addition, there is little difference in steady-state mRNA levels when both uORFs are disabled (supplementary material, figure S9). In these reporter assays, it is assumed that the low level of N-terminally extended proteoforms (relative to the main ORF) have only a minor contribution to total FLuc activity. This is supported by our own unpublished observations showing that an AUG-initiated PTEN-L extension severely reduces FLuc activity. To gain further understanding of the effect of the uORFs on downstream translation, we performed antiFLuc immunoblots from cells transfected with constructs 1–6 (figure 2.3.6b and supplementary material, figure S10). Dual luciferase assays from replicate lysates show similar FLuc activity for each construct to those shown in figure 2.3.6a and supplementary material, figure S8. Densitometry of the main ORF (FLuc) normalized to cotransfected RLuc levels indicates that, similar to the luciferase assay results, preventing initiation of the uORFs results in an almost twofold
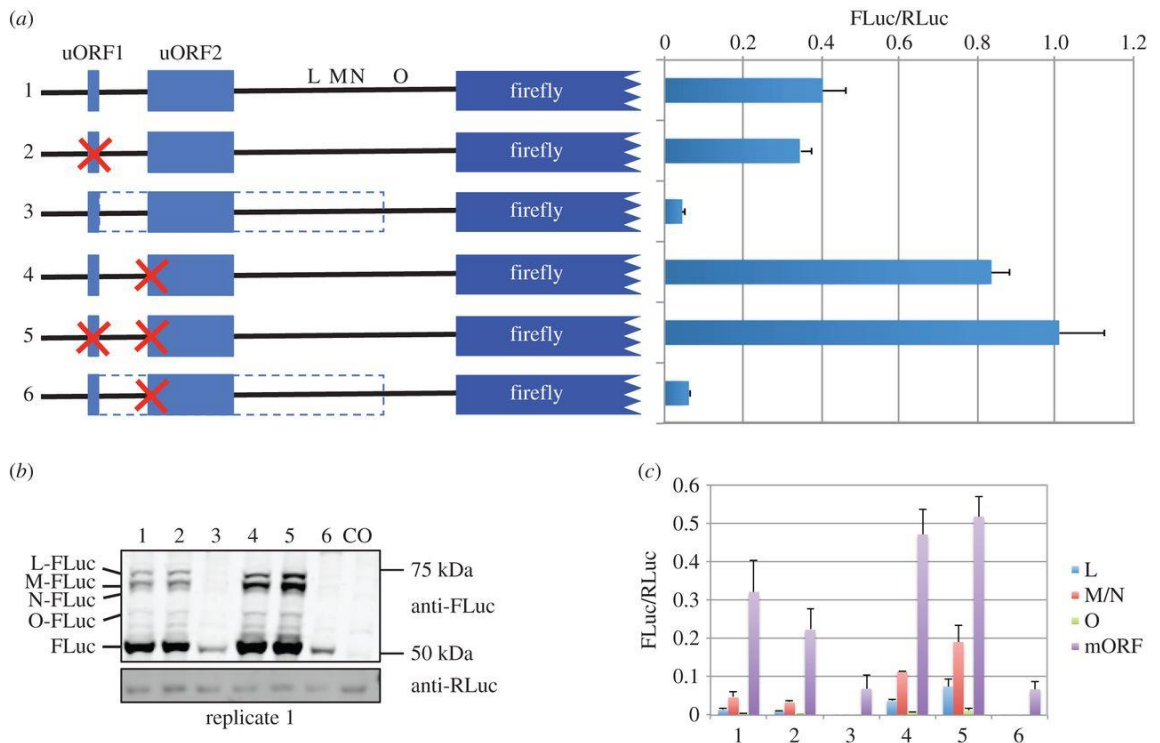
Figure 2.3.6

(a) Relative luciferase activities (FLuc/RLuc) of firefly encoding sequences fused to the wild-type or mutant *PTEN* 5' leader as indicated and cotransfected (10 : 1 ratio) with a *Renilla* expressing plasmid into HEK-293T cells. Red crosses indicate mutation of AUG start codons to non-initiating AGG codons. L, M, N and O depict the approximate site of initiation of PTEN extensions. The dashed box represents the increase in ORF length when the stop codon of uORF1 is changed to a sense codon. (b) Anti-FLuc and anti-RLuc immunoblots of cell lysates prepared from HEK-293T cells transfected with *Renilla* and firefly luciferase expressing constructs indicated in (a). The four proteoforms with extended PTEN N-termini are indicated as L, M, N and O. CO represents lysates prepared from mock-transfected cells. (c) Densitometry analysis from three biological replicates of the proteins detected by anti-FLuc and anti-RLuc in (b) and supplementary material, figure S10b. Proteoforms M and N could not be resolved sufficiently from each other for accurate densitometry analysis, so the intensity of both proteins together is determined. FLuc intensities were calculated relative to RLuc.

increase in main ORF levels (figure 2.3.6c). Therefore, even though levels of the N-terminally extended proteoforms increase when uORF2 is mutated (lanes 4 and 5, figure 2.3.6b and supplementary material, figure S10), they have only a minor contribution to overall FLuc activity. When we estimate the levels of each N-terminally extended proteoform by densitometry, we observe that mutation of uORF2 causes the levels of the L and M/N proteoforms to increase approximately

2.5-fold, the O proteoform to increase 2-fold, whereas the main ORF increases approximately 1.5-fold. Mutation of both uORFs together increases the L-proteoform approximately 5.3-fold, M/N-proteoforms approximately 4.1-fold, O-proteoform approximately 3.1-fold and the main ORF proteoform approximately 1.6-fold. This densitometry analysis also allows us to estimate the relative abundance of each proteoform under normal conditions and when ribosomes do not translate either or both uORFs. Interestingly, although all proteoforms increase when ribosomes do not translate uORF2, the ratio of N-terminally extended proteoforms relative to each other does not change (supplementary material, figure S10c). In contrast, mutation of uORF2 increases the ratio of N-terminal proteoforms relative to the main ORF proteoform (supplementary material, figure S10d) such that proteoforms M/N increase by 50% from 12% to 18% of all proteoforms. Similar approximately 50% increases were also observed for the L- and O-proteoforms (supplementary material, figure S10d). Furthermore, there is a concomitant decrease in the relative abundance of the main ORF proteoform from 83% down to 75%. Mutation of uORF1 and uORF2 together results in even further increases to the relative abundance of N-terminally extended proteoforms and further decreases the main ORF proteoform (supplementary material, figure S10d).

Because we show that translation of uORF2 can alter the ratio of extended and main ORF proteoforms and uORF2 translation seems to be dependent on efficient re-initiation after translation of uORF1, we predict that conditions which can regulate re-initiation events could alter translation of the main ORF. Increasing the phosphorylation status of eIF2 can reduce re-initiation by limiting the pool of functional (non-phosphorylated) eIF2. To gain some understanding of possible regulation of these uORFs, we transfected wild-type and 'no uAUG' firefly reporters into HEK-293T cells and then treated with sodium arsenite. Sodium arsenite activates the integrated stress response by inducing phosphorylation of eIF2. Many mRNAs (main ORFs) whose translation is resistant to eIF2 phosphorylation harbour translated uORFs (Andreev et al. 2015). If sodium arsenite decreased re-initiation on uORF2, then we would expect to see an increase in FLuc activity with the wild-type reporter; however, we observed little difference in main ORF reporter translation upon sodium arsenite treatment (supplementary material, figure S11).

2.4   Discussion

The results presented above provide strong evidence for the existence of three (or perhaps four) non-AUG-initiated proteoforms of PTEN that are expressed in human cells in addition to, and at lower levels than, the well-studied canonical AUG-initiated PTEN. The longest PTEN proteoform, PTEN-L, has previously been reported (Ivanov et al. 2011; Hopkins et al. 2013; Liang et al. 2014) and its function has been investigated in more detail in subsequent studies (Sean B. Johnston & Raines 2015; Wang et al. 2015; Malaney et al. 2013; Malaney et al. 2015; Sean B Johnston & Raines 2015). However, PTEN-M, PTEN-N and PTEN-O have not been previously described and are reported here for the first time. We show that PTEN-M initiates at an AUU codon that is completely conserved in 52 eutherian mammalian species with available sequences, whereas the other PTEN proteoforms (PTEN-L, PTEN-N and PTEN-O) initiate at CUG, the first two also completely conserved while the latter only partially conserved. CUG codons are generally better initiators than AUU codons (Peabody 1989; Ivanov et al. 2010; Loughran et al. 2012), so it is somewhat surprising that AUU-initiated PTEN-M is more abundant than CUG-initiated PTEN-L, especially, because the CUG is more 5' than the AUU. The PTEN-L CUG initiation codon Kozak context is slightly less favourable (C at -3 and G at +4) than that of the PTEN-M AUU initiation codon (A at -3). It is also possible that an as yet unidentified RNA secondary structure 3' of the AUU codon contributes to its favourable utilization for initiation (see Introduction), especially because the *PTEN* 5' leader is 70% GC rich with numerous potential stem-loops. Yet another possibility is that the abundance of the PTEN proteoforms is a reflection of differing protein stabilities rather than initiation levels, although a similar level of exogenous PTEN-L and PTEN-M expression when their non-AUGs start codons are mutated to AUGs (figure 2.3.4b and supplementary material, figure S3) would argue against this. The faint band observed in the first three lanes of figure 2.3.4b that migrates at the same position as PTEN, could be the PTEN protein whose translation has initiated at AAA. As already mentioned, all codons that differ by one nucleotide from AUG, with the exception of AAG and AGG can serve as translation start sites (Peabody 1989). This fact combined with the potential secondary and tertiary structures GC rich RNA can form, could make the position of the canonical PTEN start site, favourable for

translation initiation. It is also likely that the above mentioned faint bands are cleaved products of the N-terminally extended PTEN proteoforms (PTEN-L, -M, -N).

The identity of the major non-AUG-initiated PTEN proteoforms identified in this study contrasts with the findings of both Hopkins *et al*. (2013) and Liang *et al*. (2014), who both report only a single CUG-initiated proteoform that corresponds in our study to PTEN-L. Most of the anti-PTEN immunoblots presented by Hopkins *et al*. show a single slower migrating approximately 75 kDa PTEN proteoform which they term PTEN-Long (PTEN-L in our study). A possible reason for this discrepancy is that, in their study, proteins were separated for shorter time intervals while we purposely allowed SDS–PAGE gels to run for extended periods (see Material and methods) in an attempt to resolve as many PTEN proteoforms as technically possible. Perhaps the approximately 75 kDa PTEN-long detected by Hopkins *et al*. corresponds to a mixture of unresolved PTEN-L, PTEN-M and PTEN-N.

Liang *et al*. similarly identify a PTEN proteoform (PTEN-α) that initiates with the most 5' in-frame CUG and corresponds to our PTEN-L and Hopkins et al. PTEN-Long. Their anti-PTEN immunoblots from cells exogenously expressing a PTEN-α construct clearly show two higher molecular weight proteins in addition to PTEN. In agreement with our study, mutation of the first in-frame CUG prevents translation of the longest protein. Furthermore, even when the first and third CUG codons (PTEN-L and PTEN-N in our study) are mutated together, an extended PTEN proteoform is still apparent which is very likely to correspond to PTEN-M. A recent report showing anti-PTEN immunoblots from matched normal and tumour tissue samples clearly identifies, in addition to canonical PTEN, at least two slower migrating endogenous PTEN proteoforms expressed in a similar ratio to PTEN-L and PTEN-M in our study (Wang et al. 2015).

When multiple proteoforms are translated from a single mRNA, the efficiency of initiation at each start codon could set the ratio of proteoform steady-state levels assuming each protein has similar stability. However, this ratio may vary during conditions in which initiation efficiency is altered. Eukaryotes have developed elaborate mechanisms for the recognition of the correct initiation codon and the levels of certain initiation factors can regulate the fidelity of initiation, especially on suboptimal (non-AUG and AUG in poor context) start codons (Ivanov

et al. 2010; Loughran et al. 2012; Martin-Marcos et al. 2011). While elevated levels of eIF1 can increase the stringency of start codon selection, elevated levels of eIF5 have the opposite effect. Here we show that overexpression of either eIF1 or eIF5 had minimal effect on the steady-state levels of PTEN proteoforms (supplementary material, figure S6). This suggests that either the non-AUG initiation events in *PTEN* are refractory to normal stringency controls or the steady-state protein levels of these proteoforms are regulated tightly by rapid turnover. An alternative explanation is that because the *PTEN* 5' leader is long, the many potential out-of-frame near-cognate codons could create uORFs and thus preclude the expected effects of eIF1/5 overexpression on translation of the PTEN N-terminal extensions.

Several groups have observed that PTEN-L can downregulate the PI3K pathway in a similar manner to PTEN (Hopkins et al. 2013; Liang et al. 2014; Wang et al. 2015). *In vitro* studies comparing the catalytic activities of purified PTEN and PTEN-L reveal that both enzymes can dephosphorylate PIP3, although, interestingly, PTEN phosphatase activity can be activated by its reaction product (PIP2), whereas PTEN-L cannot and is thus constitutively active (Sean B. Johnston & Raines 2015; Sean B Johnston & Raines 2015). We tested whether the PTEN proteoforms identified in our study still retained the ability to downregulate the PI3K pathway (figure 2.3.4 and supplementary material, figure S3). All PTEN proteoforms were able to reduce AKT phosphorylation to levels similar to those of canonical PTEN and PTEN-L, suggesting that the N-terminal extensions do not have major effects on the dephosphorylation activity of PTEN proteoforms.

Hopkins *et al*. (2013) report that exogenously expressed PTEN-L is targeted for secretion from cells by a predicted N-terminal signal peptide and cleavage site. Furthermore, they also show that immediately following the predicted cleavage site is a functional cell re-entry signal similar to the HIV TAT protein (Hopkins et al. 2013). Subsequently, Wang *et al.* (Wang et al. 2015) confirmed that PTEN-L, but not PTEN, can enter cells, although whether PTEN-L can be secreted from cells was not tested. Intriguingly, both PTEN-M and PTEN-N, while lacking the predicted signal peptide, do still retain the putative cell re-entry signal (supplementary material, figure S5). We overexpressed GFP fused to both the PTEN-L (with putative secretion signal) and PTEN-M N-terminal extensions but could not detect PTEN-L-GFP in the

cell culture media after concentrating by immunoprecipitation (figure 2.3.5b). It is not so surprising that PTEN-M-GFP is not found extracellularly because it lacks the putative signal peptide, but failure to detect PTEN-L was unexpected and suggests that most PTEN-L is not secreted. It is possible that this assay was not sensitive enough to detect low levels of secreted PTEN-L-GFP, so we further tested whether a secretion signal resided in the PTEN-L extension by fusing the *PTEN* 5' leader to *Gaussia* luciferase. *Gaussia* luciferase is approximately 1000 times more sensitive than either *Renilla* or firefly luciferases (Tannous et al. 2005), yet we could not detect any extracellular luciferase activity, above background, in constructs harbouring the putative PTEN-L signal peptide (figure 2.3.5c). One explanation for the discrepancy between our PTEN localization experiments and those of previous studies is that, because canonical PTEN has been found in exosomes (Putz et al. 2012; Gabriel et al. 2013) and can homodimerize (Papa et al. 2014), we decided to make reporters that do not contain sequences encoding the canonical *PTEN* CDS. Perhaps the important targeting signals are only 'visible' in the context of the full-length PTEN proteins. Yet another possible explanation for not detecting our PTEN-L chimeras in the cell culture media could be that the efficiency of cell reentry is extremely high. However, detection of extracellular luciferase activity when the PTEN-L signal peptide is replaced with either the *Gaussia* luciferase or interleukin-2 signal peptide would argue against this possibility (figure 2.3.5c).

The functional significance of the PTEN-M N-terminal extension has yet to be determined but perhaps some insight may be gained from previous studies on PTEN-L. There are conflicting reports as to whether the PTEN-L extension has the potential to form a three-dimensional structure (Sean B. Johnston & Raines 2015; Sean B Johnston & Raines 2015) or whether it is intrinsically disordered (Malaney et al. 2013; Malaney et al. 2015). A recent elegant HDX-MS approach by Masson *et al.* (Masson et al. 2016) indicates that while most of the PTEN-L N-terminal extension is indeed intrinsically disordered, there is a potential a-helix at position 151–174 (where residue 174 is the methionine encoded by the canonical *PTEN* AUG). This peptide is protected by liposomes, suggesting an interaction with the membrane. Furthermore, this potential membrane-spanning region alters both the interfacial kinetics of PTEN-L and the protein/membrane interface, causing PTEN-L to bind

more tightly and dissociate less readily from membranes than PTEN. The mode of PTEN-L interaction with membranes is called the 'scooting' mode as opposed to the 'hopping' mode, which is characteristic of PTEN, and involves prompt association and release from the membrane (Masson et al. 2016). All N-terminally extended proteoforms described in our study possess this potential α-helix so it will be interesting to see whether these new PTEN proteoforms act in a similar manner. It is perhaps noteworthy that we could not detect PTEN-O by immunoprecipitation with PTEN antibodies, presumably because this N-terminal extension (and not L, M and N) alters protein conformation in a manner that prevents antibody access to the PTEN antigen under native conditions. This suggests possible structural differences between the PTEN-O N-terminal extension and the other PTEN proteoforms.

In our analysis of the conserved PTEN uORFs, we initially hypothesized that translation of uORF1 could reduce translation of uORF2, which, because of its size (45 codons), we expect to be severely inhibitory for downstream translation. In this way, translation of uORF1 could, in theory, have an overall positive effect on main ORF translation by reducing the number of ribosomes accessing the predicted inhibitory uORF2. However, intriguingly, increasing the number of ribosomes accessing uORF2 by mutation of uORF1 appears not to be very inhibitory under the conditions tested (figure 2.3.6). Therefore, we conclude that either uORF1 is frequently passed by leaky scanning, which seems unlikely given the evidence we described previously for uORF1 translation, or ribosomes translating uORF1 can re-initiate efficiently at uORF2. Alternatively, similar results would be observed if uORF2 were not very inhibitory (i.e. permits high level re-initiation). However, when ribosomes do not translate uORF2, downstream translation increases approximately 2.5-fold at the CUG of the L-proteoform in comparison with when uORF2 is available for translation (figure 2.3.6b,c). As one would predict from the scanning model of translation initiation, removal of both uORFs further increases the level of downstream initiation a further twofold (for L-FLuc). Importantly, the presence of uORF2 affects the ratio of N-terminally extended proteoforms relative to the main ORF, but has no effect on the ratio of N-terminally extended proteoforms to each other (supplementary material, figure S10c,d).

These reporter assay results raise the intriguing possibility that the deeply conserved uORFs in the *PTEN* 5' leader may become less inhibitory for PTEN translation under, as yet unidentified, conditions that could either decrease elongation rates, which in theory would result in ribosome accumulation along uORFs and hamper scanning, or else downregulate reinitiation. There is evidence that the canonical mTOR –S6K pathway regulates re-initiation after uORFs in plants (Schepetilnikov et al. 2013). We predict that such regulation could have dramatic effects on the abundance of N-terminally extended proteoforms, especially if initiation of both uORFs were reduced.

These findings, together with the findings on PTEN-L from other groups, could have profound implications for the interpretation of previous studies on both the catalytic activity and localization of endogenous PTEN as well as the analysis of polymorphisms within the *PTEN* 5' leader. Furthermore, the discovery of these new PTEN proteoforms could have implications for the development of PTEN-based chemotherapeutic agents.

Chapter 2 Bibliography

Andreev, D.E. et al., 2015. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression N. Sonenberg, ed. *eLife*, 4, p.e03971. Available at: https://dx.doi.org/10.7554/eLife.03971.

Arnaud, E. et al., 1999. A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. *Molecular and cellular biology*, 19(1), pp.505–514.

Baranov, P. V, Gesteland, R.F. & Atkins, J.F., 2004. P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA (New York, N.Y.)*, 10(2), pp.221–230.

Calvo, S.E., Pagliarini, D.J. & Mootha, V.K., 2009. Upstream open reading frames cause widespread reduction of protein expression and are. , 106(18), pp.7507–7512.

Chen, S.-J. et al., 2008. Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. *The Journal of biological chemistry*, 283(6), pp.3173–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18065417 [Accessed May 24, 2016].

Cigan, A.M., Feng, L. & Donahue, T.F., 1988. tRNAi(met) functions in directing the scanning ribosome to the start site of translation. *Science (New York, N.Y.)*, 242(4875), pp.93–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3051379 [Accessed May 24, 2016].

Coldwell, M.J. et al., 2004. Expression of fragments of translation initiation factor eIF4GI reveals a nuclear localisation signal within the N-terminal apoptotic cleavage fragment N-FAG. *Journal of cell science*, 117(Pt 12), pp.2545–2555.

Crooks, G.E. et al., 2004. WebLogo: a sequence logo generator. *Genome research*, 14(6), pp.1188–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15173120 [Accessed May 24, 2016].

Damme, P. et al., 2014. N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol Cell Proteomics*, 5.

Dmitriev, R.I. et al., 2013. Imaging of neurosphere oxygenation with phosphorescent probes. *Biomaterials*, 34(37), pp.9307–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24016849 [Accessed May 24, 2016].

Fritsch, C. et al., 2012. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res*, 22. Available at: http://dx.doi.org/10.1101/gr.139568.112.

Gabriel, K. et al., 2013. Regulation of the Tumor Suppressor PTEN through Exosomes: A Diagnostic Potential for Prostate Cancer. *PLoS ONE*, 8(7).

Grentzmann, G. et al., 1998. A d u a l-lu c ife r a s e r e p o r te r s y s te m fo r s tu d y in g re c o d in g s ig n a ls. , pp.479–486.

Hann, S.R. et al., 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell*, 52(2), pp.185–195.

Heringa, J., 1999. Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. *Computers & chemistry*, 23(3–4), pp.341–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10404624 [Accessed May 24, 2016].

Herr, A.J. et al., 2004. Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *The Journal of biological chemistry*, 279(12), pp.11081–11087.

Hinnebusch, A.G., 2014. The scanning mechanism of eukaryotic translation initiation. *Annual review of biochemistry*, 83, pp.779–812. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24499181 [Accessed May 24, 2016].

Hinnebusch, A.G., 2005. Translational Regulation of GCN4 and the General Amino Acid Control of Yeast *.

Hopkins, B.D. et al., 2013. A Secreted PTEN Phosphatase That Enters Cells to Alter Signaling and Survival. *Science*, 341(6144), pp.399–402. Available at: http://www.sciencemag.org/content/341/6144/399.abstract.

Hussain, T. et al., 2014. Structural Changes Enable Start Codon Recognition by the Eukaryotic Translation Initiation Complex. *Cell*, 159(3), pp.597–607.

Ingolia, N.T. et al., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. , 324(April), pp.218–223.

Ingolia, N.T., Lareau, L.F. & Weissman, J.S., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), pp.789–802.

Ivanov, I.P. et al., 2011. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Research*, 39(10), pp.4220–4234.

Ivanov, I.P. et al., 2010. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). , 1.

Jackson, R.J., Hellen, C.U.T. & Pestova, T. V, 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*, 11(2), pp.113–127. Available at: http://dx.doi.org/10.1038/nrm2838.

Jackson, R.J., Hellen, C.U.T. & Pestova, T. V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2), pp.113–127. Available at: http://www.nature.com/doifinder/10.1038/nrm2838 [Accessed May 24, 2016].

Johnston, S.B. & Raines, R.T., 2015. Catalysis by the Tumor-Suppressor Enzymes PTEN and PTEN-L. *Plos One*, 10(1), p.e0116898. Available at: http://dx.plos.org/10.1371/journal.pone.0116898.

Johnston, S.B. & Raines, R.T., 2015. Conformational stability and catalytic activity of PTEN variants linked to cancers and autism spectrum disorders. *Biochemistry*, 54(7), pp.1576–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25647146 [Accessed May 24, 2016].

Kawaji, H. et al., 2011. Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Nucleic acids research*, 39(Database issue), pp.D856-60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21075797 [Accessed May 24, 2016].

Kozak, M., 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* , 15(20), pp.8125–8148. Available at: http://nar.oxfordjournals.org/content/15/20/8125.abstract.

Kozak, M., 1989. Context Effects and Inefficient Initiation at Non-AUG Codons in Eucaryotic Cell-Free Translation Systems. , 9(11), pp.5073–5080.

Kozak, M., 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 87(21), pp.8301–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2236042 [Accessed May 24, 2016].

Kozak, M., 1980. Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. *Cell*, 22(1 Pt 1), pp.7–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7000367 [Accessed May 24, 2016].

Kozak, M., 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2), pp.187–208.

Lee, S. et al., 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *ProcNatlAcad Sci U S A*, 109. Available at:

http://dx.doi.org/10.1073/pnas.1207846109.

Li, J. et al., 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science (New York, N.Y.)*, 275(5308), pp.1943–1947.

Liang, H. et al., 2014. PTENα, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. *Cell Metabolism*, 19(5), pp.836–848. Available at: http://dx.doi.org/10.1016/j.cmet.2014.03.023.

Llácer, J.L. et al., 2015. Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Molecular cell*, 59(3), pp.399–412. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26212456 [Accessed May 24, 2016].

Loughran, G. et al., 2012. Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. , 40(7), pp.2898–2906.

Maehama, T. & Dixon, J.E., 1998. The Tumor Suppressor, PTEN/MMAC1, Dephosphorylates the Lipid Second Messenger, Phosphatidylinositol 3,4,5-Trisphosphate. *Journal of Biological Chemistry*, 273(22), pp.13375–13378.

Malaney, P., Uversky, V.N. & Davé, V., 2015. Identification of intrinsically disordered regions in PTEN and delineation of its function via a network approach. *Methods (San Diego, Calif.)*, 77–78, pp.69–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25449897 [Accessed May 24, 2016].

Malaney, P., Uversky, V.N. & Davé, V., 2013. The PTEN Long N-tail is intrinsically disordered: increased viability for PTEN therapy. *Molecular BioSystems*, 9(11), p.2877. Available at: http://xlink.rsc.org/?DOI=c3mb70267g [Accessed May 24, 2016].

Martin-Marcos, P., Cheung, Y.-N. & Hinnebusch, A.G., 2011. Functional elements in initiation factors 1, 1A, and 2β discriminate against poor AUG context and non-AUG start codons. *Molecular and cellular biology*, 31(23), pp.4814–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21930786 [Accessed May 24, 2016].

Masson, G.R. et al., 2016. The intrinsically disordered tails of PTEN and PTEN-L have distinct roles in regulating substrate specificity and membrane activity. *The Biochemical journal*, 473(2), pp.135–144.

Michel, A.M. et al., 2014. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*, 42. Available at: http://dx.doi.org/10.1093/nar/gkt1035.

Michel, A.M., Andreev, D.E. & Baranov, P. V, 2014. Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics*, 15(1), pp.1–10. Available at:

http://dx.doi.org/10.1186/s12859-014-0380-4.

Mitchell, S.F. & Lorsch, J.R., 2008. Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. *The Journal of biological chemistry*, 283(41), pp.27345–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18593708 [Accessed May 24, 2016].

Nanda, J.S. et al., 2009. eIF1 Controls Multiple Steps in Start Codon Recognition during Eukaryotic Translation Initiation. *Journal of Molecular Biology*, 394(2), pp.268–285.

Noderer, W.L. et al., 2014. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Molecular systems biology*, 10(8), p.748. Available at: http://msb.embopress.org/cgi/doi/10.15252/msb.20145136 [Accessed May 24, 2016].

Ogle, J.M. et al., 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science (New York, N.Y.)*, 292(5518), pp.897–902.

Packham, G., Brimmell, M. & Cleveland, J.L., 1997. Mammalian cells express two differently localized Bag-1 isoforms generated by alternative translation initiation. *The Biochemical journal*, 328 (Pt 3, pp.807–813.

Papa, A. et al., 2014. Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell*, 157(3), pp.595–610. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24766807 [Accessed May 24, 2016].

Peabody, D.S., 1989. Translation initiation at non-AUG triplets in mammalian cells. *Journal of Biological Chemistry*, 264(9), pp.5031–5035.

Portis, J.L., Spangrude, G.J. & Mcatee, F.J., 1994. Identification of a Sequence in the Unique 5 ' Open Reading Frame of the Gene Encoding Glycosylated Gag Which Influences the Incubation Period of Neurodegenerative Disease Induced by a Murine Retrovirus. , 68(6), pp.3879–3887.

Potapov, A.P., Triana-Alonso, F.J. & Nierhaus, K.H., 1995. Ribosomal decoding processes at codons in the A or P sites depend differently on 2'-OH groups. *The Journal of biological chemistry*, 270(30), pp.17680–17684.

Pulido, R. et al., 2014. A unified nomenclature and amino acid numbering for human PTEN. *Science signaling*, 7(332), p.pe15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24985344 [Accessed May 24, 2016].

Putz, U. et al., 2012. The Tumor Suppressor PTEN Is Exported in Exosomes and Has Phosphatase Activity in Recipient Cells. *Science Signaling*, 5(243), p.ra70-ra70.

Rajkowitsch, L. et al., 2004. Reinitiation and Recycling are Distinct Processes Occurring

Downstream of Translation Termination in Yeast. , pp.71–85.

Ramakrishnan, V., 2002. Ribosome structure and the mechanism of translation. *Cell*, 108(4), pp.557–572.

Schepetilnikov, M. et al., 2013. TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *The EMBO journal*, 32(8), pp.1087–102. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23524850 [Accessed May 24, 2016].

Shabalina, S.A. et al., 2004. Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Research*, 32(5), pp.1774–1782. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC390323/.

Simonetti, A. et al., 2008. Structure of the 30S translation initiation complex. *Nature*, 455(7211), pp.416–420.

Sonenberg, N. & Hinnebusch, A.G., 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4), pp.731–45. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3610329&tool=pmcentrez&rendertype=abstract [Accessed July 16, 2014].

Stambolic, V. et al., 2000. High Incidence of Breast and Endometrial Neoplasia Resembling Human Cowden Syndrome in pten ▨/▨ Mice. *CANCER RESEARCH*, 60, pp.3605–3611.

Starck, S.R. et al., 2012. Leucine-tRNA Initiates at CUG Start Codons for Protein Synthesis and Presentation by MHC Class I. *Science*, 336(6089), pp.1719–1723.

Steck, P.A. et al., 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature genetics*, 15(4), pp.356–362.

Svidritskiy, E. & Korostelev, A.A., 2015. Ribosome Structure Reveals Preservation of Active Sites in the Presence of a P-Site Wobble Mismatch. *Structure (London, England : 1993)*, 23(11), pp.2155–2161.

Tannous, B.A. et al., 2005. Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo. *Molecular therapy : the journal of the American Society of Gene Therapy*, 11(3), pp.435–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15727940 [Accessed May 24, 2016].

Tee, M.K. & Jaffe, R.B., 2001. A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *The Biochemical journal*, 359(Pt 1), pp.219–226.

Tikole, S. & Sankararamakrishnan, R., 2006. A survey of mRNA sequences with a non-AUG

start codon in RefSeq database. *Journal of biomolecular structure & dynamics*, 24(1), pp.33–42.

Valásek, L. et al., 2001. Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *The EMBO journal*, 20(4), pp.891–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11179233 [Accessed May 24, 2016].

Varga, E.A. et al., 2009. The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. *Genetics in medicine : official journal of the American College of Medical Genetics*, 11(2), pp.111–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19265751 [Accessed May 24, 2016].

Verhaegent, M. & Christopoulos, T.K., 2002. Recombinant Gaussia luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. *Analytical chemistry*, 74(17), pp.4378–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12236345 [Accessed May 24, 2016].

Wang, H. et al., 2015. Relevance and Therapeutic Possibility of PTEN-Long in Renal Cell Carcinoma D. Calvisi, ed. *PLoS ONE*, 10(2), p.e114250. Available at: http://dx.plos.org/10.1371/journal.pone.0114250 [Accessed May 24, 2016].

2.5 Supplementary material

Supplementary Figure 1.

Alignment of the *PTEN* 5' leader from the 5' most upstream in-frame stop codon (in human and mouse) to the main ORF AUG from 52 mammals. The human and mouse sequences were obtained from RefSeq. The sequences of *Callithrix jacchus* and *Papio hamadryas* were obtained using a BLAST search against the Transcriptome Shotgun Assembly (TSA) database. The other 48 sequences were obtained by a BLAST search against the WholeGenome Shotgun contigs (wgs) database. The initial alignment was performed with the help of the ClustalX algorithm, which was then followed by a manual re-alignment step. Insertions/deletions highlighted in red are the 3' most insertions/deletions that represent frameshifts in a given sequence. The upstream in-frame stop codon for the human sequence is the equivalent in most other organisms and is highlighted in magenta. The previously described CUG and main ORF AUG start codons are highlighted in green. Alternating red and black stripes marks codons in the previously identified N-terminal extension. Species names are on the left of each sequence. Completely conserved nucleotides are indicated by an asterisk at the bottom of the alignment. The non-AUG start sites identified in this study are underlined.
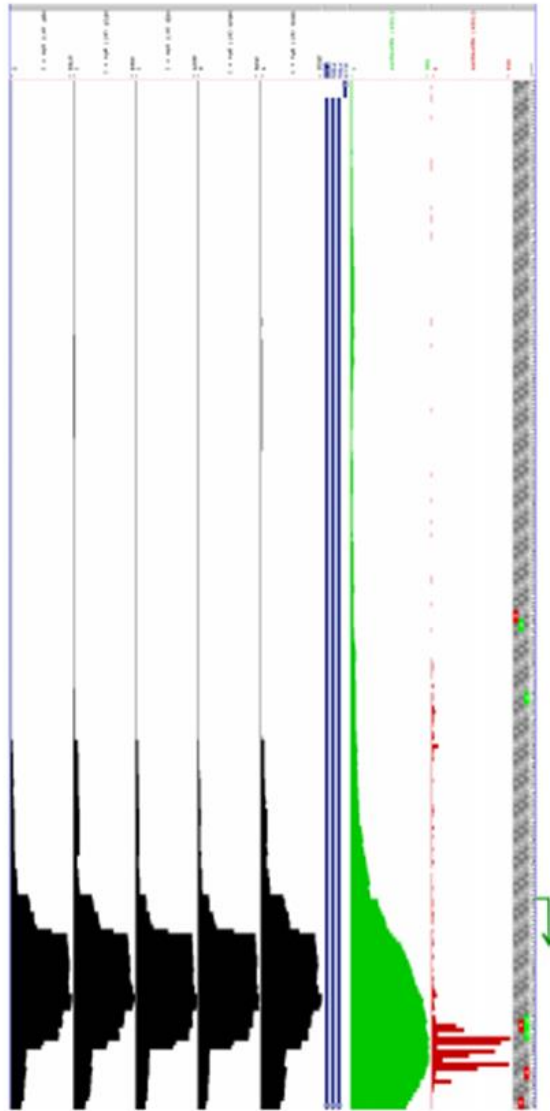
Supplementary Figure 2.

