Characterisation of alternative splice variants of *PIP5K1A* in the human metastatic prostate cancer cell line, LNCaP C4-2

Christopher J. Roberts, BSc, MSc, MRes.

Thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

March 2022

Abstract

PIP5KIα is lipid kinase of the type I phosphatidylinositol 4-phosphate 5-kinase subfamily, responsible for the generation of the majority of cellular phosphatidylinositol (4,5) bisphosphate, an important signalling molecule required for cellular membrane dynamics, vesicular transport, actin remodelling and motility. PIP5KIα functions in the growth and metastasis of prostate, breast and other tumours, and pharmacological targeting of PIP5KIα curtails these processes in mouse models. In prostate tumours there appears to be crosstalk between the PIP5KIα and androgen receptor (AR; a key driver of prostate cancer), although the molecular mechanisms and functions in disease are unclear.

Here we show that in LNCaP C4-2 prostate cancer cells CRISPR edited at the *PIP5K1A* locus, genes implicated in cell cycle regulation, K-Ras signalling, p53 signalling, metastasis, steroid hormone signalling and epithelial mesenchymal transition pathways are differentially expressed compared to wild type LNCaP C4-2 cells, highlighting the importance of PIP5KIα in these pathways in prostate cancer cells. Amongst the most significantly altered differentially expressed genes were androgen regulated genes such as *KLK3, KLK4, UGT2B15 and UGT2B17*, consistent with crosstalk between PI(4,5)P_2 and androgen signalling. Exogenous expression of a fluorescently tagged form of PIP5KIα partially reversed the changes in *KLK3* and *UGT2B17* expression in KO cells, evidencing a causal link between PIP5KIα and expression of these genes.

Previous studies have shown that the several isoforms of PIP5KIγ, a paralog of PIPI5KIα, each have unique localisation and function facilitated by their variant C-termini. While five

alternate splice variants of *PIP5K1A* have previously been identified, RNA sequencing data suggests the existence of many more. The existence in prostate and breast cancer cells of numerous alternate transcripts of *PIP5K1A*, resulting from complex alternate splicing, was confirmed by cloning and sequencing. Analysis of the conservation of such transcripts was conducted, finding conservation throughout the class Mammalia. Alternate function of a C-terminal variant isoform of PIP5KIα was investigated, finding alternate localisation compared to canonical isoforms, and a putative nuclear export signal was mapped to C-terminal sequence of this novel variant.

Finally, putative direct interaction of PIP5KIα and AR was assessed by yeast two-hybrid (2YH), finding no evidence toward the existence of such interaction. However, various positive controls for PIP5KIα function in the Y2H failed, suggesting that full length PIP5KIα is not suitable for use in this Y2H system.

Acknowledgements

Firstly, I would like to thank my primary supervisor, Professor David Heery, for the opportunity to work in his lab, and for his continuous guidance and support in this project over the past four years, and even before that during my time as a masters' student in his lab and during the application process for this PhD. Often in the course of this project, I would run unannounced, gel photo in hand, to David's office to ask for interpretation and advice, which he would generously dispense at length. This is just one small example of both his insight and his generosity. I could not have asked for a better mentor.

I would also like to thank Dr. Hilary Collins, my second supervisor, for all of her time in my training, her support and her advice throughout my PhD. Special mention is also due for her effort, especially during the past two years of restrictions, to ensure the operation of the GRRB lab.

I am grateful to the many members of the Gene regulation group, past and present. I would particularly like to thank Jonathon Whitchurch, Bismoy Mazumder, Cíntia Monteiro, Mitchell Masterson, Jonas Schönfeld, Ruby Chrisp and especially Ryan Morris. It was a pleasure to work with each of you. Thank you for your friendship. I would also like to this opportunity to apologise to Cíntia and her sister Patricia regarding maintenance of some cell lines: there was a miscommunication, but I acted badly and I'm sorry.

Thank you also to my friends who kept me going through the six-month isolation of the first lockdown: Ernie, Jurgen, Anwyn, Grom, and especially Dr. James Furness. Fauxenor didn't need you, but I did. Thank you all!

Thank you to my family, and especially my Mum and Dad, for all their love and support over the years, without which I could never have started this PhD, and would not have finished. I have hidden away from you all in Nottingham for many years and I am sorry. You all deserve more praise than these few short lines.

Thank you also to my in-laws, Fernanda and Alfredo. Much of the introductory chapter to this thesis was written in their house. Fernanda's cooking and Alfredo's *pinginhas de vinho* saw me through that disaster. *Muito obrigado para ambos.*

Last but by no means least; thank you to my partner (and Portuguese teacher), Daniela Magalhães. I was never a very good student. How many times in the last four years did you have to convince me not to quit? *Sem ti, nada disto seria possível.*

In loving memory of Ruth Whittle

By yea and nay, sir, then I swore in jest.
What is the end of study? Let me know.

-Shakespeare, Love's Labour's Lost, 1598

To know that no one before you has seen an organ you are examining, to trace relationships that have occurred to no one before, to immerse yourself in the wondrous crystalline world of the microscope, where silence reigns, circumscribed by its own horizon, a blindingly white arena - all this is so enticing that I cannot describe it.

-Nabokov, Letter to his sister Elena Sikorski, 1945

Table of Contents

Abstract.....2

Acknowledgements.....4

List of Figures.....15

Chapter 1: Introduction.....18

1.1. Prostate cancer.....19

1.1.1. The prostate and prostate development....19

1.1.2. Prostate cancer.....19

1.1.3. Treatment of Prostate cancer.....23

1.2. The Androgen Receptor.....25

1.2.1. AR structure.....25

1.2.2. AR signalling.....28

1.2.3. The role of AR in the normal prostate....33

1.3. Phosphoinositide phosphates.....35

1.3.1. Phosphoinositide phosphate structure and synthesis.....35

1.3.2. General functions of PIPs.....42

1.4. The PIP kinase family.....43

1.4.1. The phosphoinositide kinase superfamily..43

1.4.2. Structure of the PIP kinases.....46

1.5. The PI3K/AKT pathway in prostate cancer.....60

1.6. PIP5KIα in prostate cancer.....61

1.7. Aims of the projects.....63

Chapter 2: Materials and Methods.....65

2.1. Sources of materials.....66

2.1.1. General Suppliers.....66

2.1.2 Bacterial reagents.....66

2.1.3. Molecular biology reagents.....67

2.1.4. Yeast reagents.....67

2.1.5. Nuclear hormone receptor ligands.....67

2.1.6. Tissue culture reagents.....68

2.1.7. SDS-PAGE, western blotting and chemiluminescence reagents.....68

2.1.8. List of antibodies used.....69

2.2. Bacteriological preparation and culture.....70

2.2.1. Composition of solutions and media used in bacterial procedures.....70

2.2.2. Culture of *Escherichia coli* strains.....71

2.2.3. Preparation of chemically competent *E.coli* cells.....71

2.2.4. Heat shock transformation of chemically competent *E.coli*.....73

2.3. Molecular biology techniques.....74

2.3.1. Small scale plasmid DNA preparation (miniprep).....74

2.3.2. Large scale plasmid DNA preparation (maxiprep).....74

2.3.3. Quantification and qualitative assessment of DNA.....75

2.3.4. Ethanol precipitation of DNA.....76

2.3.5. Agarose gel electrophoresis.....77

2.3.6. Polymerase chain reaction (PCR).....77

2.3.7. Purification of genomic DNA from mammalian cells.....79

2.3.8. Purification of DNA fragments from agarose gel slices.....80

2.3.9. Restriction digest of DNA.....81

2.3.10. Ligation of DNA fragments.....81

2.3.11. Vaccinia Virus Topoisomerase 1b mediated cloning (TOPO cloning).....82

2.3.12. Generation of mutations by recombinant PCR.....83

2.3.13. Colony PCR screening.....83

2.3.14. Sequencing of DNA.....84

2.3.15. Oligonucleotides.....85

2.3.16. Purification of RNA from mammalian cells.....85

2.3.17. Reverse transcription of RNA to synthesise DNA.....85

2.3.18. Quantitative PCR (qPCR).....86

2.4. Yeast methods.....89

2.4.1. Composition of solutions for yeast culture.....89

2.4.2. Yeast two-hybrid plasmids.....90

2.4.3. Yeast culture conditions.....91

2.4.4. Yeast transformation.....91

2.4.5. Preparation of yeast cell free lysate....92

2.4.6. Quantitative β-galactosidase activity assay.....92

2.4.7. Pellet X-gal assay.....93

2.4.8. Modified X-gal filter lift assay.....94

2.5. Cell culture.....96

2.5.1. Maintenance of cell lines.....96

2.5.2. Cryopreservation of cell lines.....97

2.5.3. Resuscitation of cell lines.....97

2.5.4. Lipofectamine mediated transfection of adherent cells.....98

2.5.5. Confocal microscopy.....99

2.5.6. Cell viability/proliferation assays.....99

2.5.7. Florescence-activated cell sorting.....100

2.6. Biochemical techniques.....101

2.6.1. Composition of solutions used in biochemical methods.....102

2.6.2. Preparation of cellular protein extracts.....102

2.6.3. Protein concentration determination.....103

2.6.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).....104

2.7. RNA Sequencing analysis of differentially expressed genes.....104

Chapter 3: Generation and functional studies of PIP5KIα knock-out in a human metastatic prostate cancer cell line.....107

3.1. Introduction.....108

3.2. Generation of sublines with altered PIP5KIα expression.....112

3.2.1. Choice of LNCaP C4-2 subline for genome editing of *PIP5K1A*.....112

3.2.2. Generation of an LNCaP C4-2 subline expressing a truncated isoform of PIP5KIα - LNCaP C4-2 ΔN PIP5KIα.....114

3.2.3. Generation and screening of an LNCaP C4-2 subline expressing greatly reduced amounts of PIP5KIα -LNCaP C4-2 PIP5KIα KO.....118

3.3. Identity of truncated PIP5KIα expressed in the ΔN subline.....128

3.4. Altered AKT phosphorylation in PIP5K1A edited lines.....138

3.5. Subcellular localisation of N-terminally truncated PIP5KIα variants.....142

3.5.1. Distribution of endogenously expressed PIP5KIα in CRISPR edited clones.....142

3.5.2. Localisation of variants of PIP5KIα-Venus fusion proteins.....144

3.6. Effect of *PIP5K1A* editing on gene expression in LNCaP C4-2 cells.....150

3.6.1. RNA-Seq analysis.....150

3.6.2. Validation of *UGT2B10*, *UGT2B11*, *UGT2B15* *UGT2B17* and *KLK3* gene expression in

CRISPR edited clones.....169

3.7. Summary and discussion.....175

Chapter 4: Characterisation of alternative splice variants of Human *PIP5K1A* gene in metastatic prostate cancer LNCaP C4-2 cells.....179

4.1. Introduction.....180

4.2. Canonical splice variants of human *PIP5K1A*...184

4.3. Identification of alternate 5’ exons of *PIP5K1A* using RNA-Seq data sets.....190

4.4. Predicted N-terminal amino acid sequences encoded by alternate 5’ exons and Kozak sequence analysis.....200

4.5. Evolutionary conservation of putative alternate PIP5KIα N-terminal sequences.....205

4.5.1. Use of Exon 1h.....205

4.5.2. Use of Exons 1a, 1b and 1c.....209

4.5.3. Use of Exon 1e.....215

4.5.4. Conclusions on evolutionary conservation of alternate 5’ exons of *PIP5K1A*.....218

4.6. Identification of alternative splice events involving *PIP5K1A* Exons 12, 13, 14 and 15....221

4.7. Evolutionary conservation of *PIP5K1A* Exon 12-Exon 15 splice variants.....228

4.8. PCR validation of novel upstream exons in *PIP5K1A* expressed in human prostate cancer cells.....232

4.9. Cloning alternate full-length *PIP5K1A* transcripts.....238

4.10. Subcellular localisation of PIP5KIα isoforms in LNCaP C4-2 cells.....248

4.10.1. Localisation of full-length PIP5KIα isoforms.....248

4.10.2. Isoform-specific C-terminal sequences of PIP5KIα exhibit different subcellular localisations.....254

4.10.3. Cloning of a Star-PAP-mCherry expression vector.....269

4.11. Evolutionary conservation of CAG repeat variation in PIP5K1A transcripts.....271

4.12. Summary and discussion.....281

Chapter 5: Assessment of protein-protein interactions of PIP5KIα and AR in Yeast two hybrid assays..288

5.1. Introduction.....289

5.2. Initial experiments assessing PIP5KIα interaction with steroid receptor LBDs by Y2H ONPG assay.....295

5.2.1. Construction of yeast expression vectors of LexA-PIP5KIα and NLS-VP16-PIP5KIα....295

5.2.2. Expression of NLS-VP16-PIP5KIα induces growth defect in *S. cerevisiae*.....298

5.2.3. Yeast two-hybrid assay to assess interaction of PIP5KIα i2 with AR-LBD and ERα-LBD.....307

5.3. Improved protocols for cloning with Y2H vectors.....312

5.3.1. Full sequencing of the pASV3mod plasmid.312

5.3.2. Cloning AR-NTD into pASV3mod by In Vivo Assembly.....315

5.3.3. Cloning of PIP5KIα i2 CDS into pASV3mod and pBTM116mod by modified restriction site cloning protocols.....323

5.4. Assessing PIP5KIα interaction with AR-LBD and AR-NTD by Y2H X-gal assay.....330

5.5. Summary and discussion.....339

Concluding remarks.....344

Appendices.....356

Supplementary data.....356

qPCR primer validation.....356

Discovery of an alternative splice variant of *TUT1*.....362

Tables of oligonucleotide sequences.....370

References.....383

List of Figures

Figure 1: Exon map of AR gene and schematic domain organisation of the Androgen Receptor protein.

Figure 2: The hypothalamic-pituitary-gonadal axis in a post-pubertal human male.

Figure 3: Structure of Phosphatidylinositol.

Figure 4: The PI cycle and the pathways of PI and PIP interconversions

Figure 5: Distribution of PIPs at cellular membrane

Figure 6: Domain Organization of the seven PIP Kinases

Figure 7: Structure of PIP5KIα

Figure 8: The insert loop of PIP5KIα

Figure 9: The "PAM out" orientation of CAS9 nickases RNPs targeted to a genomic location to facilitate editing.

Figure 10: CAS9 nickase editing of PIP5KIA Exon 1

Figure 11: Screening of clones dual targeted at Exon 1 and at the Exon 6-intron boundary.

Figure 12: Identification of truncated PIP5KIα expressed in the ΔN subline.

Figure 13: Elevated phosphorylated AKT in LNCaP C4-2 sublines.

Figure 14: Immunofluorescence staining of endogenous PIP5KIα in WT, ΔN and KO cells.

Figure 15: Localisation of exogenously expressed PIP5KIα truncated variants.

Figure 16: Differential gene expression between PIP5KIA-edited clones and WT LNCaP C4-2.

Figure 17: Differential expression of PIP5K genes between LNCaP C4-2 WT, ΔN and KO sublines.

Figure 18: Differential expression of AR and KLK genes between LNCaP C4-2 WT, ΔN and KO sublines.

Figure 19: Differential expression of UGT2B genes between LNCaP C4-2 WT, ΔN and KO sublines.

Figure 20: (Lack of) Differential expression of GUSB gene between LNCaP C4-2 WT, ΔN and KO sublines.

Figure 21: Validation of altered expression of UGT2B and KLK3 genes in ΔN and KO sublines.

Figure 22: Reported splice variants of *PIP5K1A.*

Figure 23: Alternative splice acceptor sites at the 5' end of Exon 2 of *PIP5K1A*.

Figure 24: Alternate 5' exons of *PIP5K1A*.

Figure 25: Sequence alignment of deduced amino acid sequences for predicted alternate transcripts of PIPI5K1A.

Figure 26: Conservation of Exon 1h like encoded sequences.

Figure 27: Conservation of *PIP5K1A* Exon 1a, 1b 1c like encoded sequences.

Figure 28: *PIP5K1A* Exon 1e like encoded sequences.

Figure 29: Existence of *PIP5K1A* Exon 12-Exon 15 splice events in LNCaP C4-2 and public RNA-Seq datasets.

Figure 30: Evolutionary conservation of amino acid sequences encoded by *PIP5K1A* Exons 12 – Exon 15 like splicing events.

Figure 31: PCR validation of existence of Exon 1b and 1c in *PIP5K1A* transcripts.

Figure 32: Cloning of alternate transcripts of *PIP5K1A*.

Figure 33: Subcellular localisation of variant of PIP5KIα isoforms.

Figure 34: Multiple sequence alignments of the alternate C-terminal sequences of PIP5KIα isoforms: canonoical (PIP5KIα i2 and i4) and novel (PIP5KIα x15).

Figure 35: Subcellular localisation of variant termini of PIP5KIα isoforms.

Figure 36: Mapping of a putative NES within the C-terminus of PIP5KIα x15.

Figure 37: Alternate splice-acceptor site use at the 5' end of Exon 2 of *PIP5K1A* transcripts for various species.

Figure 38: Complex variation in 5' Exon use in full length *PIP5K1A* transcripts demonstrated in this study.

Figure 39: Sequence alignment indicating conserved potential LxxLL like motifs and a CoRNR box like motif in the three isozymes of PIP5KI.

Figure 40: Molecular cloning to generate pBTM116 PIP5KIα i4 and pASV3 PIP5KIα i4 plasmids.

Figure 41: Expression of NLS-VP16-PIP5KIα is toxic to yeast.

Figure 42: ONPG assay to assess interactions of AR LBD and ERα LBD with PIP5KIα.

Figure 43: Complete restriction map of pASV3mod.

Figure 44: In Vivo Assembly cloning of AR NTD into pASV3mod.

Figure 45: Improved protocols for cloning with Y2H vectors

Figure 46: Assessing PIP5KIα i4 interaction with AR-LBD and AR-NTD by Y2H X-gal assay.

Supplementary Figure 1: Validation of qPCR primers for UGT2B genes.

Supplementary Figure 2: Identification of a novel splice variant of *TUT1*.

Chapter 1: Introduction

1.1. Prostate cancer

1.1.1. The prostate and prostate development

The prostate is a tubuloalveolar exocrine gland forming part of the male reproductive system of most mammals. In human males, the prostate is located just below the urinary bladder, and surrounds the urethra. It is composed of a fibromuscular stroma of mesenchymal cells surrounding secretory ducts composed of basal and luminal epithelial cells with occasional neuroendocrine cells. These ducts open into the seminal colliculus (aka the verumontanum). A major function of the prostate is to secrete a nutrient rich alkaline fluid, contributing to the production of semen. This prostate secretion contributes to the viability of ejaculated spermatozoa [1], and also contains numerous peptidases such as Prostate-specific antigen (PSA), which function to degrade the network of coagulated semenogelin proteins present in ejaculate that restricts motility of the spermatozoa [2; 3].

The prenatal development and function of the prostate throughout life is dependent upon Androgen Receptor (AR; a transcription factor of the nuclear receptor superfamily)/androgen signalling as has been extensively reviewed [4; 5].

During foetal development, the steroid hormone testosterone, which is produced by the testis and circulates within the blood, is locally converted by 5-α reductase enzymes to dihydrotestosterone (DHT) in target tissues such as the prostate. DHT stimulates development of the urogenital sinus into the prostate [5], mediated by binding to the AR, which subsequently drives expression of genes necessary for differentiation into prostate tissues [6]. The postnatal maintenance of the prostate is also dependent on AR signalling: ablation of prostatic DHT levels by castration in adult male rats leads to atrophy of prostate luminal epithelium, with high rates of apoptosis, whereas the populations of basal epithelia and stromal cell are relatively unaffected [7]. Conversely, androgen supplementation in castrated mice restores prostate size and morphology [8].

The luminal epithelial cells of the prostate are AR positive, whereas the basal cells only stain weakly for AR [9]. The canonical model of prostate tissue homeostasis is one in which activated AR signalling in stromal cells of the prostate induces their secretion of growth factors, termed andromedins, which act in paracrine fashion to stimulate a subset of basal epithelial cells to begin differentiating into luminal cells. These cells differentiating into luminal cells express increased levels of AR, becoming dependent on androgen signalling for the further

differentiation and survival, and communicate with the basal cells to maintain tissue homeostasis (reviewed in [10]).

1.1.2. Prostate cancer

Prostate cancer (PCa) is one of the most common types of cancer in men, with approximately 52 thousand men diagnosed annually in the UK, and causing approximately twelve thousand deaths annually in the UK [11].

PCa arises mainly from epithelial cells, and such malignancies are thus carcinomas. There are additional types of cancer arising within the prostate, deriving from mesenchymal cells (sarcomas) [12] or neuroendocrine cells [13] of the prostate. These later types are rare (2-10% of prostate cancer) and are diverse, with many subtypes varying by cell type of origin and other clinical features. Due to the great diversity and complexity of these malignancies, they are not discussed further in this thesis.

As discussed, the prostate epithelium consists of luminal cells, which cover the surface of the ducts and secrete into the lumen, and basal cells, which sit below the luminal cells, and separate them from the basement membrane of the duct. Rare neuroendocrine cells are also interspersed amongst the basal cells. PCa was once widely presumed to originate from mature luminal cells as PCa tumours are characterized by an expansion

of luminal cells and the absence of basal cells, however, recent studies have revealed that distinct populations of stem cells in the luminal and basal layers of the epithelium can both give rise to PCa [14].

Prostate cancer involves the transformation of benign epithelial cells into malignant cells displaying numerous features characteristic of many cancers: the Hallmarks of Cancer [15]. Cancer hallmarks present in prostate cancer include: insensitivity to antiproliferative signals, loss of tumour suppressor function, self-sufficiency in growth signals, enhanced proliferative rate, enhanced cell growth, enhanced cell survival, induction of angiogenesis, enhanced motility and invasion, deregulated metabolism to support tumour growth, immune surveillance escape and unlimited replication potential (reviewed in [16]). Prostate cancer specific hallmarks are associated with the centrality of AR signalling in this disease [16]. These are:

- Enhanced AR signalling
- Androgen independence
- Castration resistance

The central role of AR in prostate cancer development will be discussed in detail below.

1.1.3. Treatment of Prostate cancer

For low risk PCa localized to the prostate, a common treatment regimen involves radical prostatectomy, radiotherapy (either external beam or brachytherapy). For localized but high risk disease, these modalities are often used in combination with androgen deprivation therapy (ADT) [17], which seeks to mitigate the driving role of AR signalling in PCa (discussed in **Section 1.2**).

There are several types of ADT: orchidectomy (now rarely used [18]); Luteinizing-hormone releasing-hormone (LHRH) agonists and LHRH antagonists (which both block pituitary gland secretion of luteinising hormone; **Section 1.2.2**).

For metastatic PCa, as well as locally recurrent PCa, LHRH agonists/antagonists and androgen receptor antagonists such as Flutamide and Bicalutamide ("first generation") or Enzulatamide ("second generation) are used. Androgen receptor antagonists such as those mentioned competitively bind to the ligand binding domain of AR (**Section 1.2.1**).

Use of both a LHRH agonist or LHRH antagonist in combination with AR antagonist is termed combined androgen blockade.

Unfortunately, for 10-20% patients suffering PCa, the disease will progress to a state known as castration resistant prostate cancer (CRPC) within 5 years [19]. CRPC is defined by

disease progression (assessed either by a continuous rise in serum levels of PSA, progression of pre-existing disease, and/or appearance of new metastases) despite ADT [20].

Alterations commonly associated with CRPC include amplification and/or mutations in AR gene, AR overexpression and expression of AR splice variants lacking the ligand binding domain (such as AR-V7) and aberrant post-translational modification of AR protein.

Treatment regimens for CRPC typically include combined androgen blockade, with the possibility of use of other anti-androgen agents such as Abiraterone, which inhibits 17α hydrolysis of pregnolone and progesterone, and thus blocks pathways that convert these steroids to testosterone [21]. Chemotherapeutic agents are also used, such as the taxanes Docetaxel and Cabazitaxel [21].

Unfortunately, CRPC can eventually evolve so that it is longer sensitive to currently available treatments. A recent analysis placed the five year survival rate for metastatic CRPC at only 28.2 % [22]. At the point that the disease becomes resistant to standard treatments, patients may be referred to clinical trials for novel agents or therapeutic regimens, or else broad-spectrum chemotherapeutic agents, though unfortunately these have low success rates [21].

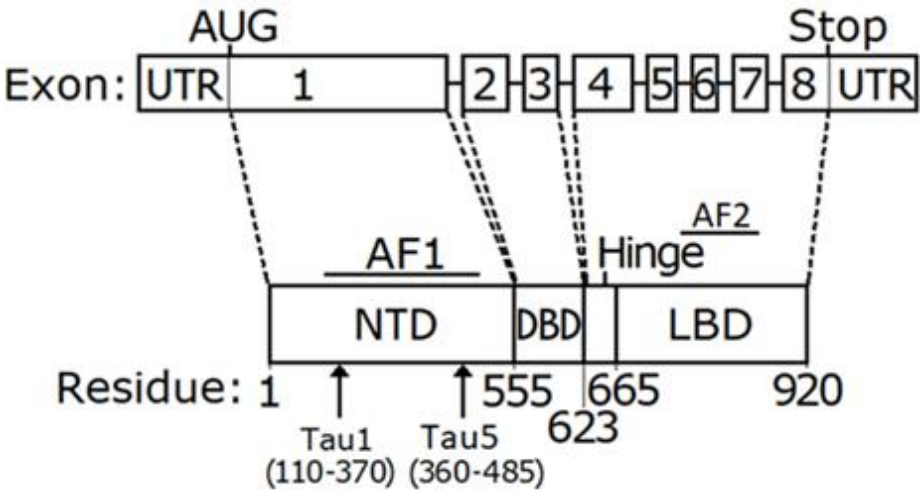
1.2. The Androgen Receptor

1.2.1 AR structure

AR is a member of the nuclear receptor (NR) superfamily, whose members function as ligand-inducible transcription factors that regulate expression of target genes in response to binding ligands specific to each receptor. AR (NR3C4, nuclear receptor subfamily 3, group C, gene 4) is a member of the oestrogen receptor (ER)-like nuclear receptor subfamily (NR subfamily 3), along with the other steroid hormone NRs: ERα, ERβ, progesterone receptor (PR), glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) and the orphan receptors, ER related (ERR) α, β and γ.

Like the other members of this subfamily, AR is composed of four main functional domains: i) the N-terminal domain (NTD), ii) the DNA binding domain (DBD), iii) the hinge region and iv) the C-terminal ligand-binding domain (LBD) (**Figure 1**).

A)



B)

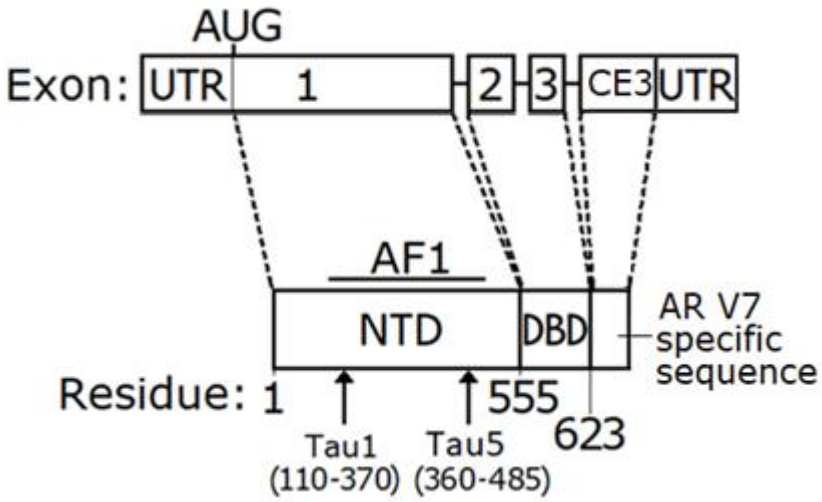


Figure 1. Exon map of *AR* gene and schematic domain organisation of the Androgen Receptor protein.

A) The canonical isoform of AR is encoded by 8 exons, located at the q11-q12 region of the X chromosome. Exon 1 contains the 5' UTR and CDS for the N-terminal Domain (NTD; residues 1-555) of AR. The NTD contains the AR Activation Function (AF) 1, which is composed of two transcriptional activation units (Tau1 and Tau5). These are primarily responsible for the transcriptional activating capacity of AR, by recruitment of p160 co-activators (Tau5), and non-p160 coactivators [23]. Exons 2 and 3 encode the DNA binding domain (DBD; residues 556-623). Upon conformational changes induced by ligand binding, the DBD binds androgen response elements.

B) An alternate isoform of AR, AR-V7 is encoded by alternate splicing, using an alternate exon, CE3, instead of the canonical Exons 4 -8. AR-V7 has an intact NTD and DBD but lacks the hinge and LBD. Numerous alternate splice variants of AR are described [24].

AR normally functions as a homodimer, though heterodimers between AR and other NRs have been reported [25; 26]. AR homodimers bind via their DBDs to androgen response elements (AREs), DNA sequences consisting of two equal hexameric half-sites, separated by a 3 base-pair spacer, with consensus sequence AGAACAnnnTGTTCT [27] though there is considerable variation in the sequences of AR binding site present in natural promoters and enhancers [28]. Such binding occurs via two zinc fingers present in the DBD, each zinc finger being composed of four cysteine residues that coordinate a zinc ion.

The NTD contains the activation function 1 (AF-1) region, a ligand independent transactivation domain composed of two discrete transcription activation sequences, Tau-1 and Tau-5, necessary for full activation of AR [29]. AF-1 mediates direct ligand-dependent interactions between the NTD and the LBD (N/C interactions) and ligand independent interactions with many co-regulators. The structure of the NTD has not yet been resolved by X-ray crystallography. This is possibly due to the NTD's intrinsically disordered structure, a feature which has

been proposed to facilitate the interactions of the NTD with its many structurally diverse binding partners [30].

The LBD forms the ligand-binding pocket of AR and also contains the activation function 2 (AF-2) region, which is important for co-regulator binding as well as mediating direct interactions between the NTD and LBD to stabilize bound androgen [31].

The hinge region is a flexible, 42 amino acid sequence linking the DBD and LBD. The hinge region contains a RKLKKL motif, which helps to facilitate DNA binding, and transactivation of AR [29]. The border of the DBD and hinge region contains a nuclear localisation signal (NLS), which is exposed upon the conformational changes that follow ligand binding, facilitating import of AR [32].

1.2.2. AR signalling

The hypothalamic-pituitary-gonadal axis

The native ligands of AR are the androgens, steroid sex hormones that bind to AR in the low nanomolar range. The biosynthesis of androgens is regulated by the hypothalamic-pituitary-gonadal axis (**Figure 2**)**.** During embryonic development, and again after onset of puberty, the hypothalamus secretes LRHR (alternatively known as gonadotropin-releasing hormone: GnRH) in a pulsatile fashion,

which stimulates secretion of luteinising hormone (LH) into circulation by the anterior pituitary gland. In males, LH in turn stimulates production of androgens, primarily testosterone, by Leydig cells within the testes. Secreted testosterone circulates in plasma bound to sex hormone binding globulin (SHBG) [33], and reaches target tissues, where it may be converted to the more potent agonist of AR, dihydrotestosterone (DHT), by 5α reductase enzymes.

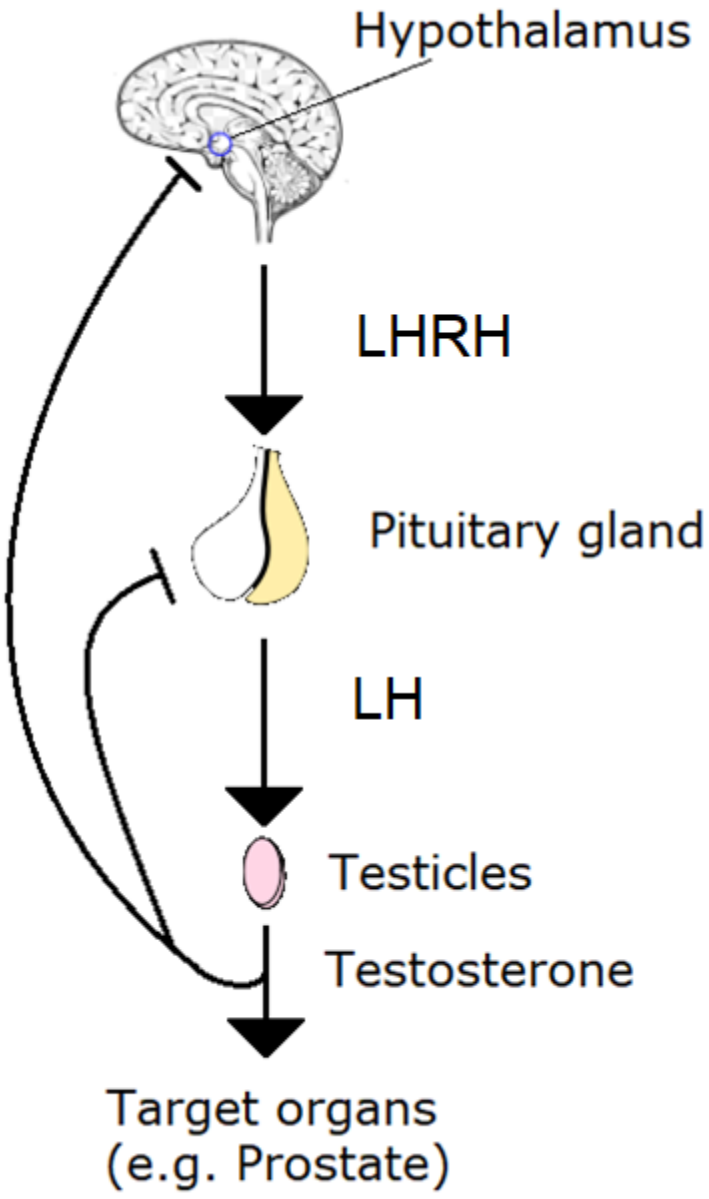


Figure 2. The hypothalamic-pituitary-gonadal axis in a post-pubertal human male.
Androgen production in the Leydig cells of the testis, principally of testosterone, is under the regulatory control of the hypothalamic-pituitary-gonadal axis. Pulsatile secretion of Luteinising hormone releasing hormone (LHRH) by the hypothalamus stimulates secretion of luteinising hormone (LH) by the pituitary gland. LH in turn (among other effects) stimulates testosterone production and secretion by the testicles. Testosterone negatively regulates both LHRH and LH secretion, and thus homeostasis of circulating testosterone levels is achieved. Circulating testosterone reaches target organs, e.g. the prostate, where it stimulates various actions. Figure adapted from original [34], published under Creative Commons 3.0 Licence [35].

DNA-binding dependent actions of AR

There are two distinct mechanisms of ligand-dependent AR signalling: DNA binding-dependent, (also known as 'canonical', 'classical' or 'genomic' signalling) and DNA binding-independent ('non-canonical', 'non-classical' or 'non-genomic' signalling).

In the absence of ligand binding, AR resides in the cytoplasm in complex with heat-shock protein 90 (HSP90) and the co-chaperones such as p23 and the immunophilins FKBP51/52, which maintain AR in a conformation conducive to ligand binding [36]. Upon androgen binding, conformational changes in AR occur which lead to dissociation of the chaperone proteins and exposure of the NLS promoting the nuclear translocation of the AR/androgen complex [36]. Androgen binding also promotes interaction between the NTD and LBD, via their

AF-1 and AF-2 regions, which stabilises ligand in the ligand-binding pocket and promotes dimerization [29].

AR binds many transcriptional co-regulators that act to either enhance (co-activator) or repress (co-repressor) transactivation of target genes. In many NRs, ligand binding induces conformational changes that complete a binding interface composed of helices 3, 4, 5, and 12 of the receptor, corresponding to the conserved AF-2 region. This surface facilitates binding to short leucine-rich LxxLL motifs (where "x" can be any amino acid) present in many transcriptional co-activators [37]. AF-2 of AR binds with high affinity to a subset of LxxLL motifs, dependent on sequences flanking the LxxLL [38], but has a strong preference for aromatic-rich motifs containing phenylalanine (F) or tryptophan (W) at either or both positions +1, +5, [39].

The AF-1 region present in the NTD also activates transcriptional activity of AR, in a ligand independent fashion [40]. AF-1 appears to be the major transcriptional activating domain, since deletion of the LBD results in an AR variant with transactivating potential equivalent to the maximal activity of full length, ligand bound AR [41]. An alternate isoform of AR, termed AR-V7 (or AR3), is encoded by an alternate splice variant, resulting from alternate terminal exon use (use of a

cryptic exon called CE3 or Exon 3b) and polyadenylation [24]. AR-V7 retains the NTD and DBD of AR, but lacks the LBD. AR-V7 also displays constitutive transcriptional activity [42].

AR has been determined to bind a large number of co-activators (reviewed in [43]). Such bound co-activators include: histone acetyl transferases (HATs) [44; 45], which catalyse addition of acetyl groups to specific residues in histone N-terminal tails, and thus promote de-condensation of chromatin [46]; chromatin remodelling complexes such as SWI/SNF [47], which stabilise nucleosomes in a conformation conducive to transcription; the Mediator complex, which facilitates recruitment of the transcriptional pre-initiation complex (PIC) [48]; and the general transcription factors and RNA polymerase II of the PIC [49; 50].

AR also interacts with many partners that oppose its activity as a transcriptional activator. These include: recruiters of histone deacetylases (HDACs), which oppose the action HATs and therefore directly oppose chromatin relaxation and transcriptional activation; negative regulators of N/C interactions, which thus oppose chromatin binding and ligand binding stability of AR; negative regulators of nuclear translocation and DNA binding; and competitive inhibitors of co-activator binding (reviewed in [51]).

DNA-binding independent actions of AR

DNA-binding independent signalling events induced by AR occur within minutes of exposure to androgen. These are brought about by AR association with protein substrates in the cytoplasm and inner leaflet of the cell membrane. AR can activate several kinase cascades, generally promoting survival and proliferation [52].

Targets of this activity include the non-receptor tyrosine kinase Src, with which AR interacts in conditions of low to physiological androgen concentration, promoting activating auto-phoshorylation of Src and thus promotes activation of the mitogenic Ras-Raf-MAPK/ERK pathway [52].

1.2.3. The role of AR in the normal prostate

During foetal development the androgen testosterone produced by the testis, and locally converted to DHT, drive the development of the urogenital sinus into the prostate [5] dependent on AR [6]. The postnatal maintenance of the prostate is also dependent on AR signalling: ablation of prostatic DHT levels by castration in adult male rats leads to atrophy of prostate luminal epithelium, with high rates of apoptosis, whereas the populations of basal epithelia and stromal cell are relatively unaffected [7]. Conversely androgen supplementation in castrated mice restores prostate size and morphology [8].

The luminal epithelial cells of the prostate are AR positive, whereas the basal cells only stain weakly for AR [9]. The canonical model of prostate tissue homeostasis is one in which activated AR signalling in stromal cells of the prostate induces their secretion of growth factors, termed andromedins, which act in paracrine fashion to stimulate a subset of basal epithelial cells to begin differentiating into luminal cells. These differentiating cells express increased levels of AR, becoming dependent on androgen signalling for the further differentiation and survival, and communicate with the basal cells to maintain tissue homeostasis (reviewed in [10]).

1.2.4. AR in prostate cancer initiation

A fundamental event in the initiation of PCa appears to be a switch in the effects of AR signalling in luminal epithelial cells, from promoting their differentiation and survival to promoting their proliferation. In this model, an early event in PCa tumorigenesis is an increase in AR expression in differentiating epithelial cells with concurrent increase in AR transregulatory activity to include genes that are otherwise not regulated by AR. In these conditions, selection then occurs for clones in which pro-proliferative and anti-apoptotic effects are promoted by androgen/AR signalling initiation (reviewed in [10]). The mechanism by which AR signalling switches from driving differentiation to driving proliferation is unclear. One common

occurrence in human PCa samples is the presence of gene fusion events between the ARE-containing promoter of TMPRSS2 to the coding sequences of members of the Ets family of transcription factors, thus AR signalling activates expression of these transcription factors, driving tumour progression [53]. A proposed alternative/supplementary mechanism is an increase in the number of genes targeted by AR due to amplification, alterations in the chromatin recognition or increased activation of AR, causing aberrant gene regulation and driving PCa [54].

1.3. Phosphoinositide phosphates

1.3.1. Phosphoinositide phosphates structure and synthesis

Phosphatidylinositol (PI) is a minor constituent (5-15% of total cellular lipid) present in various membranes within eukaryotic cells. Like other phospholipids, PI possesses two fatty acid chains each linked by an ester bond to a glycerol moiety. In PI, the head group (which is linked by a phosphate group to the glycerol moiety) is *myo*-inositol (**Figure 3 A**). *Myo*-inositol is one of nine possible stereoisomers of cylohexane-1,2,3,4,5,6-hexol (**Figure 3 B**). The hydrophobic fatty acid tails anchor PI within the lipid membrane, while the hydrophilic head group is exposed to the intracellular fluid.

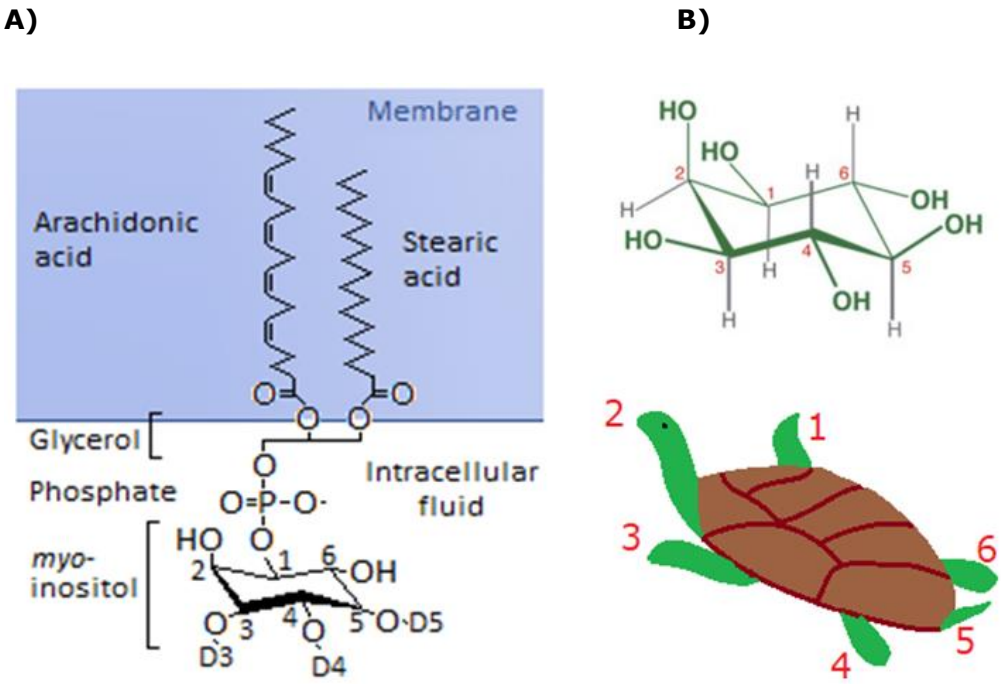


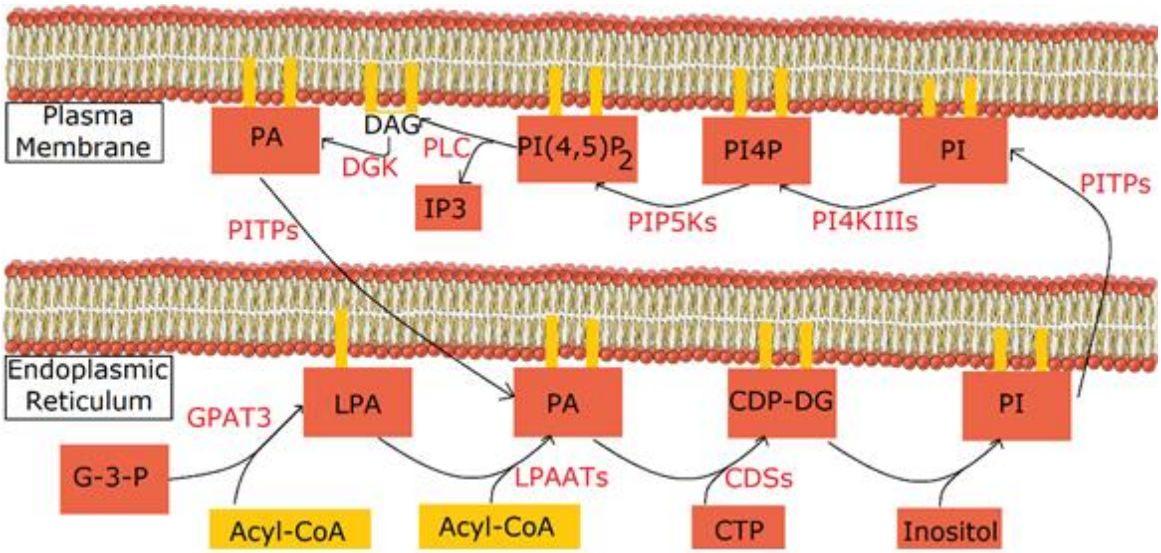
Figure 3. Structure of Phosphatidylinositol

A) The structure of Phosphatidylinositol. The D3, D4 and D5 positions of its *myo*-inositol head-group may be phosphorylated in various combinations by specific kinases. The *myo*-inositol head-group is connected by a phosphate group to diacylglycerol, made up of two fatty acid chains (typically, though not exclusively, Stearic acid and Arachidonic acid) esterified to a glycerol moiety.

B) The structure of myo-inositol, showing the orientation of the hydroxyl groups (Top). A visual mnemonic for the orientation of each hydroxyl group was proposed by Agranoff (1978), in which the axial hydroxyl is the turtle's head and the five equatorial hydroxyls are the turtle's limbs and tail. Using the D numbering convention, the D1 position, which is linked to diacylglycerol in PI, is the right forelimb. Moving counterclockwise, the axial hydroxyl serves as the turtle's head at D2, D3 is the turtle's left forelimb, D4 is the left hind limb, D5 the tail and D6 the right hind limb.

PI is synthesized *de novo* in the endoplasmic reticulum in a series of enzyme catalysed reactions, summarised in **Figure 4 A** (reviewed in [55]). Once generated, PI is shuttled from the endoplasmic reticulum to other lipid membranes by both vesicular transport and by members of the PI transfer protein family. Once within a target membrane, PI may undergo a series of reversible phosphorylation reactions of several of the hydroxyl groups of the inositol head-group. The inositol head-group of PI has five free hydroxyl groups (D2 to D6), of which D3, D4 and D5 can be reversibly phosphorylated in various combinations, generating seven phosphorylated derivatives named the phosphoinositides (PIPs): phosphatidylinositol 3-phosphate (PI3P), PI4P, PI5P, PI(3,4)P_2, PI(3,5)P_2, PI(4,5)P_2 and PI(3,4,5)P_3 (**Figure 4 B**).

A)


Plasma Membrane
PA
DAG
DGK
PLC
PI(4,5)P2
PI4P
PI
PITPs
IP3
PIP5Ks
PI4KIIIs
PITPs
Endoplasmic Reticulum
GPAT3
LPA
PA
CDP-DG
PI
G-3-P
LPAATs
CDSs
Acyl-CoA
Acyl-CoA
CTP
Inositol

B)

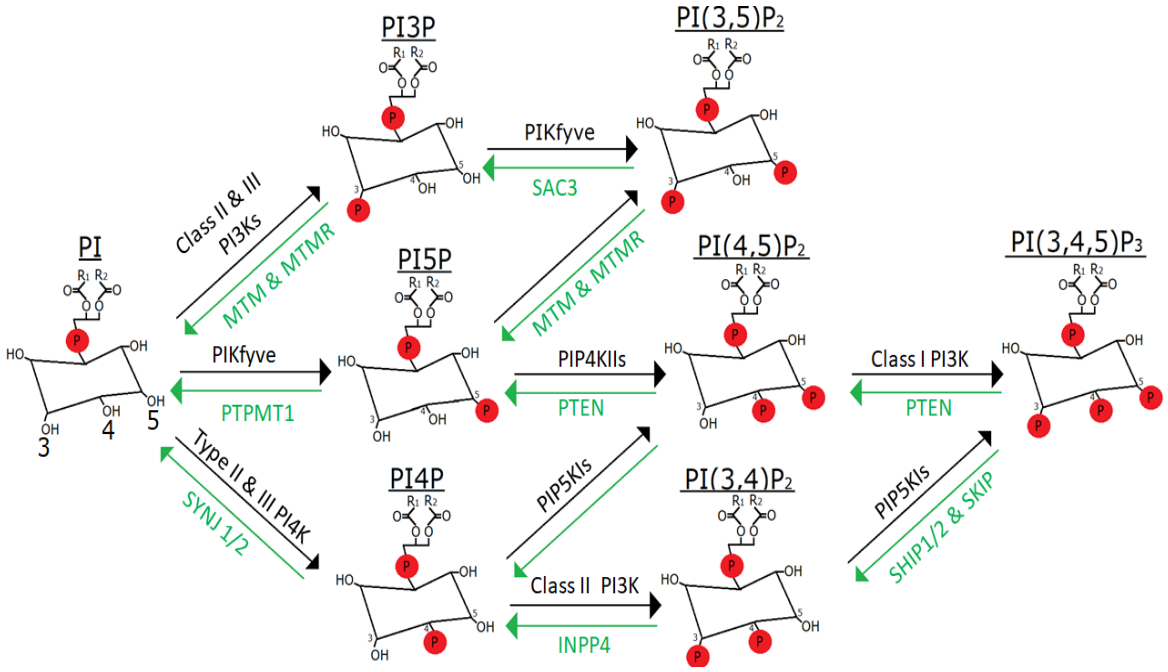

PI
PI3P
PI(3,5)P2
PI5P
PI(4,5)P2
PI(3,4,5)P3
PI4P
PI(3,4)P2
Class II & III PI3Ks
MTM & MTMR
PIKfyve
SAC3
PIKfyve
PTPMT1
PIP4KIIs
PTEN
Class I PI3K
PTEN
Type II & III PI4K
SYNJ 1/2
PIP5KIs
Class II PI3K
INPP4
PIP5KIs
SHIP1/2 & SKIP

Figure 4. The PI cycle and the pathways of PI and PIP interconversions

A) The PI cycle. Phosphatidic acid (PA) is synthesised in the endoplasmic reticulum, being converted in a series of steps into PI. PI is transported from the endoplasmic reticulum to various other membranes in the cell, where OH groups of its inositol head group can be phosphorylated. Depicted is conversion of PI to PI4P, then PI(4,5)P_2. PI(4,5)P_2 can in turn be phosphorylated to PI(3,4,5)P_3, or cleaved by phospholipase C, as depicted, to generate diacylglycerol (DAG) and inositol (1,4,5)P_3 (IP3). DAG can be phosphorylated, producing PA, which can be again shuttled to the endoplasmic reticulum to re-enter the PI cycle. *G-3-P: Glyerol 3-phosphate; LPA: lysophosphatidic acid; PA: Phosphatidic acid; CDP-DG: Cytidine diphosphate-diacylglycerol; CTP: Cytidine triphosphate; PI: Phosphatidylinositol; PI4P: Phosphatidylinositol 4-phosphate; PI(4,5)P_2: Phosphatidylinositol (4,5)-bisphosphate; IP3: Inositol (1,4,5)-trisphosphate; DAG: Diacyl glycerol; GPAT3: Glycerol-3-Phosphate Acyltransferase ; LPAATs; lysophosphatidic acidacetyl transferases; CDSs: Phosphatidate cytidylyltransferase; PI4KIIIs: Type III Phosphatidylinosito 4 kinase; PLC: Phospholipase C; DGK: Diacyl glycerol kinase. PITPs: Phosphatidylinositol transfer proteins*.

B) Schematic representation of PIP head group phosphorylation and dephosphorylation by lipid kinases and phosphatases respectively. Kinases are indicated with black text, phosphatases indicated by green text. Key sources [56; 57].

MTM :Myotubulin lipid phosphtase; MTMR Myotubulin lipid phosphtase related; SAC3: Suppressor of actin phosphatase 3; PTPMT1: Mitochondrion-localised protein tyrosine phosphatase; SYNJ: Synaptojanin; PTEN: Phosphatase and tensin homolog; INPP4: Type 1 inositol-3,4 bisphosphate phosphatase; SHIP: Src homology 2 (SH2) domain containing inositol polyphosphate 5-phosphatase; SKIP: Skeletal muscle and kidney enriched inositol phosphatase.

This interconversion of PIP species is orchestrated by multiple kinases and phosphatases, including the Type I Phosphatidylinositol phosphate kinases (PIP5KIs), which are discussed in more detail in later **Section 1.4**. Recruitment and regulation of activity of these PI modulating enzymes to specific subcellular compartments gives each PIP a unique subcellular distribution, each being found predominantly only in a subset of cellular membranes, or as for nuclear speckles, certain membraneless compartments (**Figure 5**) [58; 59].

An additional consideration to the diversity of molecular species of PIPs is the numerous combinations of fatty acid chains that can be included. While these are typically Stearic acid (C18:0) at the sn-1 position and Arachidonic acid (20:4) at the sn-2 position in mammalian cells, there is considerable diversity in the included acyl chains. Thus when referring to any one member of the PIPs by phosphorylation status of the head-group, it is important to note that one is really referring to many molecular species because of the acyl chain variations. The biological roles that this diversity plays in signalling remain largely unknown (for review see [60]).

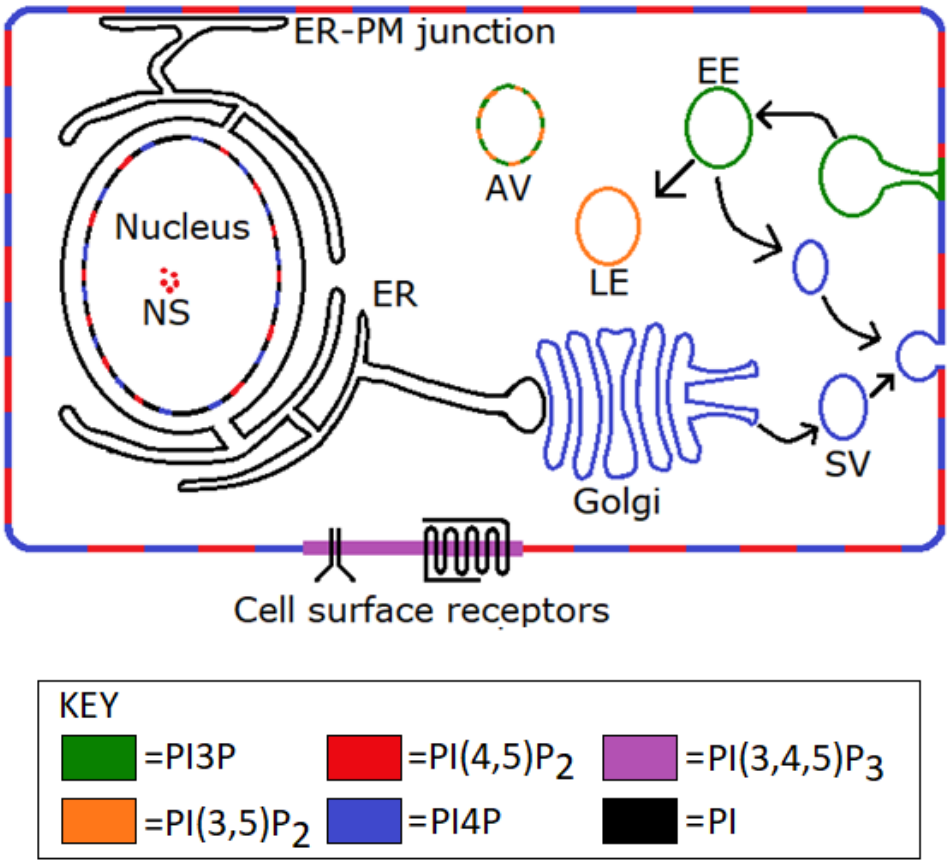


Figure 5. Distribution of PIPs at cellular membrane

Schematic of the cellular localization of PIPS with different organelles labelled, and phosphoinositide localization indicated. The plasma membrane is coloured light blue/red to indicate the presence of both PI4P and PI(4,5)P2. $PI(3,4,5)P_3$ is generated by PI3Ks, activated downstream of cell surface receptors including *G protein-coupled receptors and receptor tyrosine kinases. NS = nuclear speckles; ER = endoplasmic reticulum; SV = secretory vesicles; EE = early endosomes; LE = late endosome; AV = autophagic vacuole.*

1.3.2. General functions of PIPs

Phosphoinositides generally serve two major functions within the cell: one, as signalling molecules (secondary messengers) and two, as markers of membrane identity. These are both achieved through their recruitment and activation of effector proteins. Binding of such effector proteins is facilitated by the specific domains within the later that can bind PIP species with high affinity. These domains can typically distinguish between PIP species by the phosphorylation status of the D3 D4 and D5 positions of the inositol head-group. Such PIPs-interacting domains typically consist either of clusters of basic residues within unstructured regions or by folded domains, such as pleckstrin homology (PH) or phox homology (PX) domains (reviewed in [61]).

1.4. The PIP kinase family

1.4.1. The phosphoinositide kinase superfamily

The human genome contains 19 genes encoding the phosphoinositide kinases (PIKs), the superfamily of enzymes responsible for phosphorylation of the inositol ring of PtdIns and PIPs. Phylogenetic studies of these enzymes reveal that they fall into three major families: the PtdIns 3-kinases (PI3Ks), PtdIns 4-kinases (PI4Ks), and the PIP kinases (PIPKs) [62], the last of which includes PIP5KIα. PIKs are evolutionary well conserved, with members of each of these three families being present in all eukaryotes, including plants, fungi and Apicomplexa, suggesting the emergence of the PIPs messenger system was an early event in the evolution of eukaryotes. Vertebrates, particularly mammalian species, often possess numerous members of each PIK family, with subfamilies of several isozymes that possess the same catalytic activity. The 19 human genes encoding PIKs, their protein products and preferred catalytic activity are listed in **Table 1.**

Interestingly, several distantly related enzymes have emerged to catalyse the same reactions, as is the case for the Type II and Type III PI4Ks; while these catalyse the same reaction as each other, phylogenetic analyses groups the later subfamily with the PI3 kinases [62].

An additional subfamily of the PI3Ks is the Phosphatidylinositol 3-kinase-related kinase, which are not lipid kinases but serine/threonine protein kinases. Amongst others, this group includes the well-known ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad3-related (ATR), which are involved in the DNA damage response, and mammalian target of rapamycin (mTOR), which is activated in response to production of $PI(3,4,5)P_3$, and has myriad roles in cellular mitogenic responses, cell growth and proliferation (reviewed in [63]).

Table 1. The nineteen human kinases of the PIK superfamily

N.B: By historical accident, there is no Type I PI4K, as the enzymes first identified as such were later found to be PI3Ks, and renamed. Secondly, Type III PI4Ks and PI3Ks are grouped into one family based upon sequence homology and implied evolutionary relationship [64].

Family	Subfamily	Gene	Reaction (major)	Protein
PI3Ks and Type III PI4Ks	Type III PI4Ks	PI4KA	PI → PI4P	PI4KIIIα
		PI4KB		PI4KIIIβ
	Class I PI3Ks	PIK3CA	PI(4,5)P2 →	p110α
		PIK3CB	PI(3,4,5)P3	p110β
		PIK3CD		p110γ
		PIK3CG		p110δ
	Class II PI3Ks	PIK3C2A	PI →PI3P	PI3K-C2α
		PIK3C2B	PI4P→PI(3,4)P3	PI3K-C2β
		PIK3C2G		PI3K-C2γ
	Class III PI3Ks	PIK3C3	PI → PI3P	PIK3C3
Type II PI4Ks	Type II PI4Ks	PI4K2A	PI → PI(3,4)P2	PI4KIIα
		PI4K2B		PI4KIIβ
PIP Kinases	Type I PIPK (AKA	PIP5K1A	PI4P → PI(4,5)P2	PIP5KIα
		PIP5K1B		PIP5KIβ
		PIP5K1C		PIP5KIγ
	Type II PIPK	PIP4K2A	PI5P → PI(4,5)P2	PIP4KIIα
		PIP4K2B		PIP4KIIβ
		PIP4K2C		PIP4KIIγ
	Type III PIPK	PIKFYVE	PI3P → PI(3,5)P2	PIKfyve

1.4.2. Structure of the PIP kinases

PIP5KIα belongs to the PIP Kinase (PIPK) family, which is so named for their (preferred) catalytic activity, phosphorylating phosphoinositide mono-phosphates (PI3P, PI4P or PI5P) to produce various bisphosphates. This family shows greatest homology not to the other PI Kinases, but to the inositol polyphosphate kinases [65].

Members of this family are grouped into three subfamilies (**Figure 6**):

- Type I, the PIP5Ks, PIP5KIα, PIP5KIβ and PIP5KIγ, which primarily catalyse phosphorylation of the D5 position of the inositol ring of PI4P, thus producing $PI(4,5)P_2$. *****
- Type II, PIP4Ks, PIP4KIIα, PIP4KIIβ and PIP4KIIγ, which primarily catalyse phosphorylation of the D4 position of the inositol ring of PI4P, thus producing $PI(4,5)P_2$.
- Type III, which has a single member, PIKfyve, which primarily catalyses phosphorylation of the D5 position of PI3P, producing $PI(3,5)P_2$.

***** *It is important for when studying the literature concerning PIP5Ks to note that much of the early work in this field was conducted variously using murine and human PIP5Ks. It is unfortunate that in the naming convention then used, murine PIP5KIα referred to the orthologue of Human PIP5KIβ, and murine PIP5KIβ referred to the orthologue of Human PIP5KIα. A keen eye and reference to a protein sequence database is often needed to determine exactly which isozyme is being denoted in any particular study. This unfortunate naming convention has since fallen out of favour, and databases such as NCBI and*

UniProt now list murine PIP5KIβ as the orthologue of human PIP5KIβ, and so too for the PIP5KIαs. In this study, the modern naming convention will be used to avoid such confusion; though often perhaps excessively verbose, particular care is taken to specify any referenced gene, splice variant or isoform exactly when confusion could otherwise arise.

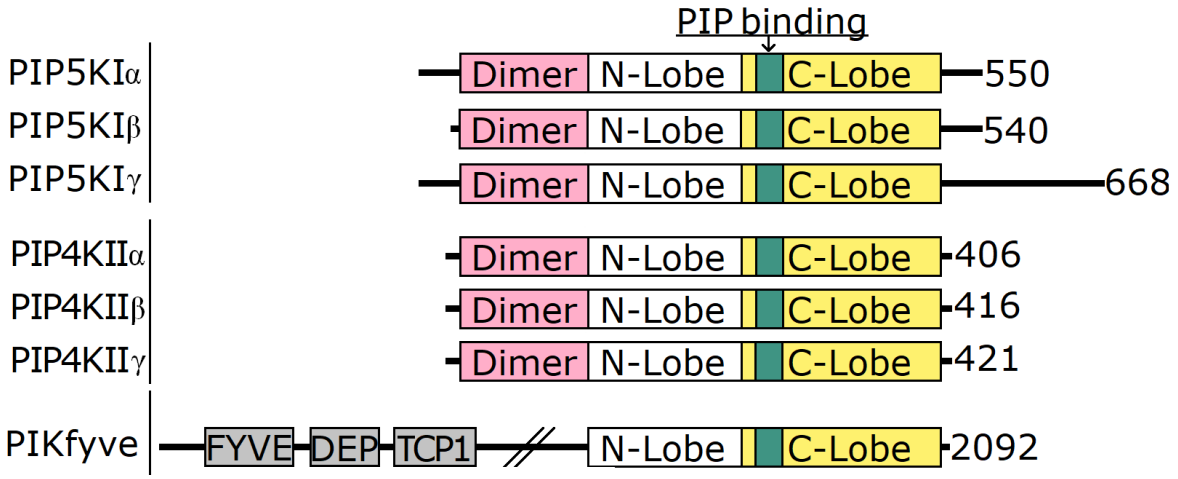


Figure 6. Domain Organization of the seven PIP Kinases
PIP5KIs and PIP4KIIs share a conserved core of their kinase domain (N-Lobe and C-Lobe), and dimerisation domain. PIKfyve also possesses the conserved kinase domain, but lack the dimerisation domain, possessing instead FYVE, DEP and TCP1 domains. A unique feature of the family is the PIP binding domain, located within the C-lobe.

Structures have been solved for the kinase domains of Human PI4KIIα, PI4KIIβ, PI4KIIγ and Zebra fish (*Danio rerio*) PIP5KIα [66; 67]. Because PIP5KIα is the primary focus of this PhD, only its structure will be described here, with reference, when informative, to the other PIPK family members. Domains, and key motifs, and their role in catalytic or other functions are described.

The crystal structure of *D. rerio* PIP5KIα reported by Hu *et al.*, [66] and the other solved PIPK structures reveal a

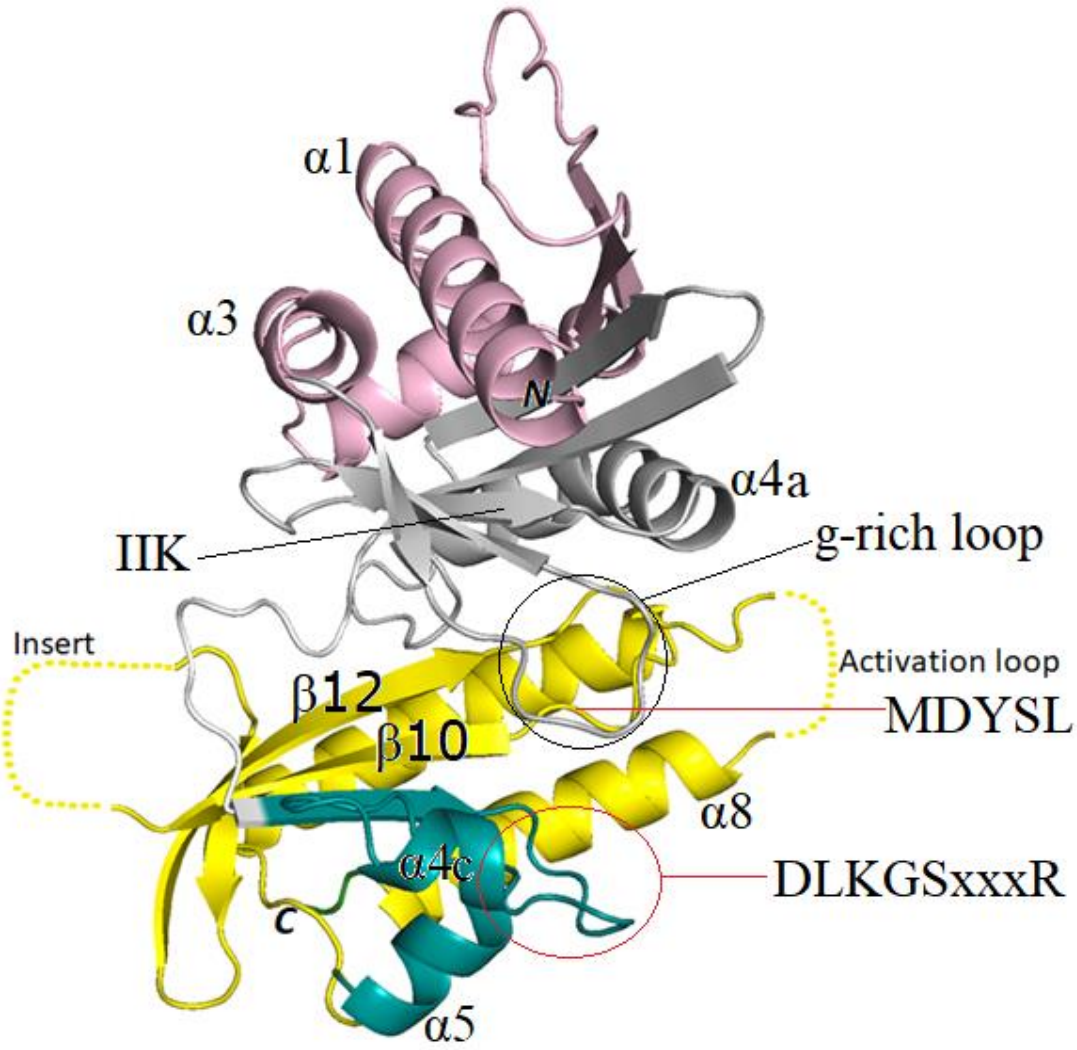
conserved core – an N-terminal dimerisation domain (in the type I and type II enzymes) and a kinase domain, constituted of an N- and C-lobe, joined by a short flexible linker, similar to protein kinases. An additional substructure within the C-lobe is the PIP binding domain, which is unique to the PIPKs, and, as the name suggests, functions in recognition and binding of the PI monophosphates. Additional conserved features within the C-lobe of the PIPKs are the activation loop and the "insert".

Flanking the conserved core in each of the Type I PIPKs are the divergent N- and C-termini: so named because these are highly-variant between the three isozymes, but show conservation within isozymes across species, leading to the suggestion that they contribute to the unique subcellular localisations of each isozyme [68]. This suggestion is further evidenced by the existence of multiple isoforms (protein products of the same gene, but with different amino acid sequences due to alternative splicing of the mRNA) of PIP5KIγ which each vary in their C-terminal sequence and display differential localisation and function [68-70].

The Type II PIPKs possess much shorter N- and C-terminal flanking regions; these will not be discussed further here. The type III PIPK, PIKfyve lacks the dimerisation domain and possess a much longer N- terminus (~1750 residues). This

contains a FYVE finger domain, which is necessary for targeting to PI3P containing membranes [71], a DEP (Dishevelled, Egl-10 and Pleckstrin domain), whose function in PIKfyve is unknown and an HSP chaperonin T-complex protein 1 like domain, which is implicated in protein-protein interactions that regulate PIPfyve's catalytic acitivty [72].

A)



B)

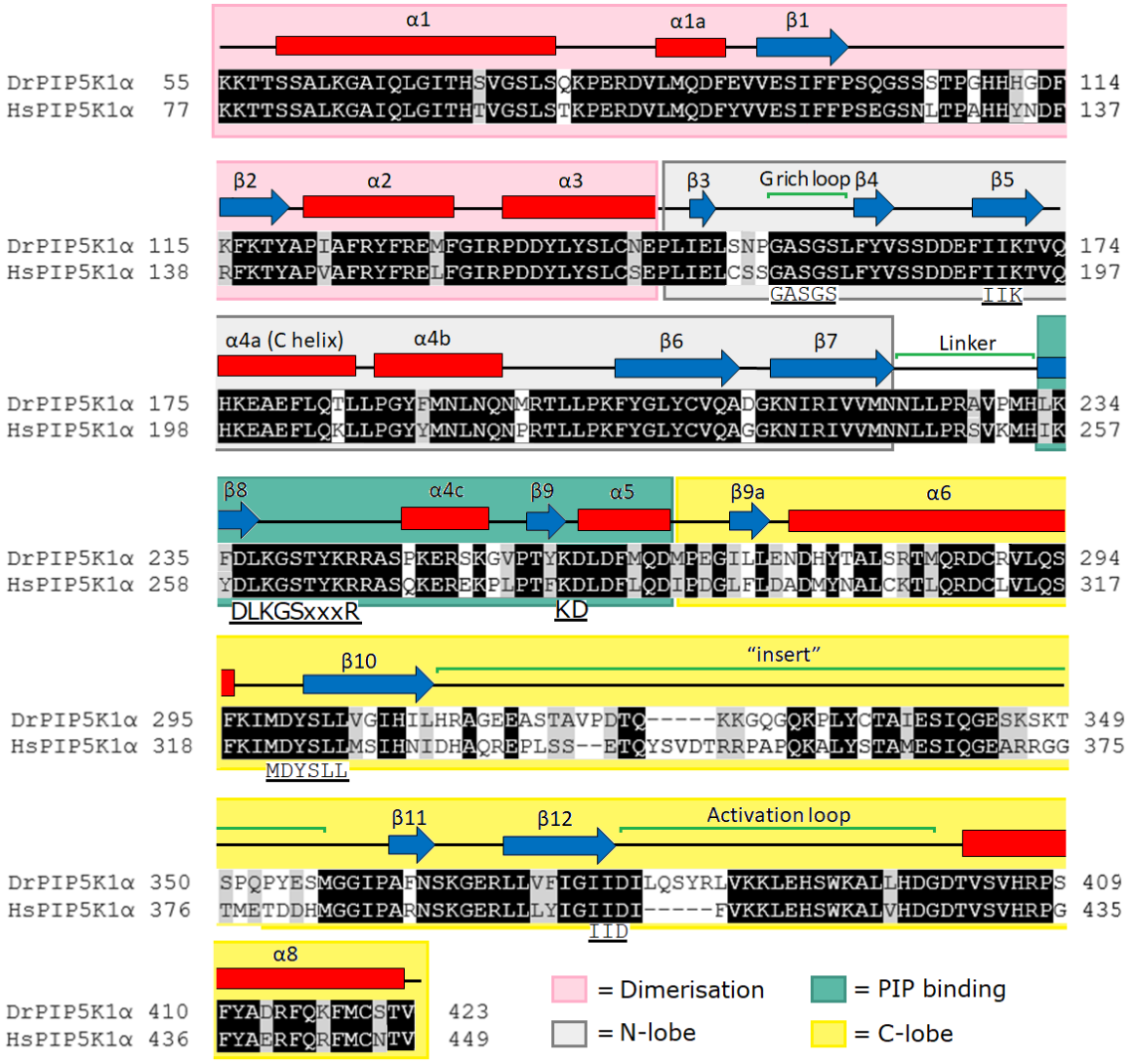


Figure 7. Structure of PIP5KIα

A) Deduced structure of *Homo sapiens* PIP5KIα, constructed by reference to the published structure of *Danio rerio* PIP5KIα [66] produced by Professor Barrie Kellam (University of Nottingham). Figure generated in PYMOL. Domains are coloured as follows: Dimerisation – Pink; N-Lobe – White; PIP binding domain – turquoise; C-lobe – Yellow. Regions that were disordered in the *Danio rerio* PIP5KIα, namely the insert and activation loop, are indicated by a dotted line. The divergent N- and C-termini, which also were not resolved, are not shown.
B) Sequence alignment of the portion of *Danio rerio* (Dr) PIP5KIα for which the crystal has been solved, with Human (Hs) PIP5KIα. Note the numbering of residues for the Human PIP5KIα is for the longest isoform, i1. Residues forming α helices, β

sheets, or functionally characterised loops in the *Danio rerio F*PIP5KIα structure are indicated. The background boxes indicate the domains to which residues belong. Key functional motifs are also indicated. Note, there is no α7 in this annotation as it is based upon the annotation of previous crystal structure of PIP4KIIβ determined by Rao *et al*, which places a seventh alpha helix within the "insert" loop.

Certain key motifs are conserved amongst all members of the PIPK family. These are:

- The consensus GxSGS motif, shortly followed by the IIK motif (177-GSG SLFYVSSDDEFIIK-194 in *H sapiens* in PIP5KIα_i1). The GxSGS motif corresponds to the Glycine–rich loop of protein kinases, and the II<u>K</u> motif corresponds to the downstream motif, VAI<u>K</u> in PKA, including a lysine that facilitates ATP binding [66]. In PIP5KIα the glycine rich loop folds down over ATP and forms multiple interactions with the γ and β phosphates of ATP. In a crystal structure of *D. rerio* PIP5KIα in which the glycine-rich loop was resolved in apo state with ATP, G154, A155, S176 and S178 each form hydrogen bonds either the β or γ phosphates of ATP [67]. These are the equivalents of *H. sapiens* PIP5KIα i1 G177, A178, S179 and S181. These interactions seemingly help to orientate the γ phosphate of ATP for phosphoryl transfer to PI4P, but are not essential for catalysis since deletion of the G-rich loop only reduces catalytic activity to 30% of the Wt

enzyme [67]. Lysine 194 of *H. sapiens* PIP5KIα i1 (K181 of PIP5KIα i2 ; K171 in *D. rerio* PIP5KIα) resides within the β-strand 5 (β5), the equivalent to β3 of PKA. The amino group of this lysine's side chain forms hydrogen bonds with the γ phosphate of ATP [73]. Mutation of this Lysine to Alanine has been reported to either completely [74] or partially (to ~10% of Wt protein[75]) abrogate kinase activity of PIP5KIα.

- The DLKGSxxxR motif (259-272 in Human PIP5KIα i1; 236-244 in *D. rerio* PIP5KIα) within the PIP binding domain. The side chain of the aspartate forms a hydrogen bond with the 2' hydroxyl of ATP. A mutation, D253N, in the equivalent position in PIP5KIγ inactivates the catalytic activity of the kinase. Homozygous inheritance of this mutation causes Lethal Congenital Contractural Syndrome 3 (LCCS3), a form of congenital arthrogryposis which is typically lethal within the neonatal period [76]. The lysine of the DLKGSxxxR motif forms a hydrogen bond with the γ phosphate of ATP. In PIP5KI's, the Lysine and the Arginine of this motif constitute a positively charged pocket for interaction with the D4 phosphate of PI4P (for PIP5KIs), allowing the D5 hydroxyl to enter a position for nucleophilic attack and phosphoryl transfer from the γ phosphate of ATP. This arrangement also functions (in co-

ordination with the activation loop) to determine the substrate preference for mono-phosphorylated phosphoinositides [73]. The three x residues within the DLKGSxxxR motif are variant between subfamilies but conserved within each family (TYK in the PIP Type I's). These contribute to substrate specificity of each family by positioning the positively charged pocket so that phosphate groups on the of PIP head-group of the wrong PIP species (i.e. PI3P, PI5P or PI(3,4)P_2 for the PIP5KIs) either cannot enter the pocket, or can only enter in a conformations in which phosphoryl transfer is unfavourable [73]. This effect is not completely precise, particularly for the PIP5KIs, so that other PIP species can enter the active site and be phosphorylated, but this occurs at much lower efficiency. Two additional residues within the PIP binding domain, 15 residues C-terminal to the DLKGSxxxR motif, are conserved as KD in the Type I and Type II PIPKs, but LD in PIPfyve. The KD motif is proposed to contribute to the preference of the Type I and Type II PIP5Ks to phosphorylate the hydroxyl adjacent to the monophosphate already present on the substrate [73].