Screenshot of the 5' end of the *PTEN* 5' leader from GWIP-viz (hg19). RNAseq alignments are shown as green histograms and the positions of ribosome A-sites inferred from alignments of ribosome protected fragments (Riboseq) are shown as red columns. The data are aggregated from 15 ribosome profiling studies carried out in several human cultured cells. CAGE analysis from the indicated cells are from the Fantom Project (Crooks et al. 2004) and are shown as black histograms. The number of reads at each nucleotide position is indicated on the y-axis. The nucleotide corresponding to the most abundant transcript site (+187 relative to RefSeq entries) is indicated with a green arrow.
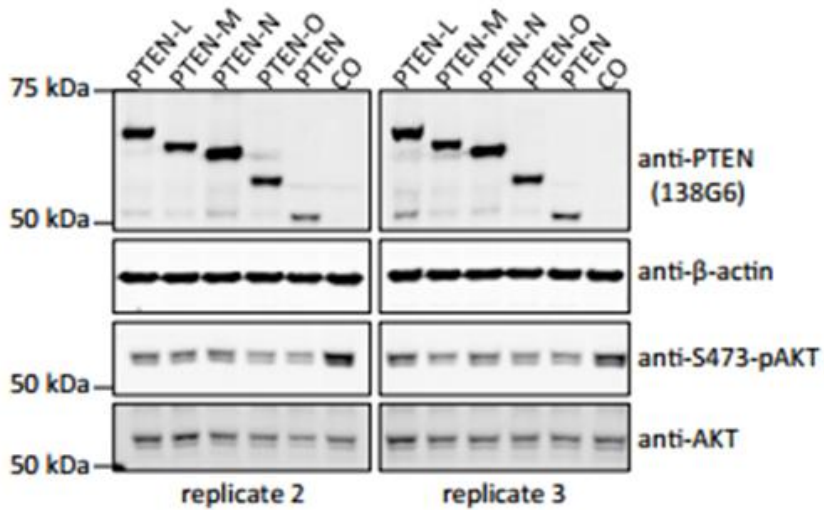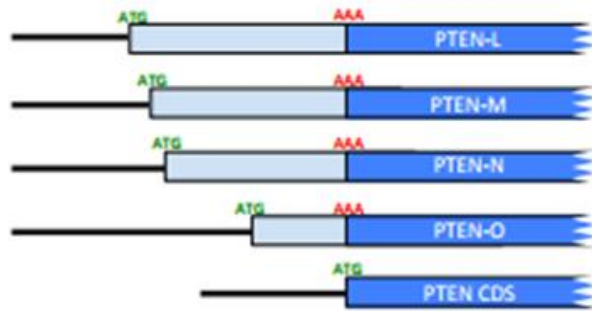
Supplementary Figure 3

Replicate immunoblots of cell lysates prepared from PTEN-null PC3 cells transfected with PTEN expressing constructs as indicated for 48 hr (serum starved for last 24 hr) and probed with antibodies against PTEN (138G6), β-actin, phospho-AKT (S473) and pan-AKT. Replicate 1 is shown in Fig. 2.3.4.
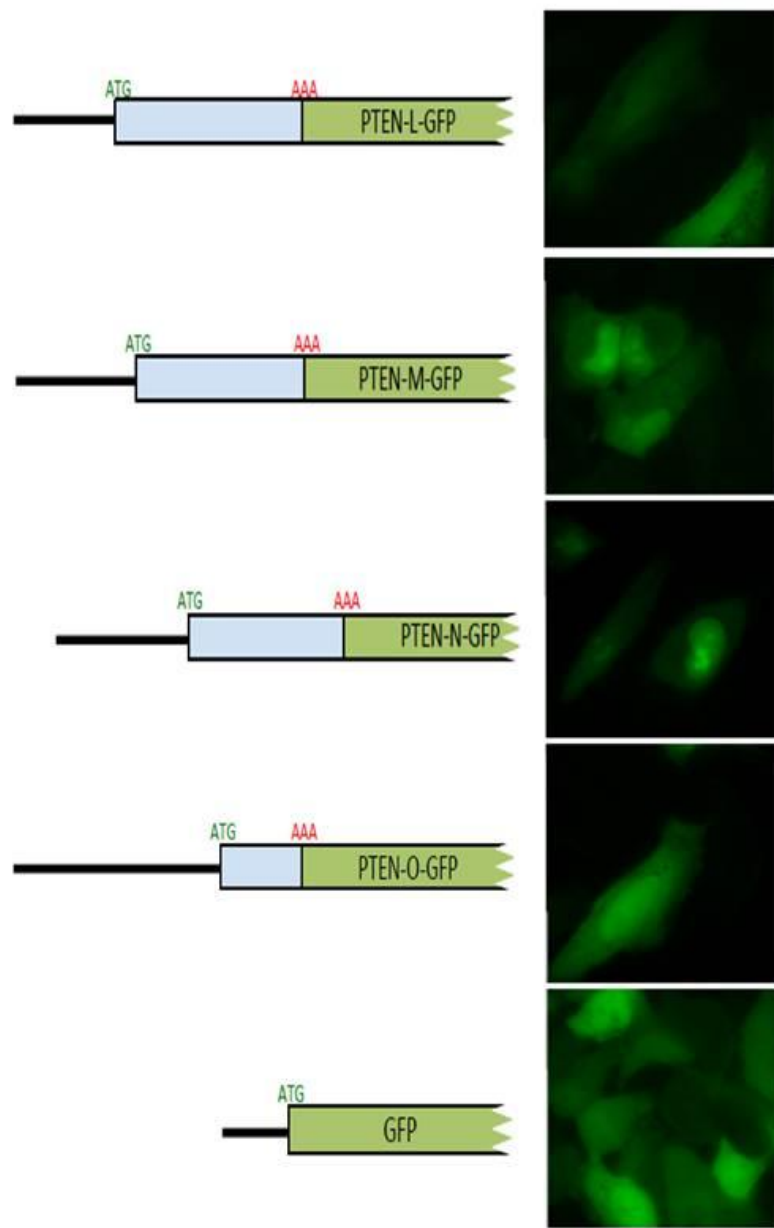
Supplementary Figure 4

Fluorescence microscopy of live HeLa cells imaged 24 hr after transfection with the GFP constructs indicated. Scale bar is in μm.
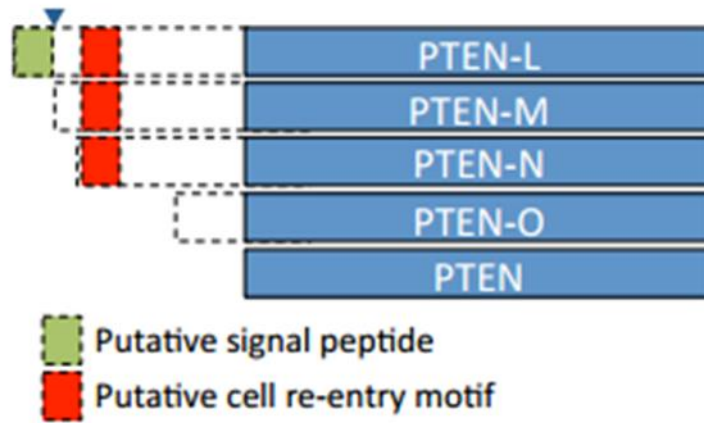
Supplementary Figure 5

Illustration of the PTEN protein and proteoforms PTEN-L, PTEN-M, PTEN-N and PTEN-O. The region corresponding to AUG initiated PTEN is shaded blue whereas the dashed box represents the non-AUG initiated extensions. The signal peptide and cell re-entry motif predicted by Hopkins et al (2013) are shown as green and red boxes respectively. The putative cleavage site is depicted as a blue triangle.
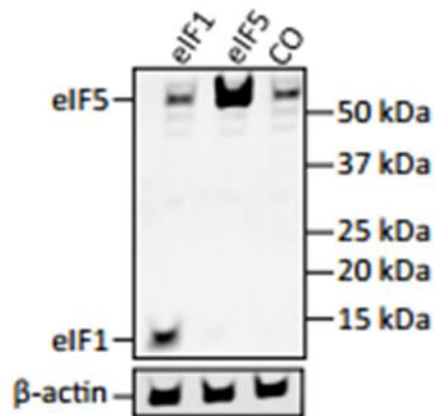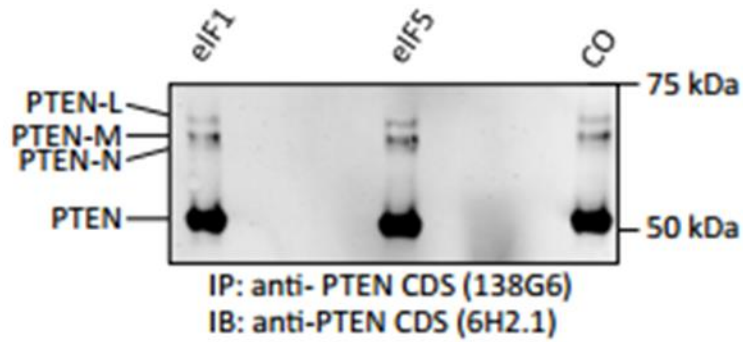
Supplementary Figure 6

Upper panel - Anti-PTEN CDS (6H2.1) immunoblot of anti-PTEN CDS (138G6) immunoprecipitates prepared from HEK-293T cells transfected with constructs overexpressing either eIF1 or eIF5. CO indicates control immunoprecipitates from cells transfected with empty vector. Lower panel – Immunoblot of lysates prepared from cells transfected for immunoprecipitation above probed with antibodies against eIF1, eIF5 and β-actin.
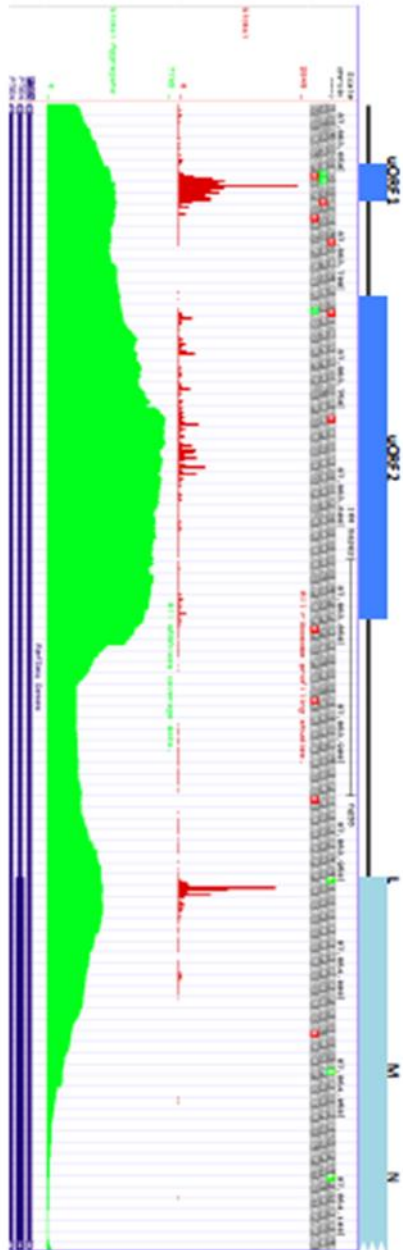
Supplementary Figure 7
Screenshot of the 5' end of the *PTEN* 5' leader from GWIP-viz (hg38) and illustration of the ORF architecture showing the two conserved AUG initiated uORFs and sites of non-AUG initiation described in this study. RNAseq alignments are shown as green histograms and the positions of ribosome A-sites inferred from alignments of ribosome protected fragments (Riboseq) are shown as red columns. The data are aggregated from 15 ribosome profiling studies carried out in several human cultured cells. The number of reads at each nucleotide position is indicated on the y-axis.
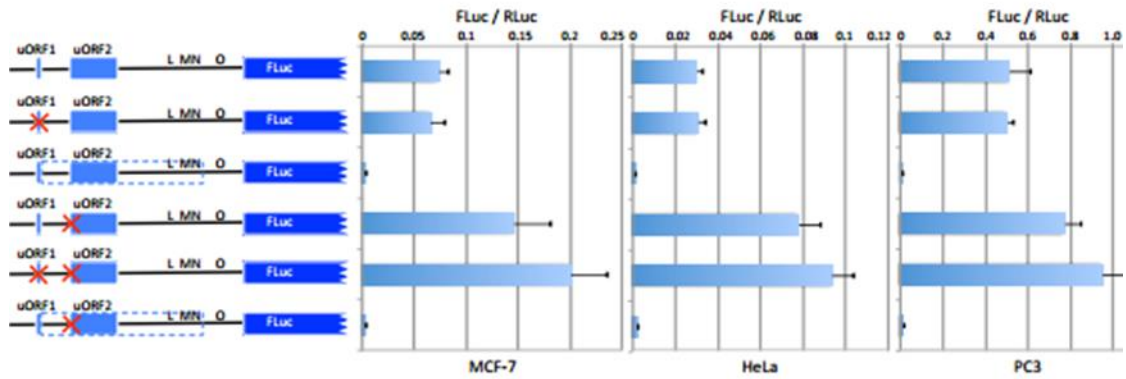
Supplementary Figure 8

Relative luciferase activities (FLuc / RLuc) of firefly encoding sequences fused to the wildtype or mutant *PTEN* 5' leader as indicated and transfected into either MCF-7, HeLa or PC3 cells as indicated. Red crosses indicate mutation of AUG start codons to non-initiating AGG codons. L, M, N and O depict the approximate site of initiation of PTEN extensions. The dashed box represents the increase in ORF length when the stop codon of uORF1 is changed to a sense codon.
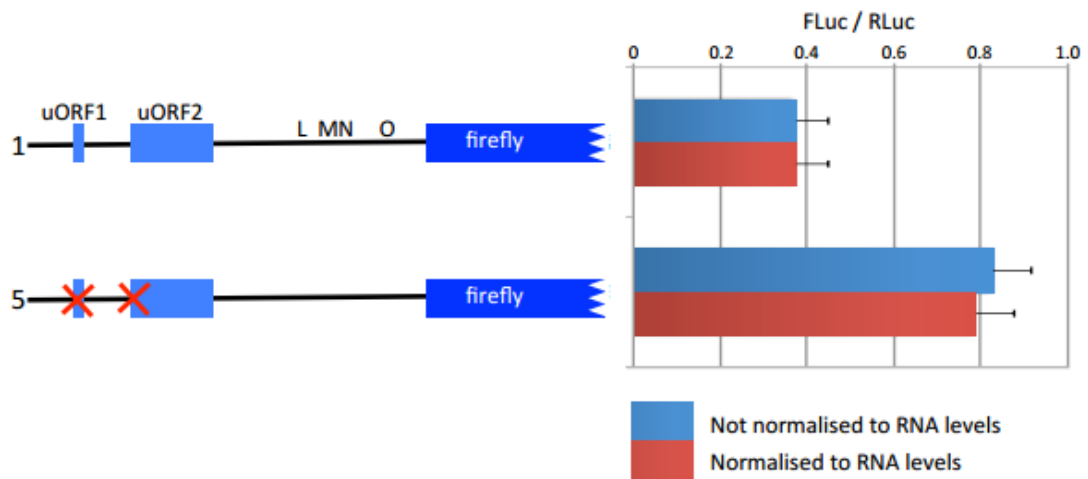
Supplementary Figure 9

Relative luciferase activities (FLuc / RLuc) of firefly encoding sequences fused to the wildtype or 'no uAUG' mutant (constructs 1 and 5 respectively from Fig. 2.3.6) *PTEN* 5' leader as indicated and transfected into HEK-293T cells. Cells were transfected in triplicate wells (6- well plate) and transfectants processed for simultaneous RNA and protein isolation (see Materials and Methods). RNA levels were determined by RT-qPCR and relative luciferase activities were either normalised to RNA levels (red) or not (blue). Red crosses indicate mutation of AUG start codons to non-initiating AGG codons. L, M, N and O depict the approximate site of initiation of PTEN extensions.
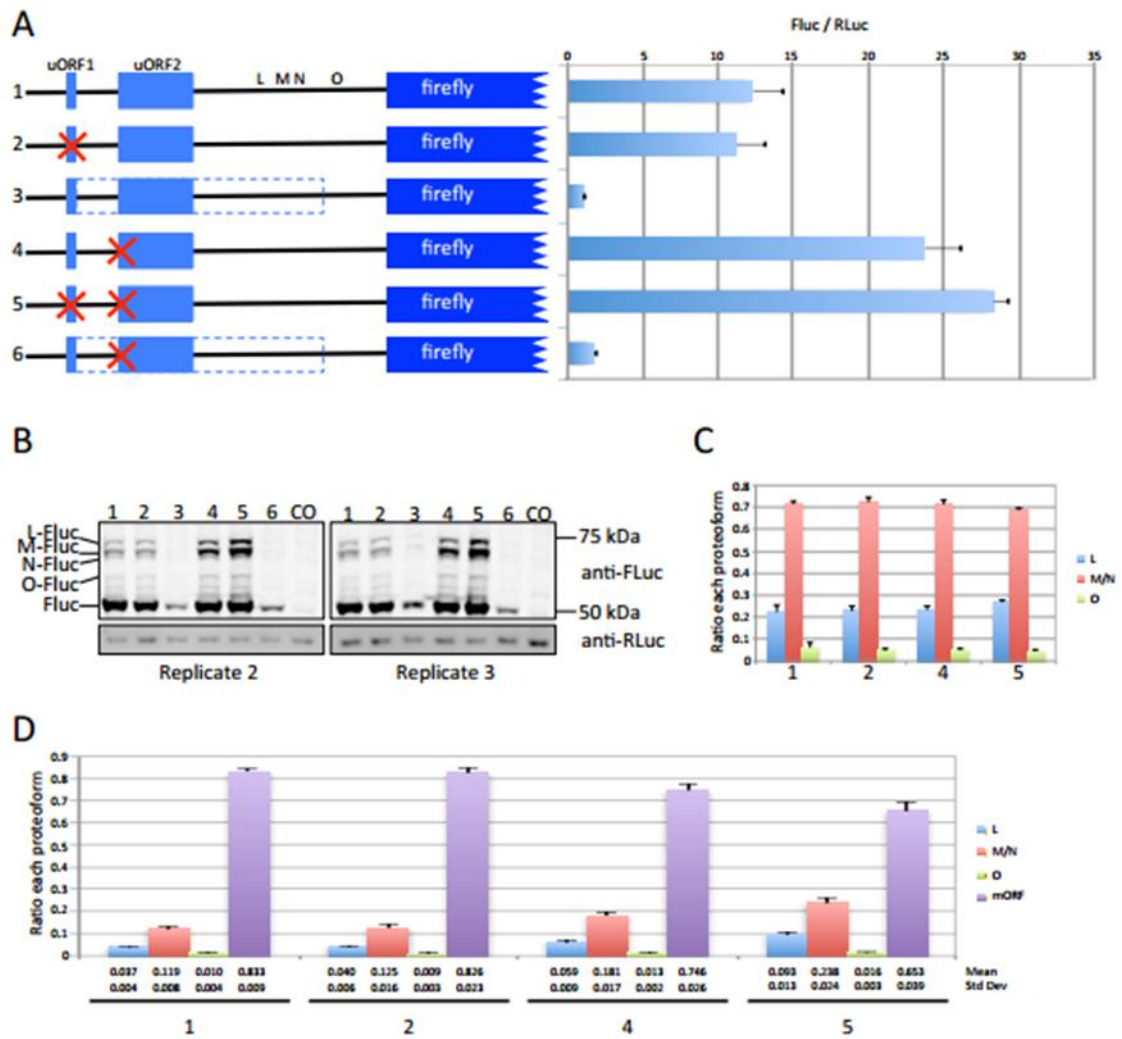
Supplementary Figure 10

A. Dual luciferase assay from cell lysates prepared from HEK-293T cells transfected (three biological replicates) with the indicated constructs. B. Replicate immunoblots of cell lysates prepared from HEK-293T cells transfected with constructs 1-6 and probed with antibodies against firefly and *Renilla* luciferases. Replicate 1 is shown in Fig. 2.3.6B. C. Within-lane densitometry analysis from three biological replicates of the proteins detected by anti-FLuc in B and Fig. 2.3.6B. Proteoforms M and N could not be resolved sufficiently from each other for accurate densitometry analysis, so the intensity of both proteins together is determined. The ratio of each proteoform is calculated as a fraction of the sum of the L, M/N and O proteoforms. D. Within-lane densitometry analysis from three biological replicates of the proteins detected by anti-FLuc in B and Fig. 2.3.6B. The ratio of each proteoform is calculated as a fraction of the sum of L, M/N, O and main ORF proteoforms.
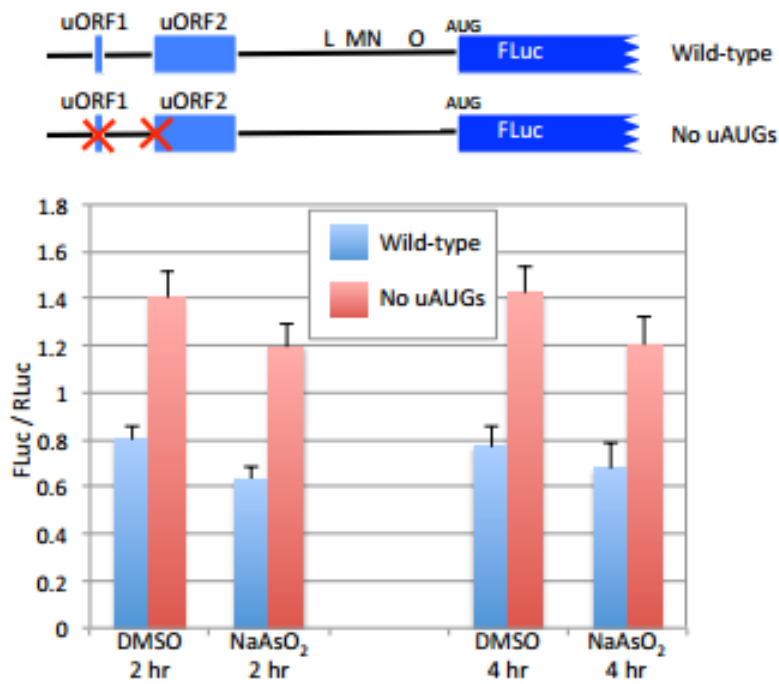
Supplementary Figure 11

Relative luciferase activities (FLuc / RLuc) of firefly encoding sequences fused to the wildtype or 'no uAUG' mutant (constructs 1 and 5 respectively from Fig. 2.3.6A) PTEN 5' leader as indicated and transfected into HEK-293T cells for 6 hr before treatment with 5 µM arsenite for either 2 hr or 4 hr as indicated.

97

Tzani, I. 2016. A study of translation regulation in cancer by application of ribosome profiling. PhD Thesis, University College Cork.

Please note that Chapter 3 (pp.98-169) is unavailable due to a restriction requested by the author.

4.1 Abstract

Polyomavirus research has played a central role in the elucidation of cancer biology (Gross 1953). Studies on murine polyomavirus have been central to the discovery of tyrosine phosphorylation (Eckhart et al. 1979) and the characterization of the PI3K pathway (Kaplan et al. 1987). Research on simian vacuolating virus 40 (SV40) has led to the characterization of the role of Rb1 in cell cycle regulation and the discovery of p53 (Lane & Crawford 1979; Linzer & Levine 1979; Reich & Levine 1982; DeCaprio 2009).

Merkel cell carcinoma (MCC) is an aggressive non melanoma skin cancer with increasing incident rates. 80% of MCCs are caused by the Merkel Cell Polyomavirus (MCV). MCV expresses a small (sT) and a large antigen through alternative splicing of a single transcript. Although both antigens are required for maintenance and proliferation of transformed cells, only the sT can transform rodent fibroblasts by dysregulating protein synthesis initiation. Ribosome profiling enables genome-wide information on protein synthesis, through sequencing of the ribosome protected fragments.

The aim of this project was to study translation regulation in MCC cells positive for MCV, at various stages of malignant transformation. To study the initial steps of sT induced carcinogenesis we established a Rat-1 cell line, which inducibly and reversibly expresses sT. This cell line failed to transform upon sT expression, possibly due to low sT levels.

To investigate the potential role of sT at the later stages of MCV induced transformation, ribosome profiling was applied on a MCC cell line positive for MCV transduced with shRNA towards sT. Knockdown of sT did not affect translation as expected, possibly because sT exerts its role at the first stages of carcinogenesis.

4.2 Introduction

4.2.1 Merkel Cell Carcinoma

Merkel cell carcinoma (MCC) is an aggressive non melanoma skin cancer originally described in 1972 (Toker 1972), which arises from the neuroendocrine mechanoreceptors Merkel cells (Moll et al. 2005; Boulais and Misery 2007; Lucarz and Brand 2007). Although relatively rare (Agelli & Clegg 2003), it has poor prognosis and increasing incident rates over the past twenty years (Fitzgerald et al. 2015; Hodgson 2005; Lyhne et al. 2011; Reichgelt & Visser 2010). Frequent occurrence of this cancer in elderly and immunosuppressed individuals suggested an infectious origin of the disease. Digital transcriptome subtraction is a method that involves *in silico* subtraction of host sequences to detect pathogens' transcripts (Feng et al. 2007). Application of this method to MCC samples identified the genome of the Merkel Cell Polyomavirus (MCV or MCPyV) in 8 out of 10 of the studied MCC tumours (Feng et al. 2008). MCV infection in MCC tumours was subsequently confirmed by other independent research groups (Kassem et al. 2008; Becker et al. 2008).

MCC is located in the dermis. Its morphological features involve round to oval nuclei, small nucleoli and syncytial appearing cytoplasm (Chang & Moore 2012). Diagnosis of MCC is based on immunohistochemical analysis. This cancer is positive for cytokeratin 20 (CK20), low molecular weight cytokeratins (CAM 5.2 or AE1/AE3) and neuron-specific enolase (NSE). Positivity for these markers along with negativity for the thyroid transcription factor 1 (TTF-1), the leukocyte common antigen (LCA) and S100, enable differential diagnosis of MCC from small cell lung cancer (SCLC), lymphoma and small cell melanoma (Grekin et al. 2016).

MCCs express neuroendocrine markers and are therefore thought to derive from Merkel cells (MC) of the skin (Grekin et al. 2016). However due to the postmitotic character of MCs, it has been suggested that MCCs could possibly derive from skin stem cells (Tilling et al. 2012), or from pro-/pre B cells (zur Hausen et al. 2013). MCs are touch receptors that have characteristics of presynaptic cells (Haeberle et al. 2004; Hitchcock et al. 2004; Nunzi et al. 2004). The origin of Merkel cells is not clear with some groups suggesting a neural crest origin (Szeder et al. 2003) and others derivation from the epidermal lineage (Morrison et al. 2009).

4.2.2   Human cancer viruses

MCV is one of the eight viruses that are known to cause tumours to date and the first polyomavirus with an etiologic role in human cancer development (Moore & Chang 2010). The viruses that are currently known to cause tumours in humans are the following.

Epstein-Barr (EBV) is a double stranded DNA herpesvirus that can cause Burkitt's lymphomas and nasopharyngeal carcinomas, lymphoproliferative disorders, Hodgkin's disease, non-Hodgkin's lymphomas and gastrointestinal lymphomas (Epstein et al. 1964). Hepatitis B virus (HBV) is a single stranded and double stranded DNA hepadenovirus that can cause hepatocellular carcinomas (Blumberg et al. 1965). The Human T-lymphotropic virus-I (HTLV-I) is a positive-strand, single-stranded RNA retrovirus that causes adult T cell leukaemia (Poiesz et al. 1980). HPV16 and HPV18, along with some other α-HPV types, which are also carcinogenic, are double stranded DNA papillomaviruses, that can cause cervical, penile, anogenital and head and neck cancers (Durst et al. 1983; Boshartb et al. 1984). The hepatitis C virus (HCV) is a positive-strand, single-stranded RNA flavivirus, which causes hepatocellular carcinomas and lymphomas (Choo et al. 1989). The Kaposi's sarcoma herpesvirus (KSHV) is a double-stranded DNA herpesvirus, which causes Kaposi's sarcoma, primary effusion lymphoma and multicentric Castleman's disease (Chang et al. 1994).

4.2.3   Polyomaviruses

The first human polyomaviruses, John Cunningham virus (JCPyV or JC) and the BK virus (BKPyV) were discovered in 1971 and they were the only members of the Polyomaviridae family to be known for decades. Due to advancements of genomic amplification technologies, which enabled the discovery of new members of this family over the last decade, there are currently 10 human polyomaviruses (HPyVs) known (White et al. 2013). The polyomavirus BK (BKV) was isolated from a kidney transplant recipient with the initials B.K. (Gardner et al. 1971) and JCV was isolated from brain tissue of a patient with progressive multifocal leukoencephalopathy (Padgett et al. 1971). The Karolinska Institute polyomavirus (Allander et al. 2007) and the Washington University polyomavirus (Gaynor et al. 2007), were both

isolated from the respiratory tract. Human polyomaviruses -6 and -7 (HPyV6 and HPyV7) were identified in normal skin (Schowalter et al. 2010). The Trichodysplasia Spinulosa-associated polyomavirus (TSV or TSPyV) has been associated with trichodysplasia spinulosa in immunocompromised patients (van der Meijden et al. 2010). The Human Polyomavirus-9 (HPyV9) was detected from an immunosuppressed kidney transplant patient (Scuda et al. 2011), the Human Polyomavirus-10 (HPyV10) was isolated from a patient with warts, hypogammaglobulinemia infections and myelokathexis (WHIM) syndrome (Buck et al. 2012) and the MW Polyomavirus (named after its discovery in Malawi) (MWPyV) was isolated from a healthy stool sample (Siebrasse et al. 2012). Although HPyVs subclinically infect the general population, they only seem to cause illnesses in immunocompromised people. Consistently with the above, HPyV6, -7, -10 and TSPyV were found more frequently on the skin of HIV positive compared to HIV negative men (Wieland et al. 2014), but HPyV6, -7, -9, 10 and MWPyV have not been associated with any pathology (Dalianis & Hirsch 2013). Through cell culture and animal studies BKPyV, JCPyV and the simian vacuolating virus 40 (SV40) have been associated with cancer (Del Valle et al. 2008; Bouvard et al. 2012), but only MCV is known to have a causative role in tumourigenesis (Feng et al. 2008).

Research on polyomaviruses began with the discovery of the murine polyomavirus (MPyV), which is able to induce tumours in newborn mice (Gross 1953) and the discovery of the SV40 (Sweet & Hilleman 1960), which can cause tumours in experimental animals. Due to the shared genomic and biochemical properties of polyomaviruses, studies on SV40 biology have served as a reference model for the understanding of the molecular physiology of the more recently discovered polyomaviruses (Gjoerup & Chang 2010).

4.2.4 Genome organization of polyomaviruses

Polyomaviruses are small, non-enveloped, double-stranded DNA viruses with 4.5-5.5kb genomes, comprised of early and late coding regions. Viral early and late genes are transcribed bidirectionally from a noncoding regulatory region (NCRR) at

the origin of DNA replication (ori) and subsequently translated from the different transcripts that were generated by alternative splicing.

The early region contains the T antigen locus, which encodes for the large and the small tumour antigens (LT and sT) in all polyomaviruses. These genes are expressed immediately after infection and encode for proteins important for the viral life cycle, by mediating viral DNA replication and regulating cellular mechanisms. A large T-antigen variant, encoded by alternative splicing, has been detected in SV40 (17k T antigen) (Zerrahn et al. 1993) and in JCV, BKV and MCV (Trowbridge & Frisque 1995; Shuda et al. 2008a; Abend et al. 2009). A middle T antigen has largely unknown functions in most polyomaviruses, but is the principal transforming protein of murine and hamster polyomaviruses (Cheng et al. 2009).

The late region encodes for three to four proteins (VP1, VP2, VP3 and VP4), through alternative splicing and for Agnoprotein. The late region proteins are expressed after viral DNA replication has initiated and are required for capsid formation. VP4 has so far only been detected in SV40 and promotes cell lysis (Daniels et al. 2007). Agnoprotein, although not found in MCV, plays a role in viral replication (White & Khalili 2004). Several polyomaviruses express miRNAs, which regulate the expression of early genes at later stages of infection (Sullivan et al. 2005).

4.2.5   Natural infection and clinical disease

Polyomaviruses have a narrow host range. Viral infections usually occur early in life, but variability is observed in the age of polyomavirus acquisition and seroprevalence for various polyomaviruses (Carter et al. 2009). Specifically for MCV, seropositivity is observed in about 50% of children up to 15 years of age and up to 80% in individuals older than 50 years old (Tolstov et al. 2009). However MCC suffering patients who are MCV positive have much higher MCV IgG titer than asymptomatic carriers (Pastrana et al. 2009).

Routes of transmission vary between human polyomaviruses, but involve the faecal-oral, oral and respiratory routes (Gjoerup & Chang 2010). Although it is not clear how polyomaviruses establish persistent infections in various organs, their

detection in blood and lymphoid tissues at high frequencies (Sharp et al. 2009), suggests that hematolyphoid cells carry and distribute the viruses.

Primary infections with human polyomaviruses are usually asymptomatic. During this latency state, there is very low viral replication. Dysfunctional host immune responses enable viral reactivation, which is characterised by active viral replication and can be detected by the presence of late transcription genes and can result in disease (Wiedinger et al. 2014). Tissue tropism, i.e. the affinity of a virus for specific tissues, and mechanisms of latency and reactivation vary among different polyomaviruses, and this is reflected in the differences observed upon reactivation, which can be either asymptomatic or symptomatic.

4.2.6 Polyomavirus life cycle

The infectious viral life cycle begins with the internalisation of virus through recognition of its VP1 protein, by specific cellular receptors, which vary between different viruses. Internalisation of BKV is mediated by the gangliosides GD1b and GT1b (Low et al. 2006). Internalisation of MCV occurs through GT1b (Erickson et al. 2009), whereas JCV is internalised through GT1b and the serotonin receptor 5HT2AR (Elphick et al. 2004). The polyomavirus is then transferred through caveolae and the endoplasmic reticulum to the nucleus, where its genome is transcribed. Initially the proteins encoded from the early region (LT and sT) are translated. LT can subsequently initiate viral DNA replication from NCRR, which contains the origin of replication. Transcription of the late region occurs after the initiation of DNA replication, through activation of the late promoter. VP1, which is the principal component of polyomavirus capsomers, assembles with VP2 and VP3 to form the viral capsid (Gjoerup & Chang 2010).

4.2.7 Molecular mechanism of polyomavirus induced transformation – SV40 as a model system

The size and complexity of the human genome, the presence of both driver and passenger mutations in cancer cells (Stratton et al. 2009), along with intratumour heterogeneity (Marusyk & Polyak 2010), render direct comparisons between

normal and tumour cells, to identify the changes that drive tumourigensesis, challenging. Viruses, with their ability to target key modules of signalling pathways, have enabled our understanding of oncogenic transformation. The small size of the polyomavirus genome has allowed for the use of genetic manipulation in a number of studies that have elucidated cellular physiology. Due to their small genome size, they rely on the cellular replication machinery, for their own DNA replication. Therefore they force S phase entry to facilitate viral genome replication. The pathways involved in DNA replication, are also involved in oncogenesis (Gjoerup & Chang 2010).

The SV40 LT is considered the major SV40 oncoprotein, because its expression is sufficient to transform rodent fibroblasts. LT's ability to interact and inactivate retinoblastoma protein (pRB) and p53, are considered key for the transforming capabilities of the virus (Pipas 1992). SV40 LT interacts with pRb, through its LXCXE motif, which is also present in MCV LT (Shuda et al. 2008a). Interaction with p53 occurs through a C-terminal site of SV40 LT. Viral DNA replication initiates when LT forms double hexamers that bind to the SV40 origin of replication. With its helicase activity SV40 LT opens the viral DNA and allows the assembly of the cellular machinery required for replication (Wright et al. 2009).

4.2.8   Genome organization of Merkel Polyomavirus

MCV has a 5.3 kb genome, with an organization similar to that of other polyomaviruses. Its characterization was guided by previous knowledge about SV40. The MCV genome has a late and an early coding region, which are divided by a 71bp NCRR, that contains the viral origin of replication (Kwun et al. 2009) (Figure 4.2.1). The early region encodes for four proteins. The large T (LT), the small T (sT) and the 57kT antigens share a common first exon and are produced through alternative splicing. The fourth protein is expressed from an alternate frame of the Large T open reading frame (ALTO) (Carter et al. 2013) (Figure 4.2.2). The late region encodes for the structural proteins VP1, VP2 and VP3, which are expressed after viral DNA replication, and assemble to form the viral structure. The late capsid proteins are expressed during the natural polyomavirus lytic infection, but not in MCV-induced tumourigenesis, during which only sT and LT are known to be

expressed. Unlike other polyomaviruses, MCV does not encode for agnoprotein (Chang and Moore, 2012; Houben et al., 2010).
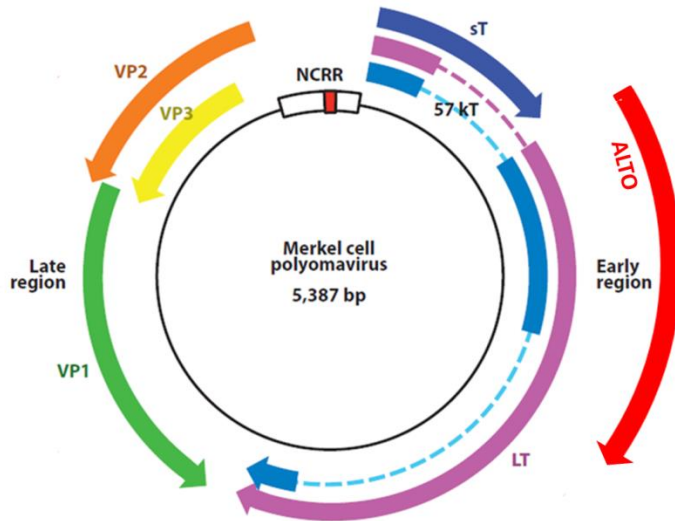
Figure 4.2.1 Genome organization of Merkel Cell Polyomavirus

The MCV genome encodes for the early (LT, sT, 57kT and ALTO) and late region proteins (VP1, VP2 and VP3). Transcription starts bidirectionally at the NCRR, which contains the viral origin of replication (red). Figure adapted from (Yuan Chang and Moore 2012, *Annu. Rev. Pathol. Mech. Dis*).
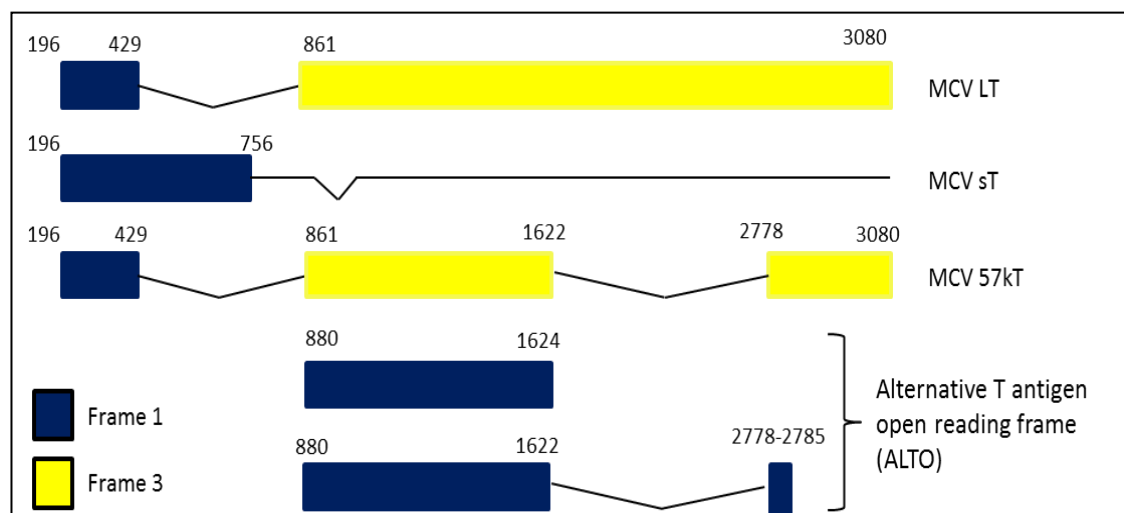
Figure 4.2.2 : Transcripts encoded by the early coding region of MCV.
More than one gene product is encoded from the early region of MCV by alternative splicing. LT, sT and 57kT are the main T antigens and share the first exon. The reading frame of the first exon of ALTO is +1 shifted in comparison to the second exon of LT.

4.2.8.1 LT MCV

The LT antigen contains a DnaJ domain at its N-terminus (Figure 4.2.3), with which it interacts with Hsc70. Disruption of this interaction by point mutations disrupts MCV viral replication *in vitro* (Kwun et al. 2009). Downstream of the DnaJ domain lays an MCV T antigen unique region (MUR), that contains binding motifs for Vamp6 and the retinoblastoma (Rb) family of proteins. LT interacts with hVam6p and sequesters it to the nucleus and thus modulates lysosomal clustering (Liu et al. 2011). Loss of the interaction between LT and Vam6p can lead to enhanced viral replication (Feng et al. 2011). SV40 LT is not known to interact with Vam6p.
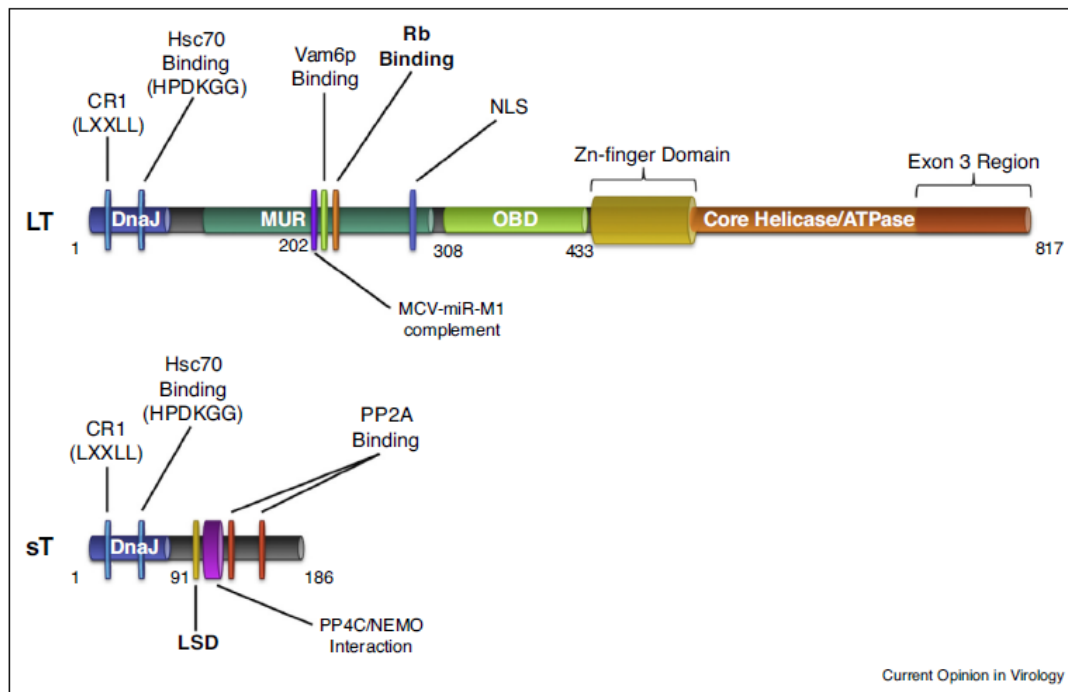
Figure 4.2.3 Protein domains and motifs within the MCV sT and LT antigens
Schematic representation of the functional domains of LT and sT antigens. The DnaJ domain is common between sT and LT. MUR contains an NLS, the binding motifs for Vam6p and Rb family members, as well as the site of complementarity for MCV-miR-M1. The C terminus of LT contains OBD and the core helicase/ATPase domain, which contains the zinc finger domain and exon 3 of 57 kT. The sT contains a DnaJ domain at its N-terminus, followed by LSD and PP2A and PP4C/NEMO binding sites. CR1: conserved region 1; RB: Retinoblastoma-binding; PP2A: Protein Phosphatase 2A; NLS: Nuclear Localisation signal; OBD: Origin Binding Domain; MUR: MCV-unique region; LSD: LT Stabilization Domain (from Wendzicki et al 2015, Current Opinion in Virology).

MCV LT contains the highly conserved (among polyomaviruses) Rb-binding motif (LXCXE), in the MUR. Similar to other polyomaviruses, MCV LT binds to Rb and this interaction leads to release of active E2F1 transcription factor from the E2F1/Rb complex (Arora et al. 2012). During the cell cycle, CDK4 and CDK6 (cyclin-dependent kinase 4 and 6) phosphorylate Rb, which releases E2F1, which subsequently promotes G1 /S phase transition (Weinberg 1995; Weinberg 1996). Therefore through its LXCXE domain MCV LT can drive cells into S phase, when DNA replication occurs. Cell growth can be rescued by ectopic expression of LT with intact Rb-binding sites, in MCV positive MCC cells, inducibly expressing shRNA targeting either all T antigens or LT only. Rb-binding MCV LT mutants fail to rescue cell growth, which highlights the importance of MCV LT interaction with LXCXE (Houben et al. 2012).