- The MDSYLL motif is in a position in the crystal structures of PIP5KIα that indicates it is equivalent to the HRD motif of PKA. The serine within this motif (S301 in *D. Rerio*

PIP5KIα; S324 in *H. Sapiens* PIP5KIα_I1) forms a hydrogen bond with the Aspartate of the DLKGSxxxR motif, binging the Aspartate into position to form a hydrogen bond with the 2' hydroxyl of ATP. The serine within the MDSYLL motif hydrogen bonds with a water molecule, which co-ordinates with a metal ion (Mn^{2+}, Mg^{2+}, or Ca^{2+}) that coordinates with the α and γ phosphates of ATP, and is necessary catalysis [73]. S301A mutation of *D. Rerio* PIP5KIα reduces catalytic activity 20 fold [67].

- The GIID motif, which immediately proceeds the activation loop. The aspartate within this motif (D378 in *Danio rerio*; D414 in *H.sapiens* PIP5KI i1) corresponds to D184 in PKA within the latter's conserved DFG motif [77]. In the crystal structures of *Danio rerio* PIP5KIα, this Aspartate coordinates with the two metal ions (including the one also co-ordinated by S301) to hold the α, β and γ phosphates of ATP in position for phosphoryl transfer [66;73]. The equivalent Aspartate is required for kinase activity of PIP5KIα [78].

An additional conserved feature of the PIPKs is termed the insert, a ~50 residue loop within the C-lobe of the kinase domain. This was not resolved in any of the published crystal

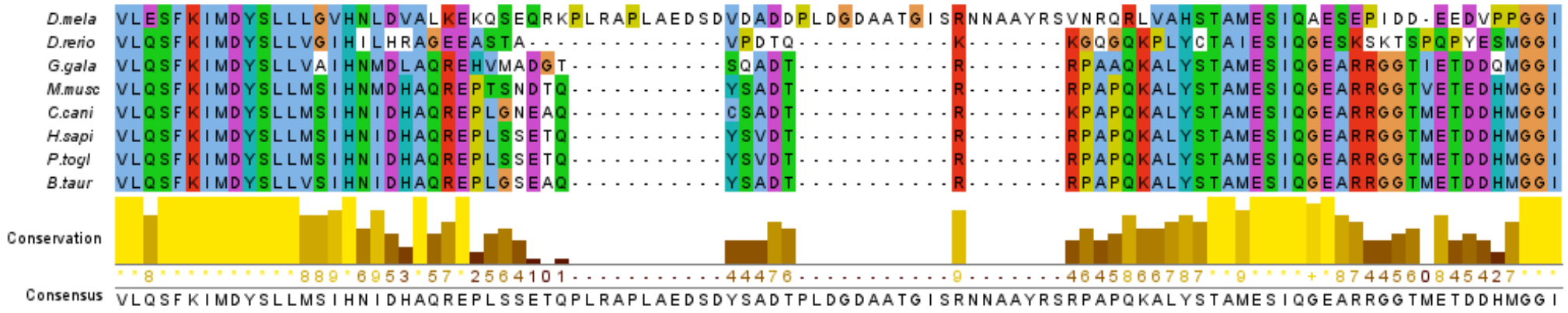
structures of *D. rerio* PIP5KIα, and is presumed to be an unstructured or partially unstructured loop [66]. Modelling of PIP5KIα's interface with the lipid membrane [79] has this insert on the face of the protein orientated into the intracellular fluid.

Similar to the N- and C-terminal flanking regions, this insert is largely divergent between isozymes but conserved within isozymes across metazoan species (**Figure 8**.), suggesting evolutionarily conserved function in each isozyme. In PIP4KIIβ, this loop was found to contain an α-helix, unique to this isozyme, which functions as an atypical nuclear localisation signal [80]. Phosphorylation of a serine residue within this insert in PIP4KIIβ also appears to function to promote its translocation to the plasma membrane [81]. A subsequence (^{356}QKALYSTAMESIQG^{369} in human PIP5Kiα i1) within this insert shows high conservation between Human PIP5KIα, PIP5KIβ and PIP5KIγ, and is also particularly well conserved between species for PIP5KIα (**Figure 8. A and B**). PhosphoSitePlus, an online repository of database of post-translational modifications identified in high throughput mass spectrometry studies [82], reports that a tyrosine residue within this subsequence, PIP5KIα Y360, has been identified as phosphorylated in 61 studies (**Figure 8 C**). This insert sequence may therefore serve an important conserved biological function, with possible

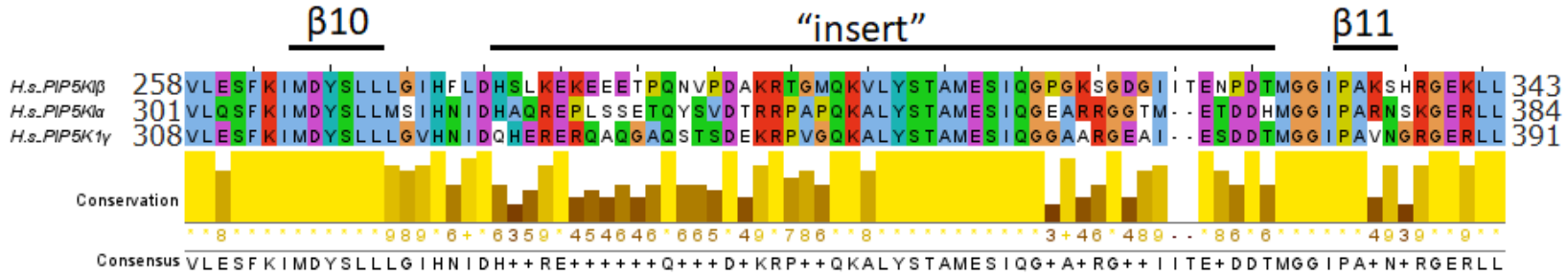
regulation by phosphorylation. Possible functions of the insert loop in the PIP5KIs have not been reported in the literature, and were not investigated further in the course of this project.

At the C-terminal end of the C-lobe is 20–30 amino acids loop referred to as the activation loop, which is responsible for substrate specificity and PIPK membrane targeting [83;84]. The activation loop of PIP kinases follows immediately from the GIIDIL sequence. The activation loop acts in concert with the PIP binding domain to determine substrate specificity of each PIPK subfamily (reviewed in [73]), and single point mutations (PIPKIIβ A381E and PIPKIβ E362A) are sufficient to swap the substrate preference of the PIP5K and PIP4K subfamilies [84]. While the activation loop structure was not resolved in the Hu *et al.*, X-ray crystallography study [66], a subsequent study using nuclear magnetic resonance found that upon interaction with a micellar surface, the unstructured loop folds into an amphiphatic loop. This anchors the enzyme to the membrane, and allows correct orientation of the enzyme against the membrane to facilitate binding of the substrate head group and subsequent catalytic activity [85].

A)


D.mela
D.rerio
G.gala
M.musc
C.cani
H.sapi
P.togl
B.taur
Conservation
Consensus
VLQSFKIMDYSLLMSIHNIDHAQREPLSSETQPLRAPLAEDSDYSADTPLDGDAATGISRNNAAYRSRPAPQKALYSTAMESIQGEARRGGTMETDDHMGGI

B)


β10
"insert"
β11
H.s_PIP5Kβ
H.s_PIP5Kα
H.s_PIP5Kγ
258
301
308
343
384
391
Conservation
Consensus
VLESFKIMDYSLLLGIHNIDH++RE++++++Q+++D+KRP++QKALYSTAMESIQG+A+RG++IITE+DDTMGGIPA+N+RGERLL

C)

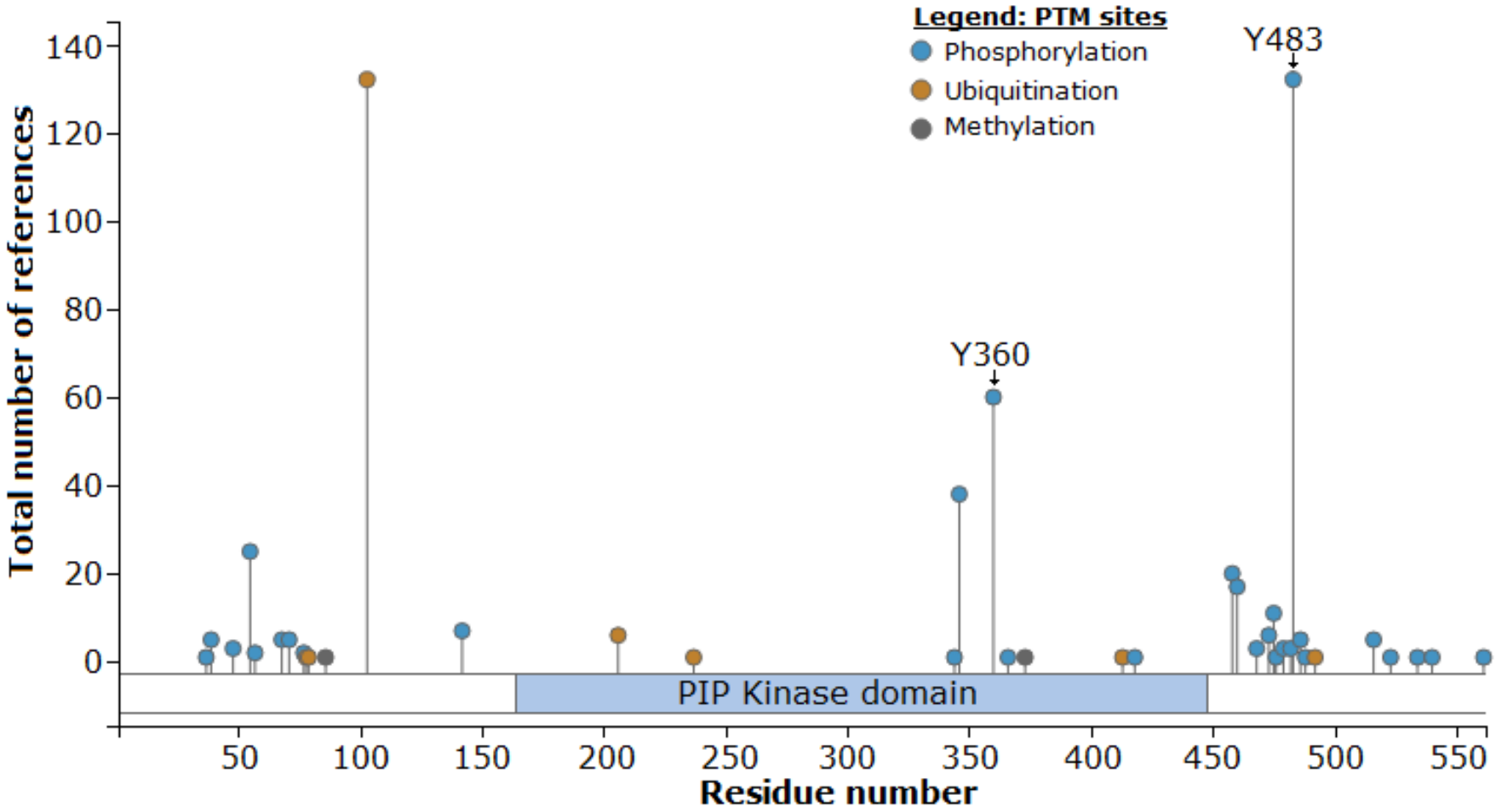

Legend: PTM sites
Phosphorylation
Ubiquitination
Methylation
Y483
Y360
PIP Kinase domain
Total number of references
Residue number
0
20
40
60
80
100
120
140
50
100
150
200
250
300
350
400
450
500
550

Figure 8. The insert loop of PIP5KIα

A) Multiple sequence alignment of the insert loop of PIP5KIα for several metazoan species.

B) Multiple sequence alignment of the insert loop of *Homo sapiens* PIP5KIα, PIP5KIβ and PIP5KIγ.

C) Post-translational modifications of PIP5KIα according to PhosphoSitePlus, which collates highthrough-put mass spectrometry studies and literature reports [86], and from which this graph is taken. Tyrosine 360 (Y360) is within the insert loop (QKALYSTAM) and has been identified as phosphorylated in 61 studies. Tyrosine 483 (Y483) is reported to be phosphorylated in 132 studies. This residue is encoded by Exon 13, and it inclusion in PIP5KIα is thus subject to alternative splicing of *PIP5K1A* transcripts.

1.5. The PI3K/AKT pathway in prostate cancer

In normal cells, production of $PI(3,4,5)P_3$ is induced transiently by growth factor stimulation, and opposed by lipid phosphatases, such as PTEN, which converts $PI(3,4,5)P_3$ back to $PI(4,5)P_2$, thus preventing continuous activation of downstream effectors. Cancer cells frequently have elevated levels of $PI(3,4,5)P_3$ due to a combination of increased signalling upstream of PI3K, mutational activation of PI3K [87-89], and loss of function of negative regulators, especially PTEN [90; 91], supporting continuous activation of the downstream effectors and the associated cancer cell phenotypes.

The vast majority of studies on the role of downstream effectors of PI3K in cancer have focused on AKT, however there are many contexts in which other PI3K-dependent signalling proteins contribute to cancer progression, independently of AKT (for review see [92]).

In prostate cancer, aberrations in the PI3K/AKT pathway are present in approximately 40% of early cases and 70% in advanced disease and loss of PTEN is present in 30% of primary and 60% of castrate-resistant prostate cancers [93]. Additionally activation of the PI3K pathway is associated with resistance to

ADT, disease progression and poor outcomes in prostate cancer [93].

1.6. PIP5KIα in prostate cancer

In 2014, Semenas *et al*., first demonstrated the clinical importance of PIP5KIα in PCa, with their observation of elevated expression of PIP5KIα (as evidenced by tissue microarray) in primary PCa tumours, and further elevated expression in metastatic PCa tumours. They also showed that PIP5KIα expression correlates with elevated AKT and AR gene expression in PCa tumour specimens. They also reported association between elevated PIP5KIα expression and poor patient prognosis [94]. The authors also demonstrated that reduction in PIP5KIα protein level in PCa cell lines, either by siRNA knockdown of *PIP5K1A*, or inhibition with compound ISA2011B (this appears not to be a classical kinase inhibitor, rather it binds PIP5KIα and induces its proteasome mediated degradation [95;96]) leads to reduced AKT activation, reduced level of AR protein, reduced proliferation, reduced invasion and reduced tumourgenicity in xenograft mice. Conversely, exogenous expression of PIP5KIα (PIP5KIα i4) in a non-malignant prostate cell line lead to increased proliferation and invasion, increased AKT activation and increased AR protein

level [94]. This study also found evidence for interaction of PIP5KIα and AR in a protein complex [94].

A later study by Sawar et al [95] demonstrated that PIP5KIa expression correlates with AR-V7 expression in patient tumour specimens. The authors also demonstrated that exogenous expression of AR-V7 in a PCa cell line increases PIP5KIα expression, and demonstrated co-immunoprecipitation of PIP5KIα, AR-V7 and Cyclin dependent kinase 1 (CDK1) in nuclei of 22RV1 PCa cells.

Thus there is growing evidence for the importance of PIP5KIα in PCa, with evidence of functional interactions of AR and PIP5KIα.

1.7. Aims of the project

The larger context of the work of which my project is a part is investigation of PIP5KIα and its functional interactions with androgen signalling pathway mediated by AR, particularly mechanisms by which PIP5KIα contributes to development and progression of PCa.

To that end, the work performed in this project follows three interconnected streams, encapsulated in the three results chapters of this thesis:

1) Chapter 3: Recently before my beginning on this project, our group utilised CRISPR-Cas9n genome editing technology in the androgen insensitive cell line LNCaP C4-2, generating multiple sublines with indels at the *PIP5K1A* locus. I set out to characterise these sublines for their expression of PIP5KIα, and to investigate effects of functional depletion of PIP5KIα on gene expression, particularly of known androgen regulated genes.

2) Chapter 4: RNA-Sequencing performed for the transcriptomes of these CRISPR-Cas9n-edited and WT LNCaP C4-2 cells lead us to observe a complex pattern of alternative splicing of *PIP5K1A* transcripts, including use of cryptic upstream exons and dual coding exons. I set out to investigate the extent of alternative splicing of *PIP5K1A* transcripts in LNCaP C4-2 and MCF7 (a breast cancer cell line) with a view

towards functional importance of such transcripts and the encoded isoforms.

3) Chapter 5: Previous work by our collaborators (JL Persson group, Umea Sweden) indicated the presence of PIP5KIα and AR in protein complexes within nuclei of PCa cell lines [95], which is interesting because it indicates one possible mechanism for functional interactions of PIP5KIα and AR in PCa. I set out to investigate the possibility of direct physical interactions of PIP5KIα and AR by use of yeast two hybrid. Building upon the work of Chapter 4, I also investigated PIP5KIα isoform specific interaction with AR and other reported interactors of PIP5KIα, K-Ras and Star-PAP.

These aims were intended to help to elucidate molecular functions of PIP5KIα and its many putative isoforms and validate it as an important vulnerability in metastatic prostate tumours.

Chapter 2

Materials and Methods

2.1. Sources of Materials

2.1.1. General Suppliers

All general laboratory chemicals were purchased from Fisher Chemicals or Sigma-Aldrich unless otherwise stated and all are analytical grade. PBS (Phosphate Buffer Saline) tablets, purchased from OXOID Ltd, were used to prepare PBS solutions. All solutions were made in double deionised water and if RNase free water was required, deionised water was purified by passage through a Purite Select/Neptune water purification system. The pH of solutions was determined with a Jenway 3510 pH meter, checking against standards of known pH. Where sterilisation was required, the materials were sterilised by filtration through 0.2 micron filter or by autoclaving for 20 mins at 120 °C and 25 lb/in^2 pressure.

2.1.2. Bacterial reagents

The bacteriological growth medium reagents tryptone and yeast extract were purchased from OXOID Ltd. The *Escherichia coli* strain DH5α (Genotype: F$^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–) [87] were purchased from Stratagene.

2.1.3. Molecular biology reagents

All enzymes and their buffers were purchased from Roche Applied Science, New England Biolabs (NEB) or Promega. 1 kb Plus DNA ladder was purchased from Invitrogen Life Sciences. Polymerase enzymes and their respective buffers and supplements were purchased from NEB. dATP, dCTP, dGTP and dTTP were purchased from Invitrogen Life Sciences. Monarch DNA gel extraction kits were purchased from NEB, Nucleospin® plasmid kits from Macherey-Nagel, and maxi-DNA kits from QIAGEN Ltd. Oligonucleotides for sequencing and PCR applications were purchased from Sigma-Aldrich.

2.1.4. Yeast reagents

The yeast strain *Saccharomyces cerevisiae* L40 (Genotype: mat-a hisΔ200trp1-901 leu2-3,112 ade2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ Gal4 was a gift from S. Hollenberg. Bacteriological-yeast nitrogen base without amino-acids, and yeast synthetic drop-out mix without Leu and Trp for preparation of yeast selective media were purchased from Sigma-Aldrich. Herring sperm DNA was purchased from Promega.

2.1.5. Nuclear hormone receptor ligands

All ligands were obtained from Sigma-Aldrich apart from mibolerone, which was supplied by Steraloids Inc.

2.1.6. Tissue culture reagents

All tissue culture supplements were purchased from Lonza Biowhittaker® and plastic-ware was obtained from Helena Biosciences and ibidi GmbH. The human metastatic prostate cancer cell line, LNCaP C4-2 was a gift from Professor Nigel Mongan (University of Nottingham).

2.1.7. SDS-PAGE, western blotting and chemiluminescence reagents

SDS-PAGE and western blot transfer apparatus were purchased from Bio-Rad. 30% (w/v) acrylamide solution was purchased from National Diagnostics. Color Prestained Protein Standard, Broad Range (11–245 kDa) was purchased from NEB. Nitrocellulose transfer membranes were purchased from Bio-Rad. Pierce™ Enhanced Chemiluminescence Western Blotting Substrate was purchased from Thermo Fisher Scientific. Bradford assay reagent was purchased from Bio-Rad and bovine serum albumin (BSA) for protein standards was purchased from First Link (UK) Ltd. The SDS-PAGE molecular weight protein marker was purchased from NEB. All other reagents such as 30% (w/v) acrylamide and N,N,N,N-tetramethylethylenediamine (TEMED), were also purchased from Sigma-Aldrich. Complete Protease Inhibitor cocktail tablets were purchased from Roche Applied Science.

2.1.8. List of antibodies used

Table 2. List of antibodies used in this project, with the application and dilution of their use, manufacturer's name and respective catalogue number.

Antibody	Immunogen	Clonality	Application	Dilution	Manufacturer/ Product Number
Rabbit anti-LexA primary antibody	*E. Coli* LexA residues 1-202	Polyclonal	Western blot	1:500	Merck Millipore Cat# 06-719
Mouse anti-VP16 primary antibody	Herpes simplex virus VP16, residues 456-490	Monoclonal	Western blot	1:500	Santa Cruz Biotechnology Cat# sc-7546
Mouse anti-β actin antibody	β actin modified N-terminal peptide, species not listed	Monoclonal	Western blot	1:500	Sigma-Aldrich Cat# A5441
Rabbit anti AKT	*M. musculus* AKT1 C-terminal sequence	Polyclonal	Western blot	1:1000	Cell Signalling Technologies Cat# 9272
Rabbit anti-Phosho-Ser473 AKT	*H. sapiens* AKT1, residues surrounding P-Ser 473	Monoclonal	Western blot	1:1000	Cell Signalling Technologies Cat#4060
Rabbit anti-Phosho-Thr308 AKT	*H. sapiens* AKT1, residues surrounding P-Thr 308	Monoclonal	Western blot	1:1000	Cell Signalling Technologies Cat# #13038
Rabbit anti-PIP5KIα	*H. sapiens* PIP5KIα (isoform1), residues centred around Gly 64	Polyclonal	Western blot	1:1000	Cell signalling Technologies Cat#9693
Chicken anti-mouse IgG-HRP secondary antibody	*M. musculus* IgG	Polyclonal	Western blot	1:5000	Santa Cruz Biotechnology Cat# sc-2954
Goat anti-Rabbit IgG-HRP secondary antibody	*O. cuniculus* IgG	Polyclonal	Western blot	1:5000	Santa Cruz Biotechnology Cat# sc-2004

2.2. Bacteriological preparation and culture

2.2.1. Composition of solutions and media used in bacterial procedures

Luria-Bertani (LB) medium: 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl. pH adjusted to 7.0 with NaOH.

Super Optimal Broth (SOB) medium: 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl**,** 10 mM $MgSO_4$, 10 mM $MgCl_2$ (filter sterilised, added immediately prior to use). pH adjusted to 7.0 with NaOH.

Super Optimal Broth with Catabolite suppression (SOC) medium: SOC was prepared identically to SOB, with the addition of 20 mM D-glucose. In order to limit formation of growth inhibitory Maillard reaction products, D-glucose was added from sterile filtered 1 M D-glucose solution after the SOB had been autoclave sterilised and allowed to cool to below 60 °C.

For solid medium plates, 2% (w/v) bacteriological agar was added to medium prior to autoclave sterilisation. Where selective antibiotics were required, these were added to molten media at ~60 °C to prevent thermal decomposition of the antibiotics. 20-25 ml of molten agar medium was poured into each Petri dish and allowed to solidify at room temperature. Plates were stored at 4 ºC for up to 1 month.

Ampicillin: 1000x stock solution: 100 mg/ml in 50% (v/v) ethanol.

Kanamycin: 1000x stock solution: 50 mg/ml in H_2O.

Chloramphenicol: 1000x stock solution: 34 mg/ml in 50% (v/v) ethanol.

2.2.2. Culture of Escherichia coli strains

The bacterial strain *E. coli* DH5α was used for all DNA manipulations. To obtain single colonies the bacteria were streaked onto agar plates (containing appropriate selective antibiotics where required) and grown at 37 °C. Once single colonies had appeared, the plates could be stored for up to one month at 4 °C. Liquid cultures were grown by placing a single colony into liquid media (with appropriate antibiotics where required) and incubated in universal screw-top tubes (Bibby Sterilin, Staffordshire, UK) or conical flasks at 37 °C in an orbital shaker at 200 rpm.

2.2.3. Preparation of chemically competent *E. coli* cells

The calcium chloride and magnesium chloride method was used to make *E. coli* DH5α cells competent for DNA transformation. A single colony from a freshly streaked SOB plate was inoculated in to 25 ml SOB medium and grown overnight at 37 ºC. The overnight grown culture was diluted

1:100 in fresh SOB medium and allowed to grow at 37 ºC until the Optical Density at 600 nm (OD_{600}) reached 0.3-0.4. The cells were incubated on ice for 20 mins, and then pelleted by centrifugation at 4 ºC, 1,500 g for 5 mins to harvest the cells. Cell pellets were resuspended in 20 ml of chilled 100 mM $MgCl_2$ and incubated on ice for 5 mins. Cells were again pelleted by centrifugation as before, and the supernatant discarded. Cell pellets were resuspended in 10 ml chilled 100 mM $CaCl_2$, and incubated on ice for 30 mins. Cells were again pelleted as before, and resuspended in 5 ml 70 mM $CaCl_2$ and 15 % glycerol solution. 200 µL aliquots of the cell suspension in microfuge tubes were snap frozen in liquid nitrogen and stored at -80°C. The transformation efficiency of the cells was checked by transforming with a known amount of control plasmid (pUC19), and then determined according to the following formula:

Transformation efficiency = Number of colonies obtained x dilution factor x 1/mass of DNA (µg)

1 µg of plasmid DNA should yield ~$1x10^6$ transformants. If the transformation efficiency of the competent cells ~$1x10^6$ or greater, and if no bacterial colonies were obtained in a no DNA control then the competent cells were deemed suitable for use.

2.2.4. Heat shock transformation of chemically competent E.coli

50 or 200 µL aliquots of competent DH5α *E.coli*, prepared as described in Section 0, were used for each transformation. When transforming previously constructed plasmids, 50 µL of competent cells were transformed with 1 ng of DNA to ensure a suitable number of successfully transformed colonies. When transforming with ligation mixtures, 5-10 µL of the ligation mix (**Section 2.3.10**) was used to transform 200 µL of competent cells.

Cells and DNA were mixed in a pre-chilled 1.5 ml microfuge tube and incubated on ice for 30 mins. The transformation mixture was then heat shocked by incubation at 42 °C for 2 mins, followed by incubation on ice for 10 mins. Transformation mixtures were made up to 1 ml with room temperature SOC media and they were incubated for 1 hr at 37 °C in an orbital shaker at 200 rpm. The cells were centrifuged at 2,500 g for 2 mins and 900 µL of the supernatant were discarded. The remaining 100 µL of the supernatant were used to resuspend the cell pellet and this cell suspension was spread on room temperature LB agar plates containing the appropriate antibiotic. The plates were then incubated at 37 °C overnight.

2.3. Molecular biology techniques

2.3.1. Small scale plasmid DNA preparation (miniprep)

Machery-Nagel NucleoSpin® plasmid mini prep kits were used for the small-scale preparation of plasmid DNA. This DNA was primarily used to screen for plasmids containing the correct DNA inserts/site directed mutations. Briefly, a single bacterial colony was picked and used to inoculate 10 ml of LB containing the appropriate antibiotic and grown overnight in an orbital shaker at 37°C, 225 rpm. 9 ml of the overnight culture were centrifuged in a microfuge tube at 4,000 g for 5 mins to pellet the cells. The remaining bacterial culture was stored at 4 °C for later use. The pelleted cells were then resuspended and subjected to alkaline lysis. The plasmid DNA was absorbed onto a silica gel membrane, washed and eluted. This technique was carried out according to manufacturer's instructions and details can be found in the manufacturer's handbook.

2.3.2. Large scale plasmid DNA preparation (maxiprep)

In order to obtain larger quantities (up to 500 µg) of high quality plasmid DNA suitable for transient transfections, the QIAGEN-tip 500 plasmid maxiprep kit was used according to manufacturer's instructions. A single colony, containing the plasmid of interest, was picked and used to inoculate 500 ml LB

containing the appropriate antibiotic. The culture was grown in an orbital shaker at 37 °C, 200 rpm overnight. Isolation of the plasmid DNA was achieved by alkaline lysis of the cells, followed by immobilization of the plasmid DNA on QIAGEN anion-exchange resin, elution and precipitation in isopropanol. DNA precipitates were resuspended in 400 μL sterile H_2O and re-precipitated in 1 ml absolute ethanol, prior to being washed in 70% (v/v) absolute ethanol, dried under vacuum to remove residual ethanol and resuspended in the appropriate volume of sterile H_2O. Details of this technique and compositions of solutions used can be found in the manufacturer's handbook.

2.3.3. Quantification and qualitative assessment of DNA

DNA quantity and quality (purity) were assessed by spectrophotometry. The concentration of DNA in a given sample was determined by optical density measurement at 260 nm (OD260) using a UV spectrophotometer (Eppendorf). DNA concentration was calculated by using the following formula:

OD260 x Dilution factor x 50 = Concentration of DNA in pg/ml

The purity of DNA in a given sample was estimated by the ratio of OD260:OD280 and the ratio of OD260:OD230. Pure DNA has an OD260/OD280 of between 1.7 and 1.9. OD260/OD280 values below this range indicate contamination

of sample with protein, phenol or other compounds that absorb strongly at ~280 nm [97].

Pure DNA has an OD260/OD230 of between 2 and 2.2. OD260/OD230 values below this range indicate the presence of contaminants such as EDTA, guanidine HCl, TRIzol or guanidine isothiocyanate, which absorb around 230 nm [97].

Additional assessments of plasmid DNA purity (presence of a single plasmid following a plasmid prep) and confirmation of the identity of the plasmid were carried out by performance of restriction digests **(Section 2.3.9)** followed by separation via agarose gel electrophoresis **(Section 2.3.5)**, comparing fragment sizes to molecular markers of known molecular weight. Semi quantitative assessment of DNA concentration was performed by comparison of DNA samples to the molecular weight markers, which are of known concentration.

2.3.4. Ethanol precipitation of DNA

Ethanol precipitation was performed on samples of DNA when their current buffer was not compatible with restriction endonucleases with which they were to be digested.

Samples of DNA were transferred to clean microfuge tubes, to which 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2.5 volumes of absolute ethanol were added to precipitate the DNA, and the samples vortexed for 10 sec. Samples were

incubated at -20 °C for 1 hour. Samples were then centrifuged at 13,000 g for 30 mins at 4 °C and the supernatant discarded. The pellet was washed in 70% (v/v) absolute ethanol, centrifuged and the supernatant discarded as before. The resulting pellet was dried under vacuum to remove residual ethanol and resuspended in an appropriate volume of sterile H_2O.

2.3.5. Agarose gel electrophoresis

DNA was size fractionated on neutral agarose gels of between 0.8 and 2% agarose (w/v) as appropriate to the size of the fragments being resolved. Gels containing 0.5 μg/ml ethidium bromide were made and run in 0.5 x TBE buffer as described in standard protocols [98]. DNA samples were loaded in a 0.5 x gel-loading buffer. Following electrophoresis, gels were visualised by trans-illumination with ultra violet light and photographed via Gel-Doc 2000 (Bio-Rad). The size (kb) and amount (ng) of DNA fragments was determined by comparison with a reference DNA ladder of known sizes and conentration.

2.3.6. Polymerase chain reaction (PCR)

PCR was used for various purposes such as to generate DNA fragments needed for construction of recombinant plasmids, to screen newly generated recombinant plasmids for the correct DNA inserts (colony PCR).

For the synthesis of DNA in which high fidelity was required, such as for use in molecular cloning and site directed mutagenesis, a DNA polymerase with proofreading (3' to 5' exonuclease) activity, Phusion® High Fidelity DNA polymerase (NEB) was used. A typical reaction mixture consisted of:

Phusion® HF 5x buffer	10.0 µl
dNTP mixture (10 mM)	1.0 µl
Forward PCR primer (10 µM)	2.5 µl
Reverse PCR primer (10 µM)	2.5 µl
DNA template (25-100 ng/µl)	1.0 µl
DMSO	1.5 µl
Phusion® High Fidelity DNA polymerase	0.5 µl
Sterile H_20	36.0 µl

For amplifications such as for screening for plasmids for inclusion of inserts, or genotyping purposes, in which high fidelity of amplicon sequence was not required, Taq DNA polymerase (NEB) was used. A typical reaction mixture consisted of:

10x Standard Taq buffer (NEB)	2.5 µl
dNTP mixture (10 mM)	0.5 µl
Forward PCR primer (10 µM)	0.5 µl
Reverse PCR primer (10 µM)	0.5 µl
DNA template (25-200 ng/µl)	1.0 µl

Taq polymerase (NEB)	0.125 µl
Sterile H_20	19.875 µl

Reactions were incubated in an Applied Biosystems 2720 thermal cycler, a 'hot start' PCR method was utilised to prevent mis-annealing of primers. Samples were initially denatured by incubating at 98 ℃, (for reactions with Phusion polymerase), or 95 ℃ (for reactions with Taq polymerase) for 3 mins before typical thermal cycling parameters as follows: denaturation, 95 ℃ for 30 seconds, annealing, 55 ℃ for 30 seconds and extension, 72 ℃ for 30 seconds per kb of target DNA, for a total of 25-35 cycles. As a final step, the samples were incubated at 72 ℃ for 10 mins. Upon completion, 10% of PCR reaction was separated by agarose gel electrophoresis to confirm that appropriately sized DNA fragments had been produced.

2.3.7. Purification of genomic DNA from mammalian cells

Genomic DNA was purified from mammalian cells using the Sigma GenElute™ Mammalian Genomic DNA Miniprep Kit, following the manufacturer's instructions. To harvest adherent cells, cells culture monolayer were trypinised to detach cells, transferred to 15 ml tubes and centrifuged at room temperature at 600 x *g* for 5 minutes. The supernatants were discarded and the pellets resuspended separately in 1 ml 1 x PBS, transferred to 2 ml microcentrifuge tubes, and the centrifugation repeated.

Again, the supernatants were removed and the genomic DNA was then purified according to the manufacturer's instructions. Genomic DNA was stored at -20°C until required.

2.3.8. Purification of DNA fragments from agarose gel slices

DNA fragments were purified from agarose gel slices using an NEB Monarch DNA gel extraction kit according to manufacturer's instructions. The DNA was bound to the silica membrane, washed and eluted in sterile, nuclease-free water (pH 7–8.5).

2.3.9. Restriction digest of DNA

Plasmid DNA and products of PCR amplification were digested with restriction endonucleases in order to generate compatible ends for cloning and to verify newly created plasmids. The reaction conditions were set up according to manufacturer's instructions. Typically, digests were set up in a 40 µL reaction volume consisting of:

DNA	x µL
restriction endonuclease (10 units/µl)	1 µL
10x buffer	3 µL
sterile H_2O	x µL, to total volume of 40 µL

Digests were normally carried out at 37 °C for 3-4 hr unless advised otherwise in the manufacturer's instructions.

2.3.10. Ligation of DNA fragments

Recombinant plasmids were created by annealing digested fragments using T4 DNA ligase. All restriction digests in this study resulted in the production of compatible cohesive ends. Relative quantities of cut vector and digested inserts were determined by spectrophotometric quantification (Section 0) and cohesive end ligation reactions were set up so the vector:insert ratio was 1:3 and 1:5. Ligations were set up as follows:

vector	x μl
insert	x μl
10x T4 DNA ligation buffer	1 μl
T4 DNA ligase (10 units/μl)	1 μl
Sterile H2O	x μl to a total volume of 10 μl

Early in the course of this project, such reaction would be incubated at 16 °C overnight. Later in the course of this project, such ligation reactions would be incubated in a thermal cycler in cycles of 10 °C for 30 secs and 30 °C for 30 secs, for 16 hours. Such conditions have been reported to increase cloning efficiency by ~6-8 fold [99]. The effect of cycling temperature on cloning efficiency was not systematically assessed in this project, though did appear to greatly increase number of colonies with correct insert.

Ligated DNA was transformed into DH5α *E. coli* (**Section 2.2.4**) and plated on LB agar plates containing appropriate antibiotics for selection. Colonies were screened by colony PCR (**Section 2.3.14**) plasmid preparations made (**Section 2.3.1**) and verified for the correct integration of gene fragment by restriction digest (**Section 2.3.9.**) and finally DNA sequencing (**Section 2.3.15.**).

2.3.11. Vaccinia Virus Topoisomerase 1b mediated cloning (TOPO cloning)

Topo cloning was performed with Zero Blunt TOPO PCR cloning kit (Invitrogen), with a modified version of the manufacturer's protocol. The modification was simply to halve all volumes of reagents used in the TOPO reaction, to economise use of the linearised pCR4 Blunt TOPO treated vector. Reactions were set up as follows:

pCR4 Blunt TOPO treated vector	0.5 μl
Salt solution (in kit)	0.5 μl
Phusion polymerase PCR product	2.5 μl

Reaction mixtures were incubated at room temperature for 15 minutes, then moved to ice, and used in transformation reactions with chemically competent DH5α *E. coli.*

2.3.12. Generation of mutations by recombinant PCR

To mutate specific nucleotides within a DNA sequence, forward and reverse oligonucleotides were designed to include the desired mutant nucleotide sequence and these used as primers for PCR. Reactions were performed to extend and incorporate the mutagenic primers, resulting in nicked circular strands of mutated plasmid DNA. PCR was carried out as described in **Section 2.3.6** with the amendment that the thermal cycling extension time was increased to 10 mins at 72 °C. PCR products were then treated with 1 µl (20 units) DpnI restriction endonuclease (NEB) to digest methylated parental DNA for 3 hours at 37 °C. Control (no oligonucleotides used) and sample reactions were separated by agarose gel electrophoresis to confirm complete digestion of parental template. The nicked, circular dsDNA was transformed into DH5α and the resulting single colonies were inoculated in cultures for small-scale DNA preparations, which were sequenced to confirm assimilation of the mutated nucleotide sequence.

2.3.13. Colony PCR screening

After transformation with ligation reaction mixtures, colonies were picked with a sterile pipette tip which was touched into 20 µl of sterile, nuclease free water in fresh 0.5 ml tubes.

These were then boiled for 5 mins. Boiled samples were centrifuged at 11,000 g for 2 mins, and 2 µl of the supernatant was used as input for a PCR reaction (**Section 2.3.6**) using TAQ DNA polymerase (NEB). Reaction products were separated by agarose gel electrophoresis and visualised, in order to screen for the appropriate DNA insert. Reactions consisted of the same reagents as previously stated with the following amendments: 1.25 units of TAQ DNA polymerase and its corresponding buffer (1×), 1.5 mM $MgCl_2$ and 2 µl of boiled bacterial colony supernatant instead of specific DNA template. Where large numbers of colonies were to be screened, multiple colonies (typically 4-6) were pooled into the same tube of sterile, nuclease free water, and processed as described above. Upon successful amplification from a pool, colonies constituting that pool were then colony PCR screened individually.

2.3.14. Sequencing of DNA

DNA samples were diluted to a concentration of 100 ng/µl for plasmid DNA, and 10 ng/µl for band extracted PCR products. Oligonucleotides primers were made up to a concentration of 3 pmol/µl .These were dispatched to Source Bioscience Ltd, Nottingham, UK, and sequenced using Big Dye v3.1 chemistry and an AB13730xl automated sequencer. Sequencing data was analysed in SnapGene Viewer software.

2.3.15. Oligonucleotides

Oligonucleotides were purchased from Sigma-Aldrich as lyophilised and desalted pellets. These were resuspended in sterile water to 100 μM concentration and stored at -20ºC.

2.3.16. Purification of RNA from mammalian cells

The QIAGEN RNeasy Mini kit was used to remove genomic DNA and isolate RNA ready for cDNA synthesis. Cells were seeded into dishes in appropriate volumes of media and allowed to attach and grow to confluency (~48 hours), after which time, media was aspirated and the cell monolayers were washed 2x in ice cold sterile PBS. After aspiration of PBS, a volume of lysis buffer supplied by the manufacture (volume varying with size of the dish and cell number as per the manufacturers' recommendations) supplemented with β-mercaptoethanol and cells were scraped and lysates collected into pre-chilled micro centrifuge tubes. Isolation of RNA was performed as descried in the manufacturers' protocol. RNA was stored at -80 °C until required. Samples were kept on ice at all times to prevent degradation of RNA transcripts.

2.3.17. Reverse transcription of RNA to synthesis DNA

In order to prepare extracted RNA for use as a template to synthesis cDNA, potential contaminating genomic DNA was removed using the QuantiTect Reverse Transcription kit

(QIAGEN Ltd), following the manufacturer's instructions.

Samples were set up as follows:

7x gDNA Wipeout buffer	2 µl
Template RNA	200-1000 ng
Sterile RNase-free water	To a total of 14 µl

Samples were mixed and incubated for 2 minutes at 42°C, then immediately replaced onto ice.

To synthesise cDNA, reverse transcription reactions were prepared using these gDNA wipeout treated samples, as follows:

Quantiscript Reverse Transcriptase	1 µl
5x Quantiscript RT buffer	4 µl
RT Primer mix	1 µl
gDNA removal reaction	14 µl
Total volume	20 µl

Reactions were incubated for 15 minutes at 42 °C, then for 3 mins at 95 °C to inactivate the Quantiscript Reverse Transcriptase enzyme. Subsequently, reactions were stored at -20 °C until needed.

2.3.18. Quantitative PCR (qPCR)

For each different mRNA transcript to be assessed, mastermixes were prepared as follows:

Sterile di H2O	9.0 µl
10 µM forward primer	0.5 µl

10 μM reverse primer	0.5 μl
Brilliant SYBR Green qPCR master mix	10.0 μl
Total volume	20.0 μl

qPCRs were performed in triplicate, upon biological triplicates, so that a total of 9 readings were obtained for each gene. Analysis of the housekeeping gene, GUSB, was also performed for every biological repeat in the same manner. Furthermore, negative controls were set up for each primer pair used, receiving either no cDNA template (receiving diH20 instead), or no RT (receiving gDNA wipeout treated RNA that did not subsequently undergo Reverse Transcriptase treatment). All samples and reagents were kept in ice until RT-qPCR reactions were ready.

After thorough mixing and distribution of the mastermixes into reaction tubes, 1 μl of cDNA at appropriate concentration was added to each well. Caps were placed on the wells prior to brief centrifugation. qPCRs were performed in a Stratagene Mx3005p qPCR machine. The reaction conditions were as follows:

No. of cycles	Time (mins:secs)	Temperature (°C)
1	10:00 (denaturation)	95
40	00:30 (denaturation) 01:00 (annealing)	95 Variable: 45-60

	01:00 (extension)	72
	1:00 (melt curve analysis)	95
1	00:30	57
	00:30	95

Analysis of the gene expression of each target gene was performed with output cycle threshold (Ct) values using ΔΔCt method. First, for each replicate within each treatment group for each gene, the ΔCt was determined:

ΔCt = Ct value – Mean Ct value for housekeeping gene for that treatment group

ΔΔCt was then determined for each replicate, relative to the mean ΔCt for the control group:

ΔΔCt = ΔCt – ΔCt for control group

Finally, normalised mRNA expression level was calculated:

Normalised mRNA expression level was calculated = $2^{-\Delta\Delta Ct}$

qPCR products were analysed by agarose gel electrophoresis to ensure production of only a single amplicon of the correct size. The dissociation curves and amplification plots of each reaction were also viewed in the MXPRO software to further confirm this.

2.4. Yeast methods

2.4.1. Composition of solutions for yeast culture

Z buffer: 60 mM $Na_2HPO_4 \bullet 7H_2O$, 40 mM $NaH_2PO_4 \bullet H20$, 10 mM KCl, 1 mM $MgSO4 \bullet 7H_2O$, 0.3% (3-mercaptoethanol, pH 7.0.

Breaking Buffer: 0.4 M KCl, 50 mM Tris-HCl, pH 7.9, 1 mM PMSF, one complete protease inhibitor cocktail tablet was added per 50 ml of buffer.

Yeast-peptone-dextrose (YPD) media: 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) D-glucose. Solidified where required by the addition of 2% bacteriological agar.

Synthetic dextrose (SD) minimal media: 0.67% (w/v) Bacto-yeast nitrogen base without amino acids (w/v), 2% D-glucose (w/v), 0.2% drop-out mix (w/v). Solidified where required by the addition of 2% bacteriological agar.

In order to limit formation of growth inhibitory Maillard reaction products, D-glucose was added from sterile filtered 40% (w/v) D-glucose solution after the YPD and SD media had been autoclave sterilised and allowed to cool to below 60 °C.

Transformation mix: 1.2 ml 50% PEG, 180 μl 1.0 M Lithium Acetate, 250 μl carrier DNA, 100 μL distilled water

2.4.2. Yeast two-hybrid plasmids

The DNA binding domain vector used in this project was pBTM116mod. pBTM116mod is a derivative of pBTM116 [100] that expresses the DBD of the bacterial repressor protein LexA and carries a TRP1 selection marker. Fusion proteins containing this DBD were generated by subcloning appropriate cDNAs in frame into the multiple cloning site. The acidic-activation domain (AAD) vector used in this study was pASV3mod, a modified version of pASV3 [101], which encodes the AAD from the herpes simplex virus protein VP16 c-terminally fused to a 10 amino acid nuclear localisation sequence (NLS) derived from the yeast protein L29, and carries a LEU2 selection marker. Fusion proteins to the NLS-AAD fusion were generated by subcloning appropriate cDNAs in frame into the multiple cloning site.

2.4.3. Yeast culture conditions

Saccharomyces cerevisiae cultures were grown in YPD medium, or where selection for uptake and maintenance of plasmids was required, in SD medium with appropriate supplements. Liquid cultures were grown by incubation in an orbital shaker at 30°C and 225 rpm for a time recommended by each protocol, or on plates by incubation at 30°C until individual colonies appeared.

2.4.4. Yeast transformation

A single colony of L40 yeast was inoculated into 25 ml YPD and incubated overnight in an orbital shaker at 30 °C and 200 rpm, until the culture reached an OD_{600} of 0.8-1.0 (approximately $2x10^7$ cells/ml). The cells were pelleted by centrifugation at 3,000 g for 5 mins at room temperature. The pellet was washed twice with distilled water and transferred to a fresh 1.5 ml tube. The cells were pelleted by centrifugation at 11,000 g for 15 second and the supernatant was removed.

Subsequently, the cells were resuspended in 250 μl 100 mM Lithium acetate and the suspension was incubated at 30 °C for 15 mins. In parallel, carrier Herring sperm DNA was boiled for 5-10 mins and quickly chilled on ice.

For each transformation, 173 μL of transformation mix was aliquoted into a fresh 1.5 ml tube. 500 ng of each transforming plasmid (~1-2 μL) and 25 μL of cell suspension was then distributed into the transformation mix, vortexed, and incubated at 30 °C for 30 mins with shaking. The cells were then heatshocked in a 42 °C water bath for 20 mins. The cells were pelleted by centrifugation at 11,000 g for 15 seconds and the supernatant discarded. The cells were resuspended in 200 μL distilled water and plated onto SD medium agar plates. The

plates were incubated at 30 °C for 3-4 days until colonies were ~3 mm in diameter.

2.4.5. Preparation of yeast cell free lysate

For each sample, a transformed yeast colony was used to inoculate 15 ml SD media lacking selective amino acids and incubated overnight at 30 °C. The overnight culture was centrifuged at 3000 g for 5 mins to pellet the cells, and washed with distilled water. The cells were transferred to a 1.5 ml tube, pelleted again and resuspended in 150 µl breaking buffer. Acid-washed glass-beads (425-600 µm) were added into the cell suspension until the beads reached a level just below the meniscus of the liquid. The sample was vortexed vigorously for 1 min and placed on ice to cool; this step was repeated four times for each sample. The sample was then centrifuged at 13,000 rpm, for 15 mins at 4 °C. The supernatant, which constitutes the yeast cell free lysate, was then collected into a fresh, pre-chilled 1.5 ml tube.

2.4.6. Quantitative β-galactosidase activity assay

β-galactosidase activity of the yeast cell-free extracts was measured using a colorimetric assay in 1 cm cuvettes. 5 to 20 µl of yeast cell free lysate was adjusted to 500 µl with Z buffer. Samples were vortexed and equilibrated at 30 °C for 5 mins. The reaction was initiated by addition of 100 µl of 4 mg/ml O-

nitrophenyl-β-D-galactopyranoside (ONPG) dissolved in Z buffer, and incubated at 30 °C. Reactions were allowed to proceed until a pale yellow colour had developed or after 30 mins, at which point the reactions were quenched by addition of 250 µL 1M Na_2CO_3. The reaction time was recorded and the OD_{420} was measured. Specific activity of the extracts was calculated using the formula [101]:

$$\text{Specificactivity (nmol/mg/min)} = \frac{\text{OD420} \times 0.85}{0.0045 \times [\text{Protein}] \times \text{Volume} \times \text{Time}}$$

Two-hybrid experiments were performed in triplicate, and reporter activities are expressed as nmol/min/mg protein. The protein concentration was determined by Bradford assay (**Section 2.6.3**).

2.4.7. Pellet X-gal assay

Pellet X-gal (PXG) assays were performed as described by Möckli and Auerbach [102]. Yeast transformants were inoculated into liquid YSM appropriately lacking amino acids (Leucine and Tryptophan) for maintenance of the Y2H plasmid vectors. Appropriate ligands (or vehicle), were added for assays investigating ligand dependent protein-protein interactions. After overnight incubation (at 30 °C with agitation in an orbital shaker at 200 rpm), the OD 600 of each culture was measured. Volumes of each culture corresponding to two absorbance units

were transferred to micro-centrifuge tubes, then pelleted by centrifugation (5 minutes at 5000 rpm in a benchtop micro-centrifuge) and the supernatant was removed from each pellet. Each pellet was lysed by freeze-thaw: 2 cycles of immersion in liquid Nitrogen for 3 minutes, incubation at 37 °C for 3 minutes. Lysed pellets were resuspended in 20 µL sterile diH_2O and transferred to wells of a flat-bottomed 96 well microplates (Techno Plastic Products), held on ice. 100 µL of X-gal solution (500 µg/mL X-gal, 0.5% (w/v) agarose, and 0.05% (v/v) β-mercaptoethanol, in sterile diH_2O) was added to each test well. The plate was wrapped in foil to prevent exposure to light, and incubated at 30 °C. After 30 minutes of incubation, reactions were stopped by addition of 100 µL 1M Na_2CO_3.

β-gal enzymatic activity was measured by recording the colour development using a LiDE400 flatbed scanner (Cannon). Colour formation was quantified with ImageJ (National Institutes of Health) software, using the 'Color Histogram' function to measure mean intensity of pixels in the blue channel for each well.

2.4.8. Modified X-gal filter lift assay

Pellet X-gal (PXG) assays were performed as described by Möckli and Auerbach [102], with modifications. Yeast transformants were inoculated into liquid YSM appropriately

lacking amino acids (Leucine and Tryptophan) for maintenance of the Y2H plasmid vectors. Appropriate ligands (or vehicle), were added for assays investigating ligand dependent protein-protein interactions. After overnight incubation (at 30 °C with agitation in an orbital shaker at 200 rpm), the OD 600 of each culture was measured. Volumes of each culture corresponding to one absorbance units were transferred to micro-centrifuge tubes, then pelleted by centrifugation (5 minutes at 5000 rpm in a benchtop micro-centrifuge) and the supernatant was removed from each pellet. Pellets were resuspended in 20 µL sterile diH_2O and transferred to separate squares of filter paper (approx. 3 cm x 3 cm). Lysis of the yeast was achieved by placing the filter paper squares into liquid nitrogen for three minutes, then removed and held at room temperature for 3 minutes. This was performed 2 times. Following lysis, 250 µL of X-gal solution (500 µg/mL X-gal, 0.5% (w/v) agarose, and 0.05% (v/v) β-mercaptoethanol, in sterile diH_2O) was applied to each filter paper square. The treated filter paper squares were then wrapped in foil to prevent exposure to light, and incubated at room temperature overnight.

Images of the treated filter paper squares were captured using a LiDE400 flatbed scanner (Cannon).

2.5. Cell culture

2.5.1. Maintenance of cell lines

The cell lines, LNCaP C4-2 and MCF7 were routinely maintained in Roswell Park memorial Institute (RPMI)-1640 containing phenol red and supplemented with 1% (v/v) 0.2 M L-glutamine, 1% Penicillin/Streptomycin antibiotics and 10% (v/v) heat inactivated fetal calf serum (FCS). The HEK 293 cell line was were routinely maintained in Dulbecco's modified Eagle's medium (DMEM), like wise supplemented.

Cultures were grown in cell cultured treaded sterile plastic ware at 37 °C in a humidified atmosphere containing 5% CO_2. The cells were passaged at 90% confluency for a maximum of 30 passages, at which point the cells were discarded. To passage adherent cells, media was aspirated and the cells washed twice with sterile 1× PBS. 1.5 ml trypsin/EDTA was added and the cell incubated for 5 mins at 37 °C, 5% CO_2. Trypsin was aspirated and the cells were resuspended in fresh medium. For HEK 293 cells, which are weakly adherent, cells were passaged without trypsinisation by taking up and expelling media from a pippette onto the cells to force them to detach from the plastic. Cells were subsequently diluted in fresh media in flasks (for maintenance), or appropriate vessels for intended experiments.

2.5.2. Cryopreservation of cell lines

Media was aspirated and the cells washed twice with sterile 1× PBS. 1.5 ml trypsin/EDTA was added and the cell incubated for 5 mins at 37 °C, 5% CO_2. Trypsin was aspirated and the cells were resuspended in 5 ml of media (for T-75 flask) and counted by diluting in a 1:1 ratio, loading onto a haemocytometer (Hawksley, BS.748) and counted under a light microscope (Nikon eclipse TS100). They were then diluted to the appropriate concentration. 400 µl (40%) of cell suspension in complete media was transferred to labelled cryovials containing 100 µl DMSO and a supplement of 500 µl FCS. The cell suspension was gently pipetted to mix the contents evenly. The vials were then stored at -20 °C until frozen solid. The cryovials were then placed at -80 °C overnight and transferred liquid nitrogen for long term storage.

2.5.3. Resuscitation of cell lines

To recover cells from liquid nitrogen storage, cells were thawed at 37 °C and transferred into 10 ml of pre-warmed media. The suspension was mixed thoroughly and the cells pelleted by centrifugation at 1100 rpm for 5 mins to remove residual DMSO which can impede growth. Cells were then resuspended in 10 ml of fresh media and incubated at 37 °C, 5% CO_2. A few hours after cell plating the media was changed to remove dead and unattached cells followed by another

change of media after 24 hours. Cells were passaged at least three times (approximately 12 days) before use in experiments.

2.5.4. Lipofectamine mediated transfection of adherent cells

Adherent cells were transfected via Lipofectamine 3000 (Invitrogen) transfection reagent, adjusting the number of cells, volumes of reagents and masses of plasmid DNA used according to the size of cell culture vessel as indicated by the manufacturer's protocol [103].

The day before transfection, adherent cells were seeded in media at ~50% confluence to reach 70-90% confluence on the following day. On the day of transfection the cell's medium was replaced with the appropriate volume of RPMI-1640 without antibiotics, and the cells were incubated for 30 mins at 37 °C, 5% CO_2. In parallel, Lipofectamine/DNA complexes were prepared, by first preparing dilutions of DNA in Optimem or antibiotics, and dilutions of Lipofectamine 2000 1:25 (v/v) in RPMI-1640 without FCS or antibiotics. Lipofectamine/DNA complexes were then prepared by mixing these two dilutions 1:1 (v/v), and incubating for 5 mins at room temperature before adding complexes dropwise to cells. Cells were incubated at 37 °C, 5% CO_2 overnight to permit transient transfection, after which time the media lacking antibiotics was replaced with the cell's usual media.

2.5.5. Confocal microscopy

2 x 10^5 cells were seeded into µ-Slide 8 well chambered coverslips and the following day, when fully adherent and at 70-90% confluence, were transfected with the appropriate plasmid DNA via Lipofectamine mediated transfection (**Section 2.5.4.Error! Reference source not found.**). 48 hours post-transfection cells were gently washed twice with PBS and fixed by incubation for 10 mins with 4% (w/v) paraformaldehyde. Cells were washed four times in PBS to remove excess fixative and incubated with membrane permeant DRAQ5 dye in PBS (1:5000) for 5 mins to stain nuclei. Cells were then washed four times in PBS and carefully drained of excess solution and 250 µl of PBS was placed into each well. Slides were stored at 4 °C, protected from light and to prevent slides from drying out, a large sheet of blue towel was soaked in water and placed in a sealed container creating a semi-humidified environment. Fluorescence images were captured using a Zeiss LSM510 Meta confocal microscope and images were processed in the LSM Zeiss Image Browser.

2.5.6. Cell viability/proliferation assays

Viable cells were quantified by a standard MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) reduction assay. Cell-mediated reduction of MTT was determined by reading absorbance at 550 nm. To measure the

effects of the modified lines on growth, cells were seeded into 96-well microtitre plates at a density of 5×10^3 cells/per well and allowed to adhere overnight. The initial density of viable cells prior to addition of inhibitors (denoted as time t=0) was determined in a control plate. Cells were cultured for a further 48-72 hrs before measurement of MTT activities. Data were presented as the average of 5 replicates per condition.

2.5.7. Florescence-activated cell sorting (FACS)

Transfected mammalian cells were isolated based on their GFP/Venus expression (therefore successful transfection), using the Beckman Coulter MoFlo XDP, 48 hours post-transfection. To prepare cells for FACS, cells were trypsinised to detach them from their vessel and from eachother, then pelleted, washed and suspended in their regular cell culture media. Cells in suspension were transferred to separate BD Falcon FACS tubes with cell-strainer caps to remove clumps of cells, and kept on ice until ready for sorting. GFP/Venus positive cells were first separated from their non-fluorescent counterparts (enrichment) into a fresh vessel, then sorted again (purification) and sorted into individual wells for clonal expansion in CRISPR-Cas9 genome editing experiments, or collected into a fresh vessel, and stored on ice until RNA extraction for qPCR experiments.

2.6. Biochemical techniques

2.6.1. Composition of solutions used in biochemical methods

Cell lysis buffer*:* 50 mM Tris-HCl pH 8, 150 mM NaCl, 1% (v/v) NP-40, 0.1% (w/v) SDS, 1 mM PMSF, 10 mM NaF, 0.1 mM Na_3VO_4, 1% (w/v) sodium deoxycholate (DOC), 1×complete protease inhibitor cocktail per 50 ml cell lysis buffer.

SDS-PAGE loading buffer (4×): 250 mM Tris-HCl pH 6.8, 40% (v/v) glycerol, 8% (w/v) SDS, 14.5 mM β-mercaptoethanol, 0.2% (w/v) bromophenol blue.

Tris-glycine-SDS-PAGE running buffer (10×): 250 mM Tris base, 2 M glycine, 0.037% (w/v) SDS pH 8.3.

Tris-glycine-SDS transfer buffer: 48 mM Tris base, 39 mM glycine, 0.037% (w/v) SDS, 20% (v/v) methanol.

Blocking solution (for western blotting): 5% (w/v) BSA in 1× PBS.

Ponceau S solution : 0.1% (w/v) Ponceau S in 5% (v/v) acetic acid. Made up to a total volume of 1 l with diH2O.

10% Ammonium persulphate (APS) solution: 10% (w/v) APS in diH2O. Stored at -20°C until use.

10% Sodium dodecyl sulphate (SDS) solution: 10% (w/v) SDS in diH2O.

2.6.2. Preparation of cellular protein extracts

Cells were harvested by transferring their culture vessels to ice and aspirating the media, followed by washing two times with ice cold PBS. PBS was aspirated and any residual liquid drained. A volume of pre-chilled lysis buffer appropriate to the size of the cell culture vessel was added to the cells and incubated on ice for 2 mins. The cells in lysis buffer were scraped and collected into a fresh 1.5 ml tube, vortexed for 10 seconds, and incubated on ice for 20 mins. Following this, samples were sonicated with a Bioruptor water bath sonicator (Diagenode) for 3 cycles, 30 seconds on / 30 seconds off, setting: High. The cellular debris was pelleted by centrifugation at 13,000 g, for 20 mins at 4 °C. The supernatant was transferred to a fresh tube and stored at -80 °C.

2.6.3. Protein concentration determination

Protein concentrations were determined using the Bradford assay method. Reactions were set up consisting of 2 µl of protein extract, 800 µl diH2O & 200 µl of Bio-Rad Bradford dye. The samples were mixed thoroughly by pipetting and left to stand at room temperature for 20 mins. The absorbance of each sample and of standards of known concentrations (0, 2, 4, 6, 8, 10, 20 µg/µl) of Bovine Serum Albumin (BSA) were read at 595 nm. The absorbance readings at 595 nm for these standards was plotted against their protein concentration in

Microsoft Excel using the linear trend line function. The equation of the linear trend line given by this function was used to determine the protein concentration of samples from their absorbance at 595 nm.

2.6.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Cell extracts were analysed by one-dimensional polyacrylamide gel electrophoresis. SDS-PAGE gels were prepared with the appropriate percentage resolving and stacking gel (**Table 3**), depending on the molecular weight of the protein being analysed. Protein extracts were denatured by boiling for 5 mins in 4× SDS-PAGE loading buffer before being resolved on SDS-PAGE gel of concentration 8-15% depending on the molecular weight of the protein. Gels were subjected to electrophoresis using a mini-gel system (Protean II, Bio-Rad) in 1× Tris-glycine-SDS-PAGE running buffer at fixed voltage of 180 V.

Table 3. Reagents and composition of SDS-PAGE gels.

Component	Resolving Gel				Stacking gel
Acrylamide %	8	10	12	15	5
ddH2O (ml)	2.3	1.9	1.6	1.1	1.4
30% acrylamide mix (ml)	1.3	1.7	2.0	2.5	0.33
1.5 M Tris pH 8.8 (ml)	1.3	1.3	1.3	1.3	-
1.0 M Tris pH 6.8 (ml)	-	-	-	-	0.25
10% (w/v) SDS (ml)	0.05	0.05	0.05	0.05	0.02
10% (w/v) APS solution (ml)	0.05	0.05	0.05	0.05	0.02
TEMED (ml)	0.003	0.002	0.002	0.002	0.002

2.7. RNA Sequencing analysis of differentially expressed genes

RNA sequencing (RNA-Seq) analysis was performed on triplicate cultures of wild type LNCaP C4-2, *PIP5K1A*-KO and *PIP5K1A*-ΔN cells grown to similar densities. Total RNA was prepared using a Qiagen RNeasy with on-column DNase digestion (**Section 2.3.16**). RNA quality was validated and quantified using nanodrop, qubit and Agilent Bioanalyzer (assisted by Dr. Cíntia Monteiro). All samples had RNA Integrity Number (RIN) above 9.7. Construction of mRNA-Seq libraries (TruSeq RNA Sample Preparation Kit v2, Illumina), library QC

validations and sequencing (PE150; >20M reads per sample) were performed by Novogene UK. Raw datasets were downloaded and adapter sequences and low quality reads (phred scores <30) were removed from the compressed fastq files using TrimGalore.

The resultant fastq files were aligned to the Ensembl annotated human genome (GRCh38.38) using STAR [104], gene expression was quantified using featureCounts [105]. These steps were initially performed by Professor Nigel Mongan, University of Nottingham),

Differentially expressed genes (DEGs) were identified by use of the IDEP web application [106] of EdgeR [107], to log transform RNA read counts, conduct clustering and run principle component analysis. The IDEP web application was used to run DESeq2 [108] using a Log_2 fold change of 1 (> or =1; <or = -1) and a false discovery rate of 0.01 as significance cut-offs. Identified DEGs were plotted in VolcaNoseR [109] (performed by Ryan Morris). Gene Set Enrichment analysis was conducted using The Molecular Signatures Database [110; 111] using HALLMARK and KEGG gene sets (performed by Ryan Morris and Prof David Heery). DEGs common to triplicates or between cell lines were identified using Venny 2.0. Selected DEGs (e.g.

KLK3, UGT2B15, UGT2B17) were validated by reverse transcription quantitative real-time PCR (RTqPCR).

Chapter 3

Generation and functional studies of PIP5KIα knock-out in a human metastatic prostate cancer cell line

3.1. Introduction

The main goal of this project was to perform investigations into the molecular functions of the human *PIP5K1A* gene, a key upstream enzyme in PIP2/PIP3/AKT signalling pathways. In particular, based on prior studies from our collaborators (JL Persson group, Umea Sweden), we were interested to explore the possible role of *PIP5K1A* in crosstalk with androgen/AR signalling in prostate cancer cells. The motivation for this is the potential finding that PIP5KIα might be an alternative actionable target in metastatic prostate cancer, given that it has functional crosstalk with androgen signalling, and that chemical or siRNA inhibition of PIP5KIα has been shown to block xenograft tumour growth in mouse models [94]. Our strategy was to apply CRISPR CAS9 genome editing to the molecular studies of PIP5KIα function in a cell line representative of metastatic prostate cancer. Another motivation is to better understand the function of this little studied enzyme in a key signalling pathway.

Prior to the beginning of this project, other researchers within the group, Miguel Hernandez-Quiles and Dr. Jonathan Whitchurch, had utilised the CRISPR-Cas9 nickase genome editing system to disrupt the *PIP5K1A* gene in the human prostate cancer cell line, LNCaP C4-2. The intention of this genome editing was to generate cell lines with disrupted

expression of PIP5KIα protein; such lines could then be used to explore functions of PIP5KIα in gene signalling networks, and any role it may play in development and progression of prostate cancer. The selection of edited clones performed by M. Hernandez-Quiles & J Whitchurch lead to the identification of one subline, named 'ΔN'. This line appeared to express an N-terminally truncated isoform of PIP5KIα, and also a reduced level of PIP5KIα expression compared to the unedited parental line. Another genome editing experiment was later performed, using a double targeting strategy, leading to identification of a second subline, 'KO', displaying greatly reduced to negligible levels of detectable PIP5KIα.

My initial goal at the outset was to complete a thorough characterisation of the edited cell lines to identify the precise allelic changes induced. This is important to understanding the downstream molecular consequences of the editing performed. While there are pros and cons to any gene editing strategy, the generation of clonal lines allows consistency between experiments, and the selection of clones based on GFP rather than antibiotic selection ensures transient expression of CAS9n in these clones. This, coupled with the greatly reduced probability of unwanted/undesired genome edits (off-targets) using the CAS9 nickase, gave us confidence in the quality of the edited clones generated. Having characterised these clones, the

next aim was to perform transcriptome profiling to assess whether loss of functional *PIP5K1A* (KO) or loss of the N-terminal domain (ΔN) resulted in any changes in gene expression or other functions of this cell line.

High quality RNA was prepared from parental the LNCaP C4-2 line (WT), as well as ΔN and KO sublines, and submitted for RNA-Seq. The RNA-Seq data enabled us to not only identify DEGs, but also to confirm the gene editing events at the *PIP5K1A* locus. It was also apparent that potentially novel transcripts of the *PIP5K1A* gene may exist that potentially produce novel protein isoforms, as explored in **Chapter 4**. Gene set enrichment analysis were performed within the group which enabled us to identify pathways that were potentially reliant on *PIP5K1A* function. Several of the pathways identified in these studies were consistent with previous reports e.g. the enrichment of K-Ras and p53 regulated genes. Importantly, analysis of the RNA-Seq data revealed significant up-regulation of androgen regulated genes, such as Kallikrein related 3 (*KLK3*) in the *PIP5K1A* edited lines. *KLK3* is an androgen-regulated gene, whose protein product, Prostate Specific Antigen (PSA), is a serum biomarker used in the diagnosis and monitoring of prostate cancer (reviewed in [112]). This analysis also found down-regulation of several UGT2B gene family members, a subset of the uridine 5'-diphospho-

glucuronosyltransferases (UGT) superfamily. UGTs catalyse addition of hydrophilic glucuronide moiety to various substrates, typically increasing their solubility for transport from tissues and ultimately their excretion by renal or hepatic mechanisms. The UGT2B genes are known to be repressed by androgen signalling, and significantly lower expression of *UGT2B15* and *UGT2B17* was observed in the *PIP5K1A* edited lines. UGT2B15 and UGT2B17 have substrate specificity for androgens, facilitating their inactivation in androgen target tissues including the prostate [113], and they also play roles in a shift towards androgen-independent signalling and development of CRPC (reviewed in [114]). Altered expression levels of these genes in the *PIP5K1A* edited lines were confirmed by qPCR.

Subsequently, preliminary rescue experiments were performed to determine the effect of exogenous expression of PIP5KIα protein in the KO line on expression levels of several androgen regulated genes. These experiments demonstrated that alterations in the expression of KLK3 (enhanced) and UGT2B17 (repressed) in KO cells could be at least partly reversed by expression of a PIP5KIα-Venus (YFP derivative) but not Venus alone, indicating a functional recue. Western blots performed to assess the phosphorylation status of AKT, a marker of AKT activity and thus AKT/mTOR pathway activation,

showed surprisingly, increased phosphorylation of AKT in the ΔN subline. These were some of the main points of investigation in the following chapter.

3.2. Generation of sublines with altered PIP5KIα expression

3.2.1. Choice of LNCaP C4-2 subline for genome editing of *PIP5K1A*

As mentioned, prior to my beginning on my PhD project, researchers within the group had conducted CRISPR-Cas9n genome editing with the aim of generating *PIP5K1A* knockout sublines (i.e. with mutation in the target gene causing loss-of-function of the gene product) to study PIP5KIα functions in gene signalling networks and in development and progression of prostate cancer. Intended applications of resulting cell lines were RNA sequencing to assess transcriptional effects of knockout, and for performance of mouse xenograft studies by our collaborators at the University of Umea. The cell line chosen for this project was the LNCaP C4-2 subline.

LNCaP C4-2 is a subline derived from LNCaP, an androgen-sensitive human prostate-adenocarcinoma line isolated from a supraclavicular lymph node metastasis from a 50-year-old Caucasian male [115]. The LNCaP C4-2 subline was generated by Wu et al., by xenograft of the LNCaP line with co-injection of a human bone fibroblast cell line into intact (and

then subsequently castrated) athymic adult male mice, followed by isolation of tumour forming cells, their re-inoculation into castrated male mice and finally isolation again of tumour forming cells: LNCaP C4-2 [116]. LNCaP C4-2 is androgen insensitive: it is tumorigenic when inoculated into castrated male mice, lacks a growth response when treated with androgens *in vitro*, and displays reduced levels of expression of androgen receptor mRNA and protein compared to its parental line. Wu et al., suggested that the development of the androgen insensitive state of the LNCaP C4-2 subline was dependent upon the cells interaction specifically with bone fibroblasts and their maintenance in castrated mice. Bone xenografts of LNCaP C4-2 subline recapitulate clinical features observed in human prostate cancer patients with bone metastases following androgen deprivation [117]. The LNCaP C4-2 subline was used for the genome editing of *PIP5K1A* because of the subline's androgen insensitivity, and its recapitulation of clinical features of human bone metastatic prostate cancer.

The attempts at genome editing of *PIP5K1A* in LNCaP C4-2 lead to the generation of the two sublines used in this study, which are named LNCaP C4-2 PIP5KIα ΔN (ΔN) and LNCaP C4-2 PIP5KIα KO (KO). ΔN was generated by Miguel Hernandez-Quiles, and its generation and preliminary characterisation is described in his Master's dissertation [118]. Because of its

importance in this PhD project, the generation of the ΔN subline is also briefly described here. CRISPR edited clones were selected by Jonathan Whitchurch; characterisation of clones was performed jointly by J Whitchurch, Hilary Collins and myself.

3.2.2. Generation of an LNCaP C4-2 subline expressing a truncated isoform of PIP5KIα - LNCaP C4-2 ΔN PIP5KIα (ΔN)

In our first attempt to edit the *PIP5K1A* gene and disrupt PIP5KIα expression in LNCaP C4-2, CRISPR guide RNAs were targeted immediately upstream of and overlapping the canonical ATG start codon (chromosome 1, position 151,118,997, GRCh38.p13 Primary Assembly) within the canonical Exon 1 of *PIP5K1A*. The intention was to delete the canonical ATG start codon of *PIP5K1A*, and thus disrupt initiation of translation of *PIP5K1A* transcripts. Another possible effect of indels in this region would be potential disruption of the 5'-UTR, which may hamper transcription or translation.

The Cas9 D10A nickase has an inactive RuvC nuclease domain, but an active HNH nuclease domain, and therefore cleaves the protospacer strand only (the strand complementary to the sgRNA's spacer). Use of the D10A nickase in the PAM-out orientation and a nick distance between 37 and 68 bp is recommended for maximum editing efficiency [119].

The strategy targeted Cas9 D10A nickase ribonucleoprotein complexes to the desired genomic location in a "PAM-out" orientation (**Figure 9**). For this, guide RNAs sgRNA 'A' and sgRNA 'B' were designed. The spacer sequences for these are listed in **Supplementary Table 1**. The distance between predicted nick sites ("nick distance") was 39 bp. Corresponding oligonucleotide pairs ('A Top' and 'A Bottom' and 'B Top' and 'B Bottom' were synthesised and annealed creating dsDNA oligonucleotides with 5'-CACC-3' (top strand) and 5'-CAAA-3' (bottom strand) overhangs on each pair to facilitate their ligation into the destination plasmid; BbsI digested pX461. pX461 is a mammalian expression vector for Cas9n(D10A)-2A-eGFP, and sgRNAs. Ligation of the annealed oligo nucleotides between BbsI sites facilitates the expression of the intended sgRNAs upon transfection into mammalian cell lines.

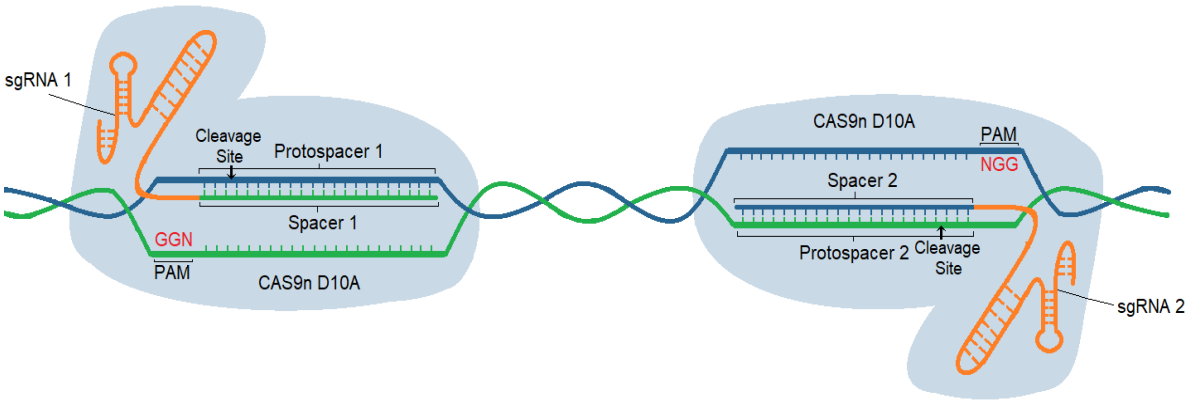


Figure 9. The "PAM out" orientation of CAS9 nickases RNPs targeted to a genomic location to facilitate editing.

LNCaP C4-2 cells were co-transfected with pX461 derivatives encoding sgRNA 'A' and sgRNA 'B' by Neon® electroporation transfection (Invitrogen). Transfected cells were sorted for eGFP expression to achieve a single cell per well in a 96 well plate. Cells were expanded to produce clonal populations, which were then screened for genomic alterations around the targeted region of *PIP5K1A* Exon 1 by PCR amplification using genomic DNA extracts and the following primer pairs as listed in **Supplementary Table 2**. Both of these pairs amplify across the targeted region. Following PCR and agarose gel electrophoresis, agarose gel slices of bands were taken, from which DNA was extracted subsequently submitted for Sanger sequencing. Lines were also screened for PIP5KIα protein expression by Western blotting of protein lysates. Preliminary PCR analysis and western blot screening were performed by M. Hernandez-Quiles; subsequent repeats of PCR and sequencing to confirm findings were carried out during the course of this PhD.