The LXCXE domain of MCV LT can also interact with survivin, an inhibitor of apoptosis (Arora et al. 2012). MCV LT contains a short nuclear localisation signal (NLS), which is frequently lost in mutations observed in MCC. Loss of this NLS results in a diffused localisation of LT in both the nucleus and the cytoplasm (Borchert et al. 2014).

In the MUR lies the site of complementarity of the MCV encoded miRNA, MCV-miR-1-5p. This miRNA is expressed at low levels in MCC tumours and although the biological relevance of this miRNA remains to be studied, its sequence suggests that it can target both the viral LT and cellular genes (Lee et al. 2012).

The C-terminus of MCV LT contains elements required for viral DNA replication. The Origin Binding Domain (OBD) recognizes and binds to the origin of replication (Harrison et al. 2011; Liu et al. 2011). Phosphorylation of two threonines upstream of the OBD (T297 and T299) can negatively affect binding to ori and thus replication (Diaz et al. 2014). The zinc-finger motif and the helicase/ATPase domain, of MCV LT are important for initiation of replication. Most MCC derived LTs studied have mutations that prematurely truncate LT and result in elimination of viral replication (Shuda et al. 2008a). Exogenous expression of an intact LT C-terminus promotes the DNA damage response (DDR) (Li et al. 2013) and can inhibit cell growth even in an established MCC cell line and also in SV40-immortalized MCC cell

lines (Cheng et al. 2013). Phosphorylation of MCV LT at S816, by ATM kinase, might contribute to the anti-proliferative effects of the LT C-terminus (Li et al. 2015) .

4.2.8.2 sT MCV

MCV sT shares its first exon with LT. Although the DnaJ domain of LT is required for replication, its role in sT is not yet defined. Similar to the DnaJ domain in SV40 sT, which does not have a role in tumorigenesis (Srinivasan et al. 1997; Boyapati et al. 2003), an intact DnaJ domain in MCV sT is not required for viral DNA replication *in vitro* (Kwun et al. 2009) or for the transformative activity of sT in Rat-1 cells (Shuda et al. 2011).

SV40 and MuPyV sT regulate the Akt-mTOR pathway, by interacting with PP2A (Pallas et al. 1990; Rodriguez-Viciana et al. 2006; Hwang et al. 2013). PP2A is a phosphatase that consists of 3 subunits. Subunit A is the scaffolding subunit, subunit B is the regulatory and subunit C is the catalytic subunit. The DnaJ domain of MCV sT, unlike sTs from other polyomaviruses, is dispensable for transformation both *in vitro* (Shuda et al. 2011) and *in vivo* (Verhaegen et al. 2015). Although it can displace the B subunit of PP2A, this activity is not sufficient for cellular transformation, most likely because of the limited number of PP2A B subunits it can displace (Kwun et al. 2015).

MCV sT forms a loop predicted to be in the opposite molecular surface of the PP2A-binding domain. Residues 91-95, which are part of this loop, form the LT-Stabilization Domain (LSD). This domain can inhibit the SCF[Fbw7] E3 ubiquitin ligase, which targets MCV LT, as well as known oncoproteins such as c-myc and cyclin E. MCV sT thus participates in tumourigenesis by inhibiting degradation of both intact and truncated MCV LT proteins as well as other SCF[Fbw7] targets (Kwun et al. 2013).

The MCV sT LSD domain interacts with the eukaryotic initiation factor 4E (eIF4E) binding protein (4E-BP1), a key regulator of eukaryotic cap-dependent translation. 4E-BP1 has four phosphorylation sites (T37, T46, T70 and S65). When 4E-BP1 is unphosphorylated or hypophosphorylated (T37 and T46), the protein can efficiently sequester eIF4E and thus prevent its interaction with other members of the translation initiation machinery and subsequently inhibit cap-dependent translation initiation. When T70 and S65 are phosphorylated, interaction of 4E-BP1

with eIF4E is abolished and cap-dependent translation can begin (Karim et al. 2001; Wang et al. 2003). MCV sT expression has been shown to promote 4E-BP1 S65 hyperphosphorylation and prevent its loss, independently of mTOR activity (Shuda et al. 2011). Interaction of MCV sT with PP2A (Shuda et al. 2011) and Fbw7 (Kwun et al. 2013) are not required for 4E-BP1 hyperphosphorylation. Expression of the phosphorylation-defective, constitutively active 4E-BP1$^{T37A/T46A}$, reverses MCV sT induced rodent fibroblast transformation, suggesting an effect of 4E-BP1 phosphorylation on sT induced transformation (Shuda et al. 2011).

MCV sT, apart from the SCF^{Fbw7} E3 ubiquitin ligase (Kwun et al. 2013), also targets and inhibits the E3 ligase activity of the anaphase-promoting complex/cyclosome (APC/C). The MCV sT-APC/C interaction results in increased CDK1/CYCB1 activity and subsequent 4E-BP1 phosphorylation (Shuda et al. 2015).

NF-kappaB is a family of transcription factors with a central role in regulation of inflammation and immunity. NF-kappaB exerts its role by translocating from the cytoplasm, where it is sequestered by the IκB kinase (IKK) complex, to the nucleus, upon inflammatory stimulation. NEMO (NF-kappaB essential modulator) is the scaffold of the IKK complex. MCV sT interacts with protein phosphatase 4C (PP4C) and/or PP2A Aβ and targets NEMO, thus preventing NF-kappaB nuclear translocation and transcription initiation (Griffiths et al. 2013).

A proteomic study (Knight et al. 2015) showed that MCV sT can also promote cell motility, migration and invasion.

4.2.9   Causal role of MCV in Merkel Cell Carcinoma development

Infection with MCV is common, but only rarely induces tumourigenesis. To test seropositivity for MCV, an enzyme-linked immunosorbent assay, designed for virus like particles (VLP), constituted by MCV VP1 and VP2, was developed. Although various control populations, including MCV negative MCC patients tested positive for the assay, their antibody titers were significantly lower than those of MCV positive MCC patients (Tolstov et al. 2009).

A range of human cancers, including hematolymphoid malignancies, cutaneous melanomas, basal cell carcinomas and neuroendocrine tumours, have

been screened for viral genome integration and MCV protein expression. However MCV was either undetectable in non-MCC tumours, or its presence could not be significantly associated with pathological features (Schmitt et al. 2011; Gheit et al. 2012; Shuda et al. 2009). Quantitive PCR studies show the presence of more than one copy of the viral genome in MCV positive MCC tumours. Non-MCC tumours that have tested positive for MCV, have 2-4 logs lower copy numbers than MCV positive MCCs (Shuda et al. 2009). MCV is detectable at low quantities even in normal samples, from the aerodigestive, the respiratory and the genitourinary tract, suggesting that it natively infects these tissues (Loyo et al. 2010).

The causative role of MCV in MCC is supported by the clonal pattern of MCV integration, which suggests viral integration prior to tumour development. In one case the same pattern of integration was observed in both the primary and metastatic tissue, further supporting the role of the virus in initiation and metastasis of the tumour (Feng et al. 2008). Although viral integration occurs at various sites of the genome, the potential importance of these sites in regulating molecular pathways required for tumourigenesis is unknown.

Tumour derived MCVs present mutations or deletions that eliminate the viral helicase activity and thus its replicative ability. Although the exact sites of these mutations differ among various MCC-isolated MCVs, they never affect the N-terminal domains, including the Rb-interacting LXCXE domain, and the sT protein. Mutations affecting the viral replicative ability, would eventually lead to viral loss, if they occurred prior to viral genome integration. Maintenance of its replicative ability, while it is integrated, could potentially lead to collisions with cellular replication forks and subsequent DNA damage, that would prevent tumour formation (Shuda et al. 2008a). The above suggests that MCV is not a virus that secondarily infects tumour cells. It is likely that viral integration and subsequent mutations that prevent viral replication, but maintain the domains vital for interaction with cellular oncoproteins, are the first steps of MCC development. The requirement for these events, could explain the rarity of MCC and its occurrence mostly in skin exposed to UV.

Immunohistochemical analysis has shown MCV T antigen expression in MCV positive MCC tumour cells (Shuda et al. 2009). MCV T antigen shRNA knockdown,

results in growth arrest and cell death, both in MCV positive MCC derived cell lines, (Houben et al., 2010) and *in vivo* (Houben et al. 2012). The above suggest that T antigen expression is required for tumourigenesis and that MCV is the infectious cause of MCV positive MCC (Houben et al., 2010).

4.2.10 Cap-dependent translation

As mentioned previously MCV sT can preserve hyperphosphorylation of 4E-BP1, which results in dysregulation of cap-dependent translation. In eukaryotic cells, translation plays an important role in cell growth and proliferation. Aberrations in translation are associated with oncogenesis, apoptosis and dysregulated cell growth (Gingras et al. 1999). All nuclear-encoded cellular mRNAs possess a 7-methyl guanosine residue, which caps their 5' ends and facilitates their binding to ribosomes and subsequent translation initiation (Richter & Sonenberg 2005). The rate limiting event in translational initiation is recruitment of the small ribosomal subunit to mRNA, which is mediated by the binding of eIF4E to the 5' cap. eIF4E is a component of the eIF4F complex, which also includes eIF4G and eIF4A (Gingras et al. 1999).

Overexpression of eIF4E has been shown to promote tumourigenesis in NIH-3T3 cells (Lazaris-Karatzas et al. 1990) and in Chinese hamster ovary (CHO) cells (De Benedetti et al. 1994). The myc oncogene is not sufficient to transform rat embryo fibroblasts, but can do so with the cooperation of eIF4E (Lazaris-Karatzas & Sonenberg 1992; De Benedetti et al. 1994). eIF4E can also stimulate the oncogenic activity of the ras oncogene (Lazaris-Karatzas et al. 1992). Knockdown of eIF4E inhibits growth of head and neck squamous carcinoma cells (Oridate et al. 2005), while its overexpression accelerates lymphomagenesis (Wendel et al. 2004). Overexpression of eIF4E in transgenic mice induces B-cell lymphomas, angiosarcomas, lung adenocarcinomas and hepatocellular adenomas with the collaboration of myc (Ruggero et al. 2004). eIF4E is overexpressed in a range of tumour cell lines (Miyagi et al. 1995) and solid tumours. Increased eIF4E expression in non-Hodgkin's lymphomas is associated with more aggressive phenotypes (Wang et al. 1999). eIF4E is overexpressed in the early stages of colon cancer development

(Rosenwald et al. 1999; Berkel et al. 2001), in bronchioloalveolar lung carninomas (Rosenwald et al. 2001), in some head and neck squamous carcinomas (Sorrells et al. 1999; Nathan et al. 1997) and in aggressive thyroid carcinomas (Wang et al. 2001). However, it appears that the effect of eIF4E overexpression on translation is not generic, instead it preferentially increases the translation rates of mRNAs with GC-rich and highly structured 5' UTRs (Koromilas et al. 1992; Graff & Zimmer 2003; Hay & Sonenberg 2004). These mRNAs are translated at low rates under normal conditions and typically encode for proto-oncoproteins (e.g. c-myc, cyclin D1 and ornithine decarboxylase) and proteins that promote angiogenesis (FGF-2, VEGF), invasion and metastasis (MMP-9 and heparanase) (Mamane et al. n.d.; De Benedetti & Graff 2004; Richter & Sonenberg 2005; Graff & Zimmer 2003). A more recent study suggests, that 4E-BP1 inactivation leads to upregulation of transcripts with 5' terminal oligopyrimidine (TOP), or TOP-like motifs. TOP mRNAs are characterized by the presence of cytidine immediately after the 5'cap, followed by an uninterrupted stretch of 4-14 pyrimidines (Thoreen et al. 2012).

The eIF4E-binding proteins (4E-BP1, 4E-BP2 and 4E-BP3) also known as PHAS (Phosphorylated Heat and Acid-Stable) are small 10-12 kD proteins that can inhibit cap-dependent translation by binding to eIF4E and thus interrupting eIF4F complex formation (Pause et al. 1994; Poulin et al. 1998). When 4E-BP1 is hypophosphorylated it binds to eIF4E and prevents its interaction with eIF4G and subsequent translational initiation. Hyperphosphorylation of 4E-BP1 disrupts the 4E-BP1- eIF4E interaction. A range of extracellular stimuli including growth factors, hormones, cytokines and mitogens induce mammalian target of rapamycin (mTOR)-mediated 4E-BP1 phosphorylation, resulting in increased translation rates. Conversely starvation and growth factor deprivation induce 4E-BP1 dephosphorylation and subsequent reduction in cap-dependent translation (Gingras et al. 1999). The phosphorylation of 4E-BP1 occurs at specific serine and threonine residues of the protein. Phosphorylation of Thr 37 and Thr 46 precedes and is required for the phosphorylation of Ser 65 and Thr 70 (Gingras et al. 2001; Gingras et al. 1999; Karim et al. 2001; Wang et al. 2003). In HEK293 cells phosphorylation of Thr 37 and Thr 46 is not inhibited by rapamycin or serum starvation. However serum addition and rapamycin have a strong effect on the

phosphorylation and dephosphorylation respectively of Ser 65 and Thr 70 (Gingras et al. 1999). 4E-BP1 has been shown to be highly phosphorylated in a series of tumours, including breast (Zhou et al. 2004; Armengol et al. 2007; Rojo et al. 2007), ovarian (Castellvi et al. 2006), melanoma (O'Reilly et al. 2009) and prostate cancer (Armengol et al. 2007).

MCV sT antigen induces hyperphosphorylation of eukaryotic initiation factor 4E (eIF4E) binding protein (4E-BP1) on serine 65, thus increasing the activity of eIF4E either by reducing turnover of hyperphosphorylated 4E-BP1 or by inhibiting 4E-BP1 dephosphorylation (Shuda et al. 2011). In MCV sT antigen-transformed cells the oncogenic phenotype can be reversed by the introduction of a constitutively active unphosphorylated 4E-BP1 protein, which implies that MCV might induce carcinogenesis by dysregulating cap-dependent translation through 4E-BP1 hyperphosphorylation (Shuda et al. 2011).

4.2.11 Aim1: Study of the effect of the MCV sT antigen on protein synthesis

MCV induced MCC tumours express both sT and LT antigens (Feng et al. 2008). Although knockdown of both antigens is required to cause cell death, sT is more highly expressed than LT and is sufficient to transform rodent fibroblasts (Rat-1) and promote serum-free proliferation of human fibroblasts (BJ-TERT) (Shuda et al. 2011). MCV sT exerts its transforming effect by inducing hyperphosphorylation of S65 4E-BP1, which results in liberation of eIF4E from 4E-BP1 and ultimately dysregulation of cap-dependent translation (Shuda et al. 2011).

The first aim was to investigate the effect of MCV sT on protein expression and its potential role in MCC tumourigenesis, by applying ribosome profiling. To create an appropriate model system we aimed to establish a tetracycline inducible system in the Rat-1 cell line that would inducibly express sT or the sT.91-95A mutant. The mutant sT was used as a negative control because of the demonstrated inability (Kwun et al. 2013) of this protein to transform Rat-1 cells. We opted for an inducible system because it can facilitate the regulated expression of MCV sT and enable the study of the initial steps of sT-induced transformation.

4.2.11.1 Tetracycline regulated mammalian expression system

The tetracycline inducible system was developed in an effort to establish regulated transcription in mammalian cells (Gossen & Bujard 1992). In the Tetracycline system we have used (tet-on), the gene of interest is cloned downstream of a CMV promoter and two tetracycline operators (TetO$_2$). A second plasmid constitutively expresses the Tet repressor which binds to the TetO$_2$ thus preventing transcription of the gene of interest. Addition of tetracycline or a stable tetracycline analogue (such as doxycycline), induces conformational changes to the Tet repressor, which can no longer bind to the TetO$_2$ sites, thus enabling transcription of the gene of interest (see Figure 4.2.4).
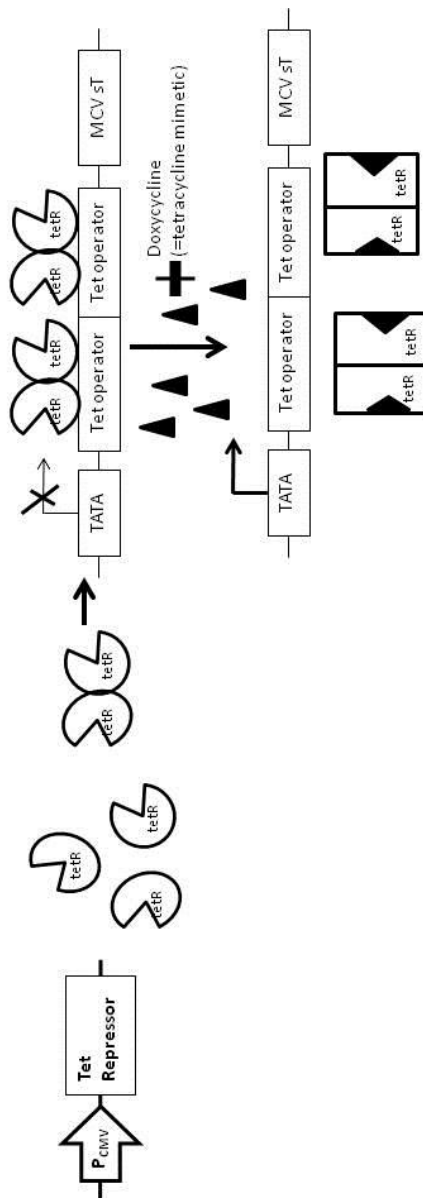
Figure 4.2.4 Schematic illustration of the components of the tetracycline inducible system. The Tet repressor (tetR) protein is constitutively expressed. TetR homodimers bind to the Tet operator (TetO$_2$) sequences, in the inducible expression vector, upstream of the gene of interest and repress its expression. When doxycycline (or any tetracycline or tetracycline analogue) is added, it binds to the Tet repressor and causes its conformational change, which results in release of the Tet repressor from the Tet operator sequences and subsequent expression of the gene of interest.

4.2.12 Aim2: Characterization of the changes in protein synthesis induced by MCV

Shuda et al (2011) have shown that deregulated cap-dependent translation has a causative role in MCV sT induced MCC tumour development. This finding implies that alterations in protein synthesis have a significant effect on MCC development and potentially progression, exactly as already suggested for non-viral induced tumours (Gingras et al. 1999).

To gain a better understanding of the alterations that occur in protein synthesis in human MCV positive MCC and their potential role in disease progression, we aimed to apply ribosome profiling on MCC derived cell lines. MKL-1 is a cell line derived from a nodal metastasis of an MCV positive MCC tumour (Rosen et al. 1987; Shuda et al. 2008b). MKL-1 cells stably transduced with sT shRNA or scrambled shRNA were analyzed using ribosome profiling.

Application of ribosome profiling in MCV induced cell lines of various stages can provide an insight into the changes of protein synthesis that occur and potentially have a causative role in tumour progression. Ribosome profiling cannot only provide quantitative information regarding gene expression, but most importantly a snapshot of *in vivo* translation (Ingolia et al. 2009; Ingolia et al. 2011).

Ribosome profiling data of MCV sT expressing cell lines were expected to show increased expression of cancer related proteins already known to be highly expressed as an effect of eIF4E overexpression (e.g. c-myc, cyclin D1, ornithine decarboxylase, FGF-2, VEGF, MMP-9 and heparanase) (Mamane et al. n.d.; Richter & Sonenberg 2005; Graff & Zimmer 2003; De Benedetti & Graff 2004). Detection of upregulation of these proteins would be used as an internal control for ribosome profiling.

4.3 Materials and methods

All chemicals used were purchased from Sigma Aldrich unless otherwise stated. All restriction enzymes were purchased from New England Biolabs (NEB, Ireland). DNA purifications, agarose gel electrophoresis, cloning, PCR and bacterial transformation were performed using standard molecular biology techniques or according to the relevant kit instructions. PCR reactions were amplified using a TProfessional Basic Gradient Thermal cycler (Biometra).

4.3.1 MTT assay for cell viability

Rat1 cells were seeded to 96 well culture plates at 25% confluency. Once they attached, the media was aspirated and replaced with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS containing various concentrations of blasticidin (1, 3, 5, 7.5 or 10µg/ml) (Invivogen, ant-bl-1), or zeocin (50, 125, 250, 500, 750 or 1000 mg/ml) (Invivogen, ant-zn-1). Each concentration of each antibiotic was assayed once every three days over a period of 15 days. Medium was replaced and fresh medium was added with the appropriate concentration of antibiotic every three days.

Stocks of 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) (Sigma, M2128-16) were prepared by dissolving MTT powder in phenol-red-free RPMI-1640 media at a concentration of 5mg/ml and stored at -20°C (protected from light). MTT was added to cells at a final concentration of 0.5mg/ml and incubated for 3 hours at 37°C, 5% CO_2. All media was aspirated carefully and purple formazan crystals were dissolved instantly upon addition of 200µl DMSO per well. The plates were then incubated at room temperature with gentle shaking for 10 minutes to solubilize the crystals. The absorbance of each well was measured at 540nm on an Infinite 2000 spectrophotometric plate reader (Tecan).

4.3.2 Cell culture

Rat-1 and HEK293T cells were maintained in DMEM supplemented with 10% FBS, 1 mM L-glutamine and 1% penicillin/streptomycin.

4.3.3 Transfections to make Rat-1 cell lines stably expressing pcDNA^{TM}6/TR or pcDNA4

Cell transfection was optimised for Rat-1 cells. 370,000 Rat-1 cells were plated in the well of a 6-well plate. After 24 hours the cells were transfected by adding 0.9 µg of the Tet repressor containing plasmid (pcDNA^{TM}6/TR), or pcDNA4 (empty vector, sT MCV or sT.91-95A cloned into pcDNA4) and 4 µl or 2 µl lipofectamine respectively, in 200 µl Optimem (Gibco).

Cell transfection in 10cm plates.

1.6×10^6 Rat-1 cells were plated in a 10 cm dish. After 24 hours the cells were transfected by adding 3µg pcDNA^{TM}6/TR and 22 µl lipofectamine in 1100 µl Optimem (Gibco).

For simultaneous transfection of pcDNA6 and pcDNA4 plasmids, 3µg of pcDNA6 and 0.5µg of pcDNA4 and 20 µl lipofectamine in 500 µl Optimem in 10ml of medium (+10% FBS) were transfected into ~90% confluent cells.

4.3.4 Antibiotic selection and cloning disks.

48 hours after transfection, cells were treated with the appropriate selection antibiotic (5 µg/ml blasticidin for plasmid pcDNA^{TM}6/TR or 500 µg/ml zeocin for plasmid pcDNA^{TM}4/TO). After a week of blasticidin treatment, the polyclonal cell culture was diluted (1:50, 1:100 and 1:200) in medium with blasticidin in 15 cm dishes, to isolate monoclones. After 10 days cells from the 1:200 dilution dish were further diluted 1:10, 1:100, 1:400 and 1:1000, to ensure monoclonality. Cells were allowed to expand and form discernible colonies for 8 days. Twelve individual colonies were trypsinised and transferred to individual wells of a 24-well plate with the use of cloning disks for monoclonal expansion. Stable integrants were assayed for the highest Firefly luciferase activity levels upon doxycycline addition.

Rat-1 cells transfected in a 10 cm dish with the pcDNA^{TM}6/TR plasmid were treated 24 hours after transfection with medium supplemented with 5 µg/ml blasticidin.

4.3.5 Cell cloning by serial dilution

Transfected cells were incubated at 37°C in 5% CO_2 in medium supplemented with the appropriate selection antibiotic (5 µg/ml blasticidin for plasmid pcDNA^{TM}6/TR or 500 µg/ml zeocin for plasmid pcDNA^{TM}4/TO) for two weeks after transfection. The polyclonal cells from the blasticidin selection step were plated at a density of 5 or 15 or 25 cells/ml in a 96-well tissue culture plate adding 200 µl per well (i.e. 1, or 3, or 5 cells per well). After seven days cells were transferred from the 5 cells/well 96-well plate to a 24-well plate.

4.3.6 Dual luciferase assay.

Stable Rat-1 monoclones were transfected using Lipofectamine 2000 reagent (Invitrogen). For each transfection the following were added to each well: 2ng of pcDNA4_FLuc, 2 ng of a Renilla luciferase expressing construct, 0.2µl lipofectamine and 25µl Optimem. The cells were incubated at 37°C in 5% CO_2 for 24 hours and then treated with doxycycline at a final concentration of 0.5 µg/ml doxycycline or DMSO for 24 hours. After the treatment cells were washed once with 1 x PBS and then lysed in 25µl of 1 x passive lysis buffer (Promega) and Firefly and *Renilla* luciferase activities were determined using the Dual Luciferase Stop & Glo® Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems).

HEK-293T cells were transfected with Lipofectamine 2000 reagent (Invitrogen), using the 1-day protocol in which suspended cells are added directly to the DNA complexes in full-area 96-well plates. For each transfection the following were added to each well: 100ng of pcDNA6 and pcDNA4_Luc plasmids at three different ratios (1:1, 4:1 or 1:4), 10ng of *Renilla* luciferase expressing plasmid plus 0.4µl Lipofectamine 2000 (Invitrogen) in 50µl Optimem (Gibco). The transfecting DNA complexes in each well were incubated with 8 x 10^4 cells suspended in 100µl DMEM + 10% FBS. Transfected cells were incubated at 37°C in 5% CO_2 for 24 hours. On the next day cells were treated for 24 hours with either doxycline at a final concentration of 0.5µg/ml or DMSO. After the treatment cells were washed once with 1 x PBS and then lysed in 25µl of 1 x passive lysis buffer and Firefly and *Renilla* luciferase activities were determined using the Dual Luciferase Stop & Glo®

Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems).

Rat-1 cells were plated at 6 x 10^4 in full area 96-well plates and allowed to attach for 4 hours. After 4 hours they were transfected with the same protocol used for HEK-293T cells.

4.3.7 Cloning

Codon optimized sT and sT.91-95A antigen sequences (provided by Moore and Chang's lab, University of Pittsburgh, Pennsylvania, USA) cloned in pcDNA6 vector, were digested with *Eco*RV and *Xho*I and directionally subcloned into an *Eco*RV and *Xho*I digested and antarctic phosphatase treated pcDNA4 vector. The Firefly luciferase gene was subcloned into a *Pst*I and *Not*I digested and antarctic phosphatase treated pcDNA4 plasmid (pcDNA4_Luc). All clones were verified by sequencing.

4.3.8 Genomic DNA extraction

Cells were pelleted and washed with PBS. The pellet was either frozen or used immediately for genomic DNA extraction. Phase separation was achieved by addition of equal volumes of PBS and phenol-chloroform-isoamyl alcohol (25:24:1) (pH=7.8-8.2)(Sigma, 77617-500ml) to the pellet, gentle mixing and centrifugation at 12,000xg for 3 minutes at 4°C. The nucleic acids containing upper aqueous phase was transferred to a fresh 1.5 ml centrifuge tube and an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to the aqueous phase and the sample was processed as above. The aqueous phase was transferred to a fresh tube and an equal volume of chloroform was added to the tube to remove traces of phenol. The sample was mixed well and centrifuged at 12,000xg for 3 minutes. The upper aqueous phase was transferred to a fresh tube and the DNA was precipitated with 10% of 3M sodium acetate (pH= 5.2M) and 3 volumes of 100% ethanol. The sample was incubated for an hour at -80°C. DNA was harvested by centrifugation at 12,000xg for 5 minutes at 4°C. The supernatant was carefully removed and the pellets were dried for 10 minutes at room temperature before resuspension in DNAse-free ddH$_2$O.

4.3.9   Soft agar colony formation assay

Two established double stable Rat1 monoclones were assessed with the soft agar colony formation assay. Rat-1 cells were trypsinised to single cells, counted, suspended in complete medium containing 0.3% low melting agarose (Sigma-Aldrich) and seeded over a 0.6% agar layer in 35 mm dishes (2,000 cells/ dish). 500 μl of fresh media containing 2 μg/ml of doxycycline or DMSO were added gently to the top of each well every 4 days. After 3 weeks, colonies were stained with crystal violet (0.01% in ethanol) and plates were photographed for soft agar colony formation assay.

4.3.10 SDS-PAGE and Immunoblot analysis

Cells were resuspended in RIPA buffer (50mM TrisHCl, pH=8, 150mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100), supplemented with protease inhibitor cocktail (Sigma, P8340). Samples were lysed on ice for 10 minutes and then centrifuge at 16,000xg at 4°C for 10 minutes. The supernatant was mixed in a 1:1 ratio with 2X Laemmli buffer and heated at 95°C for 5 minutes prior to loading. Proteins were separated using the Bio-Rad Mini-ProteanII gel electrophoresis system. Samples were run through the stacking (5% acrylamide) and resolving (15% acrylamide) gel at 100V. Separated protein samples were transferred to Whatman Protran nitrocellulose membrane using wet transfer conditions at 100V for 1 hour. The membrane was blocked in 2.5% non-fat dry milk (Marvel) in PBS containing 0.1% Tween (PBS-T) for 1 hour at RT. Blocking buffer was removed and membranes were probed with the CM8E6 primary antibody (1:250 dilution in 2.5% milk in PBS-T) overnight with gentle rocking at 4°C. Membranes were washed 3x 5minutes with PBS-T. Secondary antibody donkey anti-mouse IRDYE 800 (LI-COR) was diluted 1:5,000 in 2.5% milk in PBS-T and overlaid onto the membrane for at least 1 hour with rocking. The membrane was washed again (3x 5 minutes) with PBS-T and then a final wash in PBS. Membranes were imaged using the membrane module on the Odyssey infrared scanning system (LI-COR). The CM8E6 antibody was a kind gift from the Chang-Moore lab (University of Pittsburgh).

4.3.11 Ribosome profiling

Ribosome profiling technique was carried out according to (Ingolia et al. 2012) with some modifications. Frozen pellets of MKL-1 cells transduced with shRNA were resuspended in buffer containing 20 mM Tris–HCl (pH 7.5), 250 mM NaCl, 1.5 mM MgCl2, 1 mM DTT, 0.5% Triton X-100, 100 µg/ml cycloheximide (Sigma-Aldrich), 20 U/ml TURBO DNAse (Ambion, Waltham, MA). The buffer we used contains low magnesium (1.5 mM), since high magnesium concentrations stabilize secondary structures in mRNAs which may hamper digestion by RNAse I (RNAse I does not require divalent cations for its activity) and lowering magnesium concentration in 5 mM to 15 mM range has been shown to improve footprint resolution (Ingolia et al. 2012). Cell lysates were incubated on ice for 10 min, centrifuged at 16,000×g at 4°C for 10 min, and the supernatant was divided into two parts for ribo-seq and RNA-seq library preparation. The lysate for Riboseq was treated with RNAse I (Ambion) with 100 U per 3.1 A260 of lysate at 23°C for 50 min. The digestion was then stopped with the appropriate amount of SUPERASE inhibitor (Ambion). The treated lysate was loaded on 10–60% (m/v) sucrose density gradient containing 20 mM Tris–HCl (pH 7.5), 250 mM NaCl, 15 mM MgCl2, 1 mM DTT, 100 µg/ml cycloheximide and centrifuged in SW-41 rotor at 35,000 rpm for 3hr. Sucrose density gradients were prepared as described previously (Stone 1974). Briefly, 5.5 ml of 10% sucrose was slowly layered onto the same volume of 60% sucrose, gradient tubes were then sealed with parafilm, slowly placed horizontally for 4 hr to allow spontaneous gradient formation and then slowly returned to a vertical position. Total RNA from the fractions corresponding to 80S peak was then extracted with phenol/chloroform followed by ethanol precipitation. For the RNA-seq control, the second lysate aliquot was processed with Trizol-LS (Life Technologies, Waltham, MA) according to the manufacturer's protocol. mRNA from total RNA was isolated using the Oligotex mRNA kit (Qiagen, Netherlands). Purified mRNA was then subjected to alkaline hydrolysis as described by (Ingolia et al. 2009). Both ribo-seq and mRNA-seq samples were loaded onto a 15% denaturing urea PAGE (containing 7 M urea, and acrylamide:bisacrylamide in the ratio 19:1).

Bands corresponding to nucleic acid fragments of 28–34 nt were excised for both ribo-seq and RNA-seq samples. RNA was extracted using buffer containing 0.3 M NaOAc (pH 5.1), 1 mM EDTA, and 0.1% SDS, by overnight shaking at room temperature, followed by precipitation with one volume of isopropanol and 2 µl of GlycoBlue (Life Technologies).

The library preparation was carried out as previously described (Ingolia et al. 2012) with the following modifications. The circularization reaction was performed for 2 hr. During PCR library amplification, the temperature ramping speed was set as 2.2°C/s to reduce bias associated with GC content (Aird et al. 2011). Libraries were sequenced on an Illumina HiSeq 2000 system at the Beijing Genomics Institute (BGI).

4.4 Results: Aim1: Study of the effect of the MCV sT antigen on protein synthesis

To study the effect of MCV sT on the initial steps of Rat-1 cells transformation, we set out to establish an inducible system to express sT or its non-transforming mutant sT.91-91A, upon addition with doxycycline, which is a tet mimetic.

In the system that we used (Figure 4.4.1) the amount of Tet Repressor expressed in cells transfected with pcDNA6, determines the level of transcriptional repression of the Tet operator sequences.
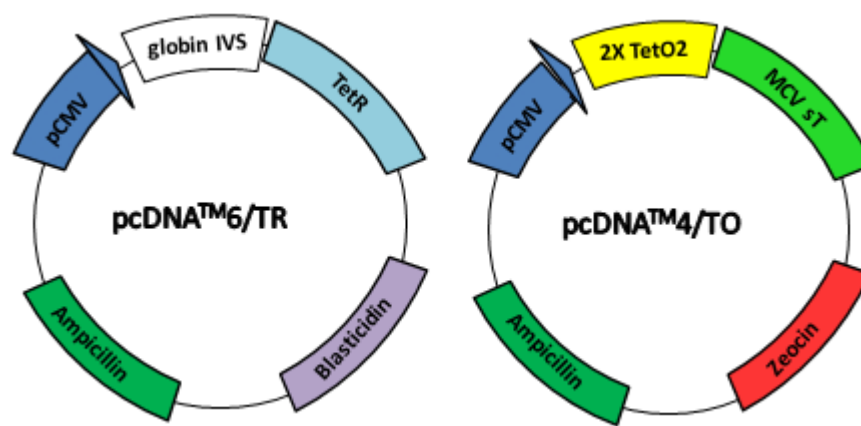
Figure 4.4.1 Schematic representation of the plasmids used in the tetracycline inducible system.
The pcDNA™6/TR plasmid encodes for the Tet repressor under the control of the CMV promoter. It contains the blasticidin resistance gene to allow selection of stable cell lines. The rabbit β-globin intron II (IVS) sequence enhances expression of TetR in mammalian cells (van Ooyen et al. 1979). In the pcDNA4 vector the gene of interest (sT or sT.A91-95A) is cloned downstream of the Tet operator. Transcription is under the control of the CMV promoter. pcDNA4 also encodes for zeocin, which can be used as a selectable marker. Both plasmids contain an ampicillin resistance gene, which enables their propagation in *E.coli* under ampicillin selective pressure.

To test the efficiency of the tetracycline inducible system the firefly luciferase gene, was cloned downstream of the Tet operator gene (pcDNA4_FLuc) and this construct was used as a monitor for Tet repressor expression. To test the transfection efficiency of Rat-1 cells and to find the optimal amount of Tet Repressor for this cell line, Rat-1 cells were transiently transfected with different ratios of the Tet repressor expressing plasmid (pcDNA6), pcDNA4_FLuc as well as an

internal control plasmid expressing *Renilla* luciferase (Rluc) and analysed by dual luciferase assays. HEK293T cells were used as positive controls, because of their high transfection efficiency. This experiment (Figure 4.4.2) showed that HEK293T cells have higher transfection efficiency than Rat-1 cells and that the ratio of pcDNA6:pcDNA4 did not have a major imjpact on the transfection efficiency.
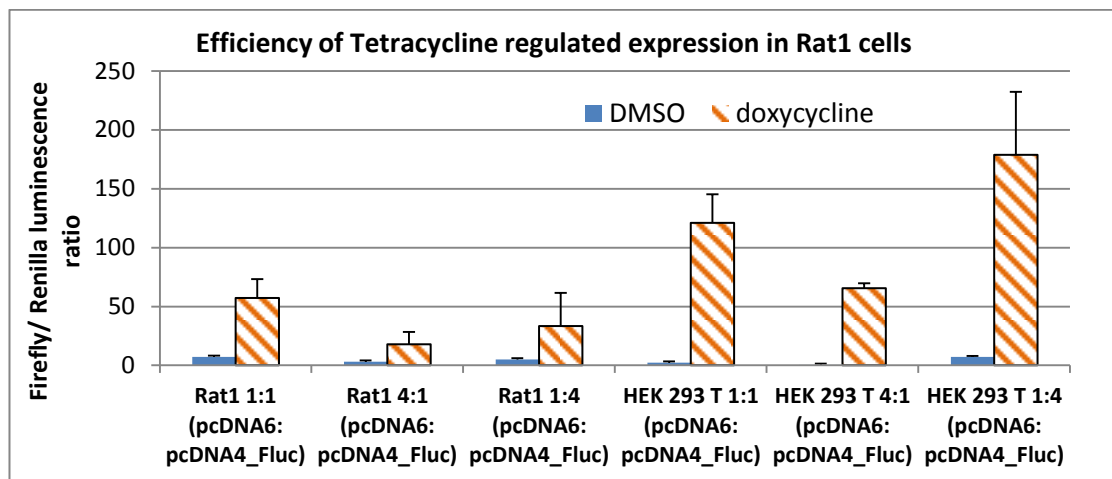
Figure 4.4.2: Efficiency of the tetracycline regulated expression system in Rat-1 and HEK-293T cells.
Rat-1 and HEK-293T cells were transiently cotransfected with both the operator containing pcDNA4_FLuc and repressor containing (pcDNA^{TM}6/TR) plasmids. The cells were cultured in the presence of doxycycline (0.5 µg/ml) or DMSO and after 24 hours, luciferase expression was measured. Firefly values were normalised for transfection using Renilla luciferase. Data represent the mean and standard deviation of one experiment done in triplicate.

To generate stable cell lines expressing the Tet repressor and sT or sT.91-95A, the minimum concentration of blasticidin and zeocin required to kill untrasfected cells, had to be determined. To ensure use of minimal antibiotic concentration, a range of concentrations suggested by the manufacturer were tested. The number of viable Rat-1 cells was quantified by a colorimetric assay with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). As shown in figure 4.4.3 the optimal antibiotic concentration was determined at 5µg/ml for blasticidin and 500 µg/ml for zeocin.