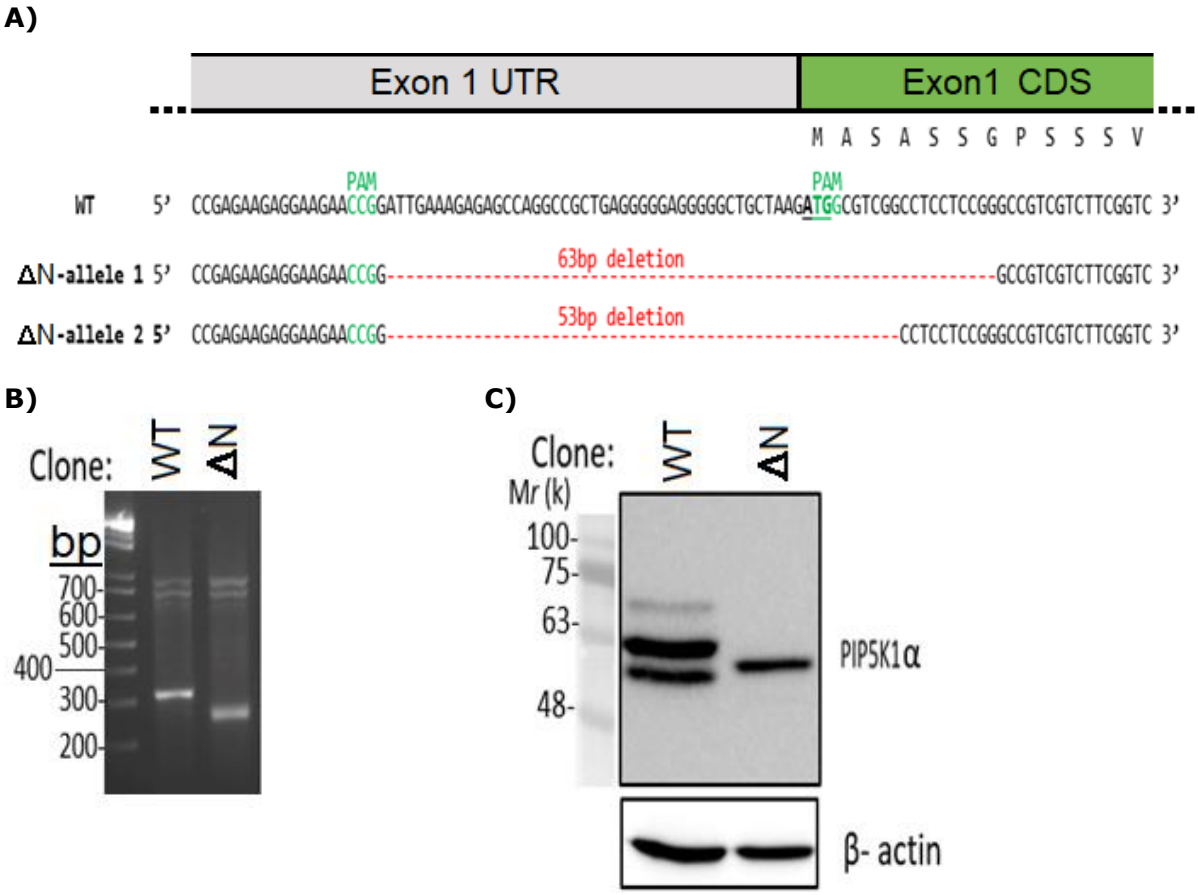


Figure 10. CAS9 nickase editing of *PIP5KIA* Exon 1
A) Sequences of alleles in the ΔN subline. The first allele contained a 63 bp deletion, and the second a 53 bp deletion with respect to the WT sequence. Both deletions spanned the canonical ATG start codon used to initiate translation of PIP5KIα.
B) Agarose gel electrophoresis of PCR amplification of gDNA extracted from the ΔN clone yields a band approximately 50-bp shorter than WT (expected amplicon for WT = 317 bp). This band was subsequently extracted, TOPO cloned and inserts sequenced, finding the sequences shown in **A**. 1% agarose; 1 kb plus ladder (Invitrogen).
C) Western blot detection of PIP5KIα expressed in the WT LNCaP C4-2 and ΔN sublines. Note that multiple bands are detected for the WT line, but a single band intermediate between the two major bands is detected for the ΔN clone.

This approach led to the identification of a subline, designated ΔN, which bears two alleles with mutations spanning

the ATG start codon of Exon 1 of *PIP5K1A* (**Figure 10. A and B**). Western blotting with an antibody against PIP5KIα in extracts from the ΔN subline detected a single band with molecular weight different to bands observed for wild type LNCaP C4-2 (**Figure 10. C**). Relative mobility calculations were performed and the molecular weight of this band was estimated to be 58.5 KDa. The possible mechanisms by which this apparently truncated isoform of PIP5KIα is expressed in the ΔN subline are discussed in **Section 3.5**.

3.2.3. Generation and screening of an LNCaP C4-2 subline expressing greatly reduced amounts of PIP5KIα - LNCaP C4-2 KO PIP5KIα (KO)

As CRISPR-nCas9 targeting of *PIP5K1A* Exon 1 alone did not result in identification of any lines without expression of PIP5KIα, a second targeting strategy was devised.

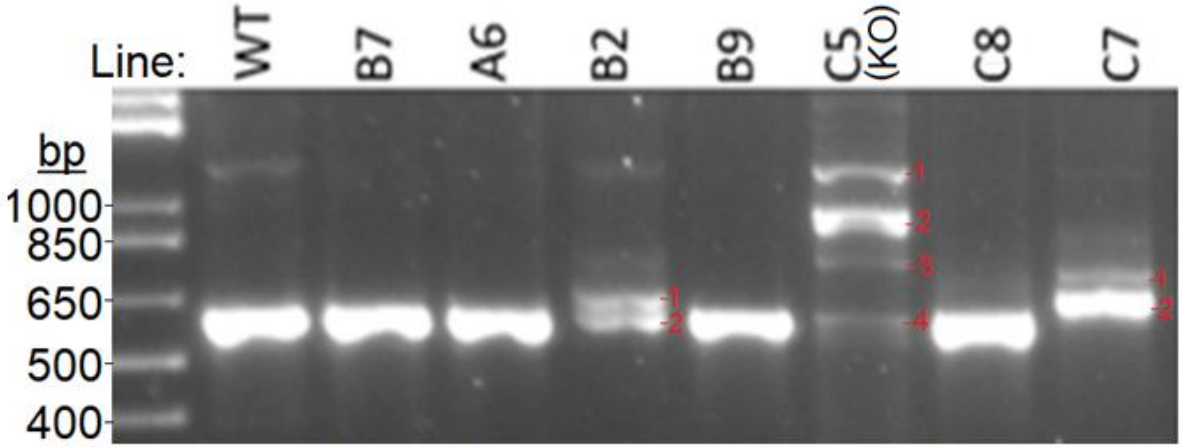
LNCaP C4-2 cells were edited at two genomic locations, with CRISPR CAS9 sgRNA pairs targeting both the ATG start codon of Exon 1 (sgRNAs 'A' and 'B'; **Supplementary Table 1**), and at the border of *PIP5K1A* Exon 6 and the subsequent intron (Chromosome 1, position 151,152,366, GRCh38.p13 Primary Assembly; sgRNAs 'C' and 'D'; **Supplementary Table 3**) .

Exon 6 encodes a portion of the dimerisation domain of PIP5KIα. sgRNAs C and D were designed to target Cas9 D10A nickase ribonucleoprotein complexes to the desired genomic

location in a "PAM-out" orientation, with a nick distance of 46 bp. Importantly, effects on translation (such as premature termination) caused by indels in Exon 6 are unlikely to be circumvented by alternative splicing because Exon 6 is included in all annotated and predicted *PIP5K1A* transcripts in the NCBI RefSeq database, and there are no reported intron spanning reads that skip Exon 6, i.e. a constitutive exon. This genomic region was targeted with the aim to introduce indels in Exon 6 leading to frameshift and thus premature termination of translation, or else to disrupt splicing events including this exon. Either of these, in combination with alterations at Exon 1 affecting transcription or translation initiation were predicted to achieve complete loss of expression of PIP5KIα protein.

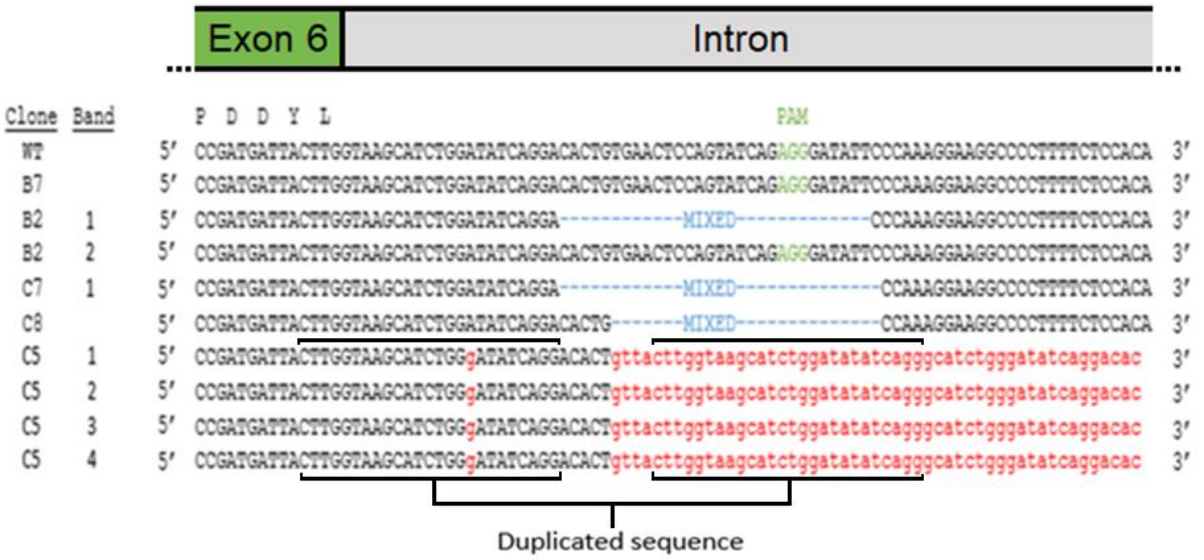
Oligonucleotide pairs ('C Top' and 'C bottom' and 'D Top' and D 'bottom' were synthesised and annealed creating dsDNA oligonucleotides with 5'-CACC-3' (top strand) and 5'-CAAA-3' (bottom strand) overhangs on each pair to facilitate their ligation into the destination plasmid; BbsI digested pX461. LNCaP C4-2 cells were co-transfected with pX461 plasmids expressing sgRNAs 'A' and 'B' and 'C' and 'D', then sorted for eGFP expression to obtain a single cell per each well of two 96 well plates.

A)

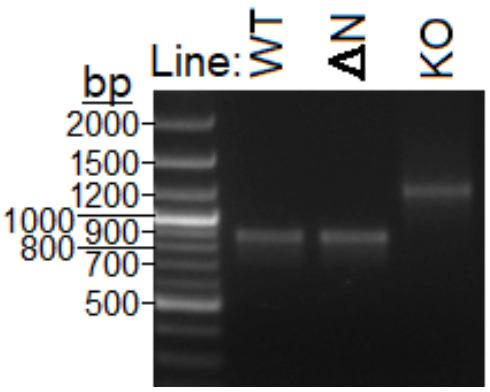


Expected amplicon for WT = 583 bp

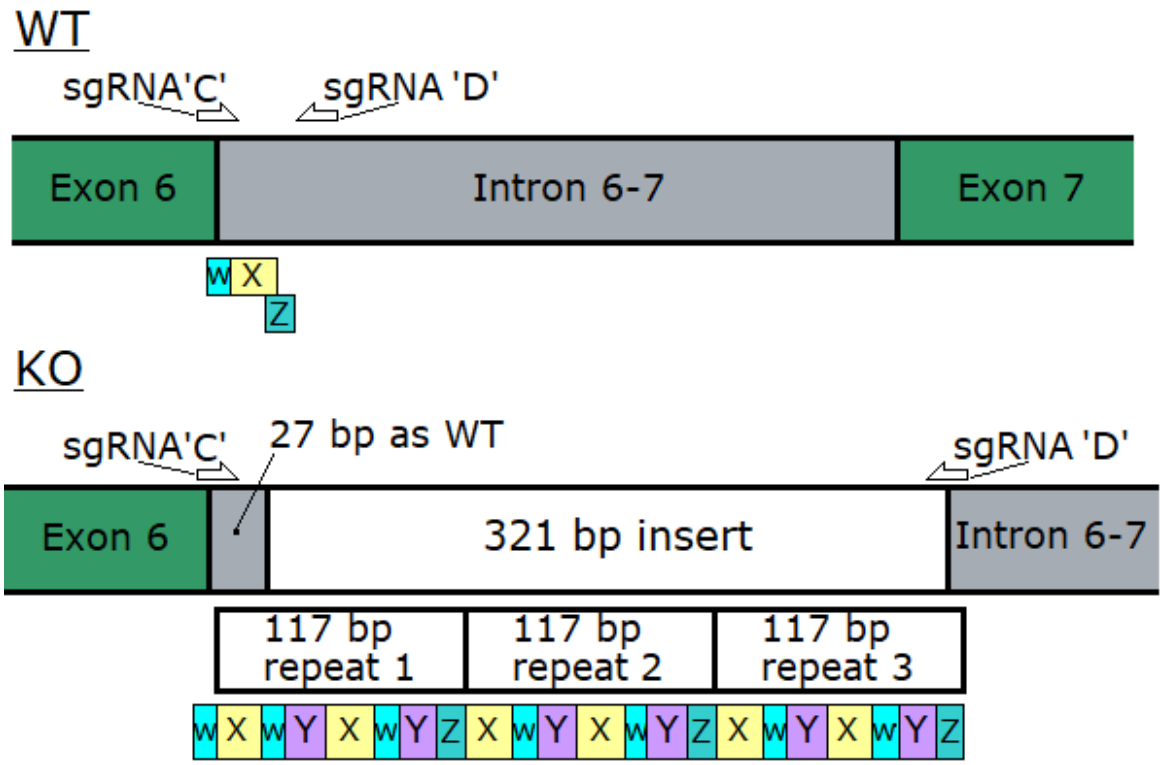
B)



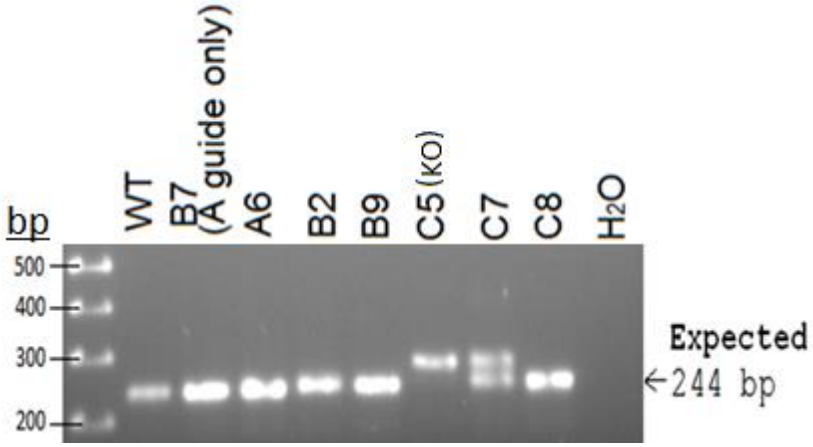
C)



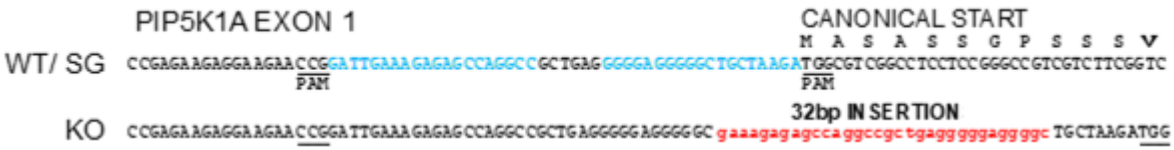
D)


WT
sgRNA'C'
sgRNA 'D'
Exon 6
Intron 6-7
Exon 7
w X
Z
KO
sgRNA'C'
27 bp as WT
sgRNA 'D'
Exon 6
321 bp insert
Intron 6-7
117 bp repeat 1
117 bp repeat 2
117 bp repeat 3
w X w Y X w Y Z X w Y X w Y Z X w Y X w Y Z

E)


bp
WT
B7 (A guide only)
A6
B2
B9
C5 (KO)
C7
C8
H2O
500
400
300
200
Expected
←244 bp

F)


PIP5K1A EXON 1
CANONICAL START
M A S A S S G P S S S V
WT/ SG
CCGAGAAGAGGAAGAACCGGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGCTGCTAAGATGGCGTCGGCCTCCTCCGGGCCGTCGTCTTCGGTC
PAM
PAM
KO
CCGAGAAGAGGAAGAACCGGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGC gaaagagagccaggccgctgaggggaggggc TGCTAAGATGG
32bp INSERTION

G)

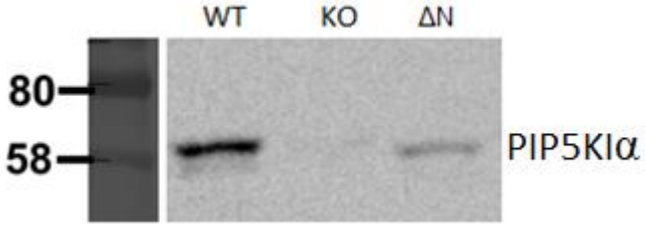

WT
KO
ΔN
80
58
PIP5KIα

Figure 11. Screening of clones dual targeted at Exon 1 and at the Exon 6-intron boundary.
A) Agarose gel electrophoresis of PCR screening for genomic alterations around *PIP5K1A* exon 6-intron border for several clones resulting after transfection and sorting. gDNA was extracted from each clonal population (subline), and PCR performed with primers targeting proximal to the PIP5KIA Exon 6- intron boundary. PCR products were resolved by agarose gel electrophoresis, extracted and subjected to sequencing. 0.9 % agarose; 1 kb plus ladder (Invitrogen).
B) Sequence alignments of bands extracted from the gel in **A**. Note, the bands for the KO (C5) subline display identical sequence, despite resolving at different sizes in the agarose gel.
C) Agarose gel electrophoresis of PCR performed for *PIP5K1A* Exon 6-intron boundary with a second primer pair amplifying a larger amplicon than for in **A**. Note that use of this primer pair did not result in multiple bands for the KO subline as in **A**. 0.8 % agarose; 1 kb plus ladder (Invitrogen).
D) Schematic of the insert at the *PIP5K1A* Exon 6- intron boundary in the KO line, compared to the wild type sequence. The KO line contains either a single allele or multiple identical alleles with a 321 bp insertion just 3' to the Exon 6 3' end. This 321 bp insertion is composed of three four sequence elements 'W', 'X', 'Y' and 'Z', derived from the wild type sequence, repeated numerous times.
E) Agarose gel electrophoresis of PCR screening for genomic alterations around *PIP5K1A* Exon 1 for several clones resulting after transfection and sorting. gDNA was extracted from each clonal population (subline), and PCR performed with primers targeting around *PIP5K1A* Exon 1 ATG translation initiation codon. PCR products were resolved by agarose gel electrophoresis, extracted and subjected to sequencing. 1 % agarose; 1 kb plus ladder (Invitrogen).
F) Sequence alignment of the KO band extracted from gel in **E** to the WT sequence flanking the ATG start codon in *PIP5K1A* Exon 1.
G) Western blot detection of PIP5KIα in the WT, KO and ΔN sublines of LNCaP C4-2.

Seven surviving lines ('A6', 'B2', 'B7', 'B9', 'C5', 'C7' and 'C8'; so named for their well position in the 96 well plate), were screened for indels around the 3' end of Exon 6 by PCR using their genomic DNA as input (primer pair in **Supplementary Table 4**) which amplify across the targeted site. Again, amplicons were resolved by agarose gel electrophoresis, gel-extracted and submitted for Sanger sequencing with the primer 5'-CCACAGTGAAGGGAGCAACC-3' (**Figure 11**).

These gel electrophoresis and sequencing results showed that several of the surviving lines possessed alleles bearing indels within the intron immediately following the 3' end of Exon 6, as follows:

- Line B2 is heterozygous, possessing one allele with wild type sequence, and one allele with a small insertion within the intron following Exon 6. The sequencing result for the band (lane B2, band 1) amplified from this edited allele yielded a chromatogram with partial mixed sequence (likely due to contamination with DNA from the lower band), so the exact length and sequence of the insert could not be determined; from the gel image, the insert appears to be 40-60 bp in length.
- Line C7 is heterozygous, possessing one allele with an approximately 90 bp insertion, and a second allele with

an approximately 60 bp insertion (based upon band heights relative to ladder). The sequencing result for the top band (lane C7, band 1), yielded a chromatogram with partial mixed sequence (again likely due to contamination with DNA from the lower band), so the exact length and sequence of insert could not be determined. The sequencing result for the lower band (lane C7, band 2), yielded a chromatogram with entirely mixed sequence, so that the sequence of the band could not be determined.

- Line C5 (subsequently named 'KO', for Knock-Out) showed amplification of four bands, which, upon extraction and sequencing, each had identical sequence. This sequence was determined to include a 25 bp repeat. The obtained sequencing reads did not extend to a length corresponding to the size of the gel bands, indicating that the full length and sequence of any inserts were as yet undetermined. The presence of four bands, and sequencing results indicating the inclusion in each of identical, repetitive sequence suggested two possibilities:
    - 1) This line possesses greater than two edited alleles (due to either chromosomal instability that is a hallmark of many cancers, or complex genome rearrangement resulting from the CRISPR editing) each including a repetitive segment with a different

number of repeats. The inclusion of the same repetitive element in each allele could result from homology directed repair of a CRIPSR induced strand break, using an already edited allele as the repair template.

- 2) This line possesses either one allele (due to chromosomal instability), or two edited alleles (with identical sequence due to homology directed repair of one allele using an already edited allele as template). If so, the observation of greater that two bands results from formation of unexpected PCR products, which can form due to the presence of repetitive elements [120].

To obtain the full sequence of the alleles present in the KO line, a new primer pair was designed (**Supplementary Table. 4**), centred around the 3' end of Exon 6. PCR with this primer pair resulted in amplification of a single band of approximately 1200 bp for the KO line. The predicted amplicon with this primer pair for the WT sequence is 863bp. This DNA was cloned into the pCR4blunt-TOPO plasmid and transformed into chemically competent DH5α *E.coli*, which were plated upon an ampicillin LB agar. A transformed colony was picked, inoculated into 10 ml ampicillin LB media, cultured overnight and its plasmid DNA extracted. This was plasmid DNA was then submitted for Sanger

sequencing with the primers 5’-TAATACGACTCACTATAGGG-3’ (“T7 promoter Fwd”), and 5’-ATTAACCCTCACTAAAG-3’ (“T3”), which bind to the plasmid backbone on opposite sides of the insert, allowing complete sequence of the insert to be obtained. The sequence around the 3’ end of Exon 6 in the allele or alleles present in the KO clone was thus determined to contain a 321 bp insert (Figure 3.2.3. D). This has a complex repeating pattern containing three 117 bp repeats (115 identical bp in each repeat, the final 2 bp of each vary between repeats), each of which is composed of four smaller elements, with the following sequences:

- Repeat W – 5’-TTACTTGGTAA-3’
- Repeat X – 5’-GCATCTGGGATATCAGGACACTG-3’
- Repeat Y – 5’-GCATCTGGATATATCAGG-3’
- Repeat Z – 5’-ACACTGTGACTAA-3’ or 5’-ACACTGTGACTCA-3’ or 5’-ACACTGTGACTCC-3’

Sequences of repeats W, X and Z (ACACTGTGACTCC) are present in the WT allele within the sequence between the protospacers targeted by sgRNA C and sgRNA D. The sequence of repeat Y seems to be derived from that of repeat X. The complex repeating pattern of the insert, with inclusion at the end of each of the 117 bp repeats of a partial protospacer sequence for sgRNA D, suggests this final sequence resulted

from multiple rounds of CRISPR Cas9n induced strand breaks and their attempted repair.

The seven selected clones were next screened for indels in *PIP5K1A* Exon 1 by PCR using their genomic DNA as input (primer pair 5'- AACGCGGGTTCTGTAAAGAG-3' and 5'-TCTTCCTCCCATCGCATATT-3'; expected amplicon = 244 bp for WT alleles). Again, amplicons were resolved by agarose gel electrophoresis, gel-extracted and submitted for Sanger sequencing (**Figure 11 E**).

For the KO subline, only one PCR product was detected, suggesting that either both alleles are identically edited, or that one allele failed to amplify, e.g. due to disruption of sequence homologous to the PCR primer pairs used. Sequencing showed this allele to contain a 32 bp insertion, upstream of the Exon 1 AUG start codon. Although this AUG start codon is intact, the 32 bp insertion is a duplication of sequence comprising a highly conserved motif 5'-GGGGAGGGGGC-3' immediately upstream of the ATG start codon (**Figure 11 F**). This motif resembles a binding site for the SP/KLF family of transcriptions factors and may be important for regulation of expression of the *PIP5K1A* gene.

The element 'W', which is present seven times in the *PIP5K1A* KO subline contains the Exon 6 splice donor site. The

repetitive inclusion of this sequence in the KO subline could feasibly disrupt splicing, leading to the lower level of *PIP5K1A* transcripts (**Section 3.6.1**) and PIP5KIα protein (**Figure 11 G**) in this subline.

The KO lines was then screened for PIP5KIα protein expression by Western Blotting of protein lysates, probing with an antibody against PIP5KIα. This indicated greatly reduced expression of PIP5KIα in the KO subline (**Figure 11 G**).

Design of sgRNAs, cell sorting and clonal expansion were performed by J. Whitchurch; western blot and genotyping were performed during this PhD.

3.3. Identity of truncated PIP5KIα expressed in the ΔN subline

We wished to determine the identity of the band detectable in the ΔN clone upon western blotting with an anti-PIP5KIα antibody. The molecular weight slightly below that of the major PIP5KIα band in the WT LNCaP C4-2 subline, suggested that the anti-PIP5KIα reactive band in the ΔN clone represents a truncated form of PIP5KIα. Estimation of the molecular weight of this band was performed by comparison to the WT bands and the molecular weight markers. Three bands are visible in the WT lane. Analysis of the RNA-Seq data for the LNCaP C4-2 line (**Chapter 4**) indicates that the major splice variants of *PIP5KIA*

expressed in this line are *PIP5K1A* v2 and *PIP5K1A* v5 (combined 80-85% of *PIP5K1A* transcripts), which encode PIP5KIα i2 (NP_003548.1) and PIP5KIα i5 (NP_001317618.1). These have a predicted molecular weight of 61.19 and 61.26 kDA respectively (computed with ExPasy [121]). Assuming protein levels of the various isoforms are proportional to the relative abundance of the respective transcripts, PIP5KIα i2 and PIP5KIα i5 would be the most abundant protein isoforms, and would explain the strongest band in the WT lane running at 61.75 kDa. PIP5KIα running at this molecular weight in SDS-PAGE is in contradiction to previous reports that PIP5KIα runs at 68 kDa [122; 123]. A weak band is detectable above the major band, at approximately 68 kDa, which may therefore correspond to the previously reported 68 kDa form. The largest reported isoform of PIP5KIα, PIP5KIα i1 (NP_001129110.1), as this has a deduced molecular weight of 62.63 kDa, and thus the small difference between it and the deduced molecular weight of i2 and i5 would not explain the ~68 kDa band. Post-translational modifications such as phosphorylation and sumoylation can effect position of a protein in a SDS page gel. Sumoylation reportedly causes an upward shift of 15-17 kDa to proteins in SDS-PAGE [124] and sumoylated PIP5KIα reportedly runs at ~90 kDa, possibly because of double sumoylation [122], thus sumoylation of PIP5KIα i2, i5 or i1 does not explain the ~68 kDa

band. Mobility shift of phosphorylated vs non-phosphorylated protein in SDS-PAGE is observed for many proteins [125]. Phosphorylation of PIP5KIα has been detected at multiple positions [86], and thus the ~68 kDa band may represent a multiply phosphorylated form.

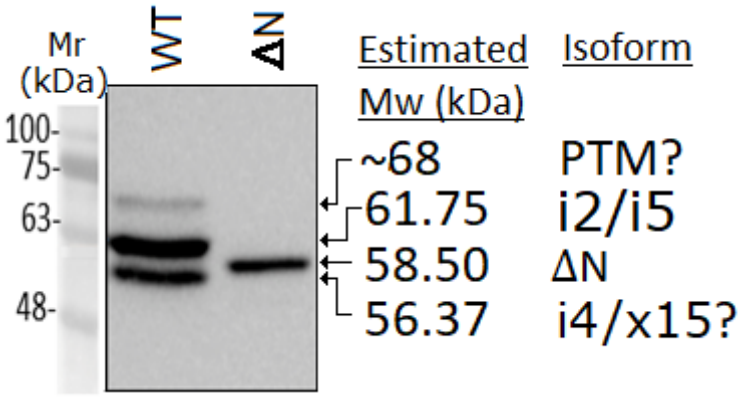
A lower band was detected in the WT lane, calculated to run at 56.37 kDa. This band may correspond to the shortest reported isoform, PIP5KIα i4 (NP_001129109.1) which has a calculated molecular weight of 56.05 kDa. Analysis of the splice variants of PIP5KIA expressed in our LNCaP C4-2 indicates the corresponding transcript (*PIP5K1A* v4) comprises only 2-5% of *PIP5K1A* transcripts (**Chapter 4**). Alternatively, the lower band could correspond to a predicted isoform, PIP5KIα x15 (XP_016857928.1) or PIP5KIα x16 (XP_016857929.1) which have a predicted molecular weight of 54.58 kDa and 54.51 kDa respectively. Like PIP5KIα i2 and i5, PIP5KIα x15 and x16 are identical to one another except for one additional Alanine residue in PIP5KIα x15. The corresponding splice variants, *PIP5KIA* x15 and *PIP5K1A* x16 combined comprise 8-10% of transcripts expressed in LNCaP C4-2 (**Chapter 4**).

The anti-PIP5KIα reactive band observed in the ΔN line was calculated to be 58.50 kDa. The deletion of the canonical AUG *PIP5K1A* start codon in PIP5KIA Exon 1 in both alleles in the ΔN

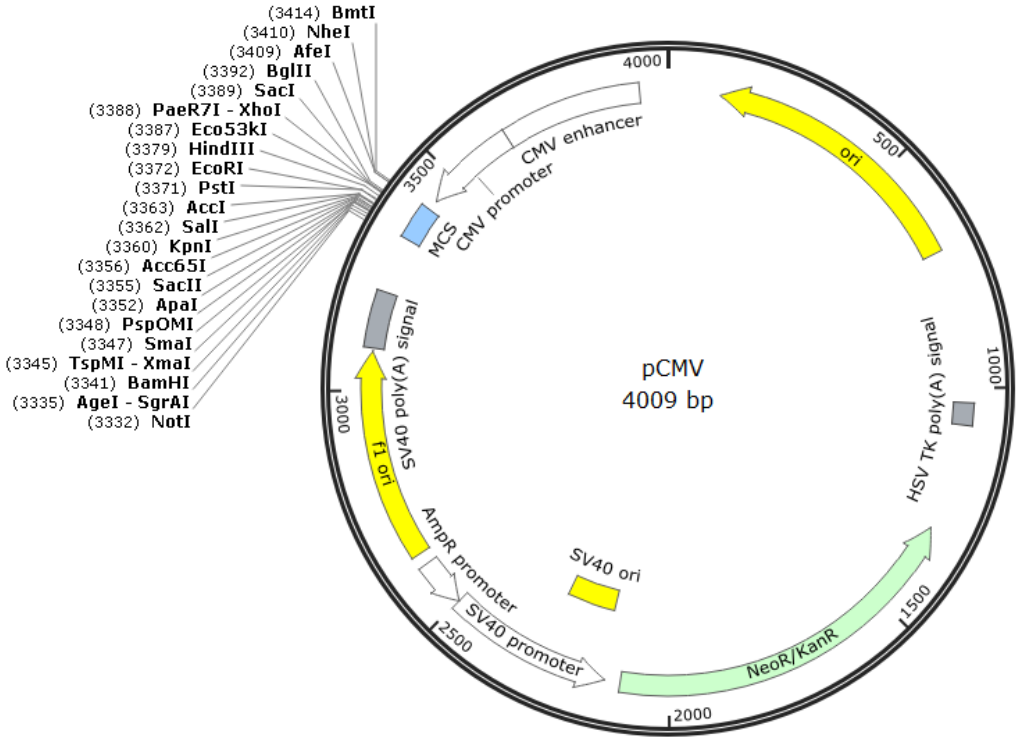
clone suggested that this band may represent a protein whose translation was initiated from alternative AUG codon, downstream of the canonical AUG of Exon 1. Such a protein would likely be translated from variants of *PIP5KIA* transcripts like *PIP5K1A* v2 or *PIP5K1A* v5, (the most abundant transcripts in the WT line) but with the deletions spanning the AUG start codon in Exon 1.

There are two candidate codons, AUG 36, and AUG 46. Proteins expressed from such transcripts, Δ36 PIP5KIα i2, and Δ46 PIP5KIα i2, have a predicted molecular weight of 57.84 and 56.90 kDa, respectively. Translation initiation from any of the subsequent AUG codons (the next being codons 97, and 200, for isoform 2), would cause too large a truncation to explain the observed molecular weight upon western blotting, and would exclude the epitope for the antibody used (Glycine 65 of PIP5KIα i1 [126]), which corresponds to Glycine 52 of PIP5KIα i2.

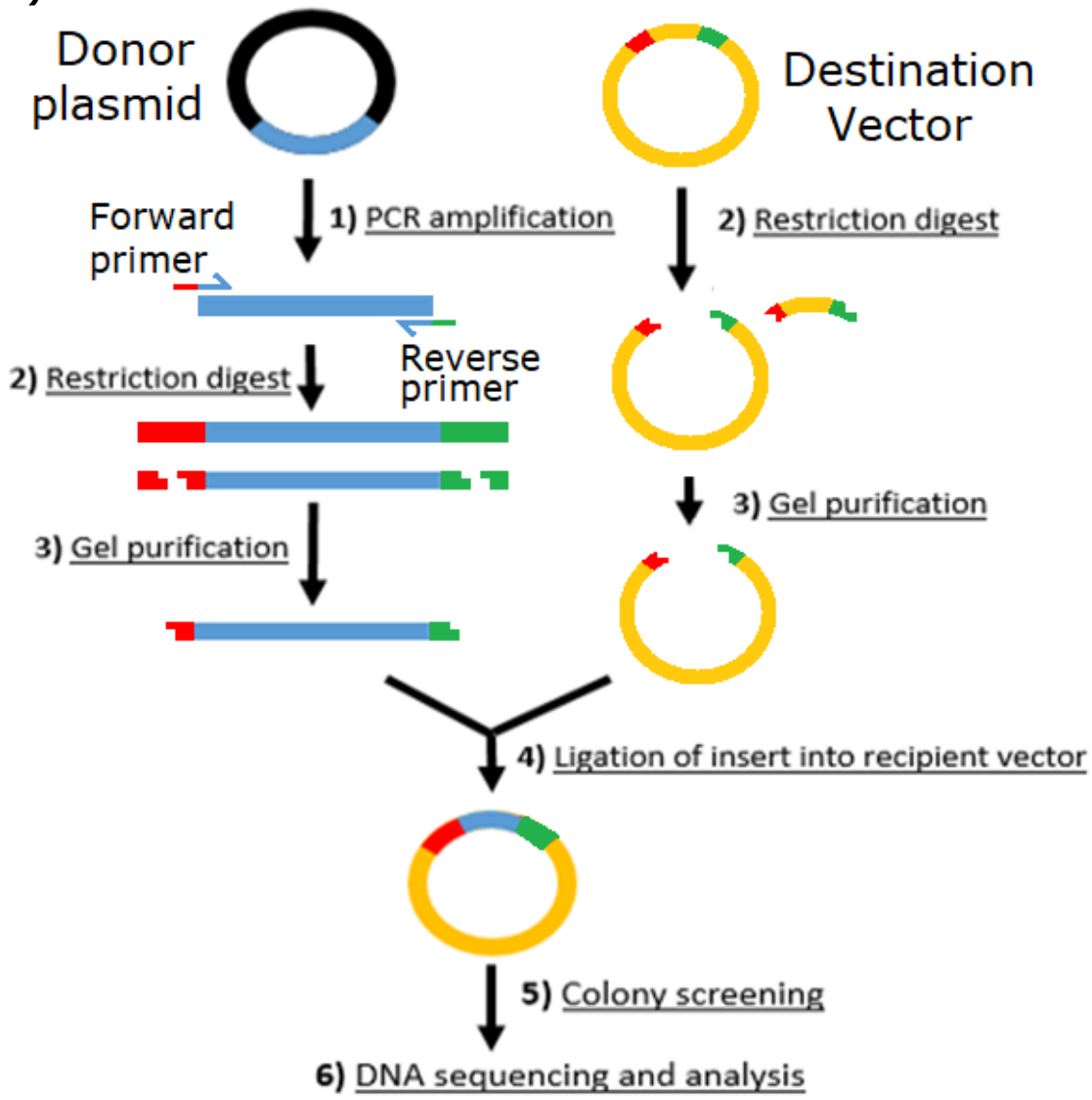
A)


Mr
(kDa)
WT
ΔN
100-
75-
63-
48-
Estimated
Mw (kDa)
Isoform
~68
PTM?
61.75
i2/i5
58.50
ΔN
56.37
i4/x15?

B)


(3414) BmtI
(3410) NheI
(3409) AfeI
(3392) BglII
(3389) SacI
(3388) PaeR7I - XhoI
(3387) Eco53kI
(3379) HindIII
(3372) EcoRI
(3371) PstI
(3363) AccI
(3362) SalI
(3360) KpnI
(3356) Acc65I
(3355) SacII
(3352) ApaI
(3348) PspOMI
(3347) SmaI
(3345) TspMI - XmaI
(3341) BamHI
(3335) AgeI - SgrAI
(3332) NotI
4000
500
1000
1500
2000
2500
3000
3500
CMV enhancer
CMV promoter
MCS
ori
SV40 poly(A) signal
f1 ori
AmpR promoter
SV40 promoter
SV40 ori
NeoR/KanR
HSV TK poly(A) signal
pCMV
4009 bp

C)


Donor plasmid
Destination Vector
Forward primer
1) PCR amplification
2) Restriction digest
Reverse primer
2) Restriction digest
3) Gel purification
3) Gel purification
4) Ligation of insert into recipient vector
5) Colony screening
6) DNA sequencing and analysis

D)

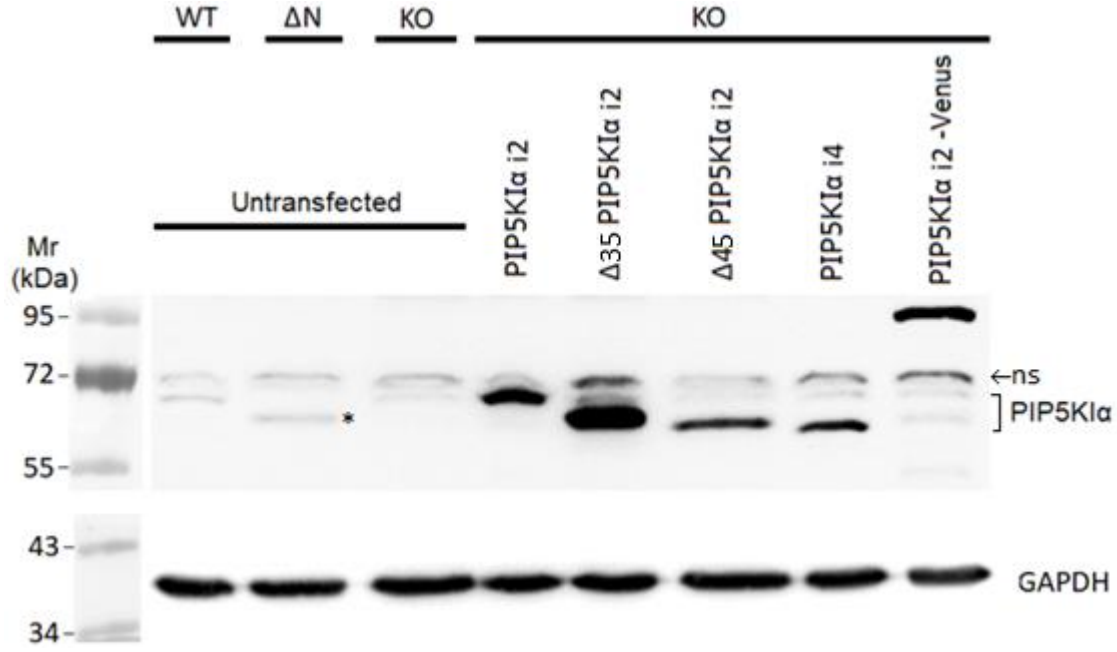

WT
ΔN
KO
KO
Untransfected
PIP5KIα i2
Δ35 PIP5KIα i2
Δ45 PIP5KIα i2
PIP5KIα i4
PIP5KIα i2 -Venus
Mr (kDa)
95-
72-
55-
43-
34-
←ns
PIP5KIα
GAPDH

Figure 12. Identification of truncated PIP5KIα expressed in the ΔN subline.