A)
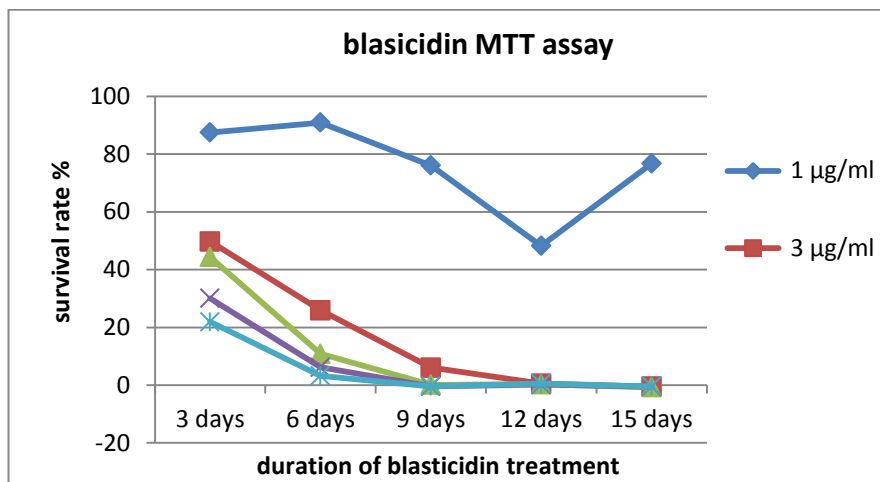
B)
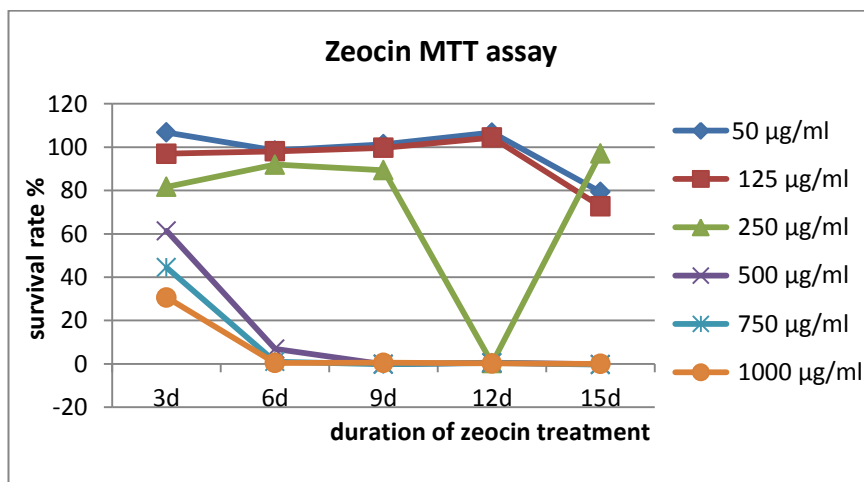
Figure 4.4.3 Viability assay to test minimum concentration of blasticidin and zeocin in untransfected Rat-1 cells. Cells were plated at 25% confluency in 12x 96 well plates (1 plate per time point per antibiotic) and treated with various concentrations for the indicated amount of time. Readings were taken after 0, 3, 6, 9, 12, or 15 days of treatment. The medium was changed every three days with fresh medium with the appropriate antibiotic concentration to compensate for loss of activity over prolonged incubation. A. Cell viability was tested at six different concentrations of blasticidin (0, 1, 3, 5, 7.5, 10 µg/ml). With 5 µg/ml of blasticidin the cell viability dropped to 44% and 10% after 3 and 6 days of treatment respectively. This concentration was used for selection after Rat-1 cells' transfection with pcDNA6/TR[©]. B. Cell viability was tested at six different concentrations of zeocin (1, 50, 125, 250, 500, 750 and 1000 µg/ml). With 500 µg/ml of zeocin the cell viability dropped to 55% and 7% after 3 and 6 days of treatment respectively, therefore this concentration was used for selection after Rat-1 cells' transfection with pcDNA4[TM]/TO vector.

Once the appropriate concentration for selection was determined, Rat-1 cells were transfected with the plasmid pcDNA^{TM}6/TR$^©$, which expresses both the Tet repressor and the blasticidin resistance gene. After treatment with 5µg/ml blasticidin for 4 weeks resistant clones were isolated, with cloning disks, expanded and screened for expression of the tet repressor by transfection of the pcDNA4_Fluc. Clones expressing the Tet repressor should only express firefly upon addition of doxycycline. Firefly should be expressed constitutively in expanded clones which do not express the tet repressor. Unfortunately firefly luciferase activity was not repressed in any of the tested clones (Figure 4.4.4), but was repressed in a positive control cell-line (T-Rex U20S) expressing the tetR.
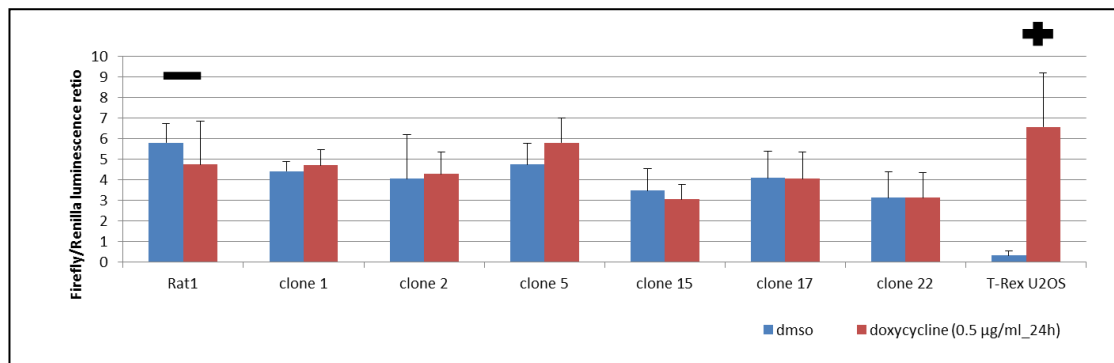
Figure 4.4.4 Activity of Tet repressor in Rat-1 cells stably transfected with pcDNA6.
Rat-1 cells stably transfected with pcDNA6 and selected with blasticidin (5µg/ml) were transiently transfected with pcDNA4_FLuc and a plasmid expressing Renilla luciferase. Untransfected Rat-1 cells and U2OS cells stably expressing the Tet repressor were used as negative and positive controls respectively.

One possible explanation for the lack of repression in the isolated Rat-1 clones is that the amount of the Tet repressor expressed in them was not sufficient to inhibit transcription of firefly luciferase in this cell line. In transient transfections high copy numbers of plasmid (in this case pcDNA4_Fluc) are delivered in the cells, while in stable transfections only a small number of the transfected plasmids gets integrated into the genome and thus low levels of proteins are produced. Another possible explanation for the observed lack of firefly repression is the total lack of Tet repressor expression due to unsuccessful integration of the Tet repressor encoding sequence into the genome. To rule out the possibility that the lack of suppression of firefly luciferase observed in the tested clones (Figure 4.4.4), was

attributed to high levels of transfected Firefly luciferase, a range of amounts of pcDNA4_FLuc were transfected in some of the clones.Figure 4.4.5 shows that there was still no repression.
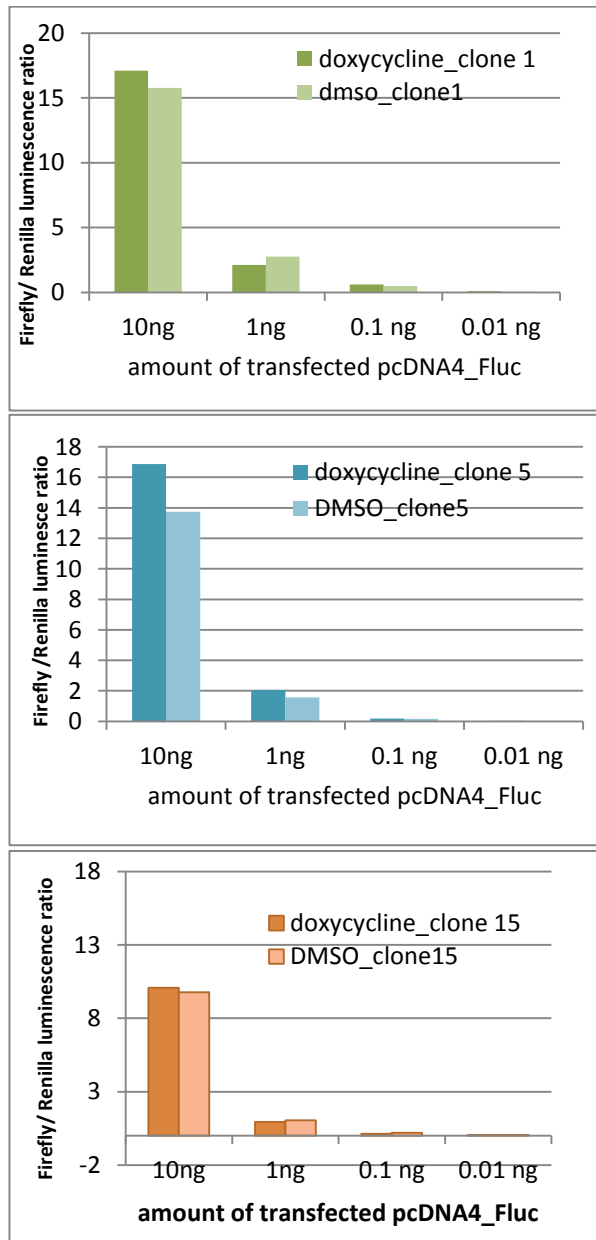
Figure 4.4.5 Efficiency of the tetracycline regulated expression system in Rat-1 cell lines stably transfected with the Tet repressor.
Three Rat-1 clones (clone 1, clone 5 and clone 15) stably transfected with the Tet repressor containing plasmid (pcDNA^{TM}6/TR), were transiently transfected with varying amounts (10ng to 0.01ng) of pcDNA4_FLuc and 2ng of a Renilla luciferase expressing plasmid, in 96 well plates.

To test if the lack of firefly luciferase suppression observed in the absence of doxycycline (Figure 4.4.4 and 4.4.5), was attributed to lack of integration of the Tet repressor sequence, we extracted genomic DNA from the clones and screened them with PCR (Figure 4.4.6) for presence of the Tet repressor. The clones screened with Tet repressor primers were negative (Figure 4.4.6). However, PCR with primers for the blasticidin resistance gene indicated the presence of the blasticidin gene thus explaining the resistance of clones to blasticidin. GAPDH was used as a positive control.
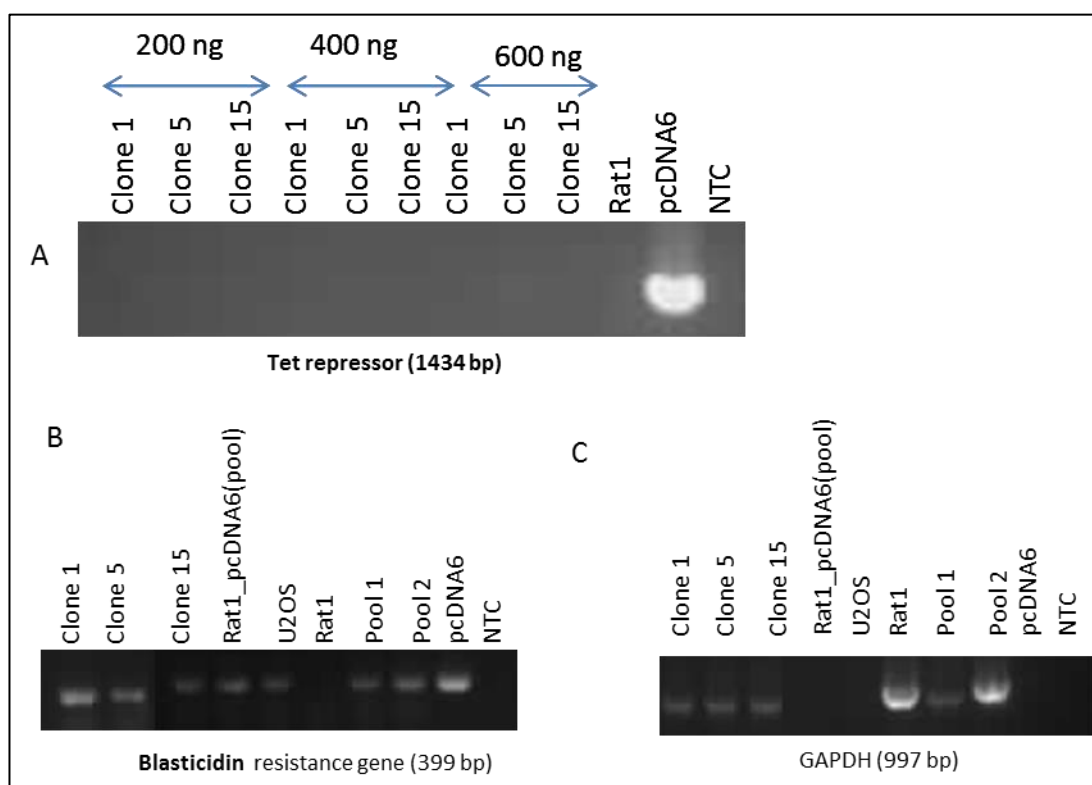
Figure 4.4.6 Detection of the Tet Repressor sequence in Rat-1 cell lines stably transfected with the pcDNA^TM6/TR vector.
A. Different amounts (200ng, 400ng and 600ng) of genomic DNA extracted from the stably transfected clones were used in PCR reactions. The Tet Repressor sequence (1434bp) was detected only in the positive control.
B. Detection of the Blasticidin resistance gene sequence and (C) the GAPDH housekeeping gene in Rat-1 cell lines stably transfected with the pcDNA^TM6/TR vector. pcDNA6: plasmid, Rat-1_pcDNA6 (pool)_1 and 2 are Rat-1 cells transfected with pcDNA^TM6/TR vector and selected with blasticidin, expected to contain a pool of Rat-1 pcDNA^TM6/TR vector stable cell lines.

Although the monoclones screened were all resistant to blasticidin, the Tet repressor sequence was not integrated to the genome (Figure 4.4.6), which

suggested that, unlike the blasticidin resistant gene, the Tet repressor gene had not been integrated into the Rat-1 cells genome.

Development of double stable cell lines is a lengthy process due to the amount of time required for selection and screening of clones. In an effort to expedite the process of stable cell line generation two different approaches were followed in parallel. In the first one (Figures 4.4.7 and 4.4.8), both plasmids were cotransfected, while in the second one (Figures 4.4.9 and 4.4.10) the plasmids were transfected successively like in the previously described experiments.

Rat-1 cells were transfected simultaneously with pcDNA^{TM}6/TR and either sT or sT.91-95A. After selection with blasticidin and zeocin, monoclonal populations were seleted by serial dilutions in 96-well plates. The activity of the Tet repressor in these clones was tested by measuring their Firefly luciferase activity after transient transfections with the pcDNA4_Fluc clone (Figure 4.4.7). This approach resulted in the identification of two clones (M_2 and sT_2) transfected with the mutant sT.91-95A and wt sT respectively, which expressed a functional Tet repressor (Figure 4.4.7).
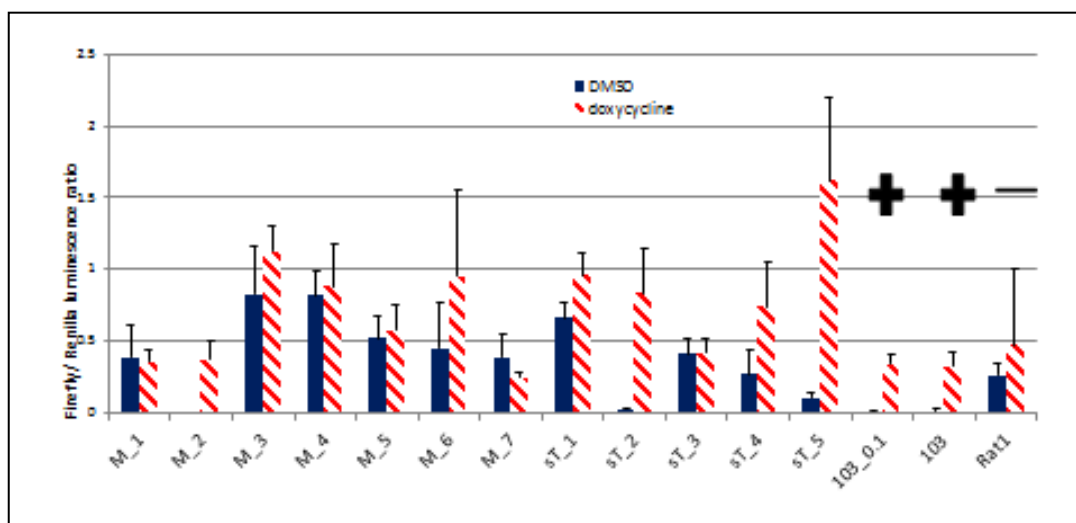
Figure 4.4.7 Dual luciferase assay for screening of Rat-1 monoclones transfected with two plasmids simultaneously (pcDNA4 and pcDNA6).
Rat-1 cells were cotransfected with the Tet repressor containing plasmid (pcDNA^{TM}6/TR) and the operator containing vector with sT (monoclones sT_1-5) or sT.91-95A (monoclones M_1-7). After simultaneous selection with blasticidin and zeocin, monoclones were screened for Tet repressor activity with a luciferase assay, as described previously. Rat-1 cells were used as negative controls. Clone 103 after a new round of serial dilutions (103_0.1) was used a positive control.

Next PCR was applied to these clones to screen for genomic integration of the Tet repressor and MCV sT (wild type or mutant). Consistently with the luciferase results of figure 4.4.7, PCR confirmed the presence of the Tet repressor (Figure 4.4.8). Although the Tet-O$_2$ and sT sequence was integrated into gDNA of the sT_2 clone, sT.91-95A was not integrated into the gDNA of the M_2 clone. Although the selection of clones generated by cotransfection of both plasmids gave promising results, the screening was not continued as the best fold induction (fold induction of M2: 0.5 and of sT_5: 1.5) was not as good as that of clone 103 (fold induction of 103:1.5), which was identified by transfection of pcDNA6^{TM}/TR by the approach of successive transfections described below (Figure 4.4.9). Furthermore, generating sequential double stable cell lines derived from a single clone would more likely result in cell lines that can induce the wild-type and the mutant sT antigens at comparable levels, thus eliminating a factor that could potentially result in phenotypic changes.
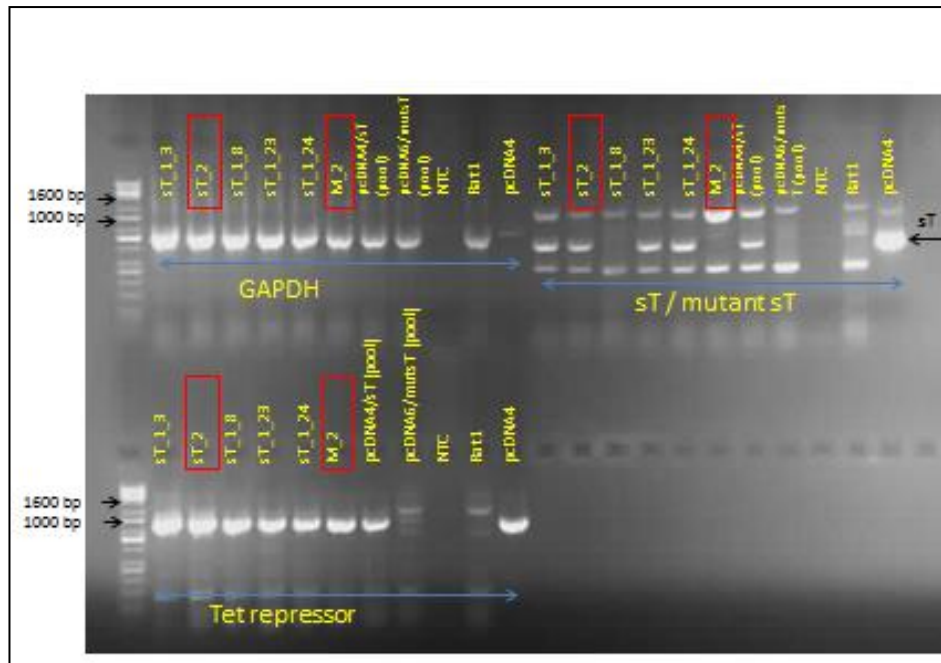
Figure 4.4.8 PCR from the genomic DNA of Rat-1 cell lines generated by cotransfection of the regulatory and response plasmids. All of the screened plasmids were positive for the Tet repressor. The set of primers (CMVF and Bghrev) used for detection of the sT antigen would generate products of 888 bp. GAPDH was used as a positive control.

Next we tried a different approach for selection. We transfected a new stock of Rat-1 cells with pcDNA6™/TR and after blasticidin selection, monoclonal populations were selected by serial dilutions in 96-well plates down to less than one cell per well. Only clones from wells that had one colony were screened. 108 monoclones were transiently transfected with the pcDNA4_Fluc plasmid and screened with dual luciferase assays for Tet repressor activity (Figure 4.4.9). Most of the screened clones showed no repression of firefly luciferase in the absence of doxycycline, however 19 clones showed varying levels of repression and induced expression of firefly luciferase in the absence and upon doxycycline addition respectively. Genomic integration of the Tet repressor and the blasticidin resistance gene was tested in these clones by PCR (Figure 4.4.10). All of the clones, except for one, were positive for genomic DNA integration of the blasticidin resistance gene and Tet repressor. Clone numbers 103 and 104 repeatedly showed the tightest regulation of firefly luciferase regulation upon doxycycline addition and both tested positive for Tet repressor and blasticidin integration (Figure 4.4.10).
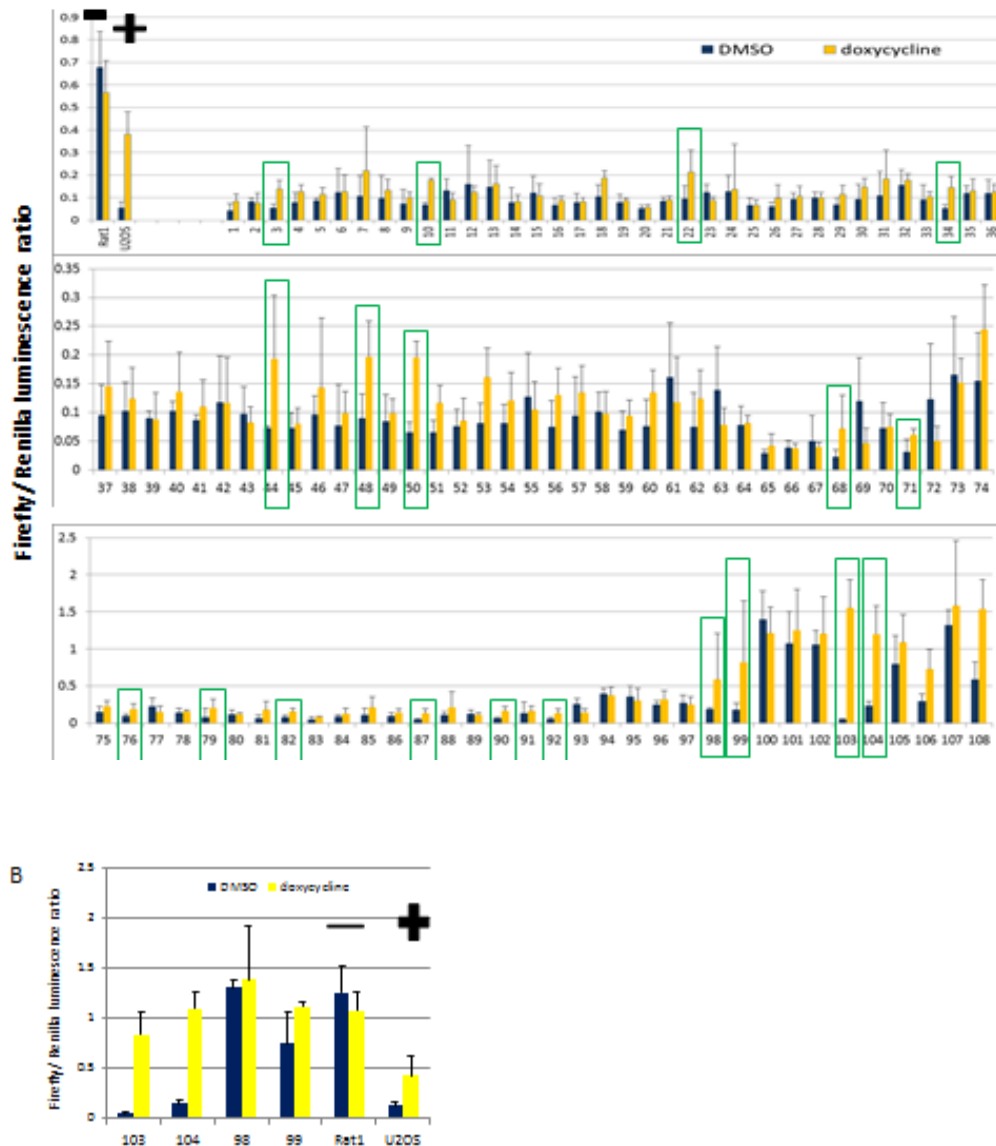
Figure 4.4.9 Dual luciferase screening of pcDNA^{TM}6/TR stably transfected Rat-1 clones. (A) Rat-1 were transiently cotransfected with both the operator containing (luciferase cloned in pcDNA4) and repressor containing (pcDNA^{TM}6/TR) plasmids. The cells were cultured in the presence of doxycycline or DMSO and after 24 hours, luciferase expression was measured. Green boxes indicate the clones that showed low background Tet expression levels and high induced expression following a 24hour treatment with 0.5 µg/ml doxycycline. (B) Selected clones that showed adequate Tet repressor activity were repeatedly assessed with the luciferase assay to confirm reproducibility. Clones 98 and 99 did not show repression of Fluc activity in absence of doxycycline. Clones 103 and 104 showed consistent induction of Firefly luciferase expression upon doxycycline addition. Repression of Firefly luciferase was better for clone 103 than for clone 104. (Firefly values were normalized for transfection efficiency using Renilla luciferase).
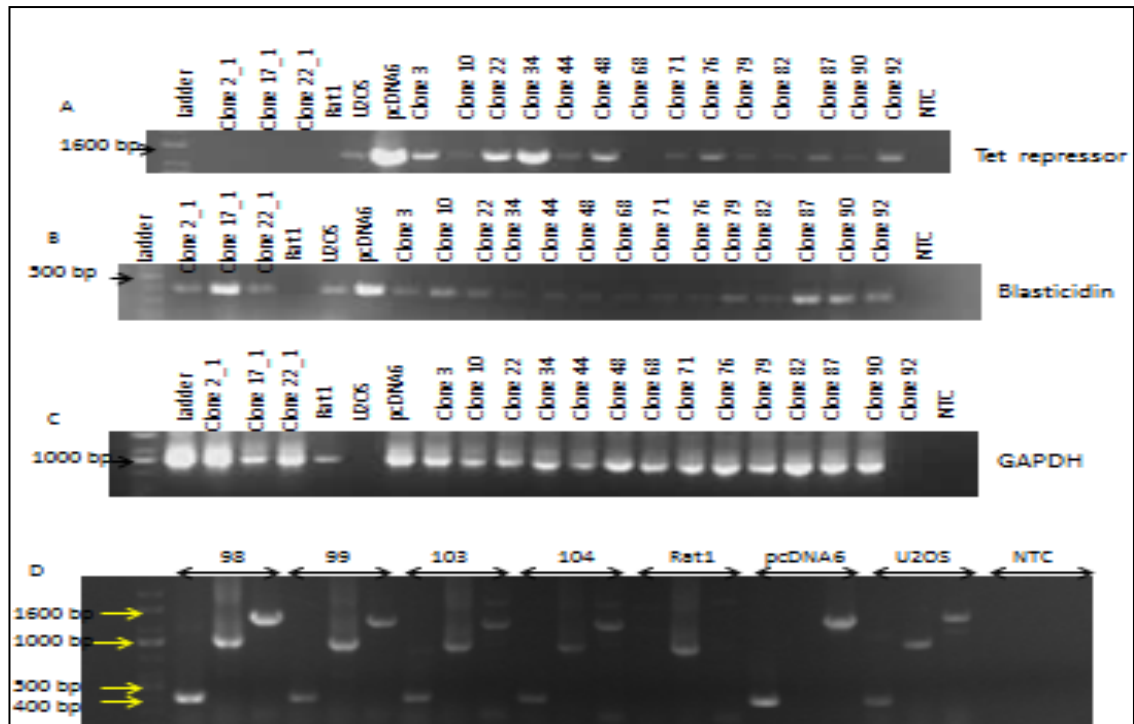
Figure 4.4.10 Genomic integration of the blasticidin resistance and the Tet repressor gene. Clones screened with the dual luciferase system for doxycycline induced Tet repressor expression were screened with PCR for genomic integration of the Tet repressor gene (B and D) and the blasticidin resistance gene(A and D).(C and D) The GAPDH housekeeping gene was used as a DNA template control.

To find the optimal conditions of doxycycline induction (concentration and duration of induction), clones 103 and 104 were transiently transfected with the pcDNA4_Fluc plasmid and treated with a range of doxycycline concentrations (0.1-1 µg/ml) for 12-47 hours. No significant differences in the expression levels of the Tet repressor were observed in the various conditions tested. Consistently with the conditions used already, treatment with 0.5 µg/ml doxycycline for 24 hours was chosen for the next experiments (Figure 4.4.11).
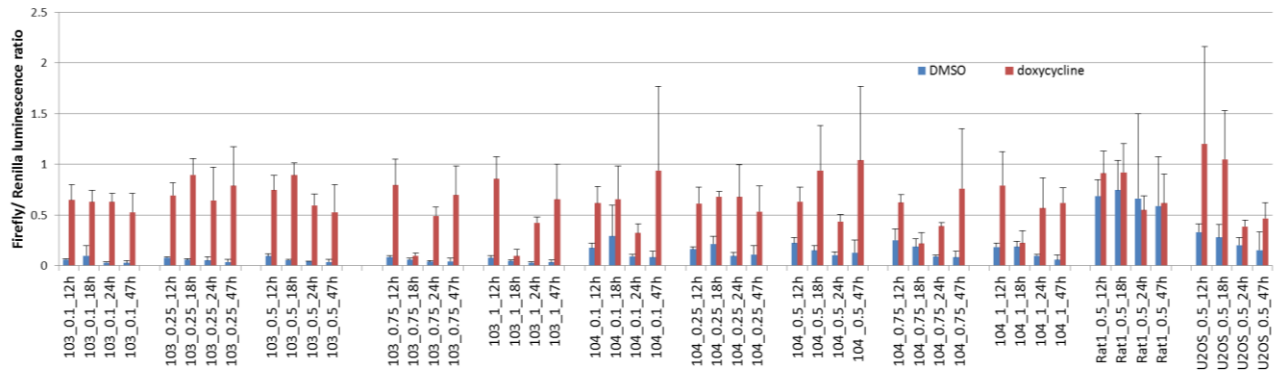
Figure 4.4.11 Conditions of doxycycline treatment for the Tet inducible system.
pcDNA™6/TR stably transfected Rat-1 clones were transiently transfected with pcDNA4_Fluc and a Renilla luciferase plasmid and treated for 12, 18, 24 or 47 hours with 0.1, 0.25, 0.5, 0.75 or 1 µg/ml doxycycline (or DMSO). Experiment was done in triplicate and error bars represent standard deviations. Firefly values were normalized for transfection efficiency using Renilla luciferase.

Clone 103 was chosen to serve as a host for inducible expression of the wild type and mutant sT, because of its high Tet repressor activity upon doxycycline induction and its low background levels. It was transfected with either the empty pcDNA™4/TO vector, or wild type MCV sT, or MCV sT.91-95A.

The transfected cells were treated with zeocin for 15 days and monoclonal populations were selected by serial dilutions in 96-well plates down to less than one cell per well. To identify clones in which MCV sT or sT.91-95A and the upstream Tet-O_2 sequence had been successfully integrated into the genome, clones were screened by PCR on genomic DNA with sT-spanning primers. Amplicons of the expected size were purified and verified by sequencing. Inducible expression of the sT or sT.91-95A genes was further confirmed by Western blotting (Figure 4.4.12) with anti-panT.

The double stable cell lines were developed in an effort to create an experimental system appropriate for the study of the effect of sT on protein synthesis by ribosome profiling. Random integration of sequences into the genome can cause insertional mutations or disrupt essential genes at the site of insertion (Woods et al. 2003). To account for any differences in gene expression in the double stable cell lines caused by random integration, as opposed to sT antigen expression, three clones were identified from each sT antigen expressing double stable cell line (Figure 4.4.12). All of them were able to express the sT antigen.

207
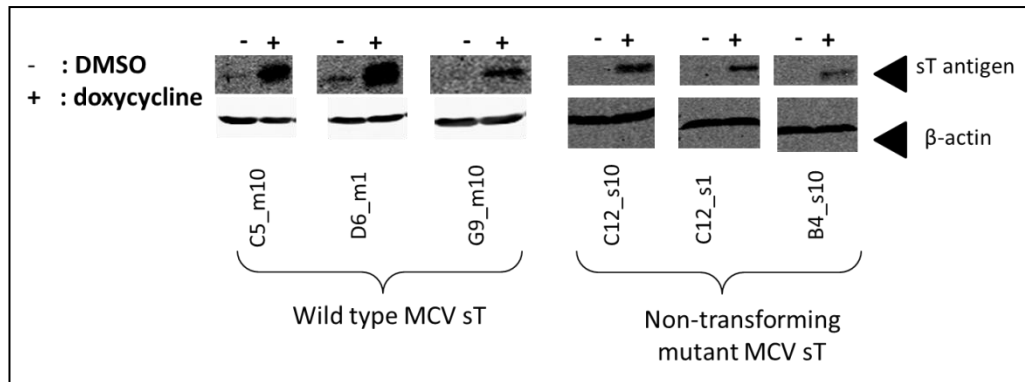
Figure 4.4.12 SDS-PAGE analysis of inducibly expressed sT and sT.91-95A antigens.
Following treatment of double stable Rat-1 cell lines with 0.5 µg/ml doxycycline (or DMSO) for 24 hours. Immunoblot analysis was performed with the mouse anti-panT antigen detecting monoclonal antibody CM8E6 (1:250).

Assessment of the transformation ability of the double stable Rat-1 cell lines

To assess the cellular transformation ability of the established double stable Rat-1 monoclones, soft agar colony formation assay was applied. The parental Rat-1 cell line and a double stable Rat-1 monoclone transfected with the pcDNA[TM]6/TR and pcDNA[TM]4/TO vectors, were used as negative controls (Figure 4.4.13). The cell line (C12_s1) expressing the mutant sT sT.91-95A antigen, had a comparable number of colonies as the wild type antigen (C5_m10), but formed bigger colonies than the wild type sT expressing cell line. Rat-1 cells transfected with the empty vector (pcDNA4) showed greater transformation efficiency than the MCV sT expressing cells.
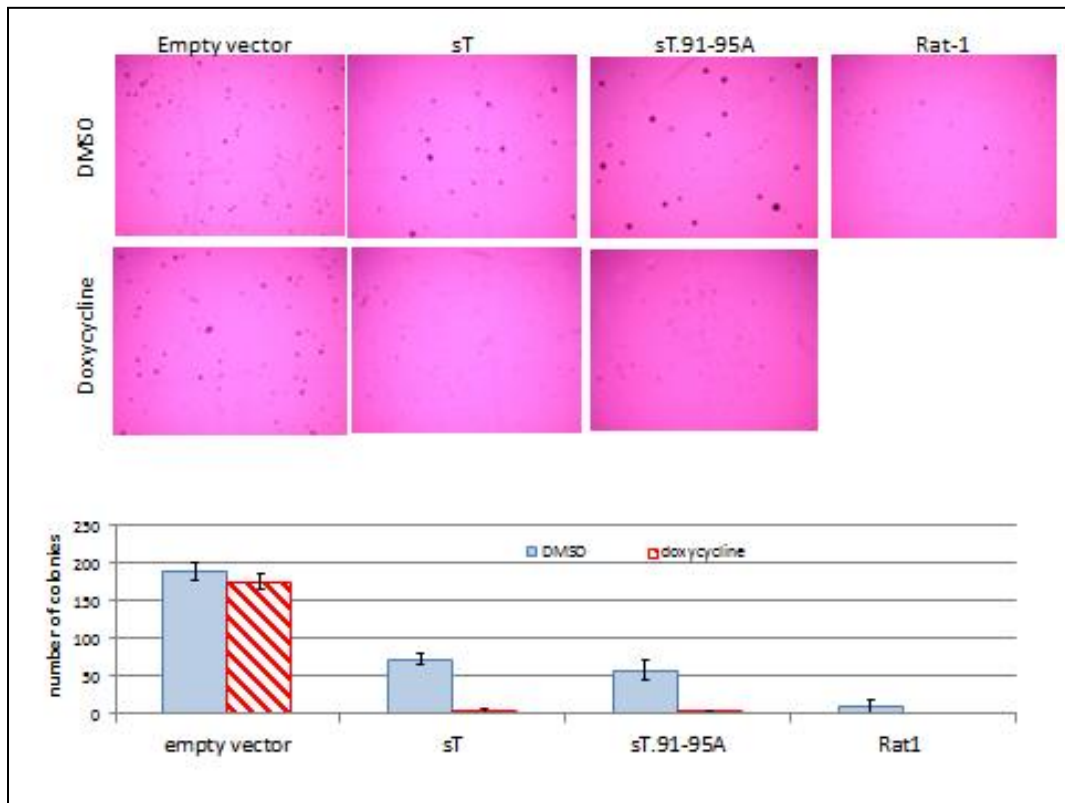
Figure 4.4.13 Induced expression of MCV sT does not transform Rat-1 cells. Soft agar assay of Rat-1 cells induced to express MCV (top). Colonies observed in 6-well triplicates were counted to determine average ± SD colonies per well (bottom).

Shuda et al (2011) had previously reported that Rat-1 cells transduced to express MCV sT were transformed. In our model system expression of wt sT failed to transform Rat-1 cells. One possible explanation for the discrepancy observed between the two systems, is that the levels of sT protein expression may be different. The levels of the induced protein can be regulated in some inducible systems, by adjustment of doxycycline levels (Shaikh & Nicholson 2006). Treating sT expressing Rat-1 cells with increasing concentrations of doxycycline did not result in altered levels of protein expression (Figure 4.4.14), in agreement with the luciferase assay in figure 4.4.11. Therefore, since the stable Rat-1 cell lines generated here could not be transformed by sT induction they could not be used for further analysis by ribosome profiling.
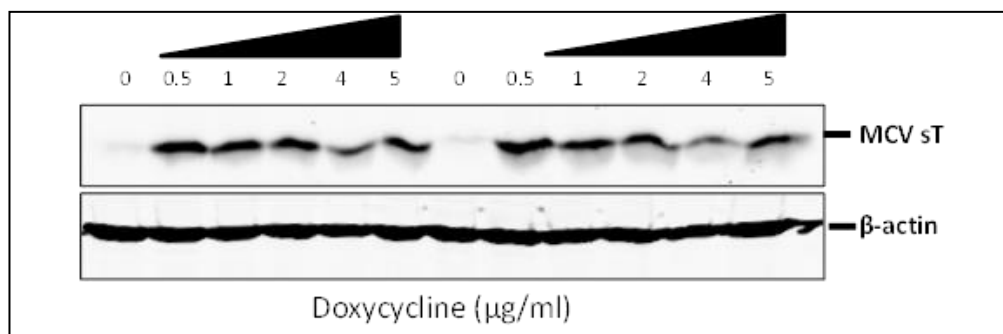
Figure 4.4.14 Induction of MCV sT in inducible Rat1 cells with various levels of doxycycline. The Rat-1 cell line stably and inducibly expressing sT (C5_m10) was treated with increasing concentrations of doxycycline for 24 hours and the lysates were analysed by western blot. No increase was obsereved in MCV sT expression with increasing amounts of doxycycline from duplicate experiments.

Ribosome profiling analysis on MKL-1 cells

MCV sT has been shown to induce and maintain hyperphosphorylation of 4E-BP1, which leads to cap-dependent translation dysregulation (Shuda et al. 2011). We wanted to study the potential effect of this interaction in tumour development and metastatic progression, by applying ribosome profiling in the MCV positive metastatic MKL-1 cell line, that was transduced with an sT targeting or control scrambled shRNAs.

First, MKL-1 cells transduced with sT and scrambled shRNAs (Figure 4.4.15 A and B) were analysed for sT knockdown by western blotting (Figure 4.4.15C). Western analysis showed decrease of the steady state levels of sT, but ribosome

profiling reads aligning to the viral genome, confirmed that silencing of MCV sT was effective (Figure 4.4.15C and 4.4.16A). A Z-score of 4 was used as an arbitrary threshold of statistical significance for differentially regulated genes (Table 2.1 and 2.2).