A) Western blot detection of PIP5KIα in the WT and ΔN sublines of LNCaP C4-2. Molecular weight estimations are indicated for each band. Isoforms predicted to be responsible for each band are indicated. These predictions were made by reference to computed molecular weights of each isoform, and assessment of major *PIP5K1A* transcripts expressed in the LNCaP C4-2 cells (**Chapter 4**).

B) Plasmid map of pCMV plasmid generated from mVenus for experiments expressing various forms of PIP5KIα.

C) Subcloning strategy. **Step 1**; PCR amplification of PIP5KIα CDS inserts, Restriction sites are included in the primer, in this case XhoI (red) and BamHI (green). **Step 2**; Restriction digestion of inserts and destination plasmid. **Step 3;** Restriction digested vector and inserts purified by gel electrophoresis and band extraction. **Step 4**; Ligation of purified and digested inserts into restriction digested vector. **Step 5;** Colony screening by colony PCR and restriction digest analysis. **Step 6;** DNA Sequencing and analysis to confirm correct insertion, without frame shift or other mutations.

D) Western blot detection of PIP5KIα in the WT, ΔN and KO sublines of LNCaP C4-2, and the KO subline transfected with expression constructs for variants of PIP5KIα. The putatively truncated variant of PIP5KIα expressed in the ΔN subline is indicated with an *. A band running at ~ 72 kDa is apparent in all lanes (indicated by ns), including the transfected KO lane. This band was not apparent in previous western blots with a different lot (Lot 1) this antibody (e.g. **Figure 3.4 A**, and **Figure 3.3.2 G**). This band was therefore determined to arise from non-specific binding with this lot (Lot 2) of the antibody.

To further investigate the identity of the band present in the ΔN clone, the PIP5KIα i4 CDS, and PIP5KIα i2 CDS, with and without the first 36 or 46 codons, were subcloned into a mammalian expression vector, pCMV (**Figure 12 B**). This vector was generated in this study, derived from mVenusN1 (a gift from Steven Vogel; Addgene plasmid #27793 [127]). To

generate pCMV, the Venus coding sequence was removed from mVenusN1 by site directed mutagenesis with the primers as listed in **Supplementary Table 5**. This left the original multiple cloning site intact, with a TAA stop codon immediately following the AgeI site. Like the parent plasmid, expression of the desired protein is achieved by cloning a CDS into the MCS, with expression in eukaryotic cells driven from a CMV immediate early promoter and enhancer, with a SV40 poly(A) signal. A Kozak motif must be included in the forward primer used for amplification of the CDS to be cloned. As per the parental plasmid, bacterial plasmid replication is driven by a ColE1 origin, and antibiotic resistance for selection of bacterial selection is performed with Kanamycin.

The PIP5KIα i4 CDS, and PIP5KIα i2 CDS, with and without the first 36 or 46 codons were subcloned into pCMV, generating mammalian expression vectors for PIP5KIα i2 (1-550); Δ36 PIP5KIα i2 (36-550); Δ46 PIP5KIα i2 (46-550) and PIP5KIα i4 (1-500). PIP5KIα i2 and PIP5KIα i4 CDS were respectively subcloned from pLPS PIP5KIA, a gift from Jenny Persson, and p15aC-3D1D-5, a gift from Sanford Simon (Addgene plasmid #121050; [128]). Additionally, a PIP5KIα i2 (1-550)-Venus expression plasmid was generated by cloning into the mVenus N1 vector, cloning with the same PCR product as used for cloning of PIP5KIα i2 CDS into pCMV. Sequencing of the

expression vectors across the PIP5KIα CDS inserts was performed, confirming the sequences contained no mutations, and in the case of the PIP5KIα i2-Venus plasmid, confirming validity of the fusion protein.

These various PIP5KIα expression plasmids were transfected into the LNCaP C4-2 PIP5KIα KO line and the lysates subjected to SDS-PAGE and western blot with an anti-PIP5KIα antibody. This experiment demonstrated that the anti-PIP5KIα reactive band observed in the ΔN clone runs at approximately the same size as Δ36 and Δ46 PIP5KIα i2 proteins, demonstrating that either of these could be the PIP5KIα isoform expressed in the ΔN clone. The resolution is not sufficient to distinguish whether the truncated protein expressed in the ΔN clone runs more closely to the Δ36 PIP5KIα of Δ46 PIP5KIα band, and so cannot definitively determine which codon is used for translation initiation. The increased density of the anti-PIP5KIα reactive band for the Δ36 PIP5KIα transfection despite consistent levels of GAPDH (a housekeeper protein; **Figure 12 D**), may result from increased transfection efficiency of the Δ36 PIP5KIα plasmid in that particular experiment. For confirmation of the identity of the PIP5KIα protein expressed in this clone, mass spectrometry of the band was considered, but not pursued.

Credit for the performance of the SDS-PAGE and western blot is owed to Ruby Chrisp.

A previous study showed that the *PIP5K1A*-related pseudogene *PIPSL*, a retrogene derived from a spliced *PIP5K1A* and *PSMD4* read-through transcript, is translated to give a functional protein that lacks lipid kinase function but retains the ubiquitin binding function of the proteasome component PSMD4 [129]. We note that in addition to other divergences in the PIP5KIα-related sequence, PIPSL lacks the N-terminal region and appears to use the codon corresponding to M36 in *PIP5K1A* CDS as its start codon [129]. While we were unable to characterise the N-terminal sequence of the truncated PIP5KIα protein in the ΔN clone unambiguously, we therefore hypothesise that M36 is the most likely candidate for the alternative translation start site in this clone.

3.4. Altered AKT phosphorylation in *PIP5K1A* edited lines

AKT is a serine threonine kinase with diverse targets and a broad spectrum of pro-growth, and pro-survival activities. The PI3K/AKT/mTOR signalling pathway is one of the most dysregulated pathways in human cancers, and functions in progression of prostate cancer to CRPC. AKT is activated by various phosphorylations, most importantly at Serine 473 and Threonine 308 [130]. There are in fact three isozymes of AKT: AKT1, AKT2 and AKT3. Threonine 308 and Serine 473 are in reference to AKT1. The three isozymes are highly homologous [131] and phosphorylation of the equivalent positions in AKT2 (Thr 309 and Ser 474) and AKT3 (Thr 305 and Ser 473) have the same activating effect in these isozymes. In the literature, the phosphorylated residues are commonly referenced simply by the residues in AKT 1. Production of $PI(3,4,5)P_3$ at cell membranes recruits AKT (and PDK1) leading to phosphorylation of AKT at Thr 308 by PDK1. Membrane recruitment of AKT also leads to its phosphorylation at Ser 473 by the mammalian target of rapamycin complex 2 (mTORC2) as well as by linked kinase (ILK) and mitogen-activated protein kinase-activated protein linase-2 (MAPKAPK2) [131].

PIP5KIα catalyses the production of $PI(4,5)P_2$ and therefore functions upstream of $PI(3,4,5)P_3$ synthesis by PI3K class I.

We therefore anticipated that depletion of PIP5KIα in the ΔN and KO sublines would result in a reduced level of phosphorylation of AKT in these cells.

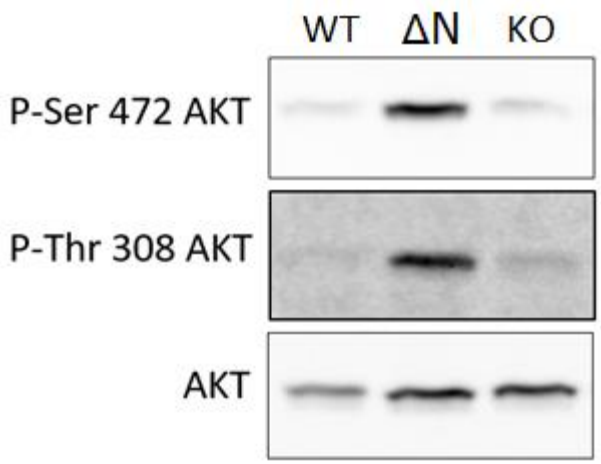
To test this, we performed western blotting with antibodies against the phosphorylated forms of AKT. Western blot against total AKT (i.e. an antibody that binds AKT regardless of phosphorylation state) was also performed as loading control. The manufactures of these antibodies state that they are not isozyme specific, and will therefore detect all three isozymes of AKT. Because of the negligible size difference between phosphorylated and unphosphorylated forms of AKT, it is not possible to blot for each form on the same membrane without serially stripping and re-probing. Since serially stripping and re-probing can lead to loss of target protein from the membrane, and therefore reduced detection, instead equalised masses of protein lysate were run in triplicate in the same SDS-PA gel, and after transfer to a membrane, the membrane was cut into three sections, one set of replicates per section. Each section was then separately probed for the different phosphorylated forms of AKT.

Low levels of phospho-Serine 473 and phospho-Threonine 308 AKT were detected in the KO subline, consistent with the level of these observed in the WT.

Surprisingly, we detected elevated levels of phospho- Serine 473 and phospho-Threonine 308 in the ΔN subline compared to the WT. The western blot for total AKT indicates that elevated phosphorylation of AKT in the ΔN subline is not explained by increased levels of total AKT protein. There was also no significant difference in expression of three *AKT* genes between the three sublines, as assessed by RNA-Seq (**Section 3.6.1; Figure 13**).

The expression of the N-terminally truncated *PIP5K1A* and high phospho AKT levels in ΔN cells compared to WT was confirmed independently by our collaborators at the University of Umea, Sweden [132]. The function of the PIP5KIα N-terminal region remains unknown and it is unclear how loss of this region correlates with enhanced AKT activation, and whether this is a direct or indirect effect. Possible mechanisms by which this occurs are currently under investigation by another PhD student in the lab, Ryan Morris.

A)



B)

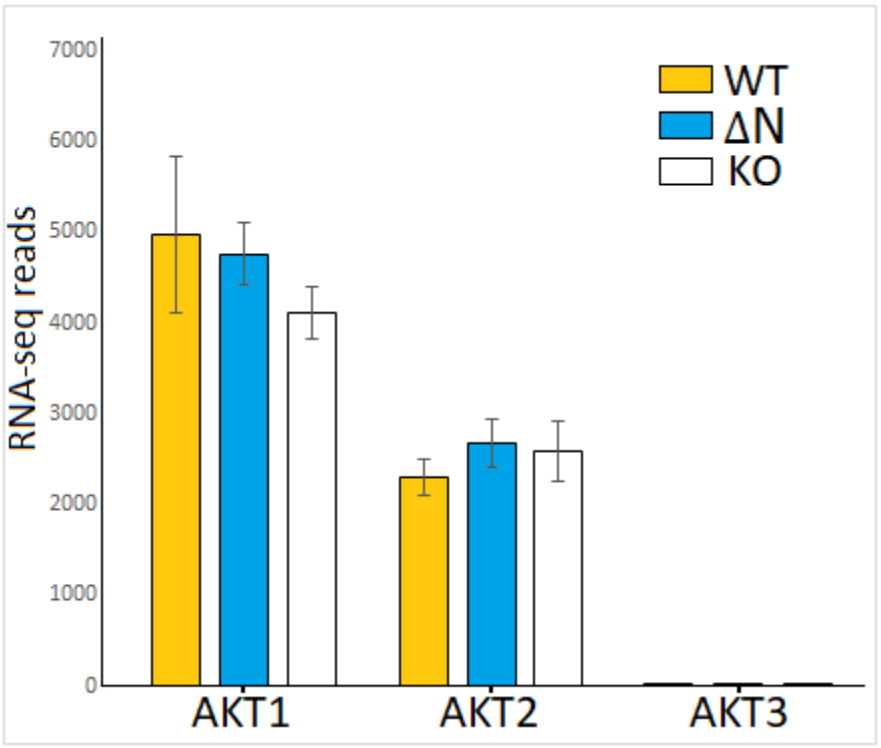


Figure 13. Elevated phosphorylated AKT in LNCaP C4-2 sublines.

A) Western blot detection of total, P-Serine 472 and P-Threonine 308 AKT in the WT, ΔN and KO sublines.

B) Mean numbers of RNA-Seq reads mapping to the three AKT genes for the LNCaP C4-2 WT, ΔN and KO sublines. Error bars represent standard deviations. n = 3.

3.5. Subcellular localisation of N-terminally truncated PIP5KIα variants

3.5.1. Distribution of endogenously expressed PIP5KIα in CRISPR edited clones

We next explored the subcellular distribution of endogenous PIP5KIα proteins in LNCaP C4-2 cells by immunofluorescence staining. The antibody used detects an epitope that is centred at G64 (of PIP5KIα i1; equivalent to G52 of PIP5KIα i2 and PIP5KIα i4), and thus will detect isoforms 2 and 4 as well as the truncated PIP5KIα protein in the ΔN cell line, as confirmed in the western blots (**Figure 12 D.**).

PIP5KIα proteins were detected both intracellularly and at the cell periphery (cell membrane). We noted that the subcellular distribution detected in C4-2 WT was influenced by cell density, as a higher proportion of the fluorescence was associated with the cell membrane when cultures were more confluent or where there were close cell-cell contacts (**Figure 14**). In particular, endogenous PIP5KIα appeared to be enriched where cell-cell contacts occurred. We also noted that the cytoplasmic PIP5KIα appeared to reveal vesicle-like structures that may be endosomes. In contrast, detection of endogenous PIP5KIα protein was strongly reduced in C4-2 KO cells, with only very low levels of staining detectable. In general, the subcellular

distribution of the N-terminally truncated PIP5KIα in ΔN cells was similar to WT, although some reduction in the intensity at cell junctions was apparent, which may be linked to the reduced expression level of the PIP5KIα ΔN protein (**Figure 14**). This suggests that the N-terminal 36 amino acids are not essential for targeting PIP5KIα to the cell membrane.

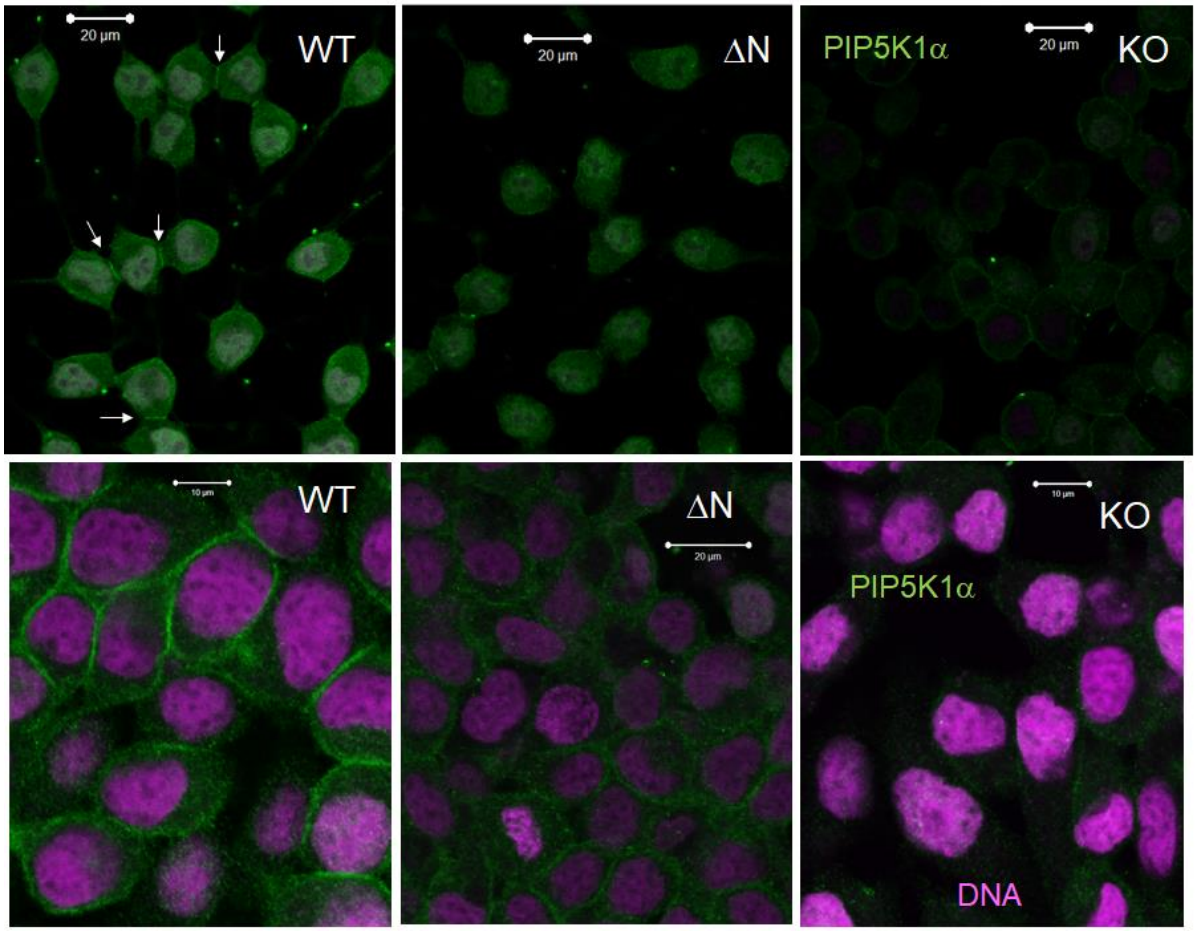


Figure 14. Immunofluorescence staining of endogenous PIP5KIα in WT, ΔN and KO cells.

Anti-PIP5KIα staining is shown in green. DNA staining (DAPI), is shown in purple. Cells were seeded at both low density (Top) and high density (Bottom), to observe differences in PIP5KIα intensity at sites of cell-cell contact.

3.5.2. Localisation of variants of PIP5KIα–Venus fusion proteins

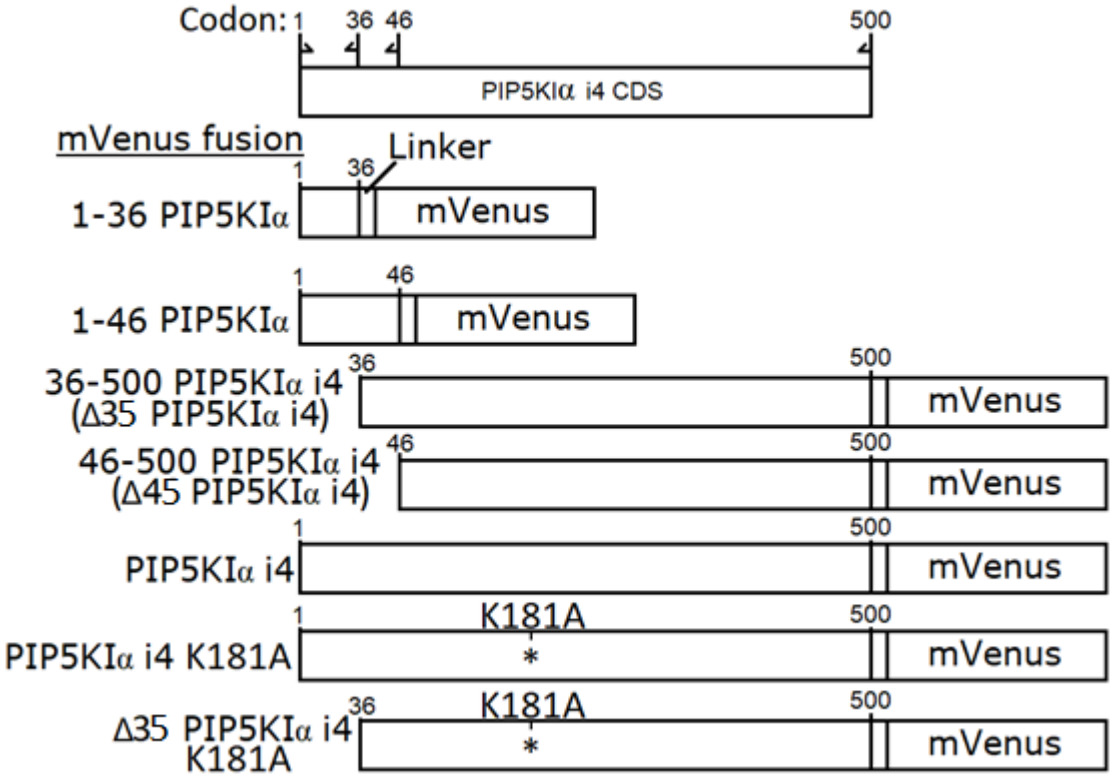
To examine how different isoforms of PIP5KIα are localised in cells, and to determine if the first 35 or 46 residues of PIP5KIα impact in subcellular localisation of PIP5KIα, molecular cloning was used to generate mammalian expression vectors encoding various truncated forms of PIP5KIα tagged with the fluorescent protein mVenus, a monomeric derivative of YFP [133].

Amplification of PIP5KIα CDS was performed with the same primers pairs as listed in **Supplementary Table 5.** Amplification was first attempted from cDNA synthesised from RNA extracts of WT LNCaP C4-2, but as this PCR failed, cloning of PIP5KIα CDSs was instead performed PIP5KIα i4 CDS in the pLPS-PIP5KIα plasmid, a gift from Jenny Perrson. CDS of PIP5KIα i4 1-500, PIP5KIα i4 36-500, PIP5KIα i4 46-500, and 1-36 PIP5KIα i4 and 1-46 PIP5KIα i4 were amplified from this plasmid, and ligated between the XhoI and BamHI sites of mVenusN1. **Figure 15 A** depicts the constructs generated. All plasmid constructs were sequenced to ensure correct sequence and orientation and that PIP5KIα and Venus CDS were in frame. The deduced amino acid sequence linking each PIP5KIα segment to mVenus is KDPPVAT. In addition, site-directed mutagenesis was used to introduce a K181A mutation into the tagged isoforms 2 and 4. The primers used for this site-directed

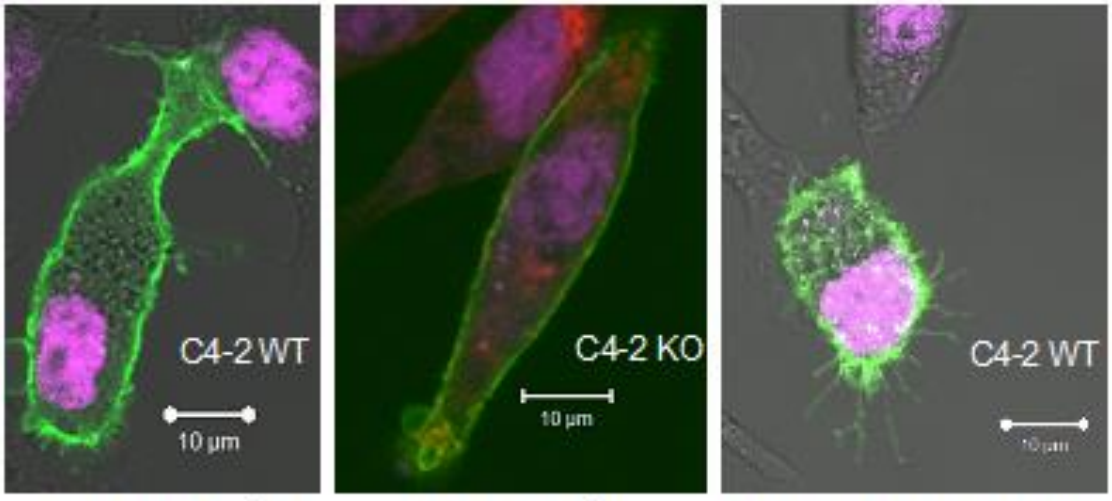
mutagenesis are listed in **Supplementary table 5**. This substitution was previously shown to result in a kinase-deficient variant of PIP5KIα [75].

These constructs were used in transient transfections of LNCaP C4-2 cells followed by confocal imaging. As shown in **Figure 15 B**, exogenously expressed Venus-tagged PIP5KIα i4 showed strong localisation to the cell membrane, including small projections and cell-cell junctions. Similar distributions were observed in WT, KO and ΔN cells. Intracellular staining to vesicle like structures, similar to that observed for endogenous PIP5KIα and some cells also displayed a highly polarised expression that co-localised with phalloidin, indicating recruitment at site of actin filamentation. No alteration in this distribution was observed due to the K181A mutation, or by deletions of the N-terminal sequences up to aa 45. Constructs consisting of Venus -tagged 1-36 or 1-46 showed a similar distribution to vector only (**Figure 15 C**).These results indicate that kinase functionality and N-terminus are not essential for intracellular targeting of PIP5KIα.

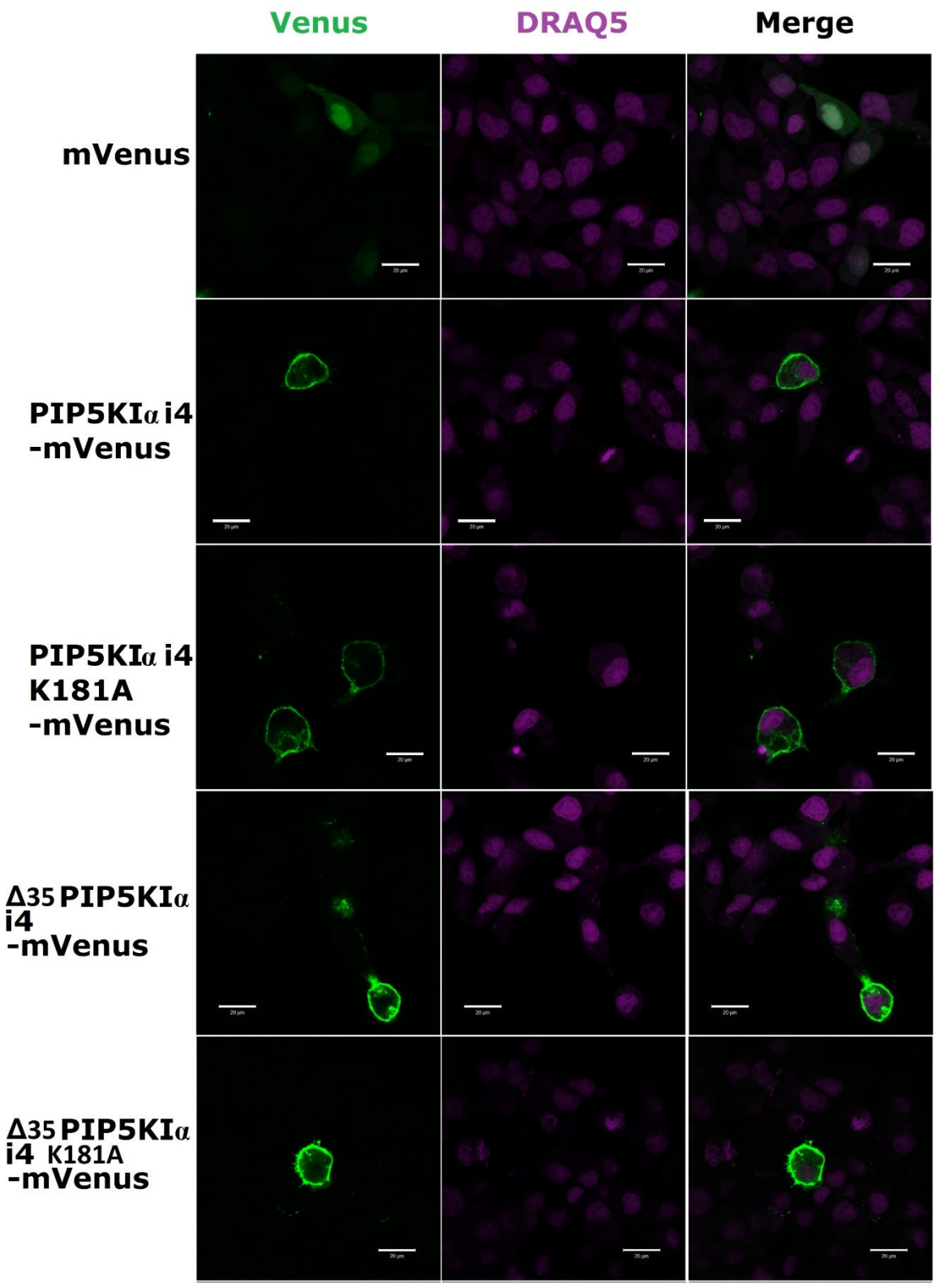
A)


Codon: 1 36 46 500
PIP5KIα i4 CDS
mVenus fusion
Linker
1-36 PIP5KIα
mVenus
1-46 PIP5KIα
mVenus
36-500 PIP5KIα i4
(Δ35 PIP5KIα i4)
mVenus
46-500 PIP5KIα i4
(Δ45 PIP5KIα i4)
mVenus
PIP5KIα i4
mVenus
K181A
PIP5KIα i4 K181A
mVenus
Δ35 PIP5KIα i4
K181A
mVenus

B)


C4-2 WT
10 µm
C4-2 KO
10 µm
C4-2 WT
PIP5KIα i4
-mVenus
PIP5KIα i4
-mVenus
PIP5KIα i4 K181A
-mVenus

C)


Venus
DRAQ5
Merge
mVenus
PIP5KIα i4 -mVenus
PIP5KIα i4 K181A -mVenus
Δ35 PIP5KIα i4 -mVenus
Δ35 PIP5KIα i4 K181A -mVenus

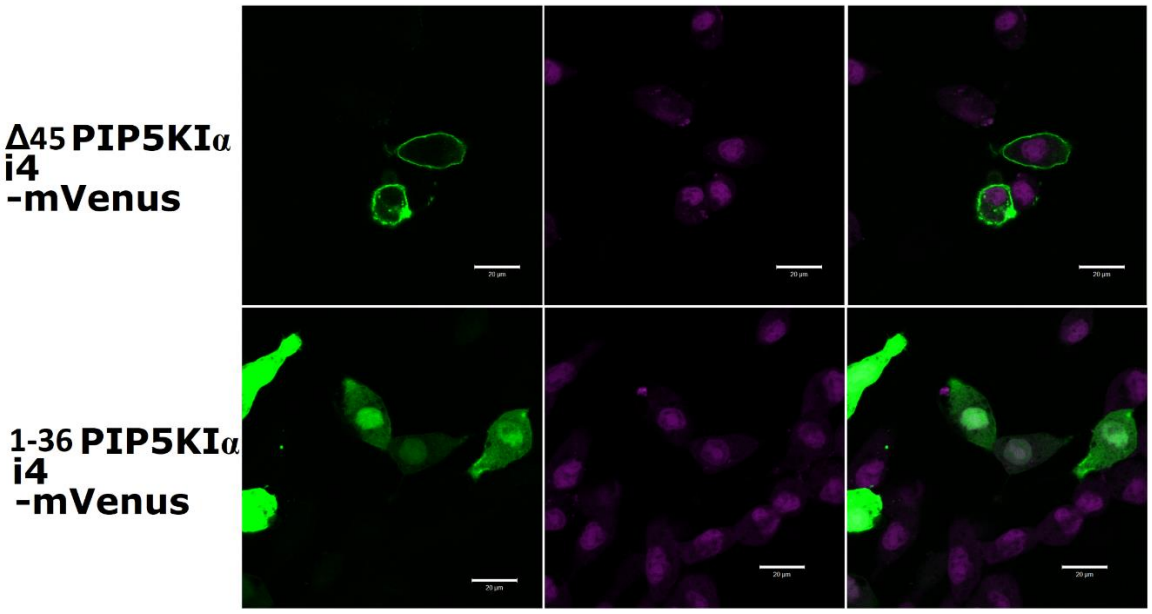

Δ45 PIP5KIα
i4
-mVenus
1-36 PIP5KIα
i4
-mVenus

Figure 15. Localisation of exogenously expressed PIP5KIα truncated variants.

A) Schematics of various PIP5KIα-mVenus fusion proteins encoded by the expression plasmids generated by molecular cloning into the mVenus vector.

B) PIP5KIα i4-mVenus or PIP5KIα i4 K181A-mVenus tagged (green) overlaid with nuclear staining (purple). The central panel was also stained with phalloidin (red) which stains actin filaments. Imaged using fluorescence microscopy. Scale bars represent 10 μm. Imaged using fluorescence microscopy.

C) mVenus-tagged PIP5KIα protein variants (green) and DNA staining (purple). Cells were the KO subline of LNCaP C4-2. Scale bares represent 20 μm. Imaged using fluorescence microscopy.

3.6. Effect of *PIP5K1A* editing on gene expression in LNCaP C4-2 cells

3.6.1. RNA-Seq analysis

To assess the impact of *PIP5K1A* disruption on gene expression, transcriptome profiling was performed with mRNA prepared from triplicate cultures of LNCaP C4-2 WT, *PIP5K1A*-KO and *PIP5K1A* ΔN cells (**Section 2.7**).

In total, we identified 818 robust DEGs (we categorise these as robust because of the stringent Log_2 fold change threshold of <=1, and a FDR of 0.01 as significance cut-offs) between KO and WT cells, of which 447 were significantly upregulated and 371 genes were downregulated. Analysis of the ΔN transcriptome compared to WT, identified almost double the number of robust DEGs (1675) of which 852 genes were upregulated and 823 showed significantly reduced expression. Volcano plots for both sublines compared to WT generated using iDEP.93 (integrated Differential Expression & Pathway Analysis tool; performed by Ryan Morris) and are shown in **Figure 16 A**. Comparison of the upregulated and downregulated DEGs in the KO and ΔN edited clones identified both common and unique sets of genes (**Figure 16 B**). Approximately 41% (152/371) of the genes downregulated in the KO line were also downregulated in ΔN cells, and similarly approximately 49% (219/447) of the upregulated DEGs were common to both cell

lines. Thus, while loss of PIP5KIα expression has effects on gene expression, additional effects may be observed due to expression of the N-terminally truncated PIP5KIα protein.

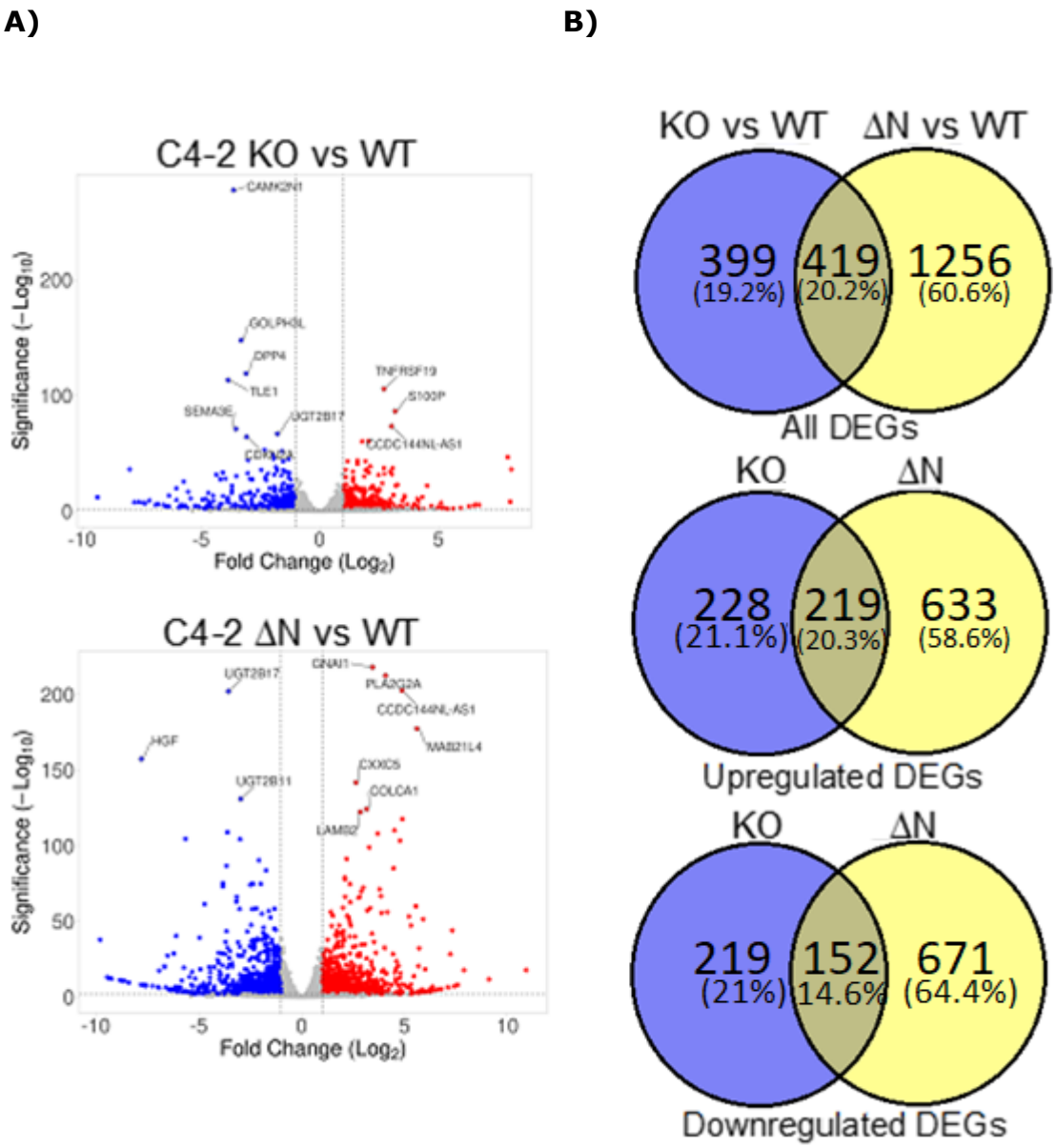


Figure 16. Differential gene expression between PIP5KIA-edited clones and WT LNCaP C4-2

A) Volcano plots showing differentially expressed genes (DEGs) identified in KO vs WT (top) or DN vs WT (bottom). Fold change expressed as Log 2Fc is represented on the X-axis, with the –Log2 significance (FDR) on the Y axis.