It had previously been reported that the mRNAs that are translationally upregulated by mTOR, have long 5'UTRs with secondary structures (Koromilas et al. 1992), or possess 5' terminal oligopyrimidine tracts (Hsieh et al. 2012). We did not observe translational upregulation in any of these categories, but only in gene SIRT7 instead (Figure 4.4.16B and Tables 4.1 and 4.2).
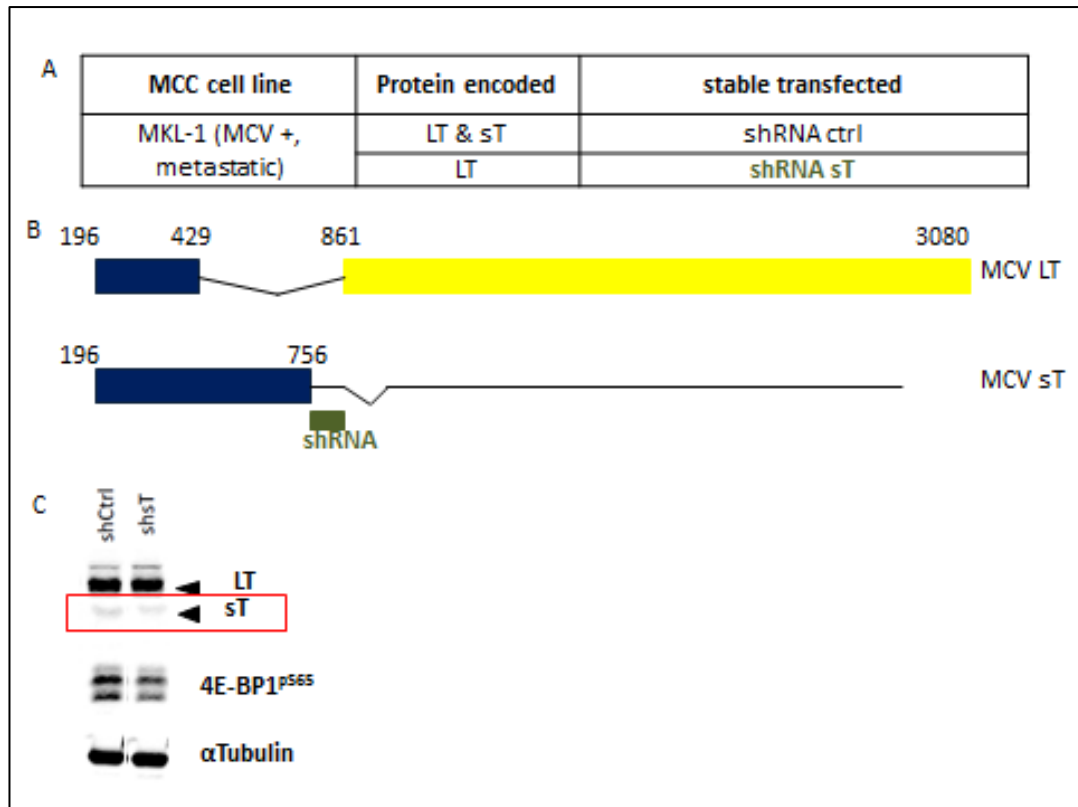
Figure 4.4.15 Knockdown of MCV sT with shRNA
(A) MKL-1 cells stably transduced with shRNA against sT or scrambled RNA
(B) Schematic representation of the sequence targeted by sT shRNA
(C) Western blot of MKL-1 cells after shRNA transduction with sT or scrambled shRNA
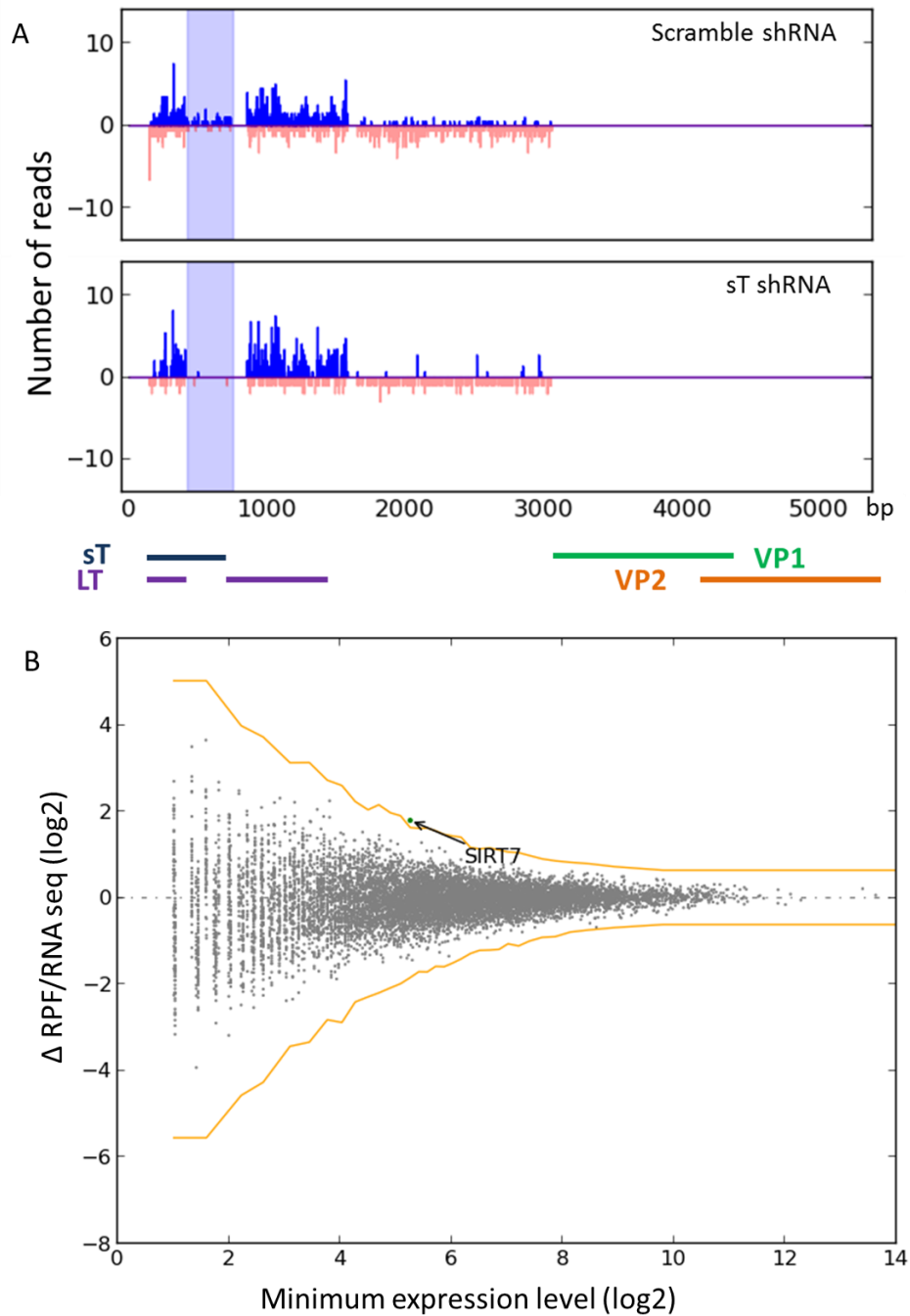
Figure 4.4.16 Analysis of differential translational regulation upon knockdown of MCV sT with shRNA.

(A)The plots show the alignments of ribo-seq (in blue) and rna-seq (in pink) reads to the positive strand of the viral genome. The MCV sT specific region is highlighted in light blue. Unlike the number of RNA-seq reads, the number of Ribo-seq reads is reduced upon transduction with sT shRNA.

(B) Differential gene expression analysis. Scatter plot compare translation efficiency (TE) between shRNA against sT or scrambled RNA. The threshold used to denote differentially expressed genes (Z-score of 4) is indicated in orange.

Table 4.4.1 List of translationally upregulated genes

Translationally upregulated genes upon MCV sT silencing	Z-score
SIRT7	4.4
C19orf10	3.4
GNMT	3.3
SMAD9	3.3
COPS7B	3.2
LGALS3BP	3.2
NQO1	3.2
GPI	3.1
TRAIP	3.1
HCFC1R1	3.1
SARS2	3.0
PCDHGB6	3.0
ATP6V0A1	3.0

Table 4.4.2 List of translationally downregulated genes.

Translationally downregulated genes upon MCV sT silencing	Z-score
TUBA1B	-3.0
SIRT1	-3.0
HIST2H2BE	-3.1
JUND	-3.1
IRX2	-3.1
SPTBN1	-3.3
RIC8B	-3.4
TMSB4X	-3.5
SPTAN1	-3.8

In ribosome profiling RNAseq is used as a control for internal RNA levels. Therefore the technique enables study of the changes that occur at the transcriptional level. As described previously (Chapter 3), differentially expressed genes were identified upon z-score transformation. Gene Ontology and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of transcriptionally upregulated and downregulated genes was performed with DAVID (Huang et al. 2009), to investigate potential over-representation of biological processes. A total of 279 transcriptionally upregulated (z $\geq$2) (Table 4.3) and 240 downregulated (z $\leq$ -2) (Table 4.4) genes, were used in GO term analysis. None of these groups revealed enrichment in any functional pathways.

4.5 Discussion

Transformation of primary human cells in tissue culture, requires minimally two oncogenes (SV40 LT and an oncogenic allele of H-ras) along with the hTERT gene that has telomerase activity (Hahn et al. 1999). Similarly, transformation of primary rodent cells normally requires at least two oncogenes (Land et al. 1983; Ruley 1983). However, Rat-1 cells are one of the few cell lines that can be transformed by a single oncogene and because of this have been extensively used in cancer studies. One theory is that inactivation of p21WAF1/CIP1, due to methylation of its promoter, might be responsible for the susceptibility of the p53 wild-type Rat-1 cells to transformation (Allan et al. 2000).

In a 2011 study from Chang and Moore's lab, it was shown that MCV sT is sufficient to transform Rat-1 fibroblasts and that it can promote 4E-BP1 hyperphosphorylation and subsequent cap-dependent translation initiation, independently of mTOR kinase activity (Shuda et al. 2011). To study the effect of MCV sT on protein synthesis at various stages of MCV-induced MCC development and progression we aimed to use ribosome profiling, because it provides insight into the dynamics of protein synthesis and regulation. To study the effect of MCV sT at the initial steps of MCV sT induced transformation we aimed to establish an inducible (Tet-on) system stably expressing MCV sT. An MCV sT (sT.91-95A) mutant, which could not transform Rat-1 cells (Kwun et al. 2013), was used as a negative control.

Tet-off systems require continuous presence of tetracycline or doxycycline, which has a half-life of 48 hours in tissue culture medium, and therefore needs to be frequently replenished, thus we opted for a Tet-on system. We used doxycycline as an inducing agent for the T-REx^{TM} system, instead of tetracycline, because it has longer half-life (48 vs 24 hours). As shown by the repression and increase of firefly expression in the absence and addition of doxycycline respectively, transfection efficiency is higher for HEK 293T, than Rat-1 cells (Figure 4.4.2), but the Trex system is functional for the Rat-1 cells too, so we decided to proceed with this system. Cotransfecting the Tet repressor and the inducible expression vector at the same time is time saving, but in the Rat-1 cells we found very few monoclones with high Tet repressor activity. Although we tried both approaches; cotransfecting and

transfecting the Tet repressor and the inducible expression vector successively, we decided to proceed with the second one. Although the exact mechanism of naked DNA integration into chromosomal DNA has not been fully elucidated, integration in mammalian cells usually occurs at random sites with nonhomologous recombination (Di Primio et al. 2005). This can result in integration of parts of the plasmid, partial deletion of the integrated or the genomic sequence or integration into heterochromatic regions. Initial transfection of Rat-1 cells with the Tet repressor plasmid and subsequent antibiotic selection resulted in clones that were resistant to blasticidn but did not express the Tet repressor (Figures 4.4.4 and 4.4.6). Failure to detect the Tet repressor by PCR (Figure 4.4.6), suggests that the Tet repressor sequence or parts of it were not integrated in the gDNA of the tested clones. Following a second round of transfections and isolation of clones through serial dilutions the clone 103 was identified. This clone was consistently showing the highest Tet repressor activity. Using this clone to make stable cell lines with sT, sT.91-95A or the empty vector, eliminates variations observed in the double stable clones attributed to differential Tet repressor integration and activity.

Numerous clones were screened in an effort to find those that would be resistant to the selection marker and would have adequately high expression of our genes of interest (either sT or sT.91-95A) and retain most of the parental phenotype, due to random integration. To eliminate the effect of phenotypes attributed to the site of integration rather than the expression of the genes of interest, three clones stably expressing either sT (C5_m10, D6_m1 and G9_m10) or sT.91-95A (B4_s10, C12_s1 and C12_s10) (Figure 4.4.12), were selected for further analysis.

In contrast to the findings of our collaborators in the Moore and Chang lab, the soft agar colony formation assay (Figure 4.4.13) showed that expression of sT inhibits cell proliferation, a result supported by the observation of increased cell death in stable cell lines inducibly expressing MCV sT upon doxycycline addition. This discrepancy could be attributed to different sT levels in the systems used in the two labs. High levels of Ras activation induces proliferative arrest *in vivo* in mammary gland, while lower levels of expression result in hyperproliferation (Sarkisian et al. 2007). Using various concentrations of doxycycline in our inducibly

expressing MCV sT cells did not result in increasing levels of sT (Figure 4.4.14), possibly because even the lowest doxycycline concentration was sufficient to saturate the tet-on system used. Lack of the transformed phenotype in these cells, regardless of the status of phosphorylation in 4E-BP1, made them inappropriate for analysis with ribosome profiling.

Although the lack of a transformed phenotype is most likely attributed to inappropriate levels of MCV sT expression, the different origin of Rat-1 cells, or different tissue culture practices, could result to Rat-1 cell lines with different genetic backgrounds, which thus respond differently to the expression of an oncogene.

To study the potential effect of MCV sT induced protein synthesis on fully transformed MCC cells, we applied ribosome profiling on an MCV positive cell line derived from a nodal metastasis (MKL-1), which was transduced with sT shRNA or scrambled RNA. Profiling results showed the expected knockdown of MCV sT in cells transduced with sT targeting shRNA. Ribosome profiling data analysis (Figure 4.4.16B) did not suggest significant MCV sT induced translational regulation. Data analysis did not reveal an overall translation upregulation, which would be expected as a result of cap dependent translation initiation. Furthermore neither TOP and TOP-like mRNA transcripts (Thoreen et al. 2012), nor GC-rich and highly structured mRNAs were upregulated (Hay & Sonenberg 2004). The only gene that was statistically significantly translationally upregulated in our data was SIRT7.

Sirt7 is a member of the sirtuins family, which consists of seven members (Sirt1-Sirt7) and is characterised by the presence of an NAD+ binding domain and a catalytic domain. Sirt7 is the least studied member of the sirtuins family. It localises mainly in the nucleolus (Michishita et al. 2005) and it is involved in lipid metabolism (Shin et al. 2013; Ryu et al. 2014; Yoshizawa et al. 2014), chromatin remodelling (Tsai et al. 2012), protein synthesis (Tsai et al. 2014) and cellular survival (Vakhrusheva et al. 2008; Kiran, Oddi, et al. 2015). Sirt7 is so far known to have a NAD-dependent deacetylase activity, but none of its currently known substrates can explain its involvement in the above mentioned cellular functions (Kiran, Anwar, et al. 2015). Sirt7 is known to be dysregulated in various types of cancers (Geng et al. 2015; Kim et al. 2013; Yu et al. 2014; Lai et al. 2013; Frye 2002; Ashraf

et al. 2006), which suggests that it could have an effect on MCC too. The mechanism causing Sirt7 induced oncogenesis is not known, although some views have been proposed (Li & Bhatia 2013). It was originally suggested that Sirt7 can promote tumourigenesis by deacetylating the tumour suppressor p53, but p53 was eventually found to be a poor substrate for Sirt7 (Vakhrusheva et al. 2008). Sirt7 induced oncogenic transformation, has been linked to its ability to deacetylate H3K18Ac (Barber et al. 2012). In a different study Sirt7 knockdown inhibited liver cancer cell growth both *in vitro* and in mouse models. Sirt7 overexpression has been suggested to induce mitotic stimulation through transcriptional inactivation of p21$^{WAF/Cip1}$ and activation of cyclin D1 in hepatocellular carcinoma (Kim et al. 2013). To confirm ribosome profiling results and study the potential role of Sirt7 in MCC, the observed Sirt7 increase should be confirmed by western analysis, in MKL-1 cells transduced with sT targeting shRNA. Experimental confirmation of these results in other MCV positive MCC cell lines (BroLi, WaGa, MS-1 and MKL-2) (Houben, Shuda, Weinkam, Schrama, et al. 2010), along with absence of the effect in MCV negative MCC cell lines (UISO, MCC13 and MaTi) (Houben, Shuda, Weinkam, Feng, et al. 2010), would suggest Sirt7 upregulation as a virally induced specific transformation mechanism in MCC cells. Myb binding protein 1a (Mybbp1a) has been proposed to be a negative regulator of Sirt7 (Karim et al. 2013). The mitotic kinase aurora b, which is a target of APC/C (Floyd et al. 2008), phosphorylates Mybbp1a at Ser1303, which is one of the sites of Mybbp1a that are found to be phosphorylated in mitotic cells (Perrera et al. 2010).

MCV sT was recently shown to inhibit APC/C, thus leading to stabilisation of Aurora b (Shuda et al. 2015). It is therefore likely that MCV sT inhibits expression of Sirt7, through its effect on APC/C, but further experimental confirmation of these interactions is required.

These results arise from the analysis of one biological replicate. At least two biological replicates with appropriate data correlation would be required, to give us sufficient confidence in the accuracy of these results. However, because these preliminary results indicate little regulation at the transcriptional or translational level upon induction of the shRNA directed toward sT, futher experiments where not attempted.

One possible explanation for the lack of translational upregulation upon the knockdown of sT, is that MKL-1 cells are fully transformed metastatic cells. It is likely that sT has different roles at various stages of MCV induced transformation and that the sT-4E-BP1 interaction is important for the initial stages of transformation, but might not be essential in fully transformed cells (Angermeyer et al. 2013; Angermeyer et al. 2014; Shuda et al. 2014).

It was recently shown (Shuda et al. 2015) that MCV sT targets and inhibits APC/C, and thus induces mitosis (Townsley & Ruderman 1998), which leads to increased CDK1/CYCB1 activity and results in 4E-BP1 phosphorylation (Shuda et al. 2015). The sT_shRNA transduced MKL-1 cells, which were analysed by ribosome profiling, were not synchronised, therefore only a small part of the total cell population was expected to be in mitosis. Lack of a synchronised population, could have masked the translational effect of CDK1 induced 4E-BP1 inactivation, which might account for the lack of translational upregulation in our ribosome profiling results.

The RNAseq expression analysis did not reveal enrichment of any molecular pathway upon downregulation of MCV sT. However sT downregulation resulted in transcriptional upregulation of genes either exclusively expressed in the nervous system (CNTN2) (fold increase: 2.73, z-score:5.55) (Uhlen et al. 2015) or predominantly expressed in neuroendocrine tissues. CHST8 (fold increase: 4.76, z-score: 5.24) is predominantly expressed in the pituitary gland (Okuda et al. 2000), CPLX2 (fold increase: 3.16, z-fold: 4.96) is mainly expressed in the brain and in endocrine tissues (Uhlen et al. 2015) and has a well-defined role in synaptic exocytosis (Kurokawa et al. 2015). STXBP1 (fold increase: 2.02, z-score: 4.14) is involved in the release of neurotransmitters (Patzke et al. 2015). The enrichment observed in genes primarily expressed in neural tissues or associated with the release of neurotransmitters, in these MCC cells that have characteristics of presynaptic cells (Haeberle et al. 2004; Hitchcock et al. 2004; Nunzi et al. 2004), suggests a possible role of MCV sT in reversing cellular differentiation, which is one of the hallmarks of cancer (Hanahan & Weinberg 2011).

In Grundhoff's lab an *in vitro* culturing system was established to facilitate the study of viral replication. In this model system the neuroectodermal tumour cell

line PFSK-1 was transfected with a synthetic MCV genome (Neumann et al. 2011). RNaseq analysis has subsequently been performed on this system, but the analysis focused primarily on the RNA transcripts of the viral genome rather than the host cell line (Theiss et al. 2015). A different group performed RNAseq analysis but focused on the mutations rather than on the levels of gene expression. This study identified mutations in TP53, RB1, PIK3CA consistently with previously available data and discovered activating mutations including HRAS, PRUNE2 and the NOTCH family genes in MCV negative tumours. RNAseq analysis of MCV positive tumours revealed a low mutation pattern (Harms et al. 2016).

The aim of this project was to study the effect of MCV sT induced protein synthesis on cellular transformation. We were not able to detect a causal role of MCV sT on Rat-1 cells' transformation in our model system. Data published following the completion of our ribosome profiling experiments in MKL-1 cells, suggested that MCV sT-induced protein synthesis dysregulation is primarily taking place in mitotic cells (Shuda et al. 2015). MCV is a newly discovered virus (Feng et al. 2008). More experiments are required to elucidate the role of each MCV oncogene in tumourigenesis and characterise the molecular mechanism of oncogenic transformation in Merkel cells.

4.6 Bibliography

Abend, J.R. et al., 2009. A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. *The Journal of General Virology*, 90(Pt 5), pp.1238–1245.

Agelli, M. & Clegg, L.X., 2003. Epidemiology of primary Merkel cell carcinoma in the United States. , (Mcc), pp.20–22.

Aird, D. et al., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome biology*, 12(2), p.R18.

Allan, L.A. et al., 2000. The p21(WAF1/CIP1) promoter is methylated in Rat-1 cells: stable restoration of p53-dependent p21(WAF1/CIP1) expression after transfection of a genomic clone containing the p21(WAF1/CIP1) gene. *Molecular and cellular biology*, 20(4), pp.1291–1298.

Allander, T. et al., 2007. Identification of a third human polyomavirus. *Journal of virology*, 81(8), pp.4130–4136.

Angermeyer, S. et al., 2013. Merkel Cell Polyomavirus-Positive Merkel Cell Carcinoma Cells Do Not Require Expression of the Viral Small T Antigen. *J Invest Dermatol*, 133(8), pp.2059–2064.

Angermeyer, S. et al., 2014. Response to Shuda et al. *J Invest Dermatol*, 134(5), pp.1481–1482.

Armengol, G. et al., 2007. 4E-binding protein 1: a key molecular "funnel factor" in human cancer with clinical implications. *Cancer research*, 67(16), pp.7551–7555.

Arora, R. et al., 2012. Survivin Is a Therapeutic Target in Merkel Cell Carcinoma. , 56.

Ashraf, N. et al., 2006. Altered sirtuin expression is associated with node-positive breast cancer. *British journal of cancer*, 95(8), pp.1056–1061.

Barber, M.F. et al., 2012. SIRT7 links H3K18 deacetylation to maintenance of oncogenic transformation. *Nature*, 487(7405), pp.114–118.

Becker, J.C. et al., 2008. MC Polyomavirus Is Frequently Present in Merkel Cell Carcinoma of European Patients. *J Invest Dermatol*, 129(1), pp.248–250.

De Benedetti, A. et al., 1994. CHO cells transformed by the translation factor elF-4E display increased c-myc expression, but require overexpression of Max for tumorigenicity. *Molecular and Cellular Differentiation*, 2(4), pp.347–371.

De Benedetti, A. & Graff, J.R., 2004. eIF-4E expression and its role in malignancies and metastases. *Oncogene*, 23(18), pp.3189–3199.

Berkel, H.J. et al., 2001. Expression of the translation initiation factor eIF4E in the polyp-cancer sequence in the colon. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 10(6), pp.663–666.

Blumberg, B.S. et al., 1965. A "New" Antigen in Leukemia Sera. *JAMA: The Journal of the American Medical Association*, 191(7), p.541. Available at: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1965.03080070 025007 [Accessed May 26, 2016].

Borchert, S. et al., 2014. High-Affinity Rb Binding, p53 Inhibition, Subcellular Localization, and Transformation by Wild-Type or Tumor-Derived Shortened Merkel Cell Polyomavirus Large T Antigens M. J. Imperiale, ed. *Journal of Virology*, 88(6), pp.3144–3160.

Boshartb, M. et al., 1984. of papillomavirus. , 3(5), pp.1151–1157.

Boulais, N. & Misery, L., 2007. Merkel cells. *Journal of the American Academy of Dermatology*, 57(1), pp.147–165.

Bouvard, V. et al., 2012. Carcinogenicity of malaria and of some polyomaviruses. *Lancet Oncology*, 13(4), pp.339–340.

Boyapati, A. et al., 2003. SV40 17KT antigen complements dnaJ mutations in large T antigen to restore transformation of primary human fibroblasts. , 315, pp.148–158.

Buck, C.B. et al., 2012. Complete genome sequence of a tenth human polyomavirus. *Journal of virology*, 86(19), p.10887. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22966183 [Accessed May 11, 2016].

Carter, J.J. et al., 2009. Association of Merkel Cell Polyomavirus – Specific Antibodies With Merkel Cell Carcinoma. , 101(21).

Carter, J.J. et al., 2013. Identi fi cation of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes. , 110(31).

Castellvi, J. et al., 2006. Phosphorylated 4E binding protein 1: a hallmark of cell

signaling that correlates with survival in ovarian cancer. *Cancer*, 107(8), pp.1801–1811.

Chang, Y. et al., 1994. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science (New York, N.Y.)*, 266(5192), pp.1865–1869.

Chang, Y. & Moore, P.S., 2012. Merkel Cell Carcinoma: A Virus-Induced Human Cancer. *Annual review of pathology*, 7, pp.123–144.

Cheng, J. et al., 2009. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens. *Seminars in cancer biology*, 19(4), pp.218–228.

Cheng, J. et al., 2013. Merkel cell polyomavirus large T antigen has growth-promoting and inhibitory activities. *Journal of virology*, 87(11), pp.6118–6126.

Choo, Q.L. et al., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science (New York, N.Y.)*, 244(4902), pp.359–362.

Dalianis, T. & Hirsch, H.H., 2013. Human polyomaviruses in disease and cancer. *Virology*, 437(2), pp.63–72.

Daniels, R., Sadowicz, D. & Hebert, D.N., 2007. A very late viral protein triggers the lytic release of SV40. *PLoS pathogens*, 3(7), p.e98.

DeCaprio, J.A., 2009. How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. *Virology*, 384(2), pp.274–284.

Diaz, J. et al., 2014. Phosphorylation of large T antigen regulates merkel cell polyomavirus replication. *Cancers*, 6(3), pp.1464–1486.

Durst, M. et al., 1983. A papillomavirus , DNA- from a cervical carcinoma and-its -prevalence in cancer biopsy samples from different geographic regions. , 80(June), pp.3812–3815.

Eckhart, W., Hutchinson, M.A. & Hunter, T., 1979. An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. *Cell*, 18(4), pp.925–933.

Elphick, G.F., Querbes, W. & Jordan, J.A., 2004. The Human Polyomavirus , JCV , Uses Serotonin Receptors to Infect Cells. , 306(November), pp.1380–1384.

Epstein, M.A., Achong, B.G. & Barr, Y.M., 1964. VIRUS PARTICLES IN CULTURED LYMPHOBLASTS FROM BURKITT'S LYMPHOMA. *The Lancet*, 283(7335), pp.702–703. Available at: http://dx.doi.org/10.1016/S0140-6736(64)91524-7.

Erickson, K.D., Garcea, R.L. & Tsai, B., 2009. Ganglioside GT1b Is a Putative Host Cell Receptor for the Merkel Cell Polyomavirus. *Journal of Virology* , 83(19), pp.10275–10279. Available at: http://jvi.asm.org/content/83/19/10275.abstract.

Feng, H. et al., 2011. Cellular and Viral Factors Regulating Merkel Cell Polyomavirus Replication. , 6(7).

Feng, H. et al., 2008. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science (New York, N.Y.)*, 319(5866), pp.1096–1100.

Feng, H. et al., 2007. Human Transcriptome Subtraction by Using Short Sequence Tags To Search for Tumor Viruses in Conjunctival Carcinoma ▯. , 81(20), pp.11332–11340.

Fitzgerald, T.L. et al., 2015. Dramatic Increase in the Incidence and Mortality from Merkel Cell Carcinoma in the United States. *The American Surgeon*, 81(8), pp.802–806.

Floyd, S., Pines, J. & Lindon, C., 2008. APC/CCdh1 Targets Aurora Kinase to Control Reorganization of the Mitotic Spindle at Anaphase. *Current Biology*, 18(21), pp.1649–1658.

Frye, R., 2002. "SIRT8" expressed in thyroid cancer is actually SIRT7. *British journal of cancer*, 87(12), p.1479.

Gardner, S.D. et al., 1971. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet (London, England)*, 1(7712), pp.1253–1257.

Gaynor, A.M. et al., 2007. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS pathogens*, 3(5), p.e64.

Geng, Q. et al., 2015. High expression of Sirt7 served as a predictor of adverse outcome in breast cancer. *International Journal of Clinical and Experimental Pathology*, 8(2), pp.1938–1945.

Gheit, T. et al., 2012. Merkel cell polyomavirus in non-small cell lung carcinomas from Chile. *Experimental and Molecular Pathology*, 93(1), pp.162–166.

Gingras, A. et al., 2001. Hierarchical phosphorylation of the translation inhibitor 4E-BP1 Hierarchical phosphorylation of the translation inhibitor 4E-BP1. , pp.2852–2864.

Gingras, A.C. et al., 1999. Regulation of 4E-BP1 phosphorylation: a novel two-step

mechanism. *Genes & development*, 13(11), pp.1422–1437.

Gjoerup, O. & Chang, Y., 2010. Update on Human Polyomaviruses and Cancer. , (10).

Gossen, M. & Bujard, H., 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), pp.5547–5551.

Graff, J.R. & Zimmer, S.G., 2003. Translational control and metastatic progression: enhanced activity of the mRNA cap-binding protein eIF-4E selectively enhances translation of metastasis-related mRNAs. *Clinical & experimental metastasis*, 20(3), pp.265–273.

Grekin, R.C. et al., 2016. NCCN Guidelines Version NCCN Guidelines Index Merkel Cell Carcinoma TOC Discussion NCCN Guidelines Version 1.2016 Panel Members Merkel Cell Carcinoma.

Griffiths, D.A. et al., 2013. Merkel Cell Polyomavirus Small T Antigen Targets the NEMO. , 87(24), pp.13853–13867.

Gross, L., 1953. A Filterable Agent, Recovered from Ak Leukemic Extracts, Causing Salivary Gland Carcinomas in C3H Mice. *Experimental Biology and Medicine* , 83(2), pp.414–421.

Haeberle, H. et al., 2004. Molecular profiling reveals synaptic release machinery in Merkel cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(40), pp.14503–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15448211 [Accessed May 31, 2016].

Hahn, W.C. et al., 1999. Creation of human tumour cells with defined genetic elements. *Nature*, 400(6743), pp.464–468.

Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, 144(5), pp.646–674.

Harms, P.W. et al., 2016. The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma Running Title: Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma.

Harrison, C.J. et al., 2011. Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin. *Journal of Molecular Biology*, 409(4), pp.529–542.

zur Hausen, A. et al., 2013. Early B-Cell Differentiation in Merkel Cell Carcinomas: Clues to Cellular Ancestry. *Cancer Research*, 73(16), pp.4982–4987. Available at: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-13-0616 [Accessed May 31, 2016].

Hay, N. & Sonenberg, N., 2004. Upstream and downstream of mTOR. *Genes & development*, 18(16), pp.1926–1945.

Hitchcock, I.S., Genever, P.G. & Cahusac, P.M.B., 2004. Essential components for a glutamatergic synapse between Merkel cell and nerve terminal in rats. *Neuroscience letters*, 362(3), pp.196–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15158013 [Accessed May 31, 2016].

Hodgson, N.C., 2005. Merkel Cell Carcinoma : Changing Incidence Trends. , (September 2004), pp.1–4.

Houben, R., Shuda, M., Weinkam, R., Feng, H., et al., 2010. Merkel Cell Polyomavirus-Infected Merkel Cell Carcinoma Cells Require Expression of Viral T Antigens Merkel Cell Polyomavirus-Infected Merkel Cell Carcinoma Cells Require Expression of Viral T Antigens ⬚.

Houben, R., Shuda, M., Weinkam, R., Schrama, D., et al., 2010. Merkel Cell Polyomavirus-Infected Merkel Cell Carcinoma Cells Require Expression of Viral T Antigens ⬚. , 84(14), pp.7064–7072.

Houben, R. et al., 2012. polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells.

Hsieh, A.C. et al., 2012. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*, 485(7396), pp.55–61.

Huang, D.W., Lempicki, R. a & Sherman, B.T., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), pp.44–57.

Hwang, J.H. et al., 2013. Protein phosphatase 2A isoforms utilizing Abeta scaffolds regulate differentiation through control of Akt protein. *The Journal of biological chemistry*, 288(44), pp.32064–32073.

Ingolia, N.T. et al., 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. , 324(April), pp.218–223.

Ingolia, N.T. et al., 2012. The ribosome profiling strategy for monitoring translation

in vivo by deep sequencing of ribosome-protected mRNA fragments.

Ingolia, N.T., Lareau, L.F. & Weissman, J.S., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), pp.789–802.

Kaplan, D.R. et al., 1987. Common elements in growth factor stimulation and oncogenic transformation: 85 kd phosphoprotein and phosphatidylinositol kinase activity. *Cell*, 50(7), pp.1021–1029.

Karim, M.F. et al., 2013. Inhibition of H3K18 deacetylation of Sirt7 by Myb-binding protein 1a (Mybbp1a). *Biochemical and biophysical research communications*, 441(1), pp.157–163.

Karim, M.M. et al., 2001. A Quantitative Molecular Model for Modulation of Mammalian Translation by the eIF4E-binding Protein 1. *Journal of Biological Chemistry* , 276(23), pp.20750–20757.

Kassem, A. et al., 2008. Frequent Detection of Merkel Cell Polyomavirus in Human Merkel Cell Carcinomas and Identification of a Unique Deletion in the VP1 Gene. , (13), pp.5009–5013.

Kim, J.K. et al., 2013. Sirtuin7 oncogenic potential in human hepatocellular carcinoma and its regulation by the tumor suppressors MiR-125a-5p and MiR-125b. *Hepatology (Baltimore, Md.)*, 57(3), pp.1055–1067.

Kiran, S., Anwar, T., et al., 2015. Sirtuin 7 in cell proliferation, stress and disease: Rise of the Seventh Sirtuin! *Cellular signalling*, 27(3), pp.673–682.

Kiran, S., Oddi, V. & Ramakrishna, G., 2015. Sirtuin 7 promotes cellular survival following genomic stress by attenuation of DNA damage, SAPK activation and p53 response. *Experimental cell research*, 331(1), pp.123–141.

Knight, L.M. et al., 2015. Merkel Cell Polyomavirus Small T Antigen Mediates Microtubule Destabilization To Promote Cell Motility and Migration. *Journal of Virology* , 89(1), pp.35–47.

Koromilas, A.E., Lazaris-Karatzas, A. & Sonenberg, N., 1992. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *The EMBO journal*, 11(11), pp.4153–4158.

Kurokawa, A. et al., 2015. Expression of the synaptic exocytosis-regulating molecule

complexin 2 in taste buds and its participation in peripheral taste transduction. *Journal of neurochemistry*, 133(6), pp.806–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25692331 [Accessed June 1, 2016].

Kwun, H.J. et al., 2013. Merkel Cell Polyomavirus Small T Antigen Controls Viral Replication and Oncoprotein Expression by Targeting the Cellular Ubiquitin Ligase SCF Fbw7. *Cell Host and Microbe*, 14(2), pp.125–135.

Kwun, H.J. et al., 2015. Restricted protein phosphatase 2A targeting by Merkel cell polyomavirus small T antigen. *Journal of virology*, 89(8), pp.4191–4200.

Kwun, H.J. et al., 2009. The Minimum Replication Origin of Merkel Cell Polyomavirus Has a Unique Large T-Antigen Loading Architecture and Requires Small T-Antigen Expression for Optimal Replication The Minimum Replication Origin of Merkel Cell Polyomavirus Has a Unique Large T-An.

Lai, C.-C. et al., 2013. Altered expression of SIRT gene family in head and neck squamous cell carcinoma. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 34(3), pp.1847–1854.

Land, H., Parada, L.F. & Weinberg, R.A., 1983. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature*, 304(5927), pp.596–602.

Lane, D.P. & Crawford, L. V, 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature*, 278(5701), pp.261–263.

Lazaris-Karatzas, A. et al., 1992. Ras mediates translation initiation factor 4E-induced malignant transformation. *Genes & development*, 6(9), pp.1631–1642.

Lazaris-Karatzas, A., Montine, K.S. & Sonenberg, N., 1990. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature*, 345(6275), pp.544–547.

Lazaris-Karatzas, A. & Sonenberg, N., 1992. The mRNA 5' cap-binding protein, eIF-4E, cooperates with v-myc or E1A in the transformation of primary rodent fibroblasts. *Molecular and Cellular Biology*, 12(3), pp.1234–1238.

Lee, S. et al., 2012. carcinomas. , 52(3), pp.272–275.

Li, J. et al., 2013. Merkel cell polyomavirus large T antigen disrupts host genomic integrity and inhibits cellular proliferation. *Journal of virology*, 87(16), pp.9173–9188.

Li, J. et al., 2015. Phosphorylation of Merkel Cell Polyomavirus Large Tumor Antigen at Serine 816 by ATM Kinase Induces Apoptosis in Host Cells. *Journal of Biological Chemistry*, 290(3), pp.1874–1884.

Li, L. & Bhatia, R., 2013. The controversial role of Sirtuins in tumorigenesis - SIRT7 joins the debate. *Cell research*, 23(1), pp.10–12.

Linzer, D.I.H. & Levine, A.J., 1979. Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell*, 17(1), pp.43–52.

Liu, X. et al., 2011. Merkel Cell Polyomavirus Large T Antigen Disrupts Lysosome Clustering by Translocating Human Vam6p from the Cytoplasm to the Nucleus. *The Journal of Biological Chemistry*, 286(19), pp.17079–17090.

Low, J.A. et al., 2006. Identification of Gangliosides GD1b and GT1b as Receptors for BK Virus. , 80(3), pp.1361–1366.

Loyo, M. et al., 2010. Quantitative detection of Merkel Cell Virus in human tissues and possible mode of transmission. *International journal of cancer. Journal international du cancer*, 126(12), pp.2991–2996.

Lucarz, A. & Brand, G., 2007. Current considerations about Merkel cells. *European journal of cell biology*, 86(5), pp.243–251.

Lyhne, D. et al., 2011. Rising incidence of Merkel cell carcinoma. , (May), pp.274–280.

Mamane, Y. et al., eIF4E - from translation to transformation. *Oncogene*, 23(18), pp.3172–3179. Available at: http://dx.doi.org/10.1038/sj.onc.1207549.

Marusyk, A. & Polyak, K., 2010. Tumor heterogeneity: causes and consequences. *Biochimica et biophysica acta*, 1805(1), p.105.

van der Meijden, E. et al., 2010. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. *PLoS pathogens*, 6(7), p.e1001024.

Michishita, E. et al., 2005. Evolutionarily Conserved and Nonconserved Cellular Localizations and Functions of Human SIRT Proteins. , 16(October), pp.4623–4635.

Miyagi, Y. et al., 1995. Elevated levels of eukaryotic translation initiation factor eIF-4E, mRNA in a broad spectrum of transformed cell lines. *Cancer letters*, 91(2),

pp.247–252.

Moll, I. et al., 2005. Human Merkel cells--aspects of cell biology, distribution and functions. *European journal of cell biology*, 84(2–3), pp.259–271.

Moore, P.S. & Chang, Y., 2010. Why do viruses cause cancer ? Highlights of the first century of human tumour virology. , 10(December).

Morrison, K.M. et al., 2009. Mammalian Merkel cells are descended from the epidermal lineage. *Developmental Biology*, 336(1), pp.76–83.

Nathan, C.A. et al., 1997. Elevated expression of eIF4E and FGF-2 isoforms during vascularization of breast carcinomas. *Oncogene*, 15(9), pp.1087–1094.

Neumann, F. et al., 2011. Replication, gene expression and particle production by a consensus merkel cell polyomavirus (mcpyv) genome. *PLoS ONE*, 6(12), p.e29112. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22216177 [Accessed May 31, 2016].

Nunzi, M.-G., Pisarek, A. & Mugnaini, E., 2004. Merkel cells, corpuscular nerve endings and free nerve endings in the mouse palatine mucosa express three subtypes of vesicular glutamate transporters. *Journal of neurocytology*, 33(3), pp.359–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15475690 [Accessed May 31, 2016].

O'Reilly, K.E. et al., 2009. Phosphorylated 4E-BP1 is associated with poor survival in melanoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(8), pp.2872–2878.

Okuda, T. et al., 2000. Molecular cloning and characterization of GalNAc 4-sulfotransferase expressed in human pituitary gland. *The Journal of biological chemistry*, 275(51), pp.40605–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11001942 [Accessed June 1, 2016].

van Ooyen, A. et al., 1979. Comparison of total sequence of a cloned rabbit beta-globin gene and its flanking regions with a homologous mouse sequence. *Science (New York, N.Y.)*, 206(4416), pp.337–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/482942 [Accessed May 14, 2016].

Oridate, N. et al., 2005. Growth inhibition of head and neck squamous carcinoma cells by small interfering RNAs targeting eIF4E or cyclin D1 alone or combined with cisplatin. *Cancer biology & therapy*, 4(3), pp.318–323.

Padgett, B.L. et al., 1971. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. *Lancet (London, England)*, 1(7712), pp.1257–1260.

Pallas, D.C. et al., 1990. Polyoma small and middle T antigens and SV40 small t antigen form stable complexes with protein phosphatase 2A. *Cell*, 60(1), pp.167–176.

Pastrana, D. V et al., 2009. Quantitation of human seroresponsiveness to Merkel cell polyomavirus. *PLoS pathogens*, 5(9), p.e1000578.

Patzke, C. et al., 2015. Analysis of conditional heterozygous STXBP1 mutations in human neurons. *The Journal of clinical investigation*, 125(9), pp.3560–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26280581 [Accessed June 1, 2016].

Pause, A. et al., 1994. Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature*, 371(6500), pp.762–767.

Perrera, C. et al., 2010. Identification of Myb-binding protein 1A (MYBBP1A) as a novel substrate for aurora B kinase. *The Journal of biological chemistry*, 285(16), pp.11775–11785.

Pipas, J.M., 1992. MINIREVIEW Common and Unique Features of T Antigens Encoded by the Polyomavirus Group. , 66(7), pp.3979–3985.

Poiesz, B.J. et al., 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. , 77(12), pp.7415–7419.

Poulin, F. et al., 1998. 4E-BP3, a New Member of the Eukaryotic Initiation Factor 4E-binding Protein Family. *Journal of Biological Chemistry* , 273(22), pp.14002–14007.

Di Primio, C. et al., 2005. Potentiation of gene targeting in human cells by expression of Saccharomyces cerevisiae Rad52. *Nucleic acids research*, 33(14), pp.4639–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16106043 [Accessed May 29, 2016].

Reich, N.C. & Levine, A.J., 1982. Specific interaction of the SV40 T antigen-cellular p53 protein complex with SV40 DNA. *Virology*, 117(1), pp.286–290.

Reichgelt, B.A. & Visser, O., 2010. Epidemiology and survival of Merkel cell carcinoma in the Netherlands . A population-based study of 808 cases in 1993 – 2007. *European Journal of Cancer*, 47(4), pp.579–585.

Richter, J.D. & Sonenberg, N., 2005. Regulation of cap-dependent translation by eIF4E inhibitory proteins. , pp.477–480.

Rodriguez-Viciana, P., Collins, C. & Fried, M., 2006. Polyoma and SV40 proteins differentially regulate PP2A to activate distinct cellular signaling pathways involved in growth control. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51), pp.19290–19295.

Rojo, F. et al., 2007. 4E-binding protein 1, a cell signaling hallmark in breast cancer that correlates with pathologic grade and prognosis. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 13(1), pp.81–89.

Rosen, S.T. et al., 1987. Establishment and characterization of a neuroendocrine skin carcinoma cell line. *Laboratory investigation; a journal of technical methods and pathology*, 56(3), pp.302–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3546933 [Accessed May 13, 2016].

Rosenwald, I.B. et al., 2001. Expression of eukaryotic translation initiation factors 4E and 2alpha is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. *Cancer*, 92(8), pp.2164–2171.

Rosenwald, I.B. et al., 1999. Upregulation of protein synthesis initiation factor eIF-4E is an early event during colon carcinogenesis. *Oncogene*, 18(15), pp.2507–2517.

Ruggero, D. et al., 2004. The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nature medicine*, 10(5), pp.484–486.

Ruley, H.E., 1983. Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. *Nature*, 304(5927), pp.602–606.

Ryu, D. et al., 2014. A SIRT7-dependent acetylation switch of GABPbeta1 controls mitochondrial function. *Cell metabolism*, 20(5), pp.856–869.

Sarkisian, C.J. et al., 2007. Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. *Nat Cell Biol*, 9(5), pp.493–

505.

Schmitt, M. et al., 2011. Human Polyomaviruses and Other Human Viruses in Neuroendocrine Tumors. *Cancer Epidemiology Biomarkers & Prevention* , 20(7), pp.1558–1561.

Schowalter, R.M. et al., 2010. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell host & microbe*, 7(6), pp.509–515.

Scuda, N. et al., 2011. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. *Journal of virology*, 85(9), pp.4586–4590.

Shaikh, S. & Nicholson, L.F.B., 2006. Optimization of the Tet-On system for inducible expression of RAGE. *Journal of biomolecular techniques : JBT*, 17(4), pp.283–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17028168 [Accessed May 13, 2016].

Sharp, C.P. et al., 2009. Reactivation and Mutation of Newly Discovered WU , KI , and Merkel Cell Carcinoma Polyomaviruses in Immunosuppressed Individuals. , 199.

Shin, J. et al., 2013. SIRT7 Represses Myc Activity to Suppress ER Stress and Prevent Fatty Liver Disease. *Cell reports*, 5(3), pp.654–665.

Shuda, M. et al., 2015. CDK1 substitutes for mTOR kinase to activate mitotic cap-dependent protein translation. *Proceedings of the National Academy of Sciences* , 112(19), pp.5875–5882.

Shuda, M. et al., 2009. Human Merkel cell polyomavirus infection I. MCV T antigen expression in Merkel cell carcinoma, lymphoid tissues and lymphoid tumors. *International journal of cancer. Journal international du cancer*, 125(6), pp.1243–1249.

Shuda, M. et al., 2011. Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. *The Journal of clinical investigation*, 121(9), pp.3623–3634.

Shuda, M. et al., 2008a. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus.

Shuda, M. et al., 2008b. T antigen mutations are a human tumor-specific signature

for Merkel cell polyomavirus. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), pp.16272–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18812503 [Accessed May 13, 2016].

Shuda, M., Chang, Y. & Moore, P.S., 2014. Merkel Cell Polyomavirus Positive Merkel Cell Carcinoma Requires Viral Small T Antigen For Cell Proliferation. *The Journal of investigative dermatology*, 134(5), pp.1479–1481.

Siebrasse, E.A. et al., 2012. Identification of MW polyomavirus, a novel polyomavirus in human stool. *Journal of virology*, 86(19), pp.10321–10326.

Sorrells, D.L. et al., 1999. Pattern of amplification and overexpression of the eukaryotic initiation factor 4E gene in solid tumor. *The Journal of surgical research*, 85(1), pp.37–42.

Srinivasan, A. et al., 1997. The Amino-Terminal Transforming Region of Simian Virus 40 Large T and Small t Antigens Functions as a J Domain. , 17(8), pp.4761–4773.

Stone, A.B., 1974. A simplified method for preparing sucrose gradients (Short Communication). *Biochemical Journal*, 137(1), pp.117–118.

Stratton, M.R., Campbell, P.J. & Futreal, P.A., 2009. The cancer genome. *Nature*, 458(7239), pp.719–724.

Sullivan, C.S. et al., 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature*, 435(7042), pp.682–686.

Sweet, B.H. & Hilleman, M.R., 1960. The Vacuolating Virus, S.V.40. *Experimental Biology and Medicine*, 105(2), pp.420–427. Available at: http://ebm.sagepub.com/lookup/doi/10.3181/00379727-105-26128 [Accessed June 6, 2016].

Szeder, V. et al., 2003. Neural crest origin of mammalian Merkel cells. *Developmental biology*, 253(2), pp.258–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12645929 [Accessed May 31, 2016].

Theiss, J.M. et al., 2015. A Comprehensive Analysis of Replicating Merkel Cell Polyomavirus Genomes Delineates the Viral Transcription Program and Suggests a Role for mcv-miR-M1 in Episomal Persistence. *PLoS pathogens*, 11(7), p.e1004974. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26218535 [Accessed May 31, 2016].

Thoreen, C.C. et al., 2012. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, 485(7396), pp.109–113.

Tilling, T. et al., 2012. Which Are the Cells of Origin in Merkel Cell Carcinoma? *Journal of Skin Cancer*, 2012, pp.1–6. Available at: http://www.hindawi.com/journals/jsc/2012/680410/ [Accessed May 30, 2016].

Toker, C., 1972. TRabecular carcinoma of the skin. *Archives of Dermatology*, 105(1), pp.107–110.

Tolstov, Y.L. et al., 2009. Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays. *International journal of cancer. Journal international du cancer*, 125(6), pp.1250–1256.

Townsley, F.M. & Ruderman, J. V, 1998. Proteolytic ratchets that control progression through mitosis. *Trends in cell biology*, 8(6), pp.238–244.

Trowbridge, P.W. & Frisque, R.J., 1995. Identification of three new JC virus proteins generated by alternative splicing of the early viral mRNA. *Journal of neurovirology*, 1(2), pp.195–206.

Tsai, Y.-C. et al., 2012. Functional proteomics establishes the interaction of SIRT7 with chromatin remodeling complexes and expands its role in regulation of RNA polymerase I transcription. *Molecular & cellular proteomics : MCP*, 11(5), pp.60–76.

Tsai, Y.-C., Greco, T.M. & Cristea, I.M., 2014. Sirtuin 7 Plays a Role in Ribosome Biogenesis and Protein Synthesis. *Molecular & Cellular Proteomics : MCP*, 13(1), pp.73–83.

Uhlen, M. et al., 2015. Tissue-based map of the human proteome. *Science*, 347(6220), pp.1260419–1260419. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1260419 [Accessed May 31, 2016].

Vakhrusheva, O. et al., 2008. Sirt7 increases stress resistance of cardiomyocytes and prevents apoptosis and inflammatory cardiomyopathy in mice. *Circulation research*, 102(6), pp.703–710.

Del Valle, L., White, M.K. & Khalili, K., 2008. Potential mechanisms of the human

polyomavirus JC in neural oncogenesis. *Journal of neuropathology and experimental neurology*, 67(8), pp.729–740.

Verhaegen, M.E. et al., 2015. Merkel Cell Polyomavirus Small T Antigen Is Oncogenic in Transgenic Mice. , pp.1415–1424.

Wang, S. et al., 2001. Expression of eukaryotic translation initiation factors 4E and 2alpha correlates with the progression of thyroid carcinoma. *Thyroid : official journal of the American Thyroid Association*, 11(12), pp.1101–1107.

Wang, S. et al., 1999. Expression of the eukaryotic translation initiation factors 4E and 2alpha in non-Hodgkin's lymphomas. *The American journal of pathology*, 155(1), pp.247–255.

Wang, X. et al., 2003. The C Terminus of Initiation Factor 4E-Binding Protein 1 Contains Multiple Regulatory Features That Influence Its Function and Phosphorylation. *Molecular and Cellular Biology* , 23(5), pp.1546–1557.

Weinberg, R.A., 1996. E2F and cell proliferation: a world turned upside down. *Cell*, 85(4), pp.457–459.

Weinberg, R.A., 1995. The retinoblastoma protein and cell cycle control. *Cell*, 81(3), pp.323–330.

Wendel, H.-G. et al., 2004. Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. *Nature*, 428(6980), pp.332–337.

White, M.K., Gordon, J. & Khalili, K., 2013. The Rapidly Expanding Family of Human Polyomaviruses : Recent Developments in Understanding Their Life Cycle and Role in Human Pathology. , 9(3).

White, M.K. & Khalili, K., 2004. Polyomaviruses and human cancer: molecular mechanisms underlying patterns of tumorigenesis. *Virology*, 324(1), pp.1–16.

Wiedinger, K., Bitsaktsis, C. & Chang, S., 2014. Reactivation of human polyomaviruses in immunocompromised states. *Journal of neurovirology*, 20(1), pp.1–8.

Wieland, U. et al., 2014. Human polyomaviruses 6, 7, 9, 10 and Trichodysplasia spinulosa-associated polyomavirus in HIV-infected men. *The Journal of general virology*, 95(Pt 4), pp.928–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24421113 [Accessed May 11, 2016].

Woods, N.-B. et al., 2003. Lentiviral vector transduction of NOD/SCID repopulating

cells results in multiple vector integrations per transduced cell: risk of insertional mutagenesis. *Blood*, 101(4), pp.1284–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12393514 [Accessed May 13, 2016].

Wright, C.M. et al., 2009. Inhibition of Simian Virus 40 replication by targeting the molecular chaperone function and ATPase activity of T antigen. *Virus Research*, 141(1), pp.71–80.

Yoshizawa, T. et al., 2014. SIRT7 controls hepatic lipid metabolism by regulating the ubiquitin-proteasome pathway. *Cell metabolism*, 19(4), pp.712–721.

Yu, H. et al., 2014. Overexpression of sirt7 exhibits oncogenic property and serves as a prognostic factor in colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 20(13), pp.3434–3445.

Zerrahn, J. et al., 1993. Independent expression of the transforming amino-terminal domain of SV40 large T antigen from an alternatively spliced third SV40 early mRNA. , 12(12), pp.4739–4746.

Zhou, X. et al., 2004. Activation of the Akt/mammalian target of rapamycin/4E-BP1 pathway by ErbB2 overexpression predicts tumor progression in breast cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 10(20), pp.6779–6788.

4.7 Appendix of Chapter 4

Table 4.7.1 Transcriptionally upregulated genes in cells transduced with shRNA for MCV sT

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
HLA-E	2.92	6.38	RAB26	1.72	3.19
KIF19	3.05	6.20	ATP8A1	1.60	3.19
CNTN2	2.73	5.55	CDK6	6.83	3.19
CHST8	4.76	5.24	VEGFA	2.48	3.14
CPLX2	3.16	4.96	SHANK3	2.14	3.12
GNS	2.09	4.93	ID4	1.70	3.12
KATNAL1	2.96	4.63	HEPACAM2	1.74	3.11
C1orf95	2.47	4.60	COL22A1	2.43	3.07
NHLH1	2.23	4.45	DOK4	4.07	3.05
CCNA1	3.56	4.43	CHGA	1.72	3.05
SIK1	3.01	4.38	UNC5A	1.82	3.05
STXBP1	2.02	4.14	SYNPO	6.27	3.04
LRRC10B	6.21	3.99	RGS16	3.42	3.04
OPTN	3.28	3.94	NDRG1	3.04	3.02
SERINC2	2.11	3.90	ZNF750	6.69	3.01
CNR1	2.40	3.83	SH3BGRL3	1.78	3.00
DYNLRB1	2.40	3.82	ZNF302	1.96	3.00
TPPP3	2.98	3.76	MGLL	4.78	2.99
MRAP2	2.95	3.72	CRAT	2.45	2.99
KIAA1462	4.94	3.70	DAAM1	1.75	2.98
DCC	3.05	3.69	FURIN	1.77	2.98
GNG8	2.12	3.69	WNT4	7.77	2.96
NID1	3.01	3.67	PTPRU	6.39	2.93
BTBD17	2.40	3.61	SOWAHA	1.81	2.92
CADM3	2.95	3.58	B3GNT4	4.60	2.92
WNT10A	1.99	3.58	KLF13	1.53	2.90
CYB5R1	1.91	3.49	TOM1L2	1.72	2.89
RGMB	1.88	3.49	MAFB	1.74	2.89
GRASP	2.88	3.49	NPTXR	1.74	2.87
SLC16A3	8.84	3.48	ABHD8	2.30	2.87
RASD2	1.86	3.46	TNFAIP1	1.75	2.86
RIMBP2	2.19	3.42	SLC7A5	2.01	2.86
NELL1	5.87	3.40	CAMK2N1	3.49	2.86
KRT20	2.36	3.39	EVI5L	2.35	2.81
CYBA	1.85	3.39	SCN1A	5.47	2.80
FAT4	3.41	3.34	OSBPL3	1.68	2.79
RGAG4	1.96	3.30	CDC42EP4	1.92	2.78
CACNA2D2	1.84	3.29	JUND	1.70	2.78
KLF4	4.51	3.28	CORO2B	2.24	2.77
SEZ6L	1.88	3.28	DDIT4	2.02	2.76
STC1	2.68	3.26	KANK4	8.27	2.74
IGDCC3	3.29	3.23	NOTCH2	1.88	2.74
HK2	11.66	3.21	MMP15	20.31	2.73
FAM105A	2.11	3.20	PSTPIP1	3.51	2.72

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
MAP1A	1.70	2.71	SYNGR3	3.64	2.46
SCG5	1.94	2.71	AHNAK	2.05	2.45
CPEB2	2.24	2.70	PTPRD	1.65	2.45
ZFPM1	1.93	2.69	BTG3	1.58	2.45
SLC12A7	7.90	2.68	PLEKHB1	2.95	2.45
GPC1	1.66	2.67	NDUFA4L2	2.70	2.45
SYT11	1.92	2.66	LDHA	1.60	2.44
GRIN2D	1.86	2.66	GAL3ST1	1.73	2.43
ADAMTS9	18.80	2.65	DYNC2H1	3.07	2.43
CCK	6.27	2.65	PCBP4	1.43	2.43
TLE6	5.42	2.65	CLDN5	2.03	2.42
FBXL16	1.66	2.64	ZMYND10	4.41	2.42
TMCC2	2.20	2.63	CLDN4	2.91	2.42
TXNIP	1.62	2.63	RNF4	1.59	2.42
IGFBP2	1.66	2.63	SNCAIP	4.38	2.41
SESN2	2.65	2.62	FBXO42	1.72	2.41
CHRNA1	1.85	2.62	TMED4	1.78	2.41
CNGA3	18.05	2.61	RIPPLY3	2.06	2.41
CD47	1.47	2.61	GPRC5C	4.70	2.41
PSKH1	1.84	2.61	PTGIR	2.45	2.40
CITED2	1.82	2.59	TIMP3	2.45	2.40
FAM134B	3.12	2.59	ALDOC	14.54	2.39
INPP5D	4.84	2.58	ATP2B4	1.52	2.38
PRKG2	2.12	2.57	PFKFB3	5.16	2.37
TSKU	1.87	2.56	KDM4B	1.57	2.37
B4GALT1	1.78	2.55	TRAF3IP1	2.41	2.36
WSCD2	3.07	2.54	DMPK	1.62	2.36
AATK	3.07	2.54	ENTPD8	3.45	2.35
NUAK1	2.09	2.53	CADPS	1.99	2.35
BHLHE22	3.76	2.52	KIAA1377	2.04	2.34
HCN2	1.85	2.52	SMTNL2	4.51	2.34
SNAP25	1.59	2.52	TMSB4X	1.50	2.33
AMOTL2	1.81	2.52	SLC36A4	1.56	2.33
CXCR4	1.60	2.51	PPP1R37	1.71	2.32
MB	3.72	2.50	M6PR	1.49	2.32
TRIOBP	2.13	2.49	MDM2	1.57	2.30
TRIM62	2.07	2.48	C4orf48	2.01	2.30
FSIP2	3.15	2.48	STMN2	2.53	2.29
THSD7B	1.77	2.47	SMPDL3A	2.53	2.29
DDR1	1.55	2.47	TCTN2	1.98	2.28
ISLR2	1.87	2.47	MAB21L1	2.34	2.27
PLD3	1.59	2.46	DRGX	1.94	2.26
PROX1	1.58	2.46	DYRK1B	2.00	2.26
JHDM1D	2.50	2.46	KCNA2	4.76	2.25

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
C11orf71	4.76	2.25	NAV1	1.49	2.13
ELAVL3	1.73	2.25	PRKAR1A	1.49	2.13
PCDH9	1.55	2.24	ARHGAP1	1.48	2.13

ACTR10	1.58	2.23	PPL	11.28	2.13
SPTSSB	2.81	2.23	PMM1	1.91	2.12
PTPRZ1	1.53	2.23	ABR	1.50	2.12
SCARB2	1.39	2.23	GAREM	2.57	2.12
CCDC40	2.29	2.22	P4HTM	1.54	2.11
MAPRE3	2.47	2.22	MYT1	1.54	2.11
TNF	5.64	2.22	RNF19A	1.68	2.11
ID3	1.53	2.22	BTG2	1.44	2.10
KSR1	1.67	2.21	ALK	1.53	2.09
ULK1	1.52	2.21	TMEM86B	4.26	2.09
HSPA12A	1.72	2.21	GATA2	4.26	2.09
ANKRD52	1.53	2.20	SLC5A8	5.11	2.09
ACPL2	3.87	2.19	PVRL1	2.19	2.09
SLC25A22	1.57	2.19	IFT57	2.18	2.08
SUN2	1.38	2.19	CMIP	1.43	2.08
PCSK2	1.53	2.19	HID1	1.48	2.08
SRXN1	2.76	2.19	OSCP1	3.01	2.08
LMO7	1.96	2.19	MAP9	1.70	2.08
CCNG2	1.50	2.18	RAB3A	1.67	2.08
AFAP1	1.56	2.18	GRK5	2.17	2.07
STK17B	1.46	2.18	CCDC104	1.88	2.07
LIMCH1	1.95	2.17	HSD17B10	1.62	2.07
KCNH3	4.51	2.17	CAMK4	1.60	2.06
PHLDA1	2.63	2.17	RHBDL3	1.47	2.06
SREBF1	1.46	2.17	MAP1B	1.43	2.05
FAM206A	1.91	2.17	EPHA5	2.60	2.05
TNFAIP2	1.93	2.17	MAPK8IP1	1.83	2.05
BAD	1.68	2.17	C1orf173	1.66	2.05
C9orf116	5.42	2.16	RHOU	2.15	2.05
IMPA1	1.52	2.16	MGC50722	3.56	2.05
RNF165	1.89	2.16	ZDHHC1	1.85	2.05
CELF3	1.69	2.16	APLP1	1.48	2.05
MAP7D2	1.91	2.15	PPP2R5B	1.66	2.05
PLCD3	2.24	2.15	RAB39B	2.59	2.04
FAM219A	1.55	2.15	MXRA7	1.85	2.04
SMIM14	1.92	2.14	MAP3K14	1.46	2.04
SEPN1	1.51	2.14	DNAL1	1.83	2.04
GNAI2	1.49	2.14	ASMTL	1.66	2.04
CDKN1A	1.46	2.14	ROBO2	1.52	2.04
TPPP	2.70	2.14	HEY1	2.14	2.03
FAM161B	2.22	2.13	C17orf97	4.08	2.03

Gene	change in RNA level	Z-score
CDHR1	2.57	2.03
BMP7	1.48	2.03
CHRNB4	1.84	2.03
BCL2L1	1.62	2.03
ZNF665	4.89	2.03
LRRIQ1	4.89	2.03
BZRAP1	4.89	2.03

ARL4D	4.89	2.03
SLC2A6	3.76	2.03
EPHA4	2.92	2.02
LOC101928120	4.81	2.00
LAD1	4.01	2.00
IQCG	4.01	2.00
IFT81	1.64	2.00
FCHO1	1.81	2.00

Table 4.7.2 Transcriptionally downregulated genes in cells transduced with shRNA for MCV sT

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
NME4	0.23	-6.59	ETS1	0.50	-2.84
FAM210B	0.33	-5.92	EVL	0.61	-2.84
PKM	0.37	-5.32	GAREML	0.54	-2.83
RELN	0.25	-4.71	TMEM74	0.60	-2.82
PCNA	0.44	-4.39	POU3F2	0.58	-2.82
TMEM98	0.45	-4.05	HEG1	0.50	-2.82
NFIB	0.47	-3.94	SKIDA1	0.22	-2.81
NDUFC1	0.27	-3.87	NFIX	0.59	-2.81
COLGALT1	0.43	-3.84	SLC2A1	0.54	-2.81
TYRO3	0.44	-3.77	SLC38A5	0.54	-2.80
MAPK6	0.43	-3.76	TUBB3	0.60	-2.80
UBE2Z	0.49	-3.73	HIST1H2BH	0.14	-2.79
LDLRAP1	0.52	-3.62	GPC3	0.58	-2.76
PXN	0.34	-3.58	ELF4	0.55	-2.74
TGIF2	0.36	-3.55	CST3	0.65	-2.70
DHCR7	0.51	-3.52	EIF4EBP1	0.49	-2.70
OAS3	0.57	-3.50	GTF3A	0.61	-2.68
FCGRT	0.32	-3.39	AUTS2	0.60	-2.68
C19orf10	0.44	-3.38	PTPRJ	0.55	-2.67
MEST	0.60	-3.35	ATF5	0.61	-2.67
PAPOLA	0.60	-3.34	EMC6	0.45	-2.67
GABRB2	0.27	-3.33	PSAP	0.60	-2.66
FBLN1	0.40	-3.32	CAMKV	0.48	-2.65
SHROOM3	0.45	-3.27	RAP1B	0.66	-2.63
CCDC3	0.54	-3.23	PIP4K2A	0.57	-2.63
PDE7A	0.56	-3.17	NGFRAP1	0.66	-2.62
WIPF3	0.46	-3.17	PIP5K1A	0.48	-2.61
SH2D5	0.39	-3.15	ANGPTL2	0.11	-2.61
TRIP13	0.50	-3.13	SZRD1	0.64	-2.61
CRTAC1	0.40	-3.13	C18orf54	0.41	-2.60
MLEC	0.55	-3.13	CCNI	0.64	-2.59
PROM1	0.40	-3.09	PHB	0.63	-2.59
TSPYL1	0.57	-3.09	FAM217B	0.60	-2.59
FTL	0.62	-3.08	FRMD6	0.48	-2.58
NEO1	0.50	-3.08	ADAM10	0.60	-2.58
CADM4	0.40	-3.03	RNF38	0.59	-2.57
NEK6	0.41	-3.01	SLC43A3	0.62	-2.56
LAPTM4B	0.56	-3.01	CHDH	0.57	-2.56
HIF1A	0.57	-3.00	PARP12	0.17	-2.56
B3GAT2	0.08	-2.96	HIPK1	0.49	-2.56
PPP3R1	0.58	-2.89	NR1D2	0.63	-2.55
CA9	0.56	-2.86	FAM195A	0.38	-2.54
ST3GAL1	0.54	-2.85	TIAM1	0.57	-2.54

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
SYT13	0.63	-2.53	FAM184B	0.53	-2.32
TSPAN7	0.54	-2.53	SNX5	0.65	-2.31
KCTD15	0.60	-2.53	EPB41	0.66	-2.31
GLCE	0.51	-2.52	PER3	0.61	-2.30
RASL10B	0.44	-2.52	PHGDH	0.70	-2.30
NIPA2	0.62	-2.51	TLE3	0.57	-2.30
C2orf44	0.52	-2.51	EPT1	0.65	-2.30
HYAL2	0.60	-2.50	LCOR	0.57	-2.29
KIAA1549	0.58	-2.50	PTDSS1	0.68	-2.29
NABP2	0.64	-2.48	HPDL	0.53	-2.29
SLITRK1	0.61	-2.48	PLEKHA2	0.51	-2.29
LIMD1	0.44	-2.47	NPW	0.51	-2.29
FOXN2	0.58	-2.47	POLR3A	0.60	-2.28
TOX3	0.34	-2.47	TBC1D1	0.55	-2.28
PYCR1	0.63	-2.46	SLITRK5	0.61	-2.27
LMOD2	0.48	-2.46	VCAN	0.51	-2.27
CADM2	0.68	-2.45	SLC1A5	0.63	-2.26
SGTA	0.64	-2.45	TFDP2	0.58	-2.25
HEMK1	0.38	-2.44	GAS2L3	0.41	-2.25
SCRT2	0.49	-2.44	SPRY2	0.16	-2.25
DDN	0.49	-2.43	SH3KBP1	0.16	-2.25
SMAD9	0.35	-2.43	ST5	0.60	-2.24
HIPK2	0.59	-2.42	LRTM2	0.60	-2.24
USH1C	0.44	-2.42	APBB2	0.64	-2.24
LRRC16A	0.44	-2.42	ICK	0.58	-2.23
PAK1	0.64	-2.41	ZNF114	0.31	-2.23
MFSD2A	0.64	-2.40	SSPN	0.31	-2.23
IFIH1	0.19	-2.39	HDAC9	0.62	-2.22
ETNK1	0.49	-2.39	SUMO3	0.64	-2.21
EME1	0.46	-2.39	CABLES2	0.64	-2.21
PDGFD	0.35	-2.39	RASAL1	0.47	-2.21
SV2B	0.32	-2.38	GLYR1	0.66	-2.21
KCTD8	0.32	-2.37	SDC2	0.53	-2.21
NOLC1	0.65	-2.36	CSRP2	0.61	-2.20
MKI67IP	0.59	-2.36	PAWR	0.47	-2.20
LBH	0.51	-2.35	C4orf32	0.62	-2.19
FAM69B	0.62	-2.35	PEX14	0.53	-2.19
CDS1	0.36	-2.34	CIB2	0.53	-2.19
BMF	0.69	-2.34	UNC5C	0.17	-2.19
RFX7	0.59	-2.33	ACACA	0.71	-2.19
AEBP1	0.29	-2.32	FAM20B	0.69	-2.19
PABPC1	0.67	-2.32	PCSK1	0.69	-2.18
ZBTB46	0.53	-2.32	KIF1C	0.49	-2.18

Gene	change in RNA level	Z-score	Gene	change in RNA level	Z-score
CNKSR2	0.49	-2.18	PMF1	0.52	-2.04
CBS	0.66	-2.18	USP2	0.61	-2.04

NME1	0.69	-2.18		LYRM1	0.57	-2.03
IFI6	0.62	-2.18		MPP5	0.57	-2.03
SLC25A53	0.32	-2.18		CDH23	0.35	-2.03
IPO7	0.66	-2.17		TMEM109	0.61	-2.03
KCNMA1	0.48	-2.17		B2M	0.71	-2.03
ATP2A2	0.66	-2.16		RHOBTB3	0.68	-2.03
VDR	0.44	-2.16		MRPL48	0.57	-2.03
ZCCHC3	0.65	-2.16		GPRC5B	0.57	-2.03
RNF114	0.67	-2.16		ABRACL	0.59	-2.02
RNF167	0.68	-2.16		SPTY2D1	0.64	-2.02
RUNX1T1	0.67	-2.15		CTDSP2	0.68	-2.02
BRD4	0.63	-2.14		FMNL3	0.57	-2.02
ADRA2C	0.67	-2.13		SNX30	0.62	-2.02
KDM8	0.33	-2.12		C3orf80	0.57	-2.02
FBN3	0.50	-2.12		SLC9A3R1	0.68	-2.02
XPO4	0.67	-2.11		MAML3	0.65	-2.02
NCAPH	0.65	-2.11		CCNB1IP1	0.64	-2.01
WWC1	0.63	-2.11		PUS10	0.25	-2.01
NME2	0.53	-2.10		PCDH17	0.25	-2.01
CBLN1	0.64	-2.10		SLC6A8	0.51	-2.01
TNPO2	0.72	-2.10		DGAT2	0.51	-2.01
LCMT2	0.60	-2.09		SLMO1	0.51	-2.01
RDH12	0.33	-2.09		KISS1R	0.64	-2.01
BEGAIN	0.33	-2.09				
TUBA1C	0.64	-2.08				
LDHB	0.67	-2.08				
SCGB2A1	0.66	-2.08				
KCNJ2	0.19	-2.07				
HMGA1	0.67	-2.07				
ZADH2	0.69	-2.07				
PCDH7	0.41	-2.06				
HIVEP2	0.24	-2.06				
ENO3	0.56	-2.06				
SRGAP2	0.61	-2.06				
ZDHHC22	0.61	-2.06				
ZDHHC20	0.66	-2.05				
TMEM5	0.61	-2.04				
RBM14	0.73	-2.04				
MRGBP	0.55	-2.04				
FAM53B	0.50	-2.04				
TSTA3	0.61	-2.04				

Summary - Conclusions

Cancer is the name given to diverse diseases, which result in millions of deaths worldwide (Ferlay et al. 2015) and whose common characteristic is uncontrolled cell proliferation. In an effort to rationalise the complexity of cancer Hanahan and Weinberg proposed the hallmarks that comprise biological capabilities that enable tumour development (Hanahan & Weinberg 2011; Hanahan & Weinberg 2000). Dysregulation of molecular pathways underlie these hallmarks of tumour progression and their study has been central to cancer biology. The PI3K/AKT/mTOR pathway has a central role in translation, cell growth and proliferation and it is frequently dysregulated in various cancers (Qiu et al. 2006; Hafner et al. 2012). In this work we aimed to elucidate the role of the PI3K pathway and protein synthesis in cancer.

Our work focused on various levels of the PI3K pathway and protein synthesis. We studied translation initiation of PTEN, which lies upstream of the PI3K pathway, and inhibits its activation.

We studied translation initiation of PTEN, which is a negative regulator of the PI3K pathway. Mutational analysis revealed at least two previously undescribed non-AUG initiated PTEN proteoforms. The most abundant extended PTEN proteoform initiates at a conserved AUU codon and extends the canonical PTEN by 146 amino acids. The studied N-terminally extended PTEN proteoforms retain the ability to downregulate the PI3K pathway. The ratio of PTEN proteoforms can be regulated by the translation of two conserved upstream open reading frames. Our findings could have profound implications for the understanding of PTEN regulation and function and could have implications for the development of PTEN-based chemotherapeutic agents.

We next focused on polyamines, which are regulated but can also regulate the PI3K pathway through a negative feedback loop. Polyamine levels are increased and their metabolism is frequently dysregulated in a range of cancers. We aimed to investigate the effect of polyamines on global transcriptional and translational regulation, by application of ribosome profiling. This technique enables the detailed study of protein synthesis. Our work confirmed what was previously known about the transcriptional and translational regulation of SAT1 and AMD1 respectively and

revealed translation of a sequence of the CHTF8 gene, previously annotated as a 3'UTR.

Gene Ontology analysis of the most translationally upregulated genes in polyamine supplemented cells showed enrichment of genes regulating RNA processing, translation, macromolecular complex subunit organization and mitosis. These results are consistent with the known role of polyamines in promoting cell growth and proliferation (Igarashi & Kashiwagi 2010; Pegg 2009). Our results show, for the first time, that high polyamine concentration induces increased translation of genes involved in DNA repair and protein catabolism. Pathway enrichment analysis of the transcriptionally upregulated genes in polyamine enriched cells showed overrepresentation of cytoskeleton associated genes. The role of polyamines in cell motility (Savarin et al. 2010) and vesicle trafficking (Kanerva et al. 2010) through direct interaction of polyamines with microtubules has already been described. Polyamines can also regulate the dynamics of actin cytoskeleton (Kucharzewska et al. 2010) and their depletion can cause disappearance of actin stress fibers (Parkkinen et al. 1997). Our study adds a new level of polyamine regulation on cytoskeleton through transcriptional upregulation of cytoskeletal genes.

Lastly we focused on translation initiation in virally induced Merkel cell carcinoma. Understanding of polyomaviruses's physiology has been critical in understanding of cellular physiology and characterization of the PI3K pathway.

Merkel cell carcinoma (MCC) is an aggressive non melanoma skin cancer with increasing incident rates. 80% of MCCs are caused by the Merkel Cell Polyomavirus (MCV). MCV expresses a small (sT) antigen which can transform rodent fibroblasts (Rat-1) by dysregulating protein synthesis initiation.

We aimed to elucidate the role of sT at the initial stages of oncogenesis, by establishing a stable Rat-1 cell line that inducibly and reversibly expresses the antigen. This cell line failed to transform upon sT expression, possibly due to low sT levels.

To investigate the potential role of sT at the later stages of MCV induced transformation, ribosome profiling was applied on a MCC cell line positive for MCV

transduced with shRNA towards sT. Knockdown of sT did not affect translation as expected, possibly because sT exerts its role at the first stages of carcinogenesis.

.

Ferlay, J. et al., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer. Journal international du cancer*, 136(5), pp.E359-86.

Hafner, C. et al., 2012. Activation of the PI3K/AKT Pathway in Merkel Cell Carcinoma K. Smalley, ed. *PLoS ONE*, 7(2), p.e31255. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3281946/.

Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, 144(5), pp.646–674.

Hanahan, D. & Weinberg, R.A., 2000. The hallmarks of cancer. *Cell*, 100(1), pp.57–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10647931 [Accessed October 3, 2016].

Qiu, W. et al., 2006. PIK3CA Mutations in Head and Neck Squamous Cell Carcinoma. *Clinical Cancer Research* , 12(5), pp.1441–1446. Available at: http://clincancerres.aacrjournals.org/content/12/5/1441.abstract.

Published papers

OPEN BIOLOGY

Research

Author for correspondence:
Gary Loughran
e-mail: g.loughran@ucc.ie

THE ROYAL SOCIETY
PUBLISHING

Systematic analysis of the *PTEN* 5′ leader identifies a major AUU initiated proteoform

Ioanna Tzani[1], Ivaylo P. Ivanov[2], Dmitri E. Andreev[1,3], Ruslan I. Dmitriev[1], Kellie A. Dean[1], Pavel V. Baranov[1], John F. Atkins[1,4] and Gary Loughran[1]

[1]School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland
[2]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA
[3]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119992, Russia
[4]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112-5330, USA

Abundant evidence for translation within the 5′ leaders of many human genes is rapidly emerging, especially, because of the advent of ribosome profiling. In most cases, it is believed that the act of translation rather than the encoded peptide is important. However, the wealth of available sequencing data in recent years allows phylogenetic detection of sequences within 5′ leaders that have emerged under coding constraint and therefore allow for the prediction of functional 5′ leader translation. Using this approach, we previously predicted a CUG-initiated, 173 amino acid N-terminal extension to the human tumour suppressor PTEN. Here, a systematic experimental analysis of translation events in the *PTEN* 5′ leader identifies at least two additional non-AUG-initiated PTEN proteoforms that are expressed in most human cell lines tested. The most abundant extended PTEN proteoform initiates at a conserved AUU codon and extends the canonical AUG-initiated PTEN by 146 amino acids. All N-terminally extended PTEN proteoforms tested retain the ability to downregulate the PI3K pathway. We also provide evidence for the translation of two conserved AUG-initiated upstream open reading frames within the *PTEN* 5′ leader that control the ratio of PTEN proteoforms.

1. Introduction

The process of translation can be described in four steps: initiation, which is usually tightly regulated; elongation; termination; and ribosome recycling [1]. In eukaryotes, the scanning model for translation initiation postulates that the small ribosomal subunit, in complex with initiation factors and Met-tRNA$_i$, binds first to the 5′ cap then scans 3′ until a suitable initiation codon is found [2]. Base-pairing interactions between the anticodon loop of the Met-tRNA$_i$ bound to the ribosome and an AUG codon in the mRNA cause the ribosome to stop scanning and set the reading frame for protein synthesis [3]. Typically, the ribosome initiates protein synthesis at the AUG codon closest to the 5′ end of the mRNA, though the efficiency of initiation is dependent on the nucleotide sequence surrounding the initiator codon with the optimal sequence known as the Kozak context [4]. The Kozak context—comprising 6 nt before and 1 nt immediately following a potential initiation codon—has significant influence on the recognition of an initiation site, through partially understood mechanisms requiring the activities of eIF1 [5–8] and eIF5 [9–11]. Using multiple sequence alignments, the consensus context in mammals was identified as GCCRCC**AUG**G [12] with the identity of the underlined nucleotides in the −3 and +4 positions (relative to the 'A' of the AUG) being the most important. However, a recent high-throughput analysis of all possible initiation contexts revealed RYMRMVAUGGC as the optimal context in human and mouse cells and additionally revealed synergistic effects of neighbouring nucleotides [13].