B) Venn diagrams created with Venny 2.0 showing overlap in DEGs identified for KO vs WT and DN vs WT. Comparisons are shown for All DEGs, upregulated DEGs and downregulated DEGs. The numbers indicate the number of genes and percentages of total are indicated.

The expression of *PIP5K1A* mRNAs was reduced in both the KO and ΔN lines, compared to WT (log2 FC of -1.5 and -0.85, respectively. Expression of *PIP5K1C*, whose protein product, PIP5KIγ, could theoretically compensate for deficiency in PI(4,5)P_2 signalling brought about by PIP5KIα depletion, was not significantly differentially expressed between the three sublines (**Figure 17**). Expression of *PIP5K1B* was likewise examined, but there was weak expression of this gene in all three sublines (WT: $\bar{x}$ = 3.6 reads; ΔN: $\bar{x}$ = 2.6 reads; KO: $\bar{x}$ = 3.6 reads; n=3).

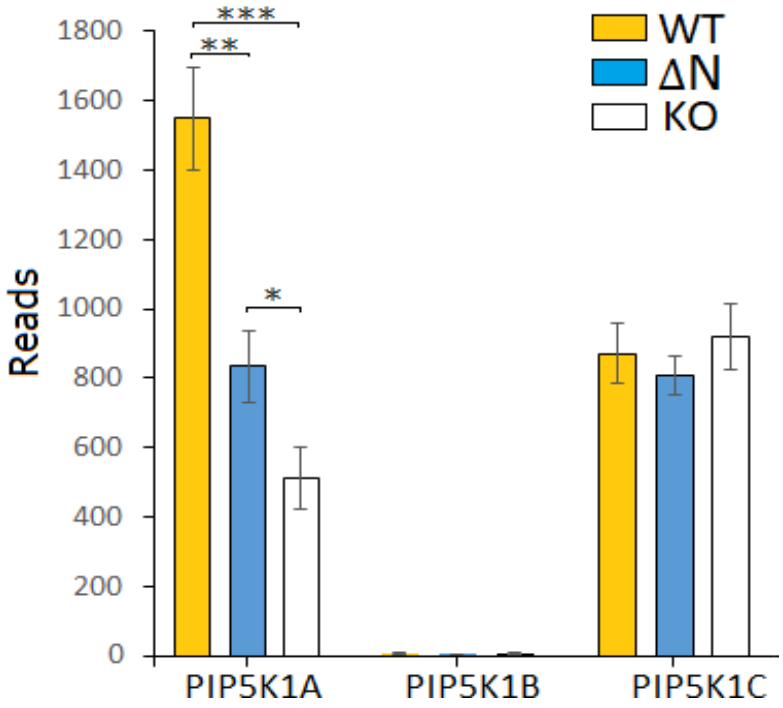


Figure 17. Differential expression of PIP5K genes between LNCaP C4-2 WT, ΔN and KO sublines.
Mean numbers of RNA-Seq reads mapping to *PIP5K1A*, *PIP5K1B* and *PIP5K1C*. Error bars indicate standard deviations. n =3. Significance for two-way t tests indicated: *: p <0.05; **: p<0.01; p<0.001.

To understand the transcriptome changes, we performed gene set enrichment analysis (GSEA/ MSigDB) to compare with Hallmark, KEGG gene sets. As shown in **Table 4** genes involved in cell cycle regulation (G2M checkpoint; mitotic spindle; E2F targets) were significantly enriched amongst the upregulated genes in both the KO and ΔN cell lines. In addition, we also noted enrichment of genes involved in steroid receptor, p53 and K-Ras signalling pathways. Interestingly, PIP5KIα has been recently reported to interact directly with mutant K-Ras and p53 proteins to mediate oncogenic signalling [134; 135].

Table 4. Gene set enrichment analysis (GSEA) for KO and ΔN sublines vs WT LNCaP C4-2.

KO vs Wt UP-regulated (HALLMARK) Converted 447 submitted identifiers into 421 NCBI (Entrez) genes			
Gene Set Name [# Genes (K)]	# Genes in Overlap (K)	p value	FDR q-value
HALLMARK G2M CHECKPOINT [200]	24	1.76 e^{-18}	8.81 e^{-17}
HALLMARK E2F TARGET [200]	20	4.4 e^{-14}	1.1 e^{-12}
HALLMARK MITOTIC SPINDLE [199]	18	4.59 e^{-12}	7.65 e^{-11}
HALLMARK ESTROGEN RESPONSE LATE[200]	11	9.41 e^{-6}	1.18 e^{-4}
HALLMARK ESTROGEN RESPONSE EARLY [200]	10	5.36 e^{-5}	4.47 e^{-4}
HALLMARK P53 PATHWAY [200]	10	5.36 e^{-5}	4.47 e^{-4}
HALLMARK GLYCOLYSIS [200]	9	2.77 e^{-4}	1.73 e^{-3}
HALLMARK KRAS SIGNALLING UP [200]	9	2.77 e^{-4}	1.73 e^{-3}
KO vs Wt DOWN-regulated (HALLMARK) Converted 371 submitted identifiers into 342 NCBI (Entrez) genes			
HALLMARK UP RESPONSE DOWN	14	2.8 e^{-11}	1.4 e^{-9}
HALLMARK GLYCOLYSIS	9	5.84 e^{-5}	7.3 e^{-4}
HALLMARK INFLAMMATORY RESPONSE	9	5.84 e^{-5}	7.3 e^{-4}
HALLMARK KRAS SIGNALLING UP	9	5.84 e^{-5}	7.3 e^{-4}
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION	8	3.33 e^{-4}	3.33 e^{-3}
HALLMARK HYPOXIA	7	1.68 e^{-3}	1.05 e^{-2}
HALLMARK KRAS SIGNALLING DOWN	7	1.68 e^{-3}	1.05 e^{-2}
HALLMARK MYOGENESIS	7	1.68 e^{-3}	1.05 e^{-2}
HALLMARK APICAL JUNCTION	6	7.46 e^{-3}	4.14 e^{-2}
KO vs Wt UP-regulated (KEGG) Converted 447 submitted identifiers into 421 NCBI (Entrez) genes			
KEGG CELL CYCLE [125]	11	9.04 e^{-8}	1.68 e^{-5}
KEGG PATHWAYS IN CANCER [325]	15	2.09 e^{-6}	1.94 e^{-4}
KEGG OOCYTE MEIOSIS [113]	8	2.61 e^{-5}	1.62 e^{-3}

KEG PROGESTERONE MEDIATED OOCYTE MATURATION [85]	6	$2.73\ e^{-4}$	$1.27\ e^{-2}$
KEGG REGULATION OF ACTIN CYTOSKELETON [213]	9	$4.38\ e^{-4}$	$1.63\ e^{-2}$
KEGG MELANOMA [71]	5	$8.97\ e^{-4}$	$2.78\ e^{-2}$
KO vs Wt DOWN-regulated (KEGG) Converted 371 submitted identifiers into 342 NCBI (Entrez) genes			
KEGG SPHINGOLIPID METABOLISM [39]	5	$1.94\ e^{-5}$	$1.29\ e^{-3}$
KEGG CALCIUM SIGNALLING PATHWAY [178]	9	$2.35\ e^{-5}$	$1.29\ e^{-3}$
KEGG PATHWAYS IN CANCER [325]	12	$2.54\ e^{-5}$	$1.29\ e^{-3}$
KEGG METABOLISM OF XENOBIOTICS BY CYTOCHROME P450 [70]	6	$2.96\ e^{-5}$	$1.29\ e^{-3}$
KEGG DRUG METABOLISM BY CYTOCHROME P450 [72]	6	$3.48\ e^{-5}$	$1.29\ e^{-3}$
KEGG ASCORBATE AND ALDARATE METABOLISM [25]	4	$5.59\ e^{-5}$	1..73 e^{-3}
KEGG STARCH AND SUCROSE METABOLISM [52]	5	$8\ e^{-5}$	$2.06\ e^{-3}$
KEGG PENTOSE AND GLUCORONATE INTERCONVERSIONS [28]	4	$8.87\ e^{-5}$	$2.06\ e^{-3}$
KEGG NEUROACTIVE LIGAND RECEPTOR INTERACTION [272]	10	$1.24\ e^{-4}$	$2.55\ e^{-3}$
KEGG ETHER LIPID METABOLISM [33]	4	$1.71\ e^{-4}$	$3.19\ e^{-3}$
ΔN vs Wt UP-regulated (HALLMARK) Converted 852 submitted identifiers into 788 NCBI (Entrez) genes			
HALLMARK G2M CHECKPOINT [200]	55	$8.66\ e^{-47}$	$4.33\ e^{-45}$
HALLMARK E2F TARGET [200]	53	$3.46\ e^{-44}$	$8.64\ e^{-43}$
HALLMARK MITOTIC SPINDLE [199]	32	$5.12\ e^{-20}$	$8.53\ e^{-19}$
HALLMARK ESTROGEN RESPONSE LATE[200]	21	$5.2\ e^{-10}$	$6.5\ e^{-9}$
HALLMARK HYPOXIA [200]	17	$4.93\ e^{-7}$	$4.93\ e^{-6}$
HALLMARK SPERMATOGENESIS [135]	13	$2.71\ e^{-6}$	$2.26\ e^{-5}$
HALLMARK ESTROGEN RESPONSE EARLY[200]	15	$1.06\ e^{-6}$	$7.56\ e^{-5}$
HALLMARK GLYCOLYSIS [200]	14	$4.43\ e^{-5}$	$2.77\ e^{-4}$

HALLMARK MTORC1 SIGNALLING [200]	13	$1.72\ e^{-4}$	$8.61\ e^{-4}$
HALLMARK MYOGENESIS [200]	13	$1.72\ e^{-4}$	$8.61\ e^{-4}$
ΔN vs Wt DOWN-regulated (HALLMARK) Converted 823 submitted identifiers into 731 NCBI (Entrez) genes			
HALLMARK KRAS SIGNALLING UP [200]	18	$3.16\ e^{-8}$	$8.63\ e^{-7}$
HALLMARK ANDROGEN RESPONSE [100]	13	$3.45\ e^{-8}$	$8.63\ e^{-7}$
HALLMARK UV RESPONSE DOWN [144]	14	$4.14\ e^{-7}$	$6.9\ e^{-6}$
HALLMARK HYPOXIA [200]	16	$8.98\ e^{-8}$	$1.12\ e^{-5}$
HALLMARK ESTROGEN RESONSE LATE [200]	15	$4.34\ e^{-6}$	$4.34\ e^{-5}$
HALLMARK ESTROGEN RESPONSE EARLY [200]	14	$1.96\ e^{-5}$	$1.63\ e^{-4}$
HALLMARK COMPLEMENT [200]	13	$8.23\ e^{-5}$	$4.57\ e^{-4}$
HALLMARK KRAS SIGNALLING DOWN [200]	13	$8.23\ e^{-5}$	$4.57\ e^{-4}$
HALLMARK TNFA SIGNALLING VIA NFKB [200]	13	$8.23\ e^{-5}$	$4.57\ e^{-4}$
HALLMARK IL2 STAT5 SIGNALLING [199]	12	$3.06\ e^{-4}$	$1.45\ e^{-3}$
ΔN vs Wt UP-regulated (KEGG) Converted 854 submitted identifiers into 788 NCBI (Entrez) genes			
KEGG CELL CYCLE [125]	23	$4.2\ e^{-16}$	$7.81\ e^{-14}$
KEGG OOCYTE MEIOSIS [113]	17	$8.25\ e^{-11}$	$7.68\ e^{-9}$
KEGG PROGESTERONE MEDIATED OOCYTE MATURATION [85]	14	$1.12\ e^{-9}$	$6.94\ e^{-8}$
KEGG P53 SIGNALLING PATHWAY [68]	10	$8.05\ e^{-7}$	$3.74\ e^{-5}$
KEGG PATHWAYS IN CACNER [325]	20	$7.83\ e^{-16}$	$2.91\ e^{-4}$
KEGG ARACHIDONIC ACID METABOLISM [58]	8	$1.66\ e^{-5}$	$5.16\ e^{-4}$
KEGG AXON GUIDANCE [129]	11	$4.89\ e^{-5}$	$1.3\ e^{-3}$
KEGG FOCAL ADHESION [199]	13	$1.64\ e^{-4}$	$3.81\ e^{-3}$
KEGG GLYCOPHOSPHOLIPID METABOLISM [77]	7	$7.81\ e^{-4}$	$1.61\ e^{-2}$
KEGG MAPK SIGNALLING PATHWAY [267]	14	$8.66\ e^{-4}$	$1.61\ e^{-2}$
ΔN vs Wt DOWN-regulated (KEGG) Converted 823 submitted identifiers into 731 NCBI (Entrez) genes			
KEGG NEUROACTIVE LIGAND RECEPTOR INTERACTION [272]	21	$3.41\ e^{-8}$	$6.34\ e^{-6}$

KEGG DRUG METABOLISM CYTOCHROME P450 [72]	10	$7.05\ e^{-7}$	$6.55\ e^{-6}$
KEGG ASCORBATE AND ALDERATE METABOLISM [25]	6	$4.6\ e^{-5}$	$2.28\ e^{-5}$
KEGG METABOLISM OF XENOBIOTICS BY CYTOCHROM P450 [70]	9	$4.9\ e^{-5}$	$2.28\ e^{-4}$
KEGG PENTOSE AND GLUCONORATE INTERCONVERSION [28]	6	$9.34\ e^{-6}$	$3.47\ e^{-4}$
KEGG FOCAL ADHESION [199]	14	$1.85\ e^{-6}$	$5.1\ e^{-4}$
KEGG RETINOL METABOLISM [64]	8	$2.04\ e^{-5}$	$5.1\ e^{-4}$
KEGG GLIOMA [65]	8	$2.28\ e^{-5}$	$5.1\ e^{-4}$
KEGG CALCIUM SIGNALLING PATHWAY [178]	13	$2.47\ e^{-5}$	$5.1\ e^{-4}$
KEGG GAP JUNCTION [90]	9	$3.85\ e^{-5}$	$7.17\ e^{-4}$

Because of the importance of the PI(3,4,5)P_3/PI3K/AKT signalling pathway in overcoming ADT and the progression of CRPC, we interrogated differential expression of genes involved in androgen signalling. AR was significantly differentially expressed between the WT and ΔN sublines (Two-tailed t-tests: t = 2.82, p = 0.048) but was not significantly differentially expressed between ΔN and KO lines (t = 2.39, p = 0.08) or WT and KO lines (t = 1.52, p = 0.20). Differential expression of splice variants of *AR* was assessed by analysis of Sashimi plots for the RNA-Seq data in Interactive Genomics Viewer (IGV). Specifically, assessment was made of differences in the ratios of spliced reads joining *AR* Exon 3 to cryptic exons (Cryptic Exon 5 and Cryptic Exon 3), as reported for transcripts encoding AR-V7 and AR-V9 [136]. Very few such reads (<3 per subline) were detected in comparison to ~150 spliced reads joining *AR* Exon 3 to Exon 4 (indicating expression of the full length *AR* splice variant) per subline.

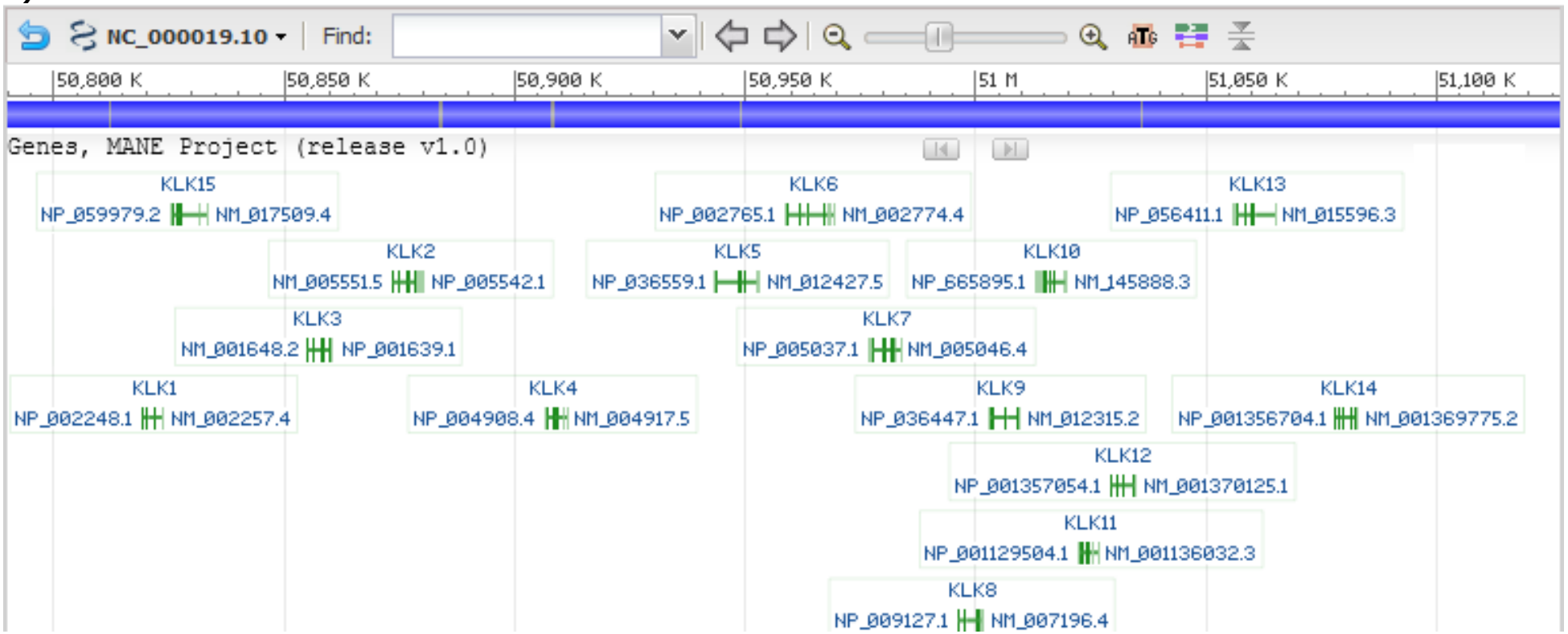
Our attention was drawn to the fact that amongst the most robustly significant DEGs in the *PIP5K1A* edited cells were several members of the Kallikrein (*KLK*) gene family, namely *KLK2*, *KLK3* and *KLK4.* These genes are clustered on chromosome 19 (**Figure 18 A**) and are known to be androgen-regulated and prostate-specific. Assessment of PSA (encoded by *KLK3*) levels in serum continues to be an important clinical

biomarker of prostate cancer, although various other non-cancer factors can also cause PSA levels in serum to rise.

The RNA-Seq data indicated significantly enhanced expression of *KLK2, KLK3* and *KLK4* in the PIP5K1A-KO cells in comparison to the WT LNCaP C4-2 cells (**Figure 18 B**). The regulation of KLK3 gene by AR/androgens is dependent on AREs in the proximal promoter and in enhancers located within long terminal repeat (LTR) sequences, and these genes are upregulated in CRPC, due to activation of AR signalling [137]. Therefore, depletion of functional PIP5KIα appears to correlate with increased expression of these key AR target genes, this provides further support for the hypothesis of a functional interplay of androgen and $PI(4,5)P_2$ signalling pathways.

However, while *KLK3* expression was also increased in the ΔN clone, curiously the expression of *KLK2* and *KLK4* was reduced in this cell line compared to WT LNCaP C4-2 cells (**Figure 18 B**). The reason for this differential effect (which needs to be validated by RTqPCR) is unknown although differential expression of these genes in response to androgens and antiandrogens in LNCaP cells has been reported previously [138]. Given that the ΔN and KO cells show only partially overlapping transcriptomes, it is likely that other DEGs may impact on *KLK2* and *KLK4* in ΔN cells.

A)


NC_000019.10
Find:
50,800 K
50,850 K
50,900 K
50,950 K
51 M
51,050 K
51,100 K
Genes, MANE Project (release v1.0)
KLK15
NP_059979.2
NM_017509.4
KLK6
NP_002765.1
NM_002774.4
KLK13
NP_056411.1
NM_015596.3
KLK2
NM_005551.5
NP_005542.1
KLK5
NP_036559.1
NM_012427.5
KLK10
NP_665895.1
NM_145888.3
KLK3
NM_001648.2
NP_001639.1
KLK7
NP_005037.1
NM_005046.4
KLK1
NP_002248.1
NM_002257.4
KLK4
NP_004908.4
NM_004917.5
KLK9
NP_036447.1
NM_012315.2
KLK14
NP_001356704.1
NM_001369775.2
KLK12
NP_001357054.1
NM_001370125.1
KLK11
NP_001129504.1
NM_001136032.3
KLK8
NP_009127.1
NM_007196.4

B)

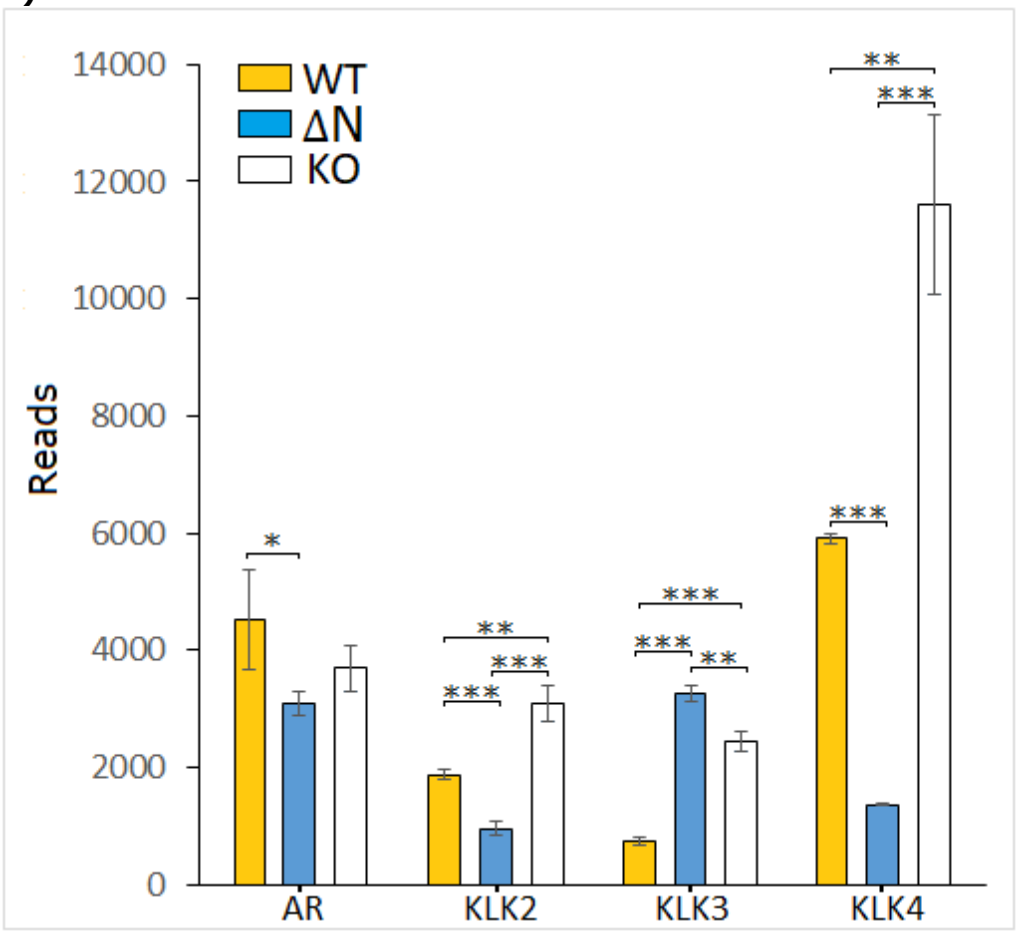


Figure 18: Differential expression of *AR* and *KLK* genes between LNCaP C4-2 WT, ΔN and KO sublines.

A) View from the NCBI Human Genome Browser showing Chromosome 19 positions 50790000 to 51125000 (GRCh38.p13 Primary Assembly), including the locus of the Kallikrien gene family.

B) Mean numbers of RNA-Seq reads mapping to *AR* and *KLK* genes for LNCaP C4-2 WT, ΔN and KO sublines. Error bars indicate standard deviations. n =3. Significance for two-way t tests indicated: *: p <0.05; **: p<0.01; p<0.001.

We next looked at expression of members of the *UGT2B* gene family, which encode UDP-glucuronosyltransferases, and were amongst the most highly significant downregulated DEGs in our analysis (**Figure 19 A**). These enzymes catalyse conjugation of hydrophilic glucuronide moiety to various substrates, typically increasing their solubility for transport from tissues and ultimately their excretion by renal or hepatic mechanisms.

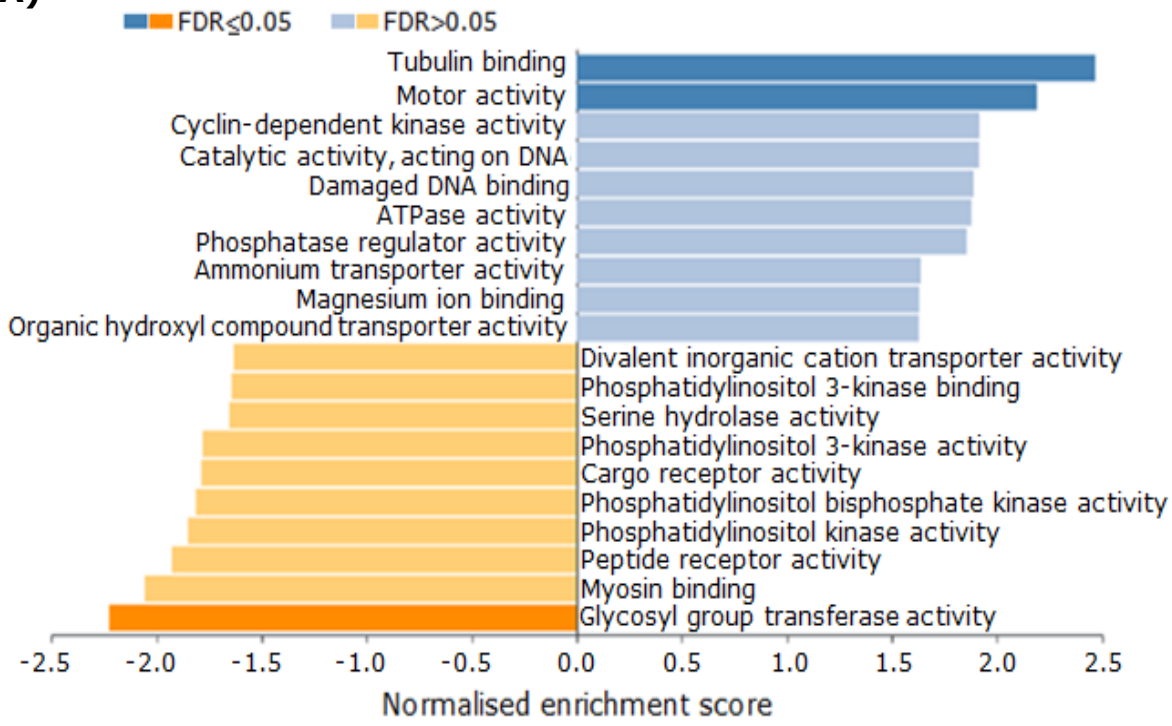
A subset of members of this family, *UGT2B15*, *UGT2B17* and *UGT2B28* are major enzymes involved in the inactivation of steroid hormones in human prostate cells [139-141]. Significantly lower expression of these genes was observed in the ΔN and KO lines relative to WT (**Figure 19 B**) suggesting that functional PIP5KIα is involved in maintaining their expression (directly or indirectly). This could also provide a plausible explanation for the enhanced expression of AR activated genes such as *KLK3*, due to enhanced availability of androgen ligand.

We also observed significantly lower expression of other members of this family, *UGT2B4*, *UGT2B10*, and *UGT2B11*. UGT2B4 is mainly expressed in the liver and catalyses glucornidation of bile acids and androgen metabolites [114]. UGT2B10 catalyses glucuronidation of various of xenobiotic compounds as well as endogenous fatty acids [114]. UGT2B11 has

poorly defined substrate preference, possibly including 3-alpha-hydroxyandrogens and fatty acids [114]. The other member of this family, *UGT2B7*, which is also implicated in prostate cancer, did not show significantly different expression between the three sublines, but had a very low level of expression (<10 reads per replicate) in each subline. Mean read numbers and p-values for two-way t-tests for each of these genes are presented in **Table 5.**

In contrast, all members of UGT2A family, *UGT2A1*, *UGT2A2* and *UGT2A3*, and all members of the *UGT1A* family showed very low expression in LNCaP C4-2, with a total of between zero and three reads per gene across all replicates (data not shown).

A)


FDR≤0.05
FDR>0.05
Tubulin binding
Motor activity
Cyclin-dependent kinase activity
Catalytic activity, acting on DNA
Damaged DNA binding
ATPase activity
Phosphatase regulator activity
Ammonium transporter activity
Magnesium ion binding
Organic hydroxyl compound transporter activity
Divalent inorganic cation transporter activity
Phosphatidylinositol 3-kinase binding
Serine hydrolase activity
Phosphatidylinositol 3-kinase activity
Cargo receptor activity
Phosphatidylinositol bisphosphate kinase activity
Phosphatidylinositol kinase activity
Peptide receptor activity
Myosin binding
Glycosyl group transferase activity
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
Normalised enrichment score

B)

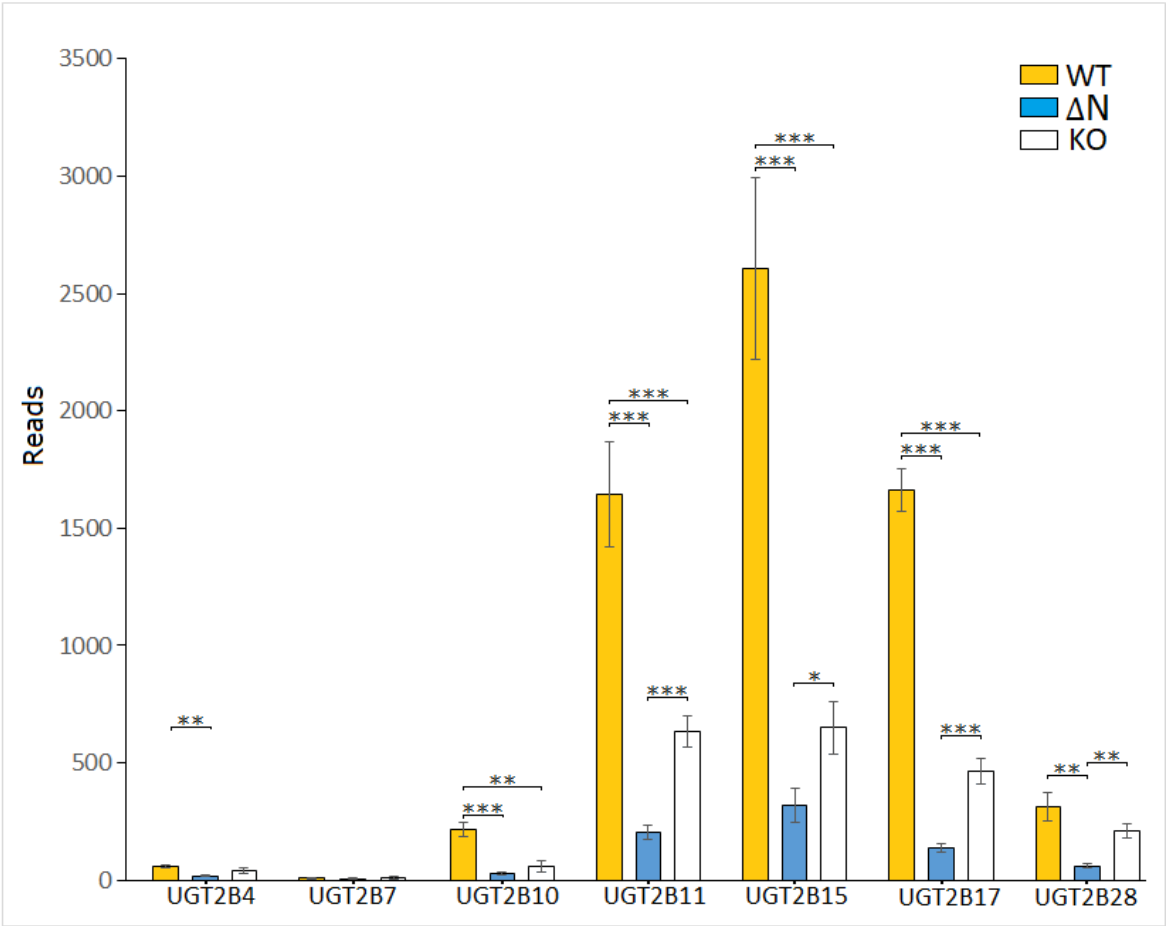

WT
ΔN
KO
Reads
3500
3000
2500
2000
1500
1000
500
0
UGT2B4
UGT2B7
UGT2B10
UGT2B11
UGT2B15
UGT2B17
UGT2B28

Figure 19. Differential expression of *UGT2B* genes between LNCaP C4-2 WT, ΔN and KO sublines.

A) Enrichment analysis performed in WebGestalt indicating the significant downregulation of glucuronosyltransferases in the ΔN cells compared to WT C4-2.

B) Mean numbers of RNA-Seq reads mapping to the *UGT2B* genes for LNCaP C4-2 WT, ΔN and KO sublines. Error bars indicate standard deviations. n =3. Significance for two-way t tests indicated: *: p <0.05; **: p<0.01; p<0.001.

The roles of UGT2B15, UGT2B17 and UGT2B28 in prostate cancer development are complex and their precise roles in are not yet fully elucidated. In general, studies suggest that higher expression of UGT2B15 and UGT2B17 protects against the development of BPH and PCa by metabolising androgens, and thus lowering their bioavailability within the prostate. While having a protective effect against development of PCa, other studies have concluded that expression of UGT2B17 is associated with higher Gleason scores and increased risk of biochemical recurrence and metastasis for PCa patients [142; 143]. A causal relationship between higher UGT2B17 expression and disease progression is suggested by findings that induced overexpression of UGT2B17, promotes cell proliferation, and progression to CRPC in mouse xenograft models of prostate cancer. These effects are reportedly associated with decreased androgen-dependent AR transcriptional activity, and elevated androgen-independent AR transcriptional activity, and are proposed to be mediated through UGT2B17 activation of c-Src kinase and subsequent phosphorylation of AR [143].

Table 5. Mean numbers of RNA-Seq reads mapping to *PIP5K, AR, KLK* and *UGT2B* genes for the WT, ΔN and KO LNCaP C4-2 sublines.

Gene	Mean read (n=3)			P-value (2 way t-test)		
	WT	ΔN	KO	WT - ΔN	ΔN-KO	WT - KO
PIP5K1A	1548.33	835.00	511.67	0.0023	0.0147	0.0005
PIP5K1B	3.67	2.67	4.67	0.5072	0.0550	0.5391
PIP5K1C	871.33	809.33	920.33	0.3540	0.1593	0.5454
AR	4523.67	3086.67	3695.67	0.0479	0.0753	0.2039
KLK2	1883.33	963.00	3090.00	0.0005	0.0004	0.0030
KLK3	736.00	3274.33	2435.67	0.0000	0.0026	0.0001
KLK4	5904.00	1366.67	11607.00	0.0000	0.0003	0.0030
UGT2B4	58.33	18.00	41.00	0.0017	0.0372	0.1191
UGT2B7	8.33	5.00	10.00	0.1161	0.3116	0.7049
UGT2B10	216.00	25.67	57.33	0.0004	0.0908	0.0021
UGT2B11	1643.67	205.00	633.00	0.0004	0.0006	0.0017
UGT2B15	2607.00	319.33	650.00	0.0005	0.0124	0.0011
UGT2B17	1664.33	138.00	466.33	0.0000	0.0006	0.0000
UGT2B28	315.33	61.33	210.33	0.0019	0.0011	0.0537
GUSB	2225.00	1951.67	2139.00	0.0939	0.2746	0.6371

3.6.2. Validation of *UGT2B10*, *UGT2B11*, *UGT2B15*, *UGT2B17* and *KLK3* gene expression in CRISPR edited clones

qPCR was performed to validate the findings of differentially expressed *UGT2B* genes (*UGT2B10, UGT2B11, UGT2B15 and UGT2B17*) and *KLK3* as observed within our RNA-Seq data for the PIP5KIA edited clones.

GUSB (ENSG00000169919) was chosen as a reference gene because our RNA-Seq data indicated robust expression, with no significant difference in read numbers between the three sublines (**Figure 20**). Two-tailed t-tests: WT - ΔN: t = 2.18797; p = 0.093905. ΔN - KO: t = -1.26474; p = 0.137316. WT – KO: t = 0.50964; p = 0.637132.

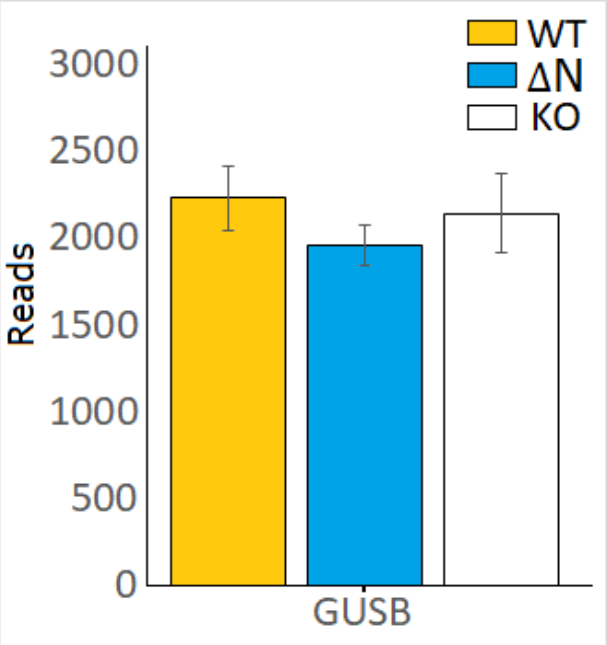


Figure 20. (Lack of) Differential expression of *GUSB* gene between LNCaP C4-2 WT, ΔN and KO sublines.
Error bars indicate standard deviations. n =3.

The primers, their design and validation for the suitability for use in qPCR are described in the appendix; **Section 7.1.1**.