Initiation can occur at most codons that differ from AUG by a single nucleotide (non-cognate or non-AUG). Seven out of the nine possible single-nucleotide substitutions at the AUG start codon of dihydrofolate reductase were functional as translation start sites in mammalian cells [14]. In all of the cases in which it was examined, the N-terminal residue of these proteins was methionine [14], suggesting that translation initiation relied on mis-pairing between the anticodon of Met-tRNA$_i$ and the non-AUG start codon in the mRNA. However, a report exists of CUG initiation by elongator Leu-tRNA functioning as initiator tRNA [15].

Initiation is the only step where an incoming aminoacyl tRNA is bound directly in the ribosomal P-site [16,17]. Unlike the A-site, where mRNA : tRNA interactions are strictly monitored by the decoding centre [18], the P-site can tolerate mismatches in the codon : anticodon duplex [19–22]. This allows the incorporation of Met-tRNA$_i$ at a wider range of codons compared with elongator Met-tRNA whose incorporation is strictly limited to AUG codons. The most favourable context for non-AUG initiation is believed to be identical to that for AUG starts [23–25]. In addition, a strong RNA secondary structure starting approximately 15 nt downstream of the non-AUG codon may significantly increase initiation efficiency [26]. Another important factor for non-AUG initiation is that it is located upstream of the most 5′ AUG codon [27].

When initiation codons occur in the 5′ leaders of transcripts they give rise to either N-terminal extensions to the main ORF or else upstream open reading frames (uORFs). It has been estimated that AUG-initiated uORFs are present in approximately half of the human protein coding genes [28]. Furthermore, ribosome profiling provided evidence for the presence of translating ribosomes on more than 200 non-AUG-initiated uORFs in yeast [29] and much more widespread non-AUG initiation in mammals [30–32]. In general, the translation of uORFs has an inhibitory effect on translation of the main protein coding ORF, because ribosomes terminating a uORF are often unable to reinitiate owing to the loss of necessary initiation factors. However, the 40S subunits of ribosomes translating short ORFs (less than 35 codons) may retain some initiation factors after termination [4,33]—although efficient re-initiation is precluded until all necessary initiation factors have been reloaded onto the 40S subunit. In most instances, there is a requirement for a ternary complex of eIF2, GTP and Met-tRNA$_i$ which is regulated by the phosphorylation status of eIF2α [34].

An earlier finding that sequences in the 5′ leaders are highly conserved and that the level of conservation globally increases towards the leader/main ORF boundaries [35] suggests that this conservation could be due, in part, to the 3′ ends of a portion of 5′ leaders encoding N-terminal extensions to the annotated AUG-initiated proteins. In total, more than 60 instances of non-AUG-initiated N-terminal extensions have been predicted or verified experimentally in mammals [36–38]. In most cases, the non-AUG initiation provides an alternative longer proteoform in addition to a proteoform resulting from initiation at a standard AUG codon downstream via a process termed 'leaky scanning'. Where alternative proteoforms are produced as a result of leaky scanning, the longer isoform frequently contains a signal for subcellular localization that is absent in the shorter form [39–43].

Previously, we performed a systematic analysis of the 5′ leaders of human GenBank RefSeq mRNAs to investigate the extent of non-AUG initiation in humans [37]. This involved analysis of codon substitution rates in pairwise alignments of human and mice orthologous sequences to identify regions of 5′ leaders evolving under the constraints of protein coding evolution. When a region within a 5′ leader evolves under such constraints, it is very likely that the encoded protein can improve an organism's fitness and is thus functional. This approach predicts a CUG-initiated, 173 amino acid, N-terminal extension within the 5′ leader of *phosphatase and tensin homologue on chromosome ten* (*PTEN*). *PTEN* is a powerful tumour suppressor gene that encodes a dual-specificity phosphatase [44,45] frequently mutated in human cancers [46] and autism spectrum disorders [47]. Its best characterized function is its ability to negatively regulate cell survival by dephosphorylating phosphatidylinositol 3,4,5 triphosphate (PIP3) and thus inhibiting phosphoinositide 3-kinase (PI3K) signalling [48].

Independently, two other groups subsequently identified the same PTEN N-terminal extension [49,50]. Here we extended our analysis of the *PTEN* 5′ leader and identify non-AUG-initiated translation that leads to the synthesis of at least two additional N-terminally extended proteoforms.

2. Material and methods

2.1. Plasmids

The *PTEN* 5′ leader was amplified by PCR from HEK-293T genomic DNA using appropriate primers (Integrated DNA Technologies) that incorporated a 5′ Hind*III* restriction site and a 3′ Bam*HI* restriction site. *PTEN* 5′ leaders were mutated by two-step PCR with appropriately designed primers. Amplicons were cloned Hind*III*/Bam*HI* into the dual luciferase plasmid p2-Luc [51] such that the *PTEN* 5′ leader replaced the *Renilla*-encoding sequences and were fused directly to the firefly-encoding sequences.

The coding sequence of PTEN was obtained as a gblock (Integrated DNA Technologies) with incorporated 5′ Hind*III* and a 3′ Xba*I* restriction site and cloned into phRL-CMV (Promega). The coding sequence of *PTEN* was also subcloned downstream of the *PTEN* 5′ leader to replace the firefly encoding sequence of the *PTEN* 5′ leader-FLuc constructs made previously.

For *PTEN* 5′ leader GFP fusions, EGFP was digested from pEGFP-N3 (Clontech) with Bam*HI* and Xba*I* restriction enzymes and cloned Bam*HI*/Xba*I* using standard cloning techniques into the *PTEN* 5′ leader firefly-encoding plasmids described above, such that the EGFP sequence replaced the firefly-encoding sequences and were fused directly to the *PTEN* 5′ leader. The EGFP AUG to AAA mutation was made by two-step PCR with appropriately designed primers.

For signal peptide–*Gaussia* luciferase fusions, amplicons generated by two-step PCR were cloned Bam*HI*/Xba*I* into pCMV-GLuc (NEB). All clones were verified by sequencing.

eIF1 and eIF5 overexpression constructs were described previously [52,53].

2.2. Cell culture

HEK-293T, MDA-MB-231, MCF-7, HeLa, HUH-7, U2OS and A172 cells were maintained in DMEM supplemented with 10% FBS, 1 mM L-glutamine and antibiotics. PC3 cells were maintained in RPMI supplemented with 10% FBS, 1 mM L-glutamine and antibiotics.

2.3. Luciferase assay

HEK-293T cells were transfected with Lipofectamine 2000 reagent (Invitrogen), using the 1 day protocol in which suspended cells are added directly to the DNA complexes in full-area 96-well plates. For each transfection, the following were added to each well: 100 ng of each firefly luciferase-expressing plasmid, 10 ng of each *Renilla* luciferase-expressing plasmid plus 0.4 µl Lipofectamine 2000 (Invitrogen) in 48.4 µl Opti-Mem (Gibco). The transfecting DNA complexes in each well were incubated with 4×10^4 cells suspended in 50 µl DMEM (RPMI for PC3 cells) plus 10% FBS. Transfected cells were incubated at 37°C in 5% CO_2 for 24 h. On the next day, cells were washed once with $1 \times$ PBS and then lysed in 25 µl of $1 \times$ passive lysis buffer (PLB; Promega) and firefly and *Renilla* luciferase activities were determined using the Dual Luciferase Stop & Glo® Reporter Assay System (Promega). Relative light units were measured on a Veritas Microplate Luminometer with two injectors (Turner Biosystems).

Firefly luciferase activity was calculated relative to the activity of the co-transfected control plasmid expressing *Renilla* luciferase (pSV40-*Renilla*). All data points were averaged, and the standard deviation calculated. Data represent the mean and standard deviation of at least three independent experiments each done in quadruplicate.

For secretion luciferase assays (see figure 5c), only *Gaussia* luciferase activities were assayed and the percentage activity in both the cell lysate (intracellular) and culture media (extracellular) calculated. For sodium arsenite treatment, HEK-293T cells were transfected as above, and sodium arsenite (5 µM) was added 6 h post-transfection for either 2 or 4 h.

2.4. Immunoblotting

Cells were transfected in six-well plates using Lipofectamine 2000 reagent, again using the 1 day protocol described above, with 1 µg of each indicated plasmid. Where FLuc- and RLuc-expressing plasmids were cotransfected (see figure 6b and electronic supplementary material, figure S10b) a ratio of 10 : 1 was used. The transfecting DNA complexes in each well were incubated with 0.8×10^6 HEK-293T cells suspended in 3 ml DMEM plus 10% FBS and incubated overnight at 37°C in 5% CO_2. Transfected cells were lysed in 100 µl $1 \times$ PLB and 10 µl removed for dual luciferase assay.

For PC3 transfections, 1.2×10^5 cells were plated in triplicate wells (12-well plates) 1 day prior to forward transfection with Lipofectamine 2000 reagent (4 µl) and 500 ng of each indicated plasmid in 500 µl of Opti-Mem. Cells were replenished with fresh RPMI media 6 h post-transfection, then after 18 h, cells were washed and treated with serum free RPMI for a further 24 h. Transfected cells were lysed in radioimmuno-precipitation assay (RIPA) buffer plus protease inhibitors (Sigma) and NaF (20 mM).

Proteins were resolved by 4–12% gradient Bis/Tris–SDS/PAGE (Bolt™: Thermo Fisher Scientific) under constant voltage (165 V) for 90 min and transferred to nitrocellulose membranes (Protran), which were incubated at 4°C overnight with primary antibodies. Immunoreactive bands were detected on membranes after incubation with appropriate fluorescently labelled secondary antibody using a LI-COR Odyssey® Infrared Imaging Scanner. Densitometry analysis was performed using Imagej software (NIH) and GraphPad Prism used for statistical analysis.

2.5. Immunoprecipitation and GFP-trap®

Cells were lysed in RIPA buffer plus protease inhibitors (Sigma), then lysates were incubated with 25 µl of protein G agarose beads (Pierce) plus anti-PTEN (138G6) overnight at 4°C with gentle rocking. The beads were washed (three times) with ice-cold RIPA buffer and then immunoprecipitated proteins removed from the beads by boiling for 5 min. in 20 µl of $2 \times$ SDS−PAGE sample buffer for electrophoresis and immunoblotting.

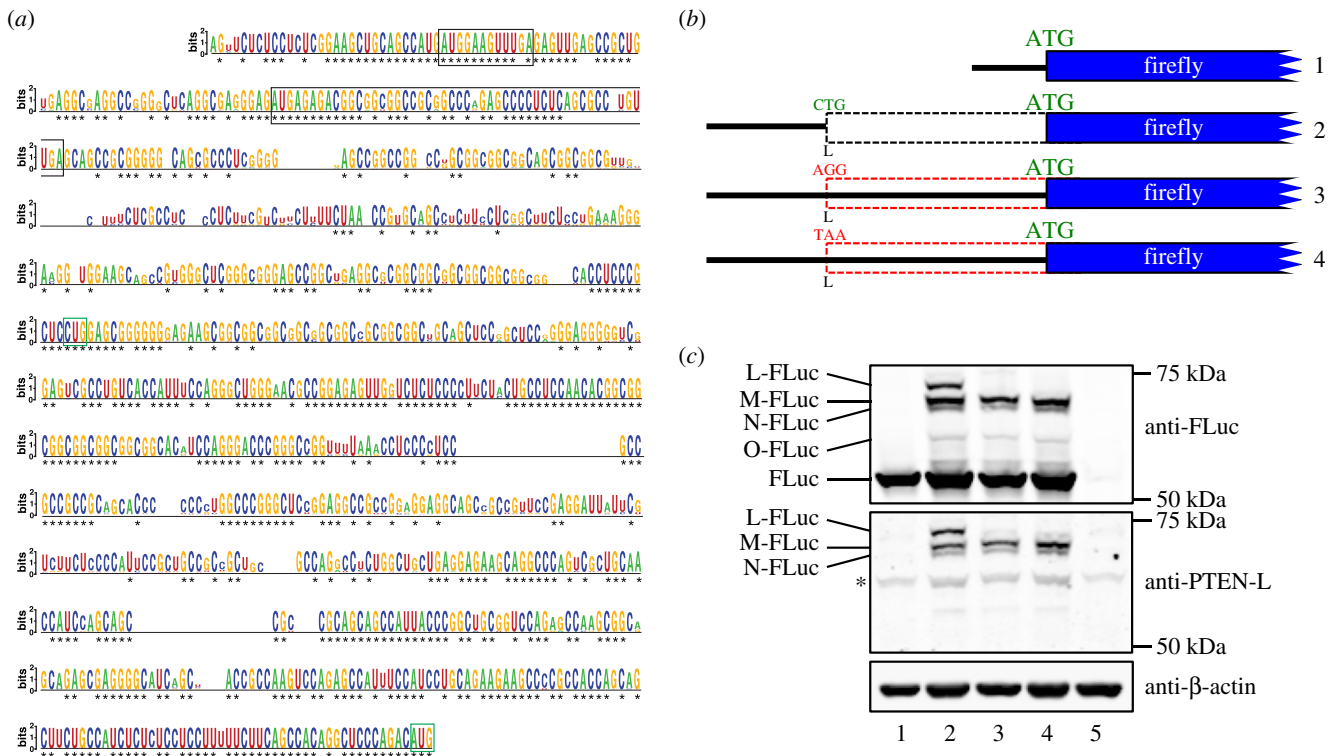GFP-trap®_A beads (ChromoTek) were equilibrated according to the manufacturer's protocol. For collection of the extracellular fractions, culture medium was centrifuged at 200g for 5 min at 4°C to remove debris. For intracellular fractions, cells were lysed in RIPA buffer as above. 10 µl of equilibrated beads were added to each fraction and incubated rotating at 4°C for 1 h. The beads were washed (three times) with ice-cold dilution buffer (10 mM Tris/Cl pH 7.5; 150 mM NaCl; 0.5 mM EDTA) and then immunoprecipitated proteins removed from the beads by boiling for 10 min in 40 µl of $2 \times$ SDS−PAGE sample buffer for electrophoresis and immunoblotting.

2.6. RT-qPCR

HEK-293T cells were transfected in triplicate wells (six-well plate) as above with either construct 1 (wild-type *PTEN* leader fused to firefly luciferase) or construct 5 (*PTEN* leader with all uAUGs mutated to AGG). 24 h post-transfection cells were removed (trypsin), divided into two aliquots of 20% and 80% and then collected by centrifugation. 20% of cells were resuspended in 20 µl $1 \times$ PLB for dual luciferase assay. RNA was isolated using Trizol reagent (Invitrogen) from the remaining 80% of cells, and 500 ng of DNAse-treated (RQ1: Promega) RNA was reverse transcribed using oligo-dT and random hexamers according to the manufacturer's instructions (Superscritpt III: Invitrogen). Reactions minus reverse transcriptase were included to control for contaminating genomic or plasmid DNA. SYBR green (Qiagen) qPCR was performed on an Applied Biosystems 7300 real-time PCR system with firefly luciferase primers (sense TGGAGAGCAACTGCATAAGG and antisense ATTCCGCGTACGTGATGTT) and a set of intron-spanning control primers for GAPDH (sense AGCCT CCCGCTTCGCTCTCT and antisense CCAGGCGCCCAA TACGACCA). Relative RNA quantitation was analysed using the Livak method $(2_T^{-\Delta\Delta C})$ and used to normalize relative luciferase activities to relative RNA levels.

2.7. Antibodies

An affinity-purified rabbit polyclonal antibody (anti-PTEN-L) directed to a predicted antigen (PRHQQLLPSLSSFFFSHR LPD) within all four extended PTEN proteoforms was prepared by GenScript. The following commercially available antibodies were also used. Mouse anti-PTEN (6H2.1; Millipore), rabbit anti-PTEN (138G6: Cell Signalling), rabbit anti-GFP (A6455: Novex), goat anti-firefly luciferase (G7451: Promega), rabbit anti-S473-phospho-AKT (D9E: Cell Signalling), mouse anti-pan-AKT (40D4: Cell Signalling), mouse anti-*Renilla* luciferase (1D5.2 Millipore), rabbit anti-eIF5 (ab85913: Abcam) and mouse anti-β-actin (AC-15: Sigma). Anti-eIF1 was a generous gift from Ariel Stanhill (Technion-Israel Institute of Technology).

Figure 1. (a) Sequence logo representation (produced with WebLogo [55]) of a multiple sequence alignment of *PTEN* 5′ leaders from 52 mammals. The alignment was generated with ClustalX [56] and corrected manually. Asterisks indicate nucleotides conserved in all 52 species. Open black boxes represent two conserved uORFs. Open green boxes indicate the main ORF AUG and previously predicted CUG initiation codons. (b) Illustration of transfected firefly luciferase constructs 1–4 used for immunoblotting. (c) Immunoblot of cell lysates prepared from HEK-293T cells transfected with firefly luciferase expressing constructs as indicated and probed with antibodies against firefly luciferase (anti-FLuc: top panel), the PTEN N-terminal extension (anti-PTEN-L: middle panel) and β-actin (bottom panel). The four proteoforms with extended PTEN N-termini are named as L, M, N and O. This nomenclature was recently proposed for novel PTEN proteoforms by Pulido *et al.* [57]. Asterisk indicates a non-specific protein that co-migrates with the O-proteoform, thus precluding its detection with anti-PTEN-L. Lane 5 contains cell lysates from mock-transfected cells.

2.8. Fluorescence microscopy

Live cell imaging was performed as described before [54] using an inverted Axiovert 200 fluorescence microscope (Zeiss), equipped with 100×/1.4 Plan Apochromat oil-immersion objective (Zeiss), pulsed excitation module (470 nm, 590 nm LEDs), bandpass filters 510–560 nm and gated CCD camera (LaVision, Biotec). Briefly, HeLa cells were seeded onto eight-well chambers pre-coated with a mixture of collagen IV and poly-D-lysine (Ibidi), allowed to attach (24 h) and forward transfected for 24 h with plasmid DNAs encoding PTEN-leader–GFP fusions as indicated. Images were processed using ImSpector software (LaVision, Biotec) and combined in Adobe Illustrator CS2.

3. Results

Previous searches for evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences predicted a CUG-initiated, 173 amino acid extension to the tumour suppressor PTEN [37]. Further phylogenetic analysis of the *PTEN* 5′ leader with additional sequence data reveals deep nucleotide conservation in mammals (figure 1a and electronic supplementary material, figure S1). Two independent groups [49,50] have recently provided experimental evidence for a human PTEN N-terminal extension that reportedly initiates at the same CUG predicted by Ivanov *et al.* [37]. Preliminary results in our laboratory indicated the existence of multiple N-terminally extended PTEN proteoforms. Here we set out to systematically investigate these multiple proteoforms and also to determine the effect, if any, of two conserved AUG-initiated uORFs on translation of these PTEN proteoforms (boxed in figure 1a).

We noted that the 5′ end of both GenBank RefSeq *PTEN* mRNA isoforms (NM_000314.6, NM_001304718.1), which have identical first exons, do not correspond to the transcription start site predicted by the Fantom Projects' CAGE analysis [58] which finds that the transcription start site is a further 187 nt 3′ of the Genbank RefSeq annotated *PTEN* mRNA 5′ end (electronic supplementary material, figure S2). A +187 transcription start site is also in agreement with mRNAseq data obtained as controls to multiple ribosome profiling experiments, available in GWIPS-viz [59] (see electronic supplementary material, figure S2) as well as with the majority of publicly available human *PTEN* expressed sequence tags. Therefore, in this study, all test constructs with the *PTEN* 5′ leader start at +187 relative to the 5′ end of GenBank RefSeq *PTEN* mRNAs.

We first transfected HEK-293T cells with plasmid DNA expressing the human *PTEN* 5′ leader fused to sequences encoding firefly luciferase (FLuc; figure 1b). Immunoblots generated from transfected cell lysates and probed with FLuc antibodies detected four slower migrating FLuc proteoforms when FLuc is preceded by the *PTEN* 5′ leader (lane 2, top panel, figure 1c). These four proteoforms are absent from cells transfected with FLuc expressing constructs lacking the *PTEN* 5′ leader (lane 1, top panel, figure 1c) indicating that the multiple proteoforms are not post-translationally modified variants of FLuc. Furthermore, the same pattern of proteoforms
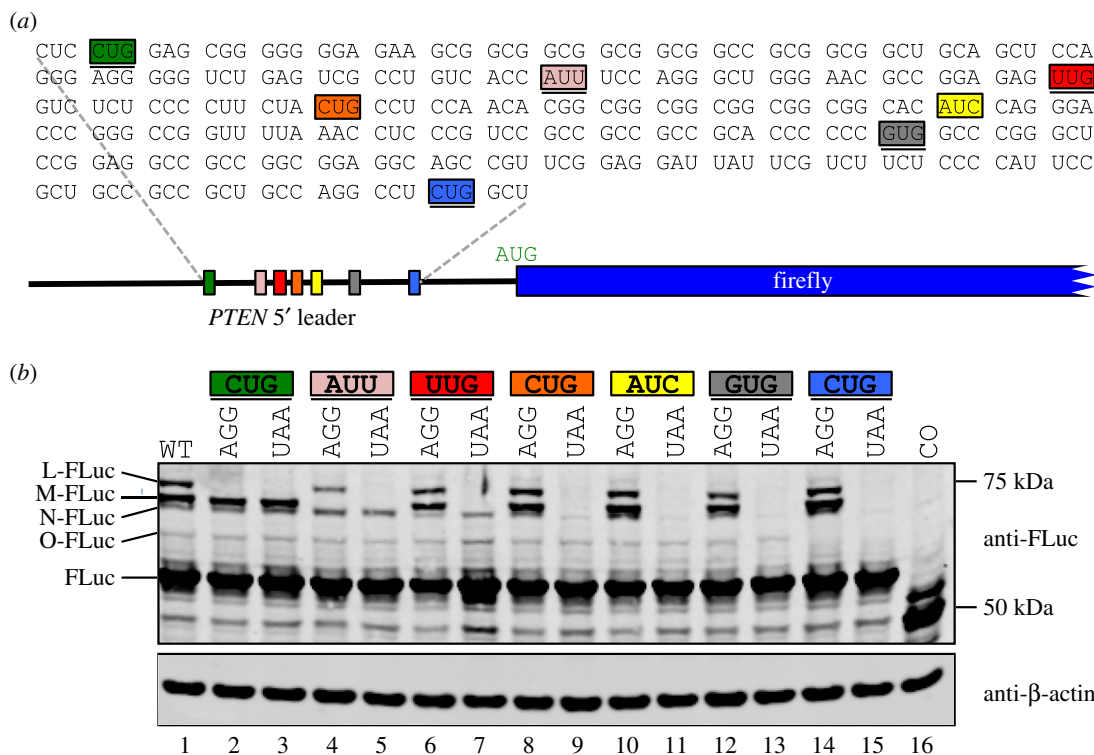
Figure 2. (a) Illustration of human *PTEN* 5′ leader fused to firefly luciferase and sequence of 5′ leader region in which potential non-AUG codons are highlighted in different colours with those in favourable Kozak context (−3 purine and/or +4 G) underlined. The previously reported CUG initiation site [37,49,50] is highlighted in green and is the most 5′ potential non-AUG initiation site. The *PTEN* CDS reading frame is indicated with spaces between the codons. (b) Anti-FLuc and anti-β-actin immunoblots of cell lysates prepared from HEK-293T cells transfected with firefly luciferase encoding sequences fused to the wild-type or non-AUG codon mutated (indicated) *PTEN* 5′ leader. The four FLuc proteoforms with extended PTEN N-termini are indicated as L-, M-, N- and O-FLuc. In the control (CO, lane 16), the main ORF (firefly) AUG is mutated to UAA.

was also detected when these same lysates were probed with a custom antibody (anti-PTEN-L) directed against a peptide predicted from sequences immediately 5′ of, and in-frame with, the *PTEN* main ORF (lane 2, middle panel, figure 1c).

These multiple proteoforms are most likely N-terminally extended variants generated by initiation at in-frame non-AUG codons within the *PTEN* 5′ leader. Alternatively, some of these proteoforms could be proteolytically cleaved variants of the previously reported [49,50] CUG-initiated 173 amino acid N-terminally extended PTEN proteoform. To investigate this latter possibility, we transfected HEK-293T cells with *PTEN* 5′ leader-FLuc expressing constructs in which this CUG was changed to either a non-initiating AGG triplet or to a UAA stop codon, that not only prevents initiation but also terminates translation from upstream initiation sites. Both anti-FLuc and anti-PTEN-L immunoblots from these transfected lysates indicate that mutation of the CUG leads to the disappearance of only the most slowly migrating proteoform, thus ruling out the possibility that some of the other proteoforms are cleavage products (lanes 3 and 4, top and middle panels, figure 1c). These observations suggest that the *PTEN* 5′ leader has the potential to generate at least four N-terminally extended proteoforms (that are within the detection limits of these experiments) and that the previously reported proteoform initiated at CUG, while likely the longest proteoform may not be the most abundant.

Because of the recent reporting of an N-terminally extended PTEN proteoform [37,49,50], a unified nomenclature for PTEN proteoforms was proposed [57] where newly identified PTEN proteins are named alphabetically as PTEN-L, PTEN-M, PTEN-N, etc. We have adopted this proposed

Table 1. Details of PTEN N-terminally extended proteoforms. Nucleotide distances from aAUG (annotated AUG of canonical PTEN) are indicated where A of the aAUG is +1.

	nt from aAUG	N-term ext AA	total nt	total AA	MW (kDa)
PTEN-L	−519	173	1728	576	64.9
PTEN-M	−438	146	1647	549	62.5
PTEN-N	−393	131	1602	534	61.0
PTEN-O	−216	72	1425	475	55.0

nomenclature and henceforth refer to these four PTEN isoforms as PTEN-L, PTEN-M, PTEN-N and PTEN-O, where PTEN-L is the presumed longest variant and is initiated at the previously reported CUG codon [37,49,50], whereas proteoform PTEN-M appears to be the most abundant (table 1 and figure 1c).

It is conceivable that PTEN-M, PTEN-N and PTEN-O are not non-AUG-initiated N-terminally extended proteoforms. Instead, PTEN-M and PTEN-N could be post-translationally modified variants of PTEN-O, whereas PTEN-N and PTEN-O could be cleavage products of PTEN-M. To address these possibilities, and to determine whether non-AUG initiation could explain the presence of these proteoforms, we made *PTEN* 5′ leader-FLuc constructs in which potential non-AUG initiation codons were systematically mutated in turn to either AGG or UAA. Potential in-frame near-cognate initiation codons from the relevant region within the *PTEN*
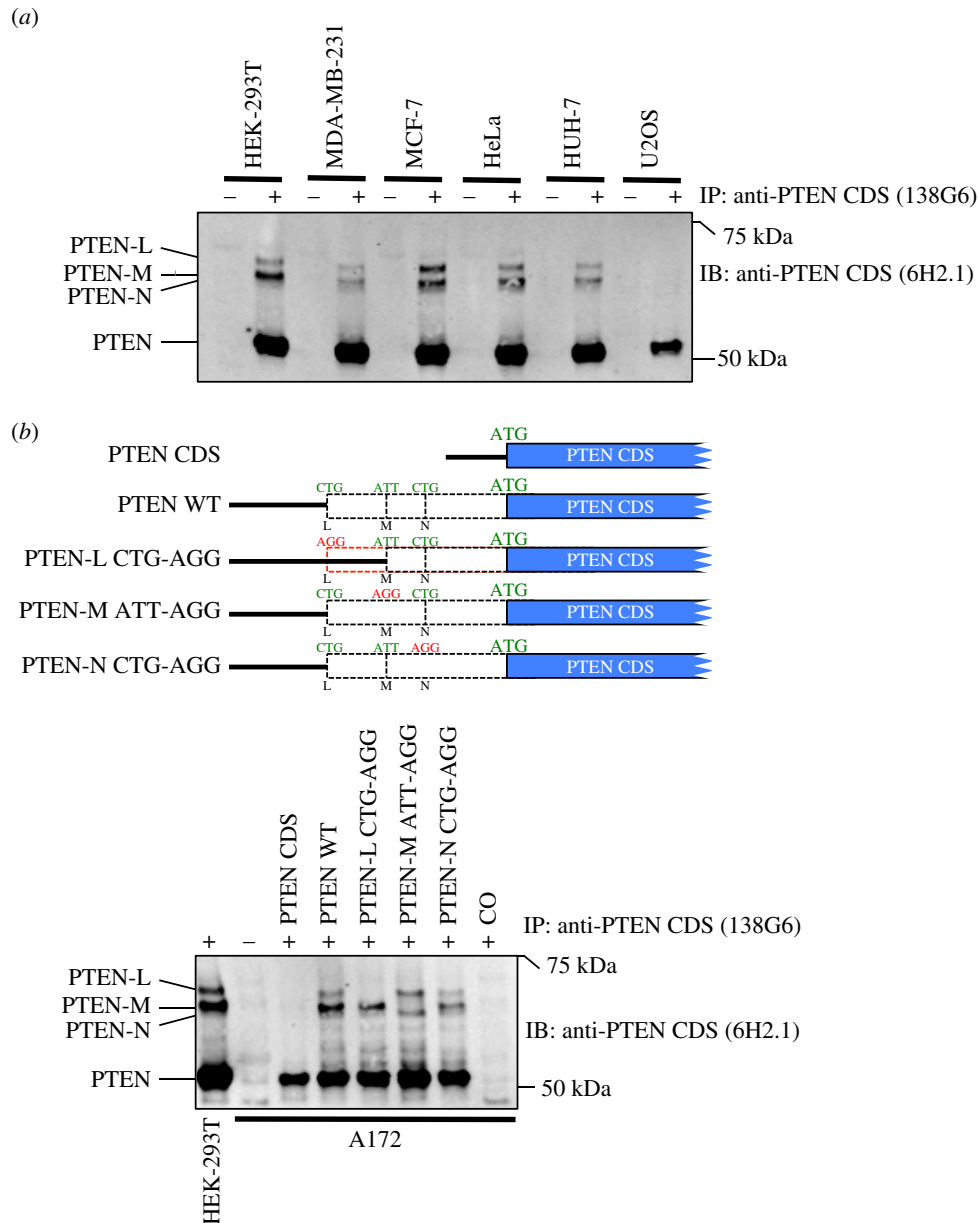
Figure 3. (a) Anti-PTEN CDS (6H2.1) immunoblot of anti-PTEN CDS (138G6) immunoprecipitates prepared from several cell lines as indicated showing detection of endogenous canonical AUG-initiated PTEN as well as the three non-AUG-initiated PTEN proteoforms PTEN-L, PTEN-M and PTEN-N. (b) Anti-PTEN CDS (6H2.1) immunoblot of anti-PTEN CDS (138G6) immunoprecipitates prepared from A172 cells (which lack endogenous PTEN) transfected with constructs expressing either wild-type PTEN or mutants of PTEN proteoforms PTEN-L, PTEN-M and PTEN-N (top panel illustration) showing that the non-AUG initiation codons of PTEN-L, PTEN-M and PTEN-N are CUG, AUU and CUG, respectively. The left-most lane shows control immunoprecipitates from HEK-293T cells indicating endogenous canonical AUG-initiated PTEN and N-terminally extended proteoforms PTEN-L, PTEN-M and PTEN-N. CO indicates control immunoprecipitates from mock-transfected cells.

5′ leader are highlighted in figure 2a and those in a favourable Kozak context (purine at −3 or G at +4) are underlined. Mutation of the most 5′ in-frame AUU codon to AGG completely abolishes synthesis of the most prevalent proteoform M-FLuc (lanes 4 and 5, figure 2b). As expected, mutation of this AUU to a termination codon causes premature termination of L-FLuc (lane 5, figure 2b). It is difficult to be certain about the nature of the initiation codon for N-FLuc, because a cross-reacting endogenous protein migrates at the same position; however, it seems likely that the second-most 5′ in-frame CUG is responsible as there is a clear decrease in N-FLuc intensity when this CUG and all 3′ non-AUGs are mutated individually to UAA (see lanes 9, 11, 13 and 15, figure 2b). Mutation of the most 3′ in-frame CUG completely abrogates expression of O-FLuc (lanes 14 and 15, figure 2b). In summary, the three minor proteoforms

(L, N and O) are all initiated at CUG, whereas the major proteoform (M) is initiated at AUU.

We next tested for the existence of endogenous human PTEN N-terminally extended proteoforms by immunoprecipitating PTEN from several different cell lines using commercially available antibodies directed against antigens within the annotated PTEN (CDS). The predicted molecular weight of PTEN is 47.2 kDa, but there are many reports indicating that the apparent molecular weight of PTEN is approximately 55 kDa. In agreement with this, immunoblots of PTEN immunocomplexes reveal the presence of an approximately 55 kDa protein in all cell lines tested (figure 3a). In addition to the canonical AUG-initiated PTEN, three slower migrating proteins are observed in all cell lines tested other than U2OS. The molecular weights of these three proteins correlate well with those predicted for PTEN

proteoforms PTEN-L, PTEN-M and PTEN-N. To determine whether the migration of these putative endogenously expressed PTEN N-terminally extended proteoforms correlate with exogenously expressed proteoforms PTEN-L, PTEN-M and PTEN-N (we could not detect any endogenous proteoform that could correspond in molecular weight to PTEN-O in these experiments), we transfected the human *PTEN* CDS fused to either wild-type or mutated *PTEN* 5′ leaders into A172 cells which do not express endogenous *PTEN*. Immunoprecipitates from transfected A172 cells were compared with immunoprecipitates from HEK-293T cells expressing endogenous *PTEN* and show that exogenously expressed proteoforms PTEN-L, PTEN-M and PTEN-N co-migrate with endogenous PTEN proteins from HEK-293T cells (compare the first and fourth lanes in figure 3*b*). Furthermore, in agreement with the FLuc reporter constructs (figure 2*b*), mutation of the first and third in-frame CUGs prevents initiation of the two minor PTEN proteoforms PTEN-L and PTEN-N, whereas mutation of the first in-frame AUU abolishes expression of the major PTEN-M proteoform (figure 3*b*). Together, these results indicate that three slower migrating PTEN proteoforms apparent in immunoprecipitates from several cell lines correlate with non-AUG-initiated proteoforms PTEN-L, PTEN-M and PTEN-N.

Several studies report that PTEN-L is an active phosphatase and retains the ability to downregulate the PI3K pathway [49,50,60–62]. To determine whether the PTEN proteoforms described here are active phosphatases, we measured the phosphorylation status of the major PI3K substrate, AKT, in PC3 cells (no endogenous PTEN expression) transfected with either wild-type or N-terminally extended PTEN proteoforms. Exogenous expression of PTEN reduced the levels of AKT phosphorylation almost twofold, and similar levels of reduction were observed for all four N-terminally extended proteoforms (figure 4 and electronic supplementary material, figure S3). Therefore, similar to previous observations for PTEN-L, the phosphatase activities of PTEN-M, PTEN-N and PTEN-O are not overtly affected by their N-terminal extensions.

N-terminal extensions often harbour signals for subcellular targeting or secretion. However, live cell imaging of AUG-initiated PTEN N-terminal extensions fused to GFP reveal diffuse cytoplasmic localization for all four PTEN–GFP fusions indistinguishable from the localization of GFP alone (electronic supplementary material, figure S4).

Hopkins *et al.* [49] suggest that the PTEN-L proteoform harbours an N-terminal signal peptide secretion signal and provide evidence that PTEN-L is secreted and can re-enter cells via a cell re-entry motif similar to the HIV TAT protein. The predicted amino acid sequences of PTEN-M and PTEN-N lack the putative secretion signal yet still retain the putative cell re-entry motif (electronic supplementary material, figure S5) reported by Hopkins *et al*. To rule out the possibility that the N-terminal extension of PTEN-M might possess a cryptic signal peptide sequence, we overexpressed, in HEK-293T cells, PTEN-L (as a control) and PTEN-M N-terminal extensions fused to GFP (same constructs as described in electronic supplementary material, figure S4). A fusion of GFP with the signal peptide sequence from *Gaussia* luciferase (GLuc) [63] was used as a positive control (figure 5*a*). Extracellular and intracellular GFP fusion proteins were concentrated by immunoprecipitation using GFP-trap® (immobilized camelid anti-GFP antibody) followed by immunoblotting with anti-GFP. Even though we immunoblotted 50% of the total extracellular fraction and only 5% of the intracellular fraction, we did not detect PTEN-
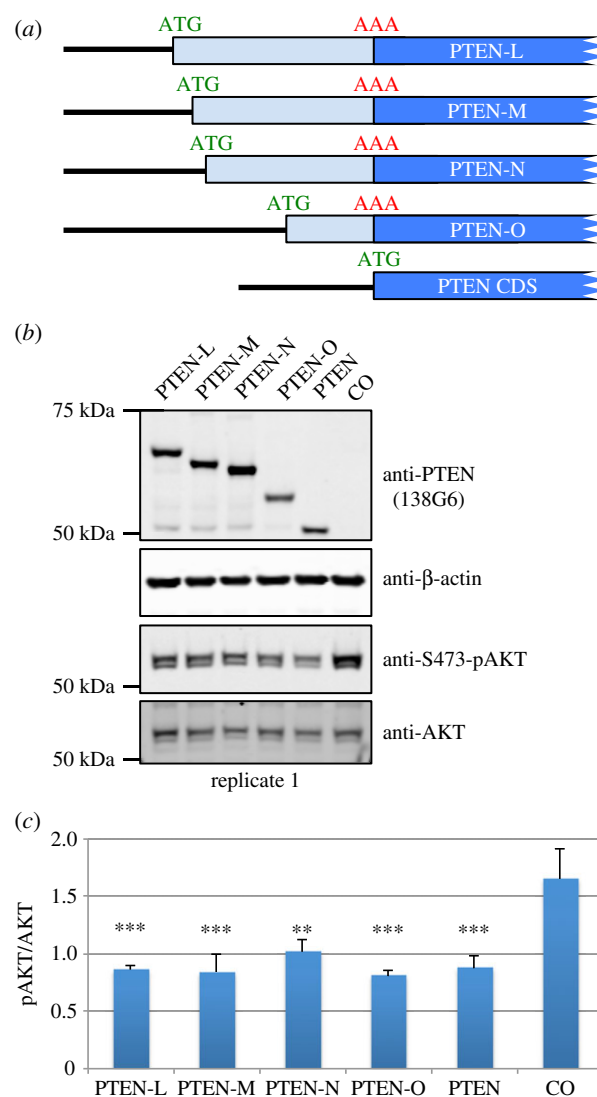
Figure 4. (*a*) Illustration shows PTEN expressing constructs transfected into PC3 cells. (*b*) Immunoblots of cell lysates prepared from PTEN-null PC3 cells transfected with PTEN expressing constructs as indicated for 48 h (serum starved for last 24 h) and probed with antibodies against PTEN (138G6), β-actin, phospho-AKT (S473) and pan-AKT. Additional replicates (replicates 2 and 3) are shown in the electronic supplementary material, figure S3. (*c*) Mean and standard deviations of relative protein intensities determined by densitometry from three biological replicates. Phospho-AKT intensities were calculated relative to pan-AKT intensities. Relative pAKT levels in lysates from cells transfected with each N-terminally extended PTEN proteoform were compared with the control sample. **$p < 0.01$, ***$p < 0.001$ by one-way ANOVA followed by Tukey's test.

L-GFP or PTEN-M-GFP in the extracellular fraction (see lanes 1 and 3, middle panel, figure 5*b*). Furthermore, when HEK-293T or U2OS cells were transfected with constructs expressing either the PTEN-L or PTEN-M N-terminal extensions, or the putative PTEN-L signal peptide alone, fused to GLuc, we failed to detect GLuc activity in the cell media at levels above background (figure 5*c*). Yet GLuc-fused downstream of the PTEN-L N-terminal extension in which the putative PTEN-L signal peptide was replaced with the signal peptide from either GLuc or interleukin-2 efficiently targeted GLuc from cells (figure 5*c*).

We next sought to ascertain whether the non-AUG initiation of the PTEN proteoforms is regulated. The selection of poor context initiation codons (including non-AUG start codons) is modulated by intracellular levels of initiation factors eIF1
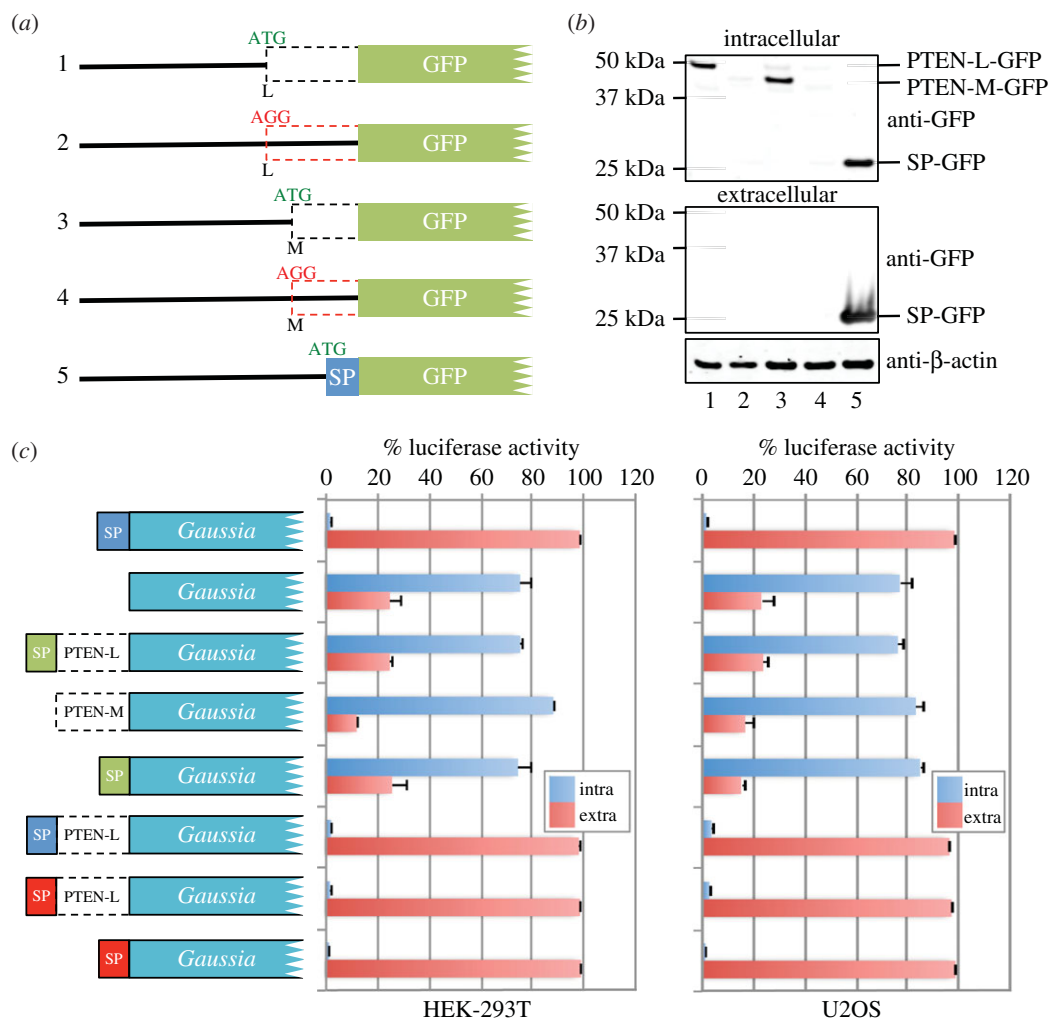
Figure 5. (a) Illustration of mutant *PTEN-L* and *PTEN-M* 5′ leader–GFP fusion constructs transfected into HEK-293T cells. SP is the secretion peptide from *Gaussia* luciferase. (b) Anti-GFP and anti-β-actin immunoblots from GFP-trap immunoprecipitates prepared from either HEK-293T cell lysates (intracellular: top panel) or culture media (extracellular: middle panel) transfected with the GFP fusion constructs shown in (a). (c) *Gaussia* luciferase assays show the percentage luciferase activity in the cell lysate (blue: intracellular) and culture media (red: extracellular) when either HEK-293T or U2OS cells (as indicated) were transfected with the constructs indicated for 22 h. The blue signal peptide (SP) is from *Gaussia* luciferase (MGVKVLFALICIAVAEAK), the putative PTEN-L SP (MERGGEAAAAAAAAAAAPGRG) is in green and the interleukin-2 (IL2) SP (MYRMQLLSCIALSLALVTNSA) is in red.

(increases stringency) and eIF5 (decreases stringency) [5–11, 52,53]. To determine whether initiation of the N-terminally extended PTEN proteoforms are regulated by altered eIF1 or eIF5 levels, we overexpressed each initiation factor separately in HEK-293T cells and then immunoprecipitated endogenous PTEN proteoforms (electronic supplementary material, figure S6). Even though eIF1 and eIF5 levels are robustly expressed, we note no discernible change in the ratio of PTEN proteoforms compared with cells transfected with an empty vector (electronic supplementary material, figure S6).

As shown in figure 1*a*, there are two conserved AUG-initiated uORFs (uORF1 and uORF2) close to the *PTEN* 5′ cap (figure 6*a* and electronic supplementary material, figure S7). The most 5′ uORF (uORF1) is only four codons long and starts with tandem AUG codons, both in good Kozak context. uORF2 is much longer (45 codons) and also starts with a good context AUG. Ribosomal profiling data compiled in GWIPS-viz [59] for *PTEN* show a large number of ribosome protected fragments aligning to uORF1 in comparison with uORF2 (electronic supplementary material, figure S7). In general, translation of uORFs has an inhibitory effect on main ORF translation although this relationship between uORFs and main ORFs is not so simple where multiple uORFs exist.

Usually, the translation of short uORFs is less inhibitory than the translation of long uORFs, because the ribosomes' ability to reinitiate after translation of ORFs more than 35 codons is normally greatly reduced (see Introduction). According to the scanning model of eukaryotic translation initiation, we would predict that the majority of scanning 43S complexes would initiate uORF1; however, because it is only four codons long, a high level of re-initiation is expected. Because re-initiation is precluded until all necessary initiation factors have reloaded onto the 40S, translation of uORF1 may favour re-initiation at ORFs more 3′ than uORF2. We set out to determine the possible role of these two uORFs on regulation of the different PTEN proteoforms by transfecting HEK-293T cells with *PTEN* 5′-leader-FLuc reporter constructs in which the uORFs were mutated and testing by dual luciferase assay (figure 6*a*) and immunoblotting (figure 6*b*).

Even though we expect a high level of uORF1 translation, mutation of the uORF1 tandem AUG codons to non-initiating AGG codons has only a minor inhibitory effect (less than 10%) on FLuc activity (compare construct 1 (wild-type *PTEN* 5′ leader) with construct 2 in figure 6*a*). Mutating uORF1 is inhibitory rather than stimulatory presumably because more ribosomes now have access to uORF2 which we predict
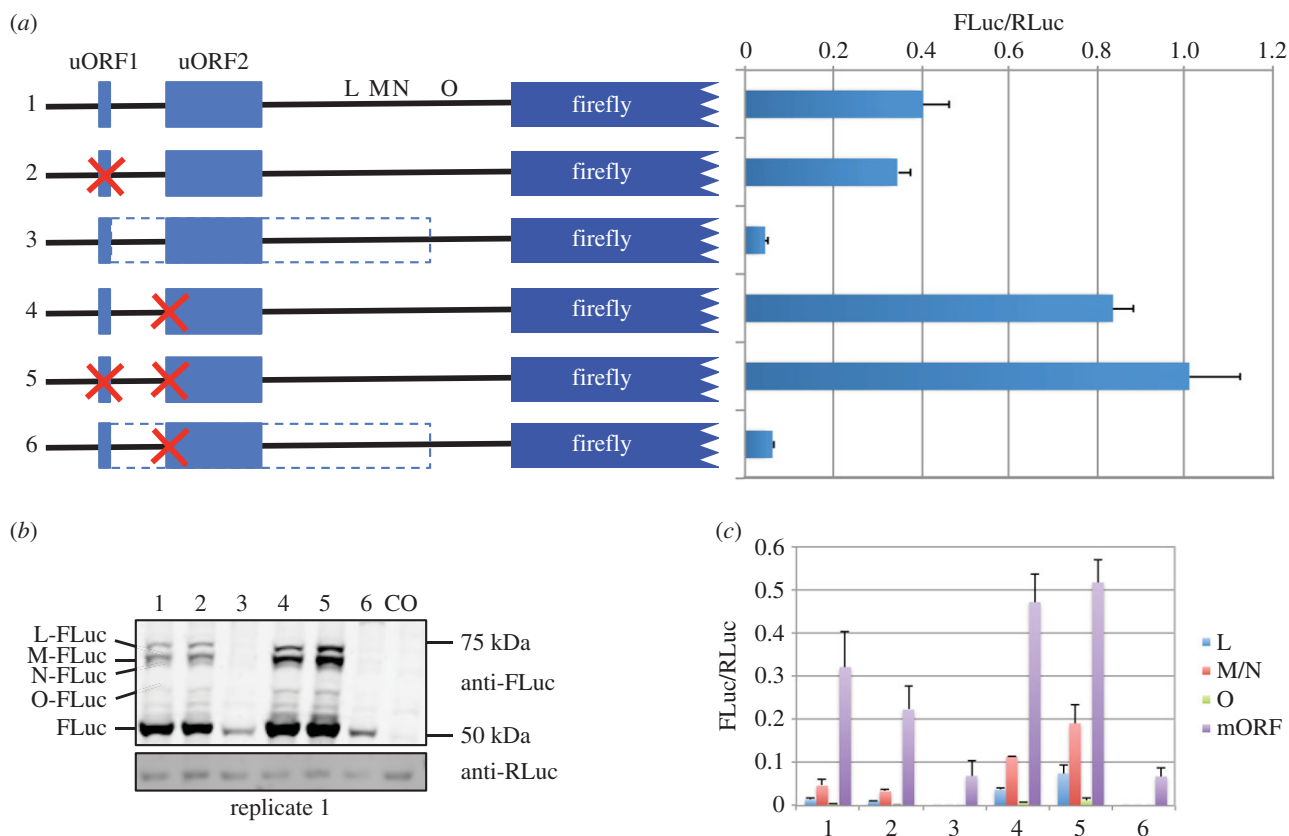
Figure 6. (a) Relative luciferase activities (FLuc/RLuc) of firefly encoding sequences fused to the wild-type or mutant *PTEN* 5' leader as indicated and cotransfected (10 : 1 ratio) with a *Renilla* expressing plasmid into HEK-293T cells. Red crosses indicate mutation of AUG start codons to non-initiating AGG codons. L, M, N and O depict the approximate site of initiation of PTEN extensions. The dashed box represents the increase in ORF length when the stop codon of uORF1 is changed to a sense codon. (b) Anti-FLuc and anti-RLuc immunoblots of cell lysates prepared from HEK-293T cells transfected with *Renilla* and firefly luciferase expressing constructs indicated in (a). The four proteoforms with extended PTEN N-termini are indicated as L, M, N and O. CO represents lysates prepared from mock-transfected cells. (c) Densitometry analysis from three biological replicates of the proteins detected by anti-FLuc and anti-RLuc in (b) and electronic supplementary material, figure S10b. Proteoforms M and N could not be resolved sufficiently from each other for accurate densitometry analysis, so the intensity of both proteins together is determined. FLuc intensities were calculated relative to RLuc.

should be inhibitory. However, the fact that the inhibition is only minor suggests that either uORF2 may not be as inhibitory as we expect or that normally ribosomes translating uORF1 can efficiently reinitiate uORF2. Another explanation could be that in the wild-type context many ribosomes do not initiate uORF1. However, mutation of the uORF1 stop codon to a sense codon (extending the length of uORF1 to 161 codons) severely diminishes FLuc activity (construct 3, figure 6a) which, in agreement with published ribosome profiling data (electronic supplementary material, figure S7), confirms that ribosomes do initiate at uORF1. A mutation that disables uORF2 translation, either alone or in combination with mutations that prevent uORF1 initiation, increased FLuc activity more than twofold (compare construct 1 (wild-type) with constructs 4 and 5 in figure 6a). This indicates that uORF2 is inhibitory and thus suggests that the level of re-initiation on uORF2 is high. Similar results were observed in several other cell lines tested, including breast carcinoma (MCF-7), prostate carcinoma (PC3) and cervical carcinoma (HeLa; electronic supplementary material, figure S8). In addition, there is little difference in steady-state mRNA levels when both uORFs are disabled (electronic supplementary material, figure S9).

In these reporter assays, it is assumed that the low level of N-terminally extended proteoforms (relative to the main ORF) have only a minor contribution to total FLuc activity. This is supported by our own unpublished observations showing that an AUG-initiated PTEN-L extension severely reduces

FLuc activity. To gain further understanding of the effect of the uORFs on downstream translation, we performed anti-FLuc immunoblots from cells transfected with constructs 1–6 (figure 6b and electronic supplementary material, figure S10). Dual luciferase assays from replicate lysates show similar FLuc activity for each construct to those shown in figure 6a and electronic supplementary material, figure S8. Densitometry of the main ORF (FLuc) normalized to cotransfected RLuc levels indicates that, similar to the luciferase assay results, preventing initiation of the uORFs results in an almost twofold increase in main ORF levels (figure 6c). Therefore, even though levels of the N-terminally extended proteoforms increase when uORF2 is mutated (lanes 4 and 5, figure 6b and electronic supplementary material, figure S10), they have only a minor contribution to overall FLuc activity. When we estimate the levels of each N-terminally extended proteoform by densitometry, we observe that mutation of uORF2 causes the levels of the L and M/N proteoforms to increase approximately 2.5-fold, the O proteoform to increase 2-fold, whereas the main ORF increases approximately 1.5-fold. Mutation of both uORFs together increases the L-proteoform approximately 5.3-fold, M/N-proteoforms approximately 4.1-fold, O-proteoform approximately 3.1-fold and the main ORF proteoform approximately 1.6-fold. This densitometry analysis also allows us to estimate the relative abundance of each proteoform under normal conditions and when ribosomes do not translate either or both uORFs. Interestingly,

although all proteoforms increase when ribosomes do not translate uORF2, the ratio of N-terminally extended proteoforms relative to each other does not change (electronic supplementary material, figure S10c). In contrast, mutation of uORF2 increases the ratio of N-terminal proteoforms relative to the main ORF proteoform (electronic supplementary material, figure S10d) such that proteoforms M/N increase by 50% from 12% to 18% of all proteoforms. Similar approximately 50% increases were also observed for the L- and O-proteoforms (electronic supplementary material, figure S10d). Furthermore, there is a concomitant decrease in the relative abundance of the main ORF proteoform from 83% down to 75%. Mutation of uORF1 and uORF2 together results in even further increases to the relative abundance of N-terminally extended proteoforms and further decreases the main ORF proteoform (electronic supplementary material, figure S10d).

Because we show that translation of uORF2 can alter the ratio of extended and main ORF proteoforms and uORF2 translation seems to be dependent on efficient re-initiation after translation of uORF1, we predict that conditions which can regulate re-initiation events could alter translation of the main ORF. Increasing the phosphorylation status of eIF2 can reduce re-initiation by limiting the pool of functional (non-phosphorylated) eIF2. To gain some understanding of possible regulation of these uORFs, we transfected wild-type and 'no uAUG' firefly reporters into HEK-293T cells and then treated with sodium arsenite. Sodium arsenite activates the integrated stress response by inducing phosphorylation of eIF2. Many mRNAs (main ORFs) whose translation is resistant to eIF2 phosphorylation harbour translated uORFs [64]. If sodium arsenite decreased re-initiation on uORF2, then we would expect to see an increase in FLuc activity with the wild-type reporter; however, we observed little difference in main ORF reporter translation upon sodium arsenite treatment (electronic supplementary material, figure S11).

4. Discussion

The results presented above provide strong evidence for the existence of three (or perhaps four) non-AUG-initiated proteoforms of PTEN that are expressed in human cells in addition to, and at lower levels than, the well-studied canonical AUG-initiated PTEN. The longest PTEN proteoform, PTEN-L, has previously been reported [37,49,50] and its function has been investigated in more detail in subsequent studies [60,61,65–67]. However, PTEN-M, PTEN-N and PTEN-O have not been previously described and are reported here for the first time. We show that PTEN-M initiates at an AUU codon that is completely conserved in 52 eutherian mammalian species with available sequences, whereas the other PTEN proteoforms (PTEN-L, PTEN-N and PTEN-O) initiate at CUG, the first two also completely conserved while the latter only partially conserved. CUG codons are generally better initiators than AUU codons [14,52,53], so it is somewhat surprising that AUU-initiated PTEN-M is more abundant than CUG-initiated PTEN-L, especially, because the CUG is more 5′ than the AUU. The PTEN-L CUG initiation codon Kozak context is slightly less favourable (C at −3 and G at +4) than that of the PTEN-M AUU initiation codon (A at −3). It is also possible that an as yet unidentified RNA secondary structure 3′ of the AUU codon contributes to its favourable utilization for initiation (see Introduction), especially because the *PTEN* 5′

leader is 70% GC rich with numerous potential stem-loops. Yet another possibility is that the abundance of the PTEN proteoforms is a reflection of differing protein stabilities rather than initiation levels, although a similar level of exogenous PTEN-L and PTEN-M expression when their non-AUGs start codons are mutated to AUGs (figure 4b and electronic supplementary material, figure S3) would argue against this.

The identity of the major non-AUG-initiated PTEN proteoforms identified in this study contrasts with the findings of both Hopkins *et al.* [49] and Liang *et al.* [50], who both report only a single CUG-initiated proteoform that corresponds in our study to PTEN-L. Most of the anti-PTEN immunoblots presented by Hopkins *et al.* show a single slower migrating approximately 75 kDa PTEN proteoform which they term PTEN-Long (PTEN-L in our study). A possible reason for this discrepancy is that, in their study, proteins were separated for shorter time intervals while we purposely allowed SDS−PAGE gels to run for extended periods (see Material and methods) in an attempt to resolve as many PTEN proteoforms as technically possible. Perhaps the approximately 75 kDa PTEN-long detected by Hopkins *et al.* corresponds to a mixture of unresolved PTEN-L, PTEN-M and PTEN-N.

Liang *et al.* similarly identify a PTEN proteoform (PTEN-α) that initiates with the most 5′ in-frame CUG and corresponds to our PTEN-L and Hopkins *et al.* PTEN-Long. Their anti-PTEN immunoblots from cells exogenously expressing a PTEN-α construct clearly show two higher molecular weight proteins in addition to PTEN. In agreement with our study, mutation of the first in-frame CUG prevents translation of the longest protein. Furthermore, even when the first and third CUG codons (PTEN-L and PTEN-N in our study) are mutated together, an extended PTEN proteoform is still apparent which is very likely to correspond to PTEN-M. A recent report showing anti-PTEN immunoblots from matched normal and tumour tissue samples clearly identifies, in addition to canonical PTEN, at least two slower migrating endogenous PTEN proteoforms expressed in a similar ratio to PTEN-L and PTEN-M in our study [61].

When multiple proteoforms are translated from a single mRNA, the efficiency of initiation at each start codon could set the ratio of proteoform steady-state levels assuming each protein has similar stability. However, this ratio may vary during conditions in which initiation efficiency is altered. Eukaryotes have developed elaborate mechanisms for the recognition of the correct initiation codon and the levels of certain initiation factors can regulate the fidelity of initiation, especially on suboptimal (non-AUG and AUG in poor context) start codons [52,53,68]. While elevated levels of eIF1 can increase the stringency of start codon selection, elevated levels of eIF5 have the opposite effect. Here we show that overexpression of either eIF1 or eIF5 had minimal effect on the steady-state levels of PTEN proteoforms (electronic supplementary material, figure S6). This suggests that either the non-AUG initiation events in *PTEN* are refractory to normal stringency controls or the steady-state protein levels of these proteoforms are regulated tightly by rapid turnover. An alternative explanation is that because the *PTEN* 5′ leader is long, the many potential out-of-frame near-cognate codons could create uORFs and thus preclude the expected effects of eIF1/5 overexpression on translation of the PTEN N-terminal extensions.

Several groups have observed that PTEN-L can downregulate the PI3K pathway in a similar manner to PTEN [49,50,61]. *In vitro* studies comparing the catalytic activities of purified

PTEN and PTEN-L reveal that both enzymes can dephosphorylate PIP3, although, interestingly, PTEN phosphatase activity can be activated by its reaction product (PIP2), whereas PTEN-L cannot and is thus constitutively active [60,67]. We tested whether the PTEN proteoforms identified in our study still retained the ability to downregulate the PI3K pathway (figure 4 and electronic supplementary material, figure S3). All PTEN proteoforms were able to reduce AKT phosphorylation to levels similar to those of canonical PTEN and PTEN-L, suggesting that the N-terminal extensions do not have major effects on the dephosphorylation activity of PTEN proteoforms.

Hopkins *et al.* [49] report that exogenously expressed PTEN-L is targeted for secretion from cells by a predicted N-terminal signal peptide and cleavage site. Furthermore, they also show that immediately following the predicted cleavage site is a functional cell re-entry signal similar to the HIV TAT protein [49]. Subsequently, Wang *et al.* [61] confirmed that PTEN-L, but not PTEN, can enter cells, although whether PTEN-L can be secreted from cells was not tested. Intriguingly, both PTEN-M and PTEN-N, while lacking the predicted signal peptide, do still retain the putative cell re-entry signal (electronic supplementary material, figure S5). We overexpressed GFP fused to both the PTEN-L (with putative secretion signal) and PTEN-M N-terminal extensions but could not detect PTEN-L-GFP in the cell culture media after concentrating by immunoprecipitation (figure 5*b*). It is not so surprising that PTEN-M-GFP is not found extracellularly because it lacks the putative signal peptide, but failure to detect PTEN-L was unexpected and suggests that most PTEN-L is not secreted. It is possible that this assay was not sensitive enough to detect low levels of secreted PTEN-L-GFP, so we further tested whether a secretion signal resided in the PTEN-L extension by fusing the *PTEN* 5′ leader to *Gaussia* luciferase. *Gaussia* luciferase is approximately 1000 times more sensitive than either *Renilla* or firefly luciferases [69], yet we could not detect any extracellular luciferase activity, above background, in constructs harbouring the putative PTEN-L signal peptide (figure 5*c*). One explanation for the discrepancy between our PTEN localization experiments and those of previous studies is that, because canonical PTEN has been found in exosomes [70,71] and can homodimerize [72], we decided to make reporters that do not contain sequences encoding the canonical *PTEN* CDS. Perhaps the important targeting signals are only 'visible' in the context of the full-length PTEN proteins. Yet another possible explanation for not detecting our PTEN-L chimeras in the cell culture media could be that the efficiency of cell re-entry is extremely high. However, detection of extracellular luciferase activity when the PTEN-L signal peptide is replaced with either the *Gaussia* luciferase or interleukin-2 signal peptide would argue against this possibility (figure 5*c*).

The functional significance of the PTEN-M N-terminal extension has yet to be determined but perhaps some insight may be gained from previous studies on PTEN-L. There are conflicting reports as to whether the PTEN-L extension has the potential to form a three-dimensional structure [60,67] or whether it is intrinsically disordered [65,66]. A recent elegant HDX-MS approach by Masson *et al.* [62] indicates that while most of the PTEN-L N-terminal extension is indeed intrinsically disordered, there is a potential α-helix at position 151–174 (where residue 174 is the methionine encoded by the canonical *PTEN* AUG). This peptide is protected by liposomes, suggesting an interaction with the membrane. Furthermore, this potential membrane-spanning region alters both the interfacial kinetics

of PTEN-L and the protein/membrane interface, causing PTEN-L to function on membranes in a 'scooting' mode rather than a 'hopping' mode that is characteristic of PTEN [62]. All N-terminally extended proteoforms described in our study possess this potential α-helix so it will be interesting to see whether these new PTEN proteoforms act in a similar manner. It is perhaps noteworthy that we could not detect PTEN-O by immunoprecipitation with PTEN antibodies, presumably because this N-terminal extension (and not L, M and N) alters protein conformation in a manner that prevents antibody access to the PTEN antigen under native conditions. This suggests possible structural differences between the PTEN-O N-terminal extension and the other PTEN proteoforms.

In our analysis of the conserved *PTEN* uORFs, we initially hypothesized that translation of uORF1 could reduce translation of uORF2, which, because of its size (45 codons), we expect to be severely inhibitory for downstream translation. In this way, translation of uORF1 could, in theory, have an overall positive effect on main ORF translation by reducing the number of ribosomes accessing the predicted inhibitory uORF2. However, intriguingly, increasing the number of ribosomes accessing uORF2 by mutation of uORF1 appears not to be very inhibitory under the conditions tested (figure 6). Therefore, we conclude that either uORF1 is frequently passed by leaky scanning, which seems unlikely given the evidence we described previously for uORF1 translation, or ribosomes translating uORF1 can re-initiate efficiently at uORF2. Alternatively, similar results would be observed if uORF2 were not very inhibitory (i.e. permits high level re-initiation). However, when ribosomes do not translate uORF2, downstream translation increases approximately 2.5-fold at the CUG of the L-proteoform in comparison with when uORF2 is available for translation (figure 6*b,c*). As one would predict from the scanning model of translation initiation, removal of both uORFs further increases the level of downstream initiation a further twofold (for L-FLuc). Importantly, the presence of uORF2 affects the ratio of N-terminally extended proteoforms relative to the main ORF, but has no effect on the ratio of N-terminally extended proteoforms to each other (electronic supplementary material, figure S10*c,d*).

These reporter assay results raise the intriguing possibility that the deeply conserved uORFs in the *PTEN* 5′ leader may become less inhibitory for PTEN translation under, as yet unidentified, conditions that could either decrease elongation rates, which in theory would result in ribosome accumulation along uORFs and hamper scanning, or else downregulate re-initiation. There is evidence that the canonical mTOR–S6K pathway regulates re-initiation after uORFs in plants [73]. We predict that such regulation could have dramatic effects on the abundance of N-terminally extended proteoforms, especially if initiation of both uORFs were reduced.

These findings, together with the findings on PTEN-L from other groups, could have profound implications for the interpretation of previous studies on both the catalytic activity and localization of endogenous PTEN as well as the analysis of polymorphisms within the *PTEN* 5′ leader. Furthermore, the discovery of these new PTEN proteoforms could have implications for the development of PTEN-based chemotherapeutic agents.

References

1. Sonenberg N, Hinnebusch AG. 2009 Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745. (doi:10.1016/j.cell.2009.01.042)

2. Kozak M. 1980 Evaluation of the 'scanning model' for initiation of protein synthesis in eukaryotes. *Cell* **22**, 7–8. (doi:10.1016/0092-8674(80)90148-8)

3. Cigan A, Feng L, Donahue T. 1988 tRNAi(met) functions in directing the scanning ribosome to the start site of translation. *Science* **242**, 93–97. (doi:10.1126/science.3051379)

4. Kozak M. 1999 Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208. (doi:10.1016/S0378-1119(99)00210-3)

5. Mitchell SF, Lorsch JR. 2008 Should I stay or should I go? Eukaryotic translation initiation factors 1 and 1A control start codon recognition. *J. Biol. Chem.* **283**, 27 345–27 349. (doi:10.1074/jbc.R800031200)

6. Jackson RJ, Hellen CUT, Pestova TV. 2010 The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127. (doi:10.1038/nrm2838)

7. Llácer JL, Hussain T, Marler L, Aitken CE, Thakur A, Lorsch JR, Hinnebusch AG, Ramakrishnan V. 2015 Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* **59**, 399–412. (doi:10.1016/j.molcel.2015.06.033)

8. Hussain T, Llácer JL, Fernández IS, Munoz A, Martin-Marcos P, Savva CG, Lorsch JR, Hinnebusch AG, Ramakrishnan V. 2014 Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597–607. (doi:10.1016/j.cell.2014.10.001)

9. Nanda JS, Cheung YN, Takacs JE, Martin-Marcos P, Saini AK, Hinnebusch AG, Lorsch JR. 2009 eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *J. Mol. Biol.* **394**, 268–285. (doi:10.1016/j.jmb.2009.09.017)

10. Valášek L, Phan L, Schoenfeld LW, Valášková V, Hinnebusch AG. 2001 Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *EMBO J.* **20**, 891–904. (doi:10.1093/emboj/20.4.891)

11. Hinnebusch AG. 2014 The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812. (doi:10.1146/annurev-biochem-060713-035802)

12. Kozak M. 1987 An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148. (doi:10.1093/nar/15.20.8125)

13. Noderer WL, Flockhart RJ, Bhaduri A, Diaz de Arce AJ, Zhang J, Khavari PA, Wang CL. 2014 Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol. Syst. Biol.* **10**, 748. (doi:10.15252/msb.20145136)

14. Peabody DS. 1989 Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**, 5031–5035.

15. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, Shastri N. 2012 Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719–1723. (doi:10.1126/science.1220270)

16. Ramakrishnan V. 2002 Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572. (doi:10.1016/S0092-8674(02)00619-0)

17. Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP. 2008 Structure of the 30S translation initiation complex. *Nature* **455**, 416–420. (doi:10.1038/nature07192)

18. Ogle JM, Brodersen DE, Clemons WM, Tarry MJ, Carter AP, Ramakrishnan V. 2001 Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902. (doi:10.1126/science.1060612)

19. Potapov AP, Triana-Alonso FJ, Nierhaus KH. 1995 Ribosomal decoding processes at codons in the A or P sites depend differently on 2'-OH groups. *J. Biol. Chem.* **270**, 17 680–17 684. (doi:10.1074/jbc.270.30.17680)

20. Baranov PV, Gesteland RF, Atkins JF. 2004 P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* **10**, 221–230. (doi:10.1261/rna.5122604)

21. Svidritskiy E, Korostelev AA. 2015 Ribosome structure reveals preservation of active sites in the presence of a P-site wobble mismatch. *Structure* **23**, 2155–2161. (doi:10.1016/j.str.2015.08.011)

22. Herr AJ, Wills NM, Nelson CC, Gesteland RF, Atkins JF. 2004 Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *J. Biol. Chem.* **279**, 11 081–11 087. (doi:10.1074/jbc.M311491200)

23. Chen SJ, Lin G, Chang KJ, Yeh LS, Wang CC. 2008 Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. *J. Biol. Chem.* **283**, 3173–3180. (doi:10.1074/jbc.M706968200)

24. Kozak M. 1989 Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol. Cell Biol.* **9**, 5073–5080. (doi:10.1128/MCB.9.11.5073)

25. Portis JL, Spangrude GJ, Mcatee FJ. 1994 Identification of a sequence in the unique 5' open reading frame of the gene encoding glycosylated gag which influences the incubation period of neurodegenerative disease induced by a murine retrovirus. *J. Virol.* **68**, 3879–3887.

26. Kozak M. 1990 Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl Acad. Sci. USA* **87**, 8301–8305. (doi:10.1073/pnas.87.21.8301)

27. Michel AM, Andreev DE, Baranov PV. 2014 Computational approach for calculating the probability of eukaryotic translation initiation from ribo-seq data that takes into account leaky scanning. *BMC Bioinformatics* **15**, 380. (doi:10.1186/s12859-014-0380-4)

28. Calvo SE, Pagliarini DJ, Mootha VK. 2009 Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106**, 7507–7512. (doi:10.1073/pnas.0810916106)

29. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009 Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223. (doi:10.1126/science.1168978)

30. Ingolia NT, Lareau LF, Weissman JS. 2011 Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802. (doi:10.1016/j.cell.2011.10.002)

31. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B. 2012 Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424–E2432. (doi:10.1073/pnas.1207846109)

32. Fritsch C *et al.* 2012 Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* **22**, 2208–2218. (doi:10.1101/gr.139568.112)

33. Rajkowitsch L, Vilela C, Berthelot K, Ramirez CV, McCarthy JEG. 2004 Reinitiation and recycling are distinct processes occurring downstream of translation termination in yeast. *J. Mol. Biol.* **335**, 71–85. (doi:10.1016/j.jmb.2003.10.049)

34. Hinnebusch AG. 2005 Translational regulation of GCN4 and the general amino acid control of yeast. *Annu. Rev. Microbiol.* **59**, 407–450. (doi:10.1146/annurev.micro.59.031805.133833)

35. Shabalina SA, Ogurtsov AA, Rogozin IB, Koonin EV, Lipman DJ. 2004 Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. *Nucleic Acids Res.* **32**, 1774–1782. (doi:10.1093/nar/gkh313)

36. Tikole S, Sankararamakrishnan R. 2006 A survey of mRNA sequences with a non-AUG start codon in

RefSeq database. *J. Biomol. Struct. Dyn.* **24**, 33–42. (doi:10.1080/07391102.2006.10507096)

37. Ivanov IP, Firth AE, Michel AM, Atkins JF, Baranov PV. 2011 Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* **39**, 4220–4234. (doi:10.1093/nar/gkr007)

38. Van Damme P, Gawron D, Van Criekinge W, Menschaert G. 2014 N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol. Cell Proteomics* **13**, 1245–1261. (doi:10.1074/mcp.M113.036442)

39. Arnaud E, Touriol C, Boutonnet C, Gensac MC, Vagner S, Prats H, Prats AC. 1999 A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. *Mol. Cell Biol.* **19**, 505–514. (doi:10.1128/MCB.19.1.505)

40. Tee MK, Jaffe RB. 2001 A precursor form of vascular endothelial growth factor arises by initiation from an upstream in-frame CUG codon. *Biochem. J.* **359**, 219–226. (doi:10.1042/bj3590219)

41. Packham G, Brimmell M, Cleveland JL. 1997 Mammalian cells express two differently localized Bag-1 isoforms generated by alternative translation initiation. *Biochem. J.* **328**, 807–813. (doi:10.1042/bj3280807)

42. Hann SR, King MW, Bentley DL, Anderson CW, Eisenman RN. 1988 A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52**, 185–195. (doi:10.1016/0092-8674(88)90507-7)

43. Coldwell MJ, Hashemzadeh-Bonehi L, Hinton TM, Morley SJ, Pain VM. 2004 Expression of fragments of translation initiation factor eIF4GI reveals a nuclear localisation signal within the N-terminal apoptotic cleavage fragment N-FAG. *J. Cell Sci.* **117**, 2545–2555. (doi:10.1242/jcs.01106)

44. Li J et al. 1997 PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943–1947. (doi:10.1126/science.275.5308.1943)

45. Steck PA et al. 1997 Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.* **15**, 356–362. (doi:10.1038/ng0497-356)

46. Stambolic V, Tsao M, Macpherson D, Suzuki A, Chapman WB, Mak TW. 2000 High incidence of breast and endometrial neoplasia resembling human Cowden syndrome in pten+/− mice. *Cancer Res.* **60**, 3605–3611.

47. Varga EA, Pastore M, Prior T, Herman GE. 2009 The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. *Genet.*

Med. **11**, 111–117. (doi:10.1097/GIM.0b013e31818fd762)

48. Maehama T, Dixon JE. 1998 The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 13 375–13 378. (doi:10.1074/jbc.273.22.13375)

49. Hopkins BD et al. 2013 A secreted PTEN phosphatase that enters cells to alter signaling and survival. *Science* **341**, 399–402. (doi:10.1126/science.1234907)

50. Liang H et al. 2014 PTENα, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. *Cell Metab.* **19**, 836–848. (doi:10.1016/j.cmet.2014.03.023)

51. Grentzmann G, Ingram JA, Kelly PJ, Gesteland RF, Atkins JF. 1998 A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479–486. (doi:10.1017/S1355838298971576)

52. Ivanov IP, Loughran G, Sachs MS, Atkins JF. 2010 Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proc. Natl Acad. Sci. USA* **107**, 18 056–18 060. (doi:10.1073/pnas.1009269107)

53. Loughran G, Sachs MS, Atkins JF, Ivanov IP. 2012 Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Res.* **40**, 2898–2906. (doi:10.1093/nar/gkr1192)

54. Dmitriev RI, Zhdanov AV, Nolan YM, Papkovsky DB. 2013 Imaging of neurosphere oxygenation with phosphorescent probes. *Biomaterials* **34**, 9307–9317. (doi:10.1016/j.biomaterials.2013.08.065)

55. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004 WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190. (doi:10.1101/gr.849004)

56. Heringa J. 1999 Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. *Comput. Chem.* **23**, 341–364. (doi:10.1016/S0097-8485(99)00012-1)

57. Pulido R et al. 2014 A unified nomenclature and amino acid numbering for human PTEN. *Sci. Signal.* **7**, ppe15. (doi:10.1126/scisignal.2005560)

58. Kawaji H et al. 2011 Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Nucleic Acids Res.* **39**, D856–D860. (doi:10.1093/nar/gkq1112)

59. Michel AM et al. 2014 GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.* **42**, D859–D864. (doi:10.1093/nar/gkt1035)

60. Johnston SB, Raines RT. 2015 Catalysis by the tumor-suppressor enzymes PTEN and PTEN-L. *PLoS ONE* **10**, e0116898. (doi:10.1371/journal.pone.0116898)

61. Wang H et al. 2015 Relevance and therapeutic possibility of PTEN-long in renal cell carcinoma.

PLoS ONE **10**, e114250. (doi:10.1371/journal.pone.0114250)

62. Masson GR, Perisic O, Burke JE, Williams RL. 2015 The intrinsically disordered tails of PTEN and PTEN-L have distinct roles in regulating substrate specificity and membrane activity. *Biochem. J.* **473**, 135–144. (doi:10.1042/BJ20150931)

63. Verhaegent M, Christopoulos TK. 2002 Recombinant *Gaussia* luciferase: overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. *Anal. Chem.* **74**, 4378–4385. (doi:10.1021/ac025742k)

64. Andreev DE et al. 2015 Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. *Elife* **4**, e03971. (doi:10.7554/eLife.03971)

65. Malaney P, Uversky VN, Davé V. 2013 The PTEN Long N-tail is intrinsically disordered: increased viability for PTEN therapy. *Mol. Biosyst.* **9**, 2877–2888. (doi:10.1039/c3mb70267g)

66. Malaney P, Uversky VN, Davé V. 2015 Identification of intrinsically disordered regions in PTEN and delineation of its function via a network approach. *Methods* **77–78**, 69–74. (doi:10.1016/j.ymeth.2014.10.005)

67. Johnston SB, Raines RT. 2015 Conformational stability and catalytic activity of PTEN variants linked to cancers and autism spectrum disorders. *Biochemistry* **54**, 1576–1582. (doi:10.1021/acs.biochem.5b00028)

68. Martin-Marcos P, Cheung Y-N, Hinnebusch AG. 2011 Functional elements in initiation factors 1, 1A, and 2β discriminate against poor AUG context and non-AUG start codons. *Mol. Cell Biol.* **31**, 4814–4831. (doi:10.1128/MCB.05819-11)

69. Tannous BA, Kim D-E, Fernandez JL, Weissleder R, Breakefield XO. 2005 Codon-optimized *Gaussia* luciferase cDNA for mammalian gene expression in culture and *in vivo*. *Mol. Ther.* **11**, 435–443. (doi:10.1016/j.ymthe.2004.10.016)

70. Putz U, Howitt J, Doan A, Goh C-P, Low L-H, Silke J, Tan S-S. 2012 The tumor suppressor PTEN is exported in exosomes and has phosphatase activity in recipient cells. *Sci. Signal.* **5**, ra70. (doi:10.1126/scisignal.2003084)

71. Gabriel K, Ingram A, Austin R, Kapoor A, Tang D, Majeed F, Qureshi T, Al-Nedawi K. 2013 Regulation of the tumor suppressor PTEN through exosomes: a diagnostic potential for prostate cancer. *PLoS ONE* **8**, e70047. (doi:10.1371/journal.pone.0070047)

72. Papa A et al. 2014 Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. *Cell* **157**, 595–610. (doi:10.1016/j.cell.2014.03.027)

73. Schepetilnikov M, Dimitrova M, Mancera-Martínez E, Geldreich A, Keller M, Ryabova LA. 2013 TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *EMBO J.* **32**, 1087–1102. (doi:10.1038/emboj.2013.61)