RNA was extracted from the LNCaP C4-2 WT, ΔN and KO sublines, and cDNA synthesised as described in **Section 2.3.16** and **Section 2.3.17**. qPCRs were performed, for *UGT2B10, UGT2B11, UGT2B15, UGT2B17* and *KLK3*, and normalised expression values calculated by reference to *GUSB*, as described in **Section 2.3.18**.

qPCRs confirmed significantly lower expression of *UGT2B11, UGT2B15* and *UGT2B17* in the ΔN and KO sublines relative to WT (**Figure 21 A**). The expression levels for *UGT2B10* are not presented because these were highly variable between biological replicates, possibly because of the lower expression of this gene in all three sublines as indicated by the RNA-Seq data (**Figure 19 B**). qPCRs also confirmed significantly higher *KLK3* expression in the ΔN subline relative to WT. Expression of *KLK3* was higher in the KO subline than in the WT, but not significantly so.

To investigate mechanistic relationship between depletion of PIP5KIα and altered expression of the *UGT2B* genes and *KLK3*, a rescue experiment was performed. KO cells were transfected with either the PIP5KIα i4-mVenus or mVenus

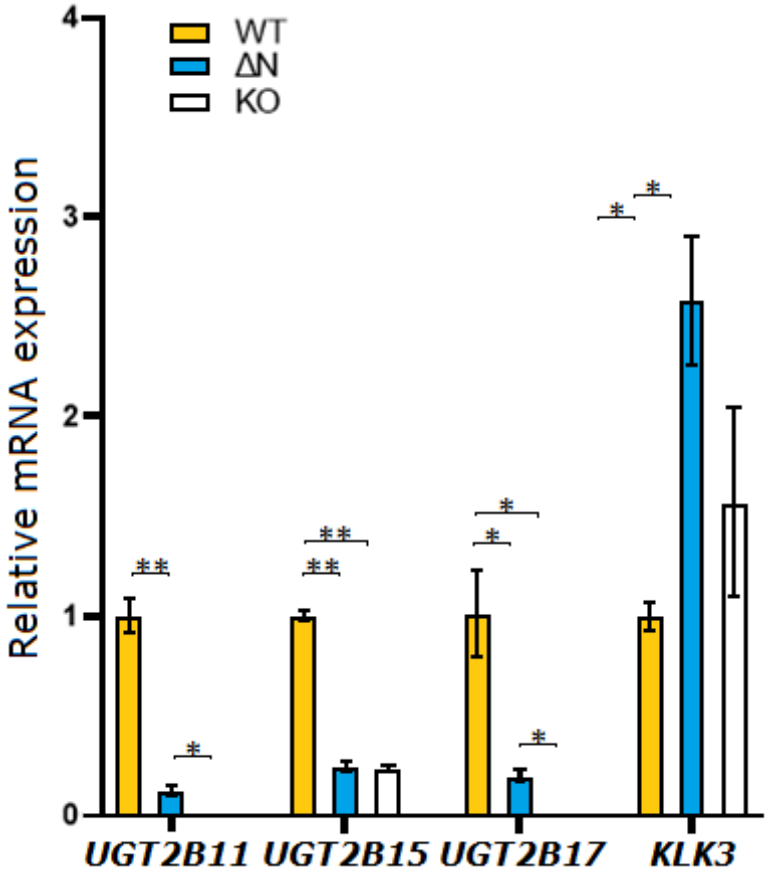
expression plasmid by Lipofectamine 3000 mediated transfection, and after 48 hours were trypsinised and subjected to FACS (**Section 2.5.7**), sorting in two rounds (first enrichment, then purification) for mVenus fluorescence. Addition of GFP tags to PIP5KIα is not known to interfere with the enzymatic or other functions of this protein, thus the mVenus (a derivative of GFP) tagged version was used to facilitate cell sorting of transfected cells.

The transfection efficiency with the PIP5KIα i4-mVenus and mVenus plasmids were ~4% and ~12% respectively, which is not untypical as these cells are known to be difficult to transfect. A total of 30192 cells transfected with PIP5KIα i4-mVenus, and 120263 cells expressing mVenus were collected. RNA was extracted and equal masses (200 ng) were used for cDNA synthesis.

qPCRs were performed for *UGT2B15, UGT2B17* and *KLK3*. These showed significantly higher expression of *UGT2B17* and significantly lower expression of *KLK3* in KO cells expressing the PIP5KIα i4-mVenus fusion than the mVenus expressing controls (**Figure 21 B**). *UGT2B15* expression was not significantly different between the treatment groups. There was not sufficient time to perform biological replicates of this last experiment.

RTqPCR analysis were performed to assess the effect of restoration of functional PIP5KIα on the expression levels of *UGT2B15, UGT2B17* and *KLK3*. As shown in **Figure 21 B**, expression of PIP5KIα i4-mVenus in KO cells strongly reduced the expression of KLK3 in comparison to mVenus transfected controls, indicating that PIP5KIα function influences expression of this AR target gene. Similarly, PIP5KIα i4-mVenus expression in KO cells elevated the expression of *UGT2B17*, reversing the effect seen upon loss of PIP5KIα in these cells. We did not however observe a similar effect on *UGT2B15* expression. Unfortunately, there was insufficient time to perform biological replicates of this experiment, or to enhance transfection efficiency. In future approaches a lentiviral delivery system would be a far more efficient strategy to perform rescue experiments. Nonetheless, with the caveat that further replications are required, these data provide a proof of principle that the LNCaP C4-2 KO subline line combined with exogenous expression vectors will be a useful model to study the molecular function of PIP5KIα isoforms at the cellular level.

A)


WT
ΔN
KO
Relative mRNA expression
4
3
2
1
0
**
*
**
**
*
*
*
*
UGT2B11
UGT2B15
UGT2B17
KLK3

B)

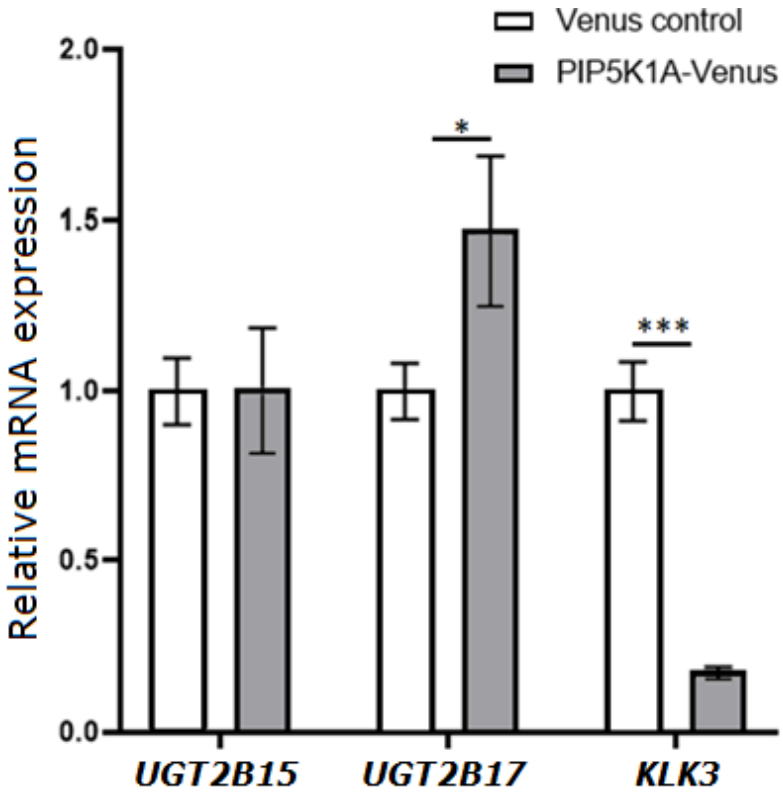

Venus control
PIP5K1A-Venus
Relative mRNA expression
2.0
1.5
1.0
0.5
0.0
*
***
UGT2B15
UGT2B17
KLK3

Figure 21. Validation of altered exression of *UGT2B* and *KLK3* genes in ΔN and KO sublines.

A) mRNA expression of *UGT2B11, UGT2B15, UGT2B17* and *KLK3* in LNCaP C4-2 ΔN and KO cells relative to WT. Statisitcal analysis performed using two-way t-test. Error bars indicate standard deviations. n =3. *: p <0.05; ***: p<0.001.

B) Relative mRNA expression of *UGT2B15, UGT2B17* and *KLK3* in LNCaP C4-2 KO cells transfected with plasmids encoding PIP5KIα i2-mVenus or mVenus control, and sorted for mVenus fluoresence. Data is for a single independent experiment. Error bars indicate standard deviations. n =3. Statisitcal analysis performed using two-way t-test. *: p <0.05; ***: p<0.001.

3.7. Summary and discussion

In summary, we have successfully applied CRISPR CAS9 nickase genome editing to target *PIP5K1A* in the human metastatic prostate cancer cell line LNCaP C4-2, leading to generation of two sublines, termed ΔN and KO.

We used PCR genotyping and sequence analysis to characterise the nature of indels in the edited alleles and confirmed that ΔN has genomic deletions in the canonical Exon 1 of *PIP5K1A*, spanning the ATG translation initiation codon. We provided strong evidence from western blots, cloned cDNA expression vectors, RNA-Seq and bioinformatic analyses that the ΔN subline expresses an N-terminally truncated form of PIP5KIα, most probably lacking the first 35 residues and expressed at a reduced level compared to the WT parental line. A second line with greatly reduced expression of functional PIP5KIα (KO) was generated by a dual targeting approach. This clone possesses insertions in Exon 1 just 5' to the ATG translation initiation codon of *PIP5K1A*, and at the junction of an internal exon (Exon 6) and its subsequent intron, feasibly disrupting splicing and leading to a reduction in transcript levels. KO expresses greatly reduced levels of PIP5KIα as confirmed by RNA-Seq, western blots, RTqPCR, and immunohistochemical staining. These two clones constitute

novel *in vitro* models to study the molecular functions of PIP5KIα and its role in signalling pathways relevant to prostate cancer.

In studies not reported here, both the KO and ΔN sublines show defects in a number of functions including reduced proliferation in MTT/MTS assays, alterations in cell cycle fitness, and reduced performance in a variety of cell growth assays such as colony formation assays, migration assays (Ryan Morris & DM Heery Pers. Comm). We also detected alterations in phalloidin staining of the ΔN and KO cells compared to WT, consistent with the function of PIP5KIα in actin polymerisation and cell motility (HM Collins & DM Heery Pers Comm). Similar results with regard to cell proliferation were obtained independently for the ΔN cell line by our collaborators in the JL Persson lab [132]. Importantly, while LNCaP C4-2 cells established tumours in nude mice, this was greatly compromised in the C4-2 ΔN cells [132]. In addition to reduced growth, our laboratory also has also evidence from ongoing studies that PI4P, $PI(4,5)P_2$ and $PI(3,4,5)P_3$ levels may be altered in the edited cells as indicated by sensory probes (R Morris & DMH unpublished).

A surprising finding was that deletion of the N-terminal region of PIP5KIα appears to correlate with an increase in activation of AKT, a major downstream effector of PI3K signalling. Studies

are underway to investigate this, although given the low transfection efficiency of C4-2 cells, it has been difficult to recapitulate this effect on AKT using exogenously expressed mutants of PIP5KIα proteins lacking the N-terminus. Viral delivery methods may need to be explored to improve this.

An important focus of this study was to assess the effect of functional depletion of PIP5KIα on gene expression, which we performed by RNA-Seq. To our knowledge, RNA-Seq studies have not previously been applied to study PIP5KIα function, despite several groups providing evidence for a nuclear role of $PI(4,5)P_2$ and PIP5KIα [59; 135]. Using stringent cut-off values we identified 818 and 1675 DEGs in the KO and ΔN lines respectively compared to unedited C4-2 cells. We noted a substantial commonality in the upregulated genes (41%) and downregulated gene (49%) cohorts, although clearly more genes were altered in the ΔN cells, possibly linked to the super-activated AKT pathway. Gene set enrichment analyses highlighted upregulated genes linked to the cell cycle including G2M checkpoints, E2F target genes and mitotic spindle, consistent with the proliferation and cell cycle defects observed. We also noted enrichment of steroid receptor, K-Ras signalling and p53 regulated genes. Bioinformatic studies currently underway indicate a strong representation of androgen regulated genes in the DEGs, supporting the evidence for a

functional intersection of PI(4,5)P_2 signalling and AR signalling (R Morris & DMH unpublished). This was validated here by preliminary evidence showing that the expression of androgen regulated genes such as *KLK3* (androgen activated) and *UGT2B17* (androgen repressed) is altered in the edited cell lines. Moreover the expression defects can be restored by rescue with an mVenus(GFP) tagged PIP5KIα protein. This provides a model system in which we can perform mutagenesis and functional studies on exogenously expressed isoforms of PIP5KIα to explore the functions of isoform specific domains or, for example the requirement of the lipid kinase activity for restoration of androgen signalling. It remains to be established whether the effects we observe are secondary or whether they are due to kinase activity, protein-protein interactions or possible nuclear functions of PIP5KIα proteins.

In subsequent chapters, I will go on to explore novel transcripts of the human *PIP5K1A* gene identified through our RNA-Seq data which potentially expand the sequence diversity of the N- and C-terminal domains of PIP5KIα isoforms. I will also describe preliminary efforts to explore protein-protein interactions of PIP5KIα in yeast two-hybrid assays.

Chapter 4

Characterisation of alternative splice variants of Human *PIP5K1A* gene in metastatic prostate cancer LNCaP C4-2 cells

4.1. Introduction

As described in detail in **Section 1.8**, multiple mRNA transcripts may be generated from any given gene by using alternative promoters, alternative termination sites and alternative splicing of exonic sequences within the primary transcripts. This latter process (alternative splicing) is thought to occur for 95% of human genes that contain multiple exons [144]. Alternative splicing increases the diversity of the set of encoded amino acid sequences (protein isoforms) from a given gene, and thus expands the protein functions available to cells.

Transcripts of *PIP5K1A, PIP5K1B and PIP5K1C*, which encode the three isozymes of the members of the Type I PIP5K subfamily (PIP5KIα, PIP5KIβ and PIP5KIγ, respectively), are known to undergo alternative splicing, with multiple variant transcripts identified for each in both human and murine cells [145; 146].

The existence of at least five PIP5KIγ isoforms, encoded by splice variants of PIP5K1C transcripts, has been confirmed at the protein level by Western blot. These isoforms vary from each other another in their C-terminal amino acid sequences as a result of complex alternate splicing of final exons of PIP5K1C, conferring each isoform with distinct subcellular localisations and cellular functions [69].

Three splice variants of the human *PIP5K1A* gene were identified and cloned more than 20 years ago by Loijens and Anderson [145]. Two additional splice variant were later listed in the NCBI gene database [147]. While three splice variants of *PIP5K1A* have previously been isolated and cloned, there is currently only one study investigating any functional differences between the PIP5KIα isoforms [148].

In the course of our characterisation of *PIP5K1A* CRISPR CAS9 gene editing (**Chapter 3**) in a prostate cancer cell line, LNCaP C4-2, total mRNA sequencing of derived edited lines was performed. Analysis of this data revealed a small number of reads mapping to genomic positions upstream of the canonical *PIP5K1A* Exon 1 (note: canonical in the sense that the sequence of this Exon matches the 5' sequence of *PIP5K1A* transcripts reported by Loijens and Anderson [145]. The 3' portions of these intron-spanning reads mapped to *PIP5K1A* Exon 2, suggesting that the 5' portions of these reads represent alternate exons to the canonical Exon 1. Further, analysis of the NCBI list of predicted transcripts of *PIP5K1A*, and the aggregate intron-spanning read data upon which these predicted transcripts are based, indicated the existence of these additional putative 5' exons, as well as yet more alternate exons of *PIP5K1A* between the canonical Exon 1 and Exon 2. These later putative exons were not detected in our own RNA-Seq data set.

Within both RNA-Seq data sets, it was also observed that a substantial portion of intron-spanning reads exiting span from the 3' end of Exon 12 to the 5' end of Exon 15, i.e. these reads likely represent transcripts in which Exon 13 and Exon 14 are spliced out by Exon skipping type alternative splicing. Deduced translations of *PIP5K1A* transcripts containing this Exon 12-Exon 15 splice event indicate an altered C-terminal sequence as a result of a frame shift of the terminal exons, Exon 15 and Exon 16.

The C-terminus of two of the canonical isoforms of PIP5KIα (NCBI annotated isoforms 1 and 2) have previously been reported to be essential for their nuclear speckle localisation and complex formation with the poly(A) polymerase Star-PAP and the $PI(4,5)P_2$ activated protein kinase CKIα. This activity reportedly regulates mRNA levels of a sub-set of genes involved in oxidative stress response [149]. The C-terminus of PIP5KIα is further implicated in nuclear localisation by the finding that Lysine 490 of PIP5KIα isoform 4 (equivalent to Lysine 539 of isoform 2 and Lysine 551 of isoform 1) is sumoylated, facilitating nucleolar localisation and interaction with H3K9me3 and heterochromatin protein 1α [122].

In-light of these observations of alternative splicing events in *PIP5K1A* transcripts, the reported functional differences

between c-terminal variants of PIP5KIγ, and the reported role of the C-terminus of canonical PIP5KIα isoforms in nuclear localisation, I set out to:

1) Determine existence of alternate exons and alternate splicing of *PIP5K1A* transcripts in LNCaP C4-2 cells.
    - As experimental work for this began in February/March of 2020, the first nationwide lockdown due to the spread of SARS-CoV2 meant our laboratory was not accessible for six months. This time was therefore spent analysing the evolutionary conservation of these alternate exons and alternate transcripts of *PIP5K1A*, to add to experimental validation of their existence in human cell lines.
2) Study subcellular localisation of YFP Venus-tagged PIP5KIα isoforms to explore potential functions of differential C-terminal sequences.

4.2. Canonical splice variants of human *PIP5K1A*

Three variant transcripts of the *PIP5K1A* gene, *PIP5K1A*_v1, *PIP5K1A*_v2 and *PIP5K1A*_v4 (aka NM_001135638.2, NM_003557.3 and NM_001135637.2, respectively) were previously characterised from a human foetal brain cDNA library by Lojens and Anderson [145]. The sequences of these transcripts were later mapped to an approximately 50 kb region at Chromosome 1q21.3 [150]. Analysis of these three reported transcript sequences, in light of the genomic mapping, indicates the existence of 16 exons, with the variations in transcript sequences arising from alternative splicing by variable inclusion of the third, tenth or thirteenth exons (Exon 3, 10, 13). A splicing map for the above reported "canonical" transcripts of *PIP5K1A* is presented in **Figure 22**.

PIP5K1A Exon 3 encodes a 12-residue sequence (^{41}VLEARQDSYISL^{52} of PIP5KIα i1) within the divergent N-terminal domain (i.e. divergent between PIP5KIα, PIP5KIβ and PIP5KIγ isoforms). BLASTn searches with the CDS for the N-terminus of Human PIP5KIα i1 (i.e. with the Exon 3 sequence included) indicate that Exon 3-like sequence is only included in PIP5K1A transcripts of primates. No species representative of any other order of mammals possess Exon 3-like sequence in their PIP5KIα transcripts.

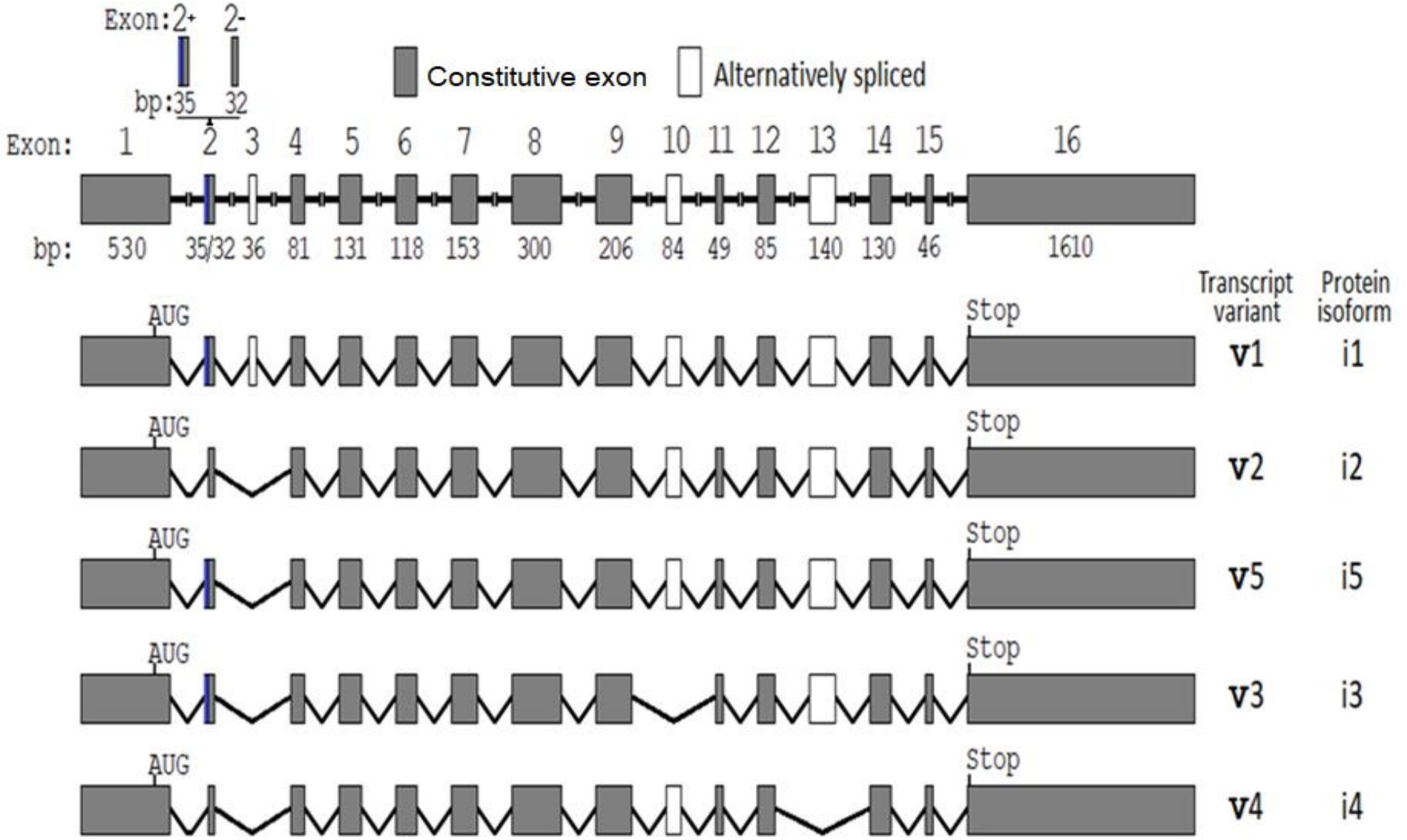

Exon: 2+ 2-
bp: 35 32
Constitutive exon
Alternatively spliced
Exon: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
bp: 530 35/32 36 81 131 118 153 300 206 84 49 85 140 130 46 1610
Transcript variant
Protein isoform
AUG
Stop
v1
i1
v2
i2
v5
i5
v3
i3
v4
i4

Figure 22. Reported splice variants of *PIP5K1A*.
Splice maps of the five annotated *PIP5K1A* splice variants listed in the NCBI database. Transcripts are designated with their NCBI variant and isoform number (v and I, respectively). v1 = NM_001135638.2; v2 = NM_003557.3 ; v3 = NM_001135636.2 ; v4 = NM_001135637.2; v5= NM_001330689.2. Constitutive exons are shown in grey, alternatively spliced exons in white.

Exon 13 encodes a 50-residue sequence (455-504 of PIP5KIα i1) within the isozyme divergent C-terminal domain. Any function attributable to these residues, or the functional significance of their inclusion/exclusion is unknown. Interestingly this sequence appears to be targeted at multiple residues for phosphorylation [86], including a site, Tyr 483, which has been detected as phosphorylated in 132 high through-put studies, and is within a sequence predicted for phosphorylation by insulin receptor (SCITYQPSV; [151])

A fourth splice variant, *PIP5K1A*_v3 (NM_001135636.2) is reported in the NCBI gene database. This variant lacks the 84 nucleotide sequence from *PIP5K1A* Exon 10, which encodes 28 residues (382-410) spanning a portion of the C-lobe and activation loop of the kinase domain. Despite the NM_ prefix indicating manual curation, the existence of *PIP5K1A*_v3 is supported by no published reports, nor by the NCBI aggregate intron-spanning read data, since the later contains no reads representing a *PIP5K1A* Exon 9 to Exon 11 splice event.

A fifth variant, *PIP5K1A*_v5 (NM_001330689.2), is almost identical to *PIP5K1A*_v2 (NM_003557.3) except for the presence in *PIP5K1A*_v5 of an additional CAG, codon at splice junction of the Exon 1 & 2. As a result, the deduced amino acid sequences of the encoded isoforms (PIP5KIα i2 and PIP5KIα i5) vary by one additional alanine (encoded by a GCA codon; the G contributed from Exon 1, the CA from the CAG inclusion) at residue 29. This variation in human *PIP5K1A* transcripts has previously been identified in a study of the occurrence of NAGNAG tandem splice-acceptors in the human transcriptome [152], with the presence of a CAGCAG tandem acceptor splice site at the 5' end of Exon 2 appearing to create alternate splice-acceptor choice during splicing (**Figure 23**). Two variants of Exon 2 are thus distinguished in this text as 'Exon 2+' (CAG present, 35 nucleotides), and 'Exon 2-' (CAG absent, 32 nucleotides).

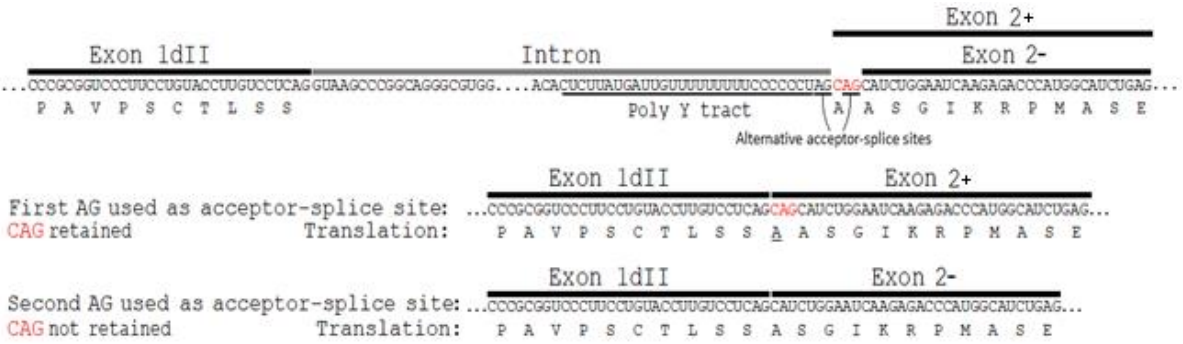
Within our RNA-Seq data set for WT LNCaP C4-2, of those reads joining *PIP5K1A* Exon 2 to the canonical Exon 1, reads with the additional CAG included (joining Exon 2+) vs reads with the CAG absent (joining to Exon 2-) occur at a ratio of 0.41:1 (reads joining Exon 2+ $\bar{x}$= 23.33; reads CAG absent $\bar{x}$ = 56.67; n = 3). In the NCBI database of intron spanning reads, reads from Exon 1d joining Exon 2+ =1693 read, vs reads from Exon 1d joining Exon 2- = 3549, a ratio of 0.48:1. These data

both suggest bias of resolution of splice events involving Exon 2 towards exclusion of the CAG sequence.

Consistent with this, sequence context of the second AG splice-acceptor is a closer match to consensus $(Y)_{20}$NCAGgt for 3' acceptor-splice sites of metazoan introns (**Figure 23 B**).

The evolutionary conservation of this phenomenon in *PIP5K1A* transcripts was briefly explored among several diverse orders of mammals. As described in more detail in **Section 4.11**, it was found that there is conservation of both the genomic sequence at the equivalent of *PIP5K1A* Exon 2 and existence ofscripts that differ in inclusion of a CAG sequence at this position, supporting the model that the sequence at the Exon 2 border functions as a tandem splice-acceptor. Any functional effect of variable inclusion of this CAG in *PIP5K1A* transcripts, or of corresponding variable inclusion of the Alanine residue in the encoded protein isoforms, remains unknown.

A)



B)



Figure 23. Alternative splice acceptor sites at the 5' end of Exon 2 of *PIP5K1A*

A) Putative mechanism for the observed variable inclusion of CAG trinucleotide sequence at the Exon 1 - Exon 2 boundary in *PIP5K1A* transcripts, e.g. in *PIP5K1A*_v2 vs *PIP5K1A*_v5. The AGCAG sequence at the 5' end of Exon 2 in *PIP5K1A* preRNAs contains two AG dinucleotides, both of which can function as splice-acceptor. Two sequences can therefore result after splicing, one containing an additional CAG nucleotide sequence, which leads to inclusion of an additional alanine in the deduced amino acid sequence.

B) Comparisons of consensus splice site sequences (top row) to the alternative splice sites at the 5' end of *PIP5K1A* Exon 2 (bottom row).

4.3. Identification of alternate 5' exons of *PIP5K1A* using RNA-Seq data sets

As part of our investigations into gene expression changes in response to *PIP5K1A* gene disruption, RNA sequencing was performed with total RNA extracted from the human prostate cancer cell line LNCaP C4-2 (**Chapter 3**). In addition to reads mapping to the 16 canonical exons of *PIP5K1A*, reads were detected with sequences mapping to three genomic positions between 400-5000 bp 5' to the canonical *PIP5K1A* Exon 1 (**Figure 24 A**). Intron spanning reads adjoined these to one another, and to *PIP5K1A* Exon 2, skipping the canonical Exon 1. This indicated that the 5' positions to which these reads map represent exons of *PIP5K1A,* spliced together in some *PIP5K1A* transcripts*,* and used alternatively to the *PIP5K1A* canonical Exon 1. These putative exons of *PIP5K1A* are henceforth referred to in this text as Exons '1a', '1b' and '1c', with the canonical Exon 1 referred to as '1d'. Exon 1d was later determined to contain a rarely excluded cryptic intron (**Section 4.8** and **Section 4.9**), and so henceforth the term '1di' is used for the 5' 236 bp portion of this exon, and 'Cryptic Intron 1dii' or '1dii' used for the 3' 294 bp portion that acts as a cryptic intron, with Exon '1d' referring to the entire 530 bp exon.

Genomic positions and sequences of these, and other putative exons of *PIP5K1A* identified in this study are listed in **Table 6**.

The NCBI Gene database of confirmed and predicted human *PIP5K1A* transcripts (NCBI Homo sapiens Annotation release 109.20201120, Gene ID: 8394) was next consulted, looking at predicted transcripts of *PIP5K1A* for inclusion of exons that are not included in published/annotated *PIP5K1A* transcripts. This database includes transcripts (those with XM_ prefixes) that are computationally predicted based upon aggregate RNA-Seq intron spanning read data. Several such predicted transcripts (XM_024450103.1, XM_006711568.4, XM_011510043.3 and XM_024450129.1) contain the putative Exon 1c that had been identified within our LNCaP C4-2 RNA-Seq dataset. No such predicted transcripts containing the putative Exons 1a or 1b were present, however a number of intron spanning reads adjoining the genomic position of putative Exons 1a to 1b (10 reads), and 1b to 1c (26 reads) are recorded (**Figure 24 B**), suggesting existence of *PIP5K1A* transcripts containing these exons.

Comparison of numbers of intron spanning reads for Exon 1c-Exon 2 vs Exon 1d-Exon 2 in our LNCaP C4-2 data indicates that the former splice event occurs far more rarely than the latter, with a ratio of 0.092 to 1 (reads spanning Exon 1c- Exon

2: $\bar{x}$ = 7.33; reads spanning Exon 1dii-2: $\bar{x}$ =80.00; n =3). The ratio of Exon 1c-Exon 2 vs Exon 1d-Exon 2 is lower in the NCBI intron spanning read data: reads spanning Exon 1c-Exon 2 = 36; reads spanning Exon 1dii-Exon 2 = 5242, a ratio of 0.007 to 1.

Within the NCBI database, four other putative exons contiguous with *PIP5K1A* Exon 2 were present. These putative exons are henceforth referred to as Exons '1e', '1f', '1g' and '1h' (**Figure 24 C**). Similarly to Exon 2, which has alternate splice-acceptor sites at its 5' end, Exon 1h appears to have alternate splice donor sites at its 3' end: there are two distinct types of reads exiting Exon 1h, those in which an additional 9 bp (GTAAGAGAG) mapping to chromosome 1 151215992-151216000 are present, and those in which this 9 bp sequence is absent. The 9 bp longer exon represented in the former reads is henceforth referred to as Exon '1h+', and the shorter, represented by the later reads, as Exon '1h-'. When referred to in total without distinguishing type, '1h' is used.

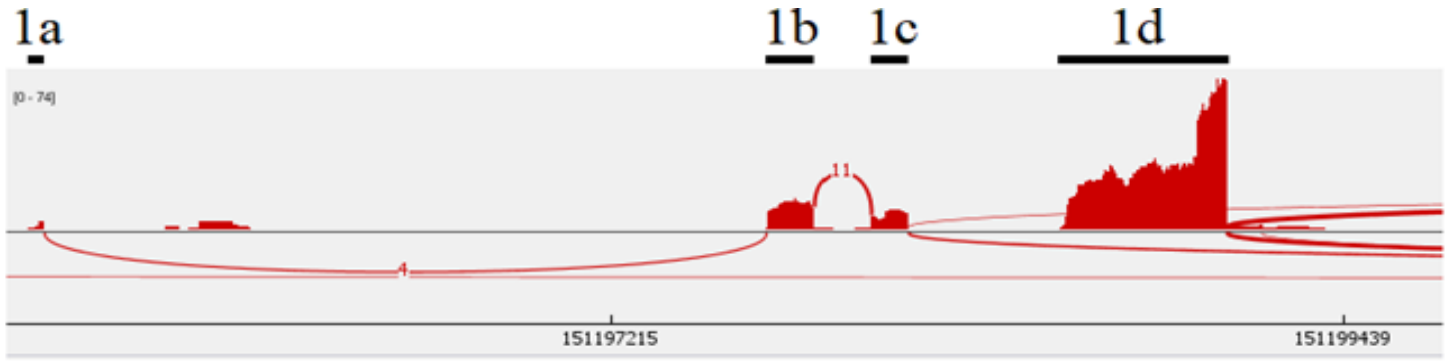
Intron spanning reads with sequences mapping to the positions of Exons 1e, 1f, 1g or 1h were not detected in our own RNA-Seq data for LNCaP C-42. Within the NCBI database of intron spanning reads, reads joining these putative exons to *PIP5K1A* Exon 2 are rare compared to those joining the

canonical Exon 1d to Exon 2. Reads joining Exon 2 to: Exon 1c = 36 (0.64 % of reads entering Exon 2); Exon 1d = 5242 (92.68 %); Exon 1e = 227 (4.01 %); Exon 1f = 23 (0.40 %); Exon 1g = 76 (1.3 %); Exon 1h = 52 (0.92 %).

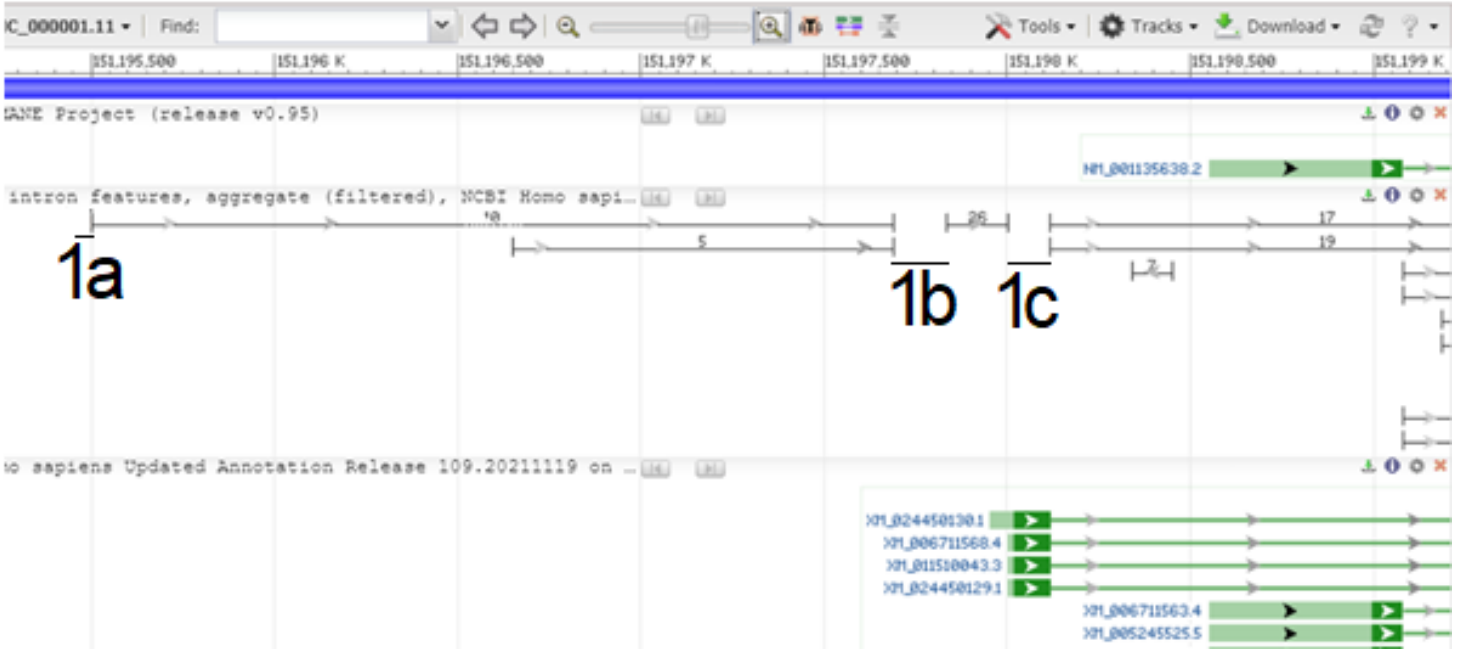
Additionally, there are intron spanning reads connecting Exon 1g to Exon 1h (5 reads, 6.17 % of reads exiting Exon 1g), and reads connecting Exon 1d to Exon 1g (67 reads, 1.25 % of reads exiting Exon 1d) and to Exon 1h (41 reads, 0.77 % of reads exiting Exon 1d). Therefore, it appears that the 5’ ends of both Exons 1g and 1h contain functional splice-acceptor sites, though evidently *PIP5K1A* transcripts in which these are used are far rarer than those transcripts in which Exon 1d is spliced directly to Exon 2.

A summary of the different splicing events involving the alternate 5’ Exons of *PIP5K1A* is presented in **Figure 24 D**.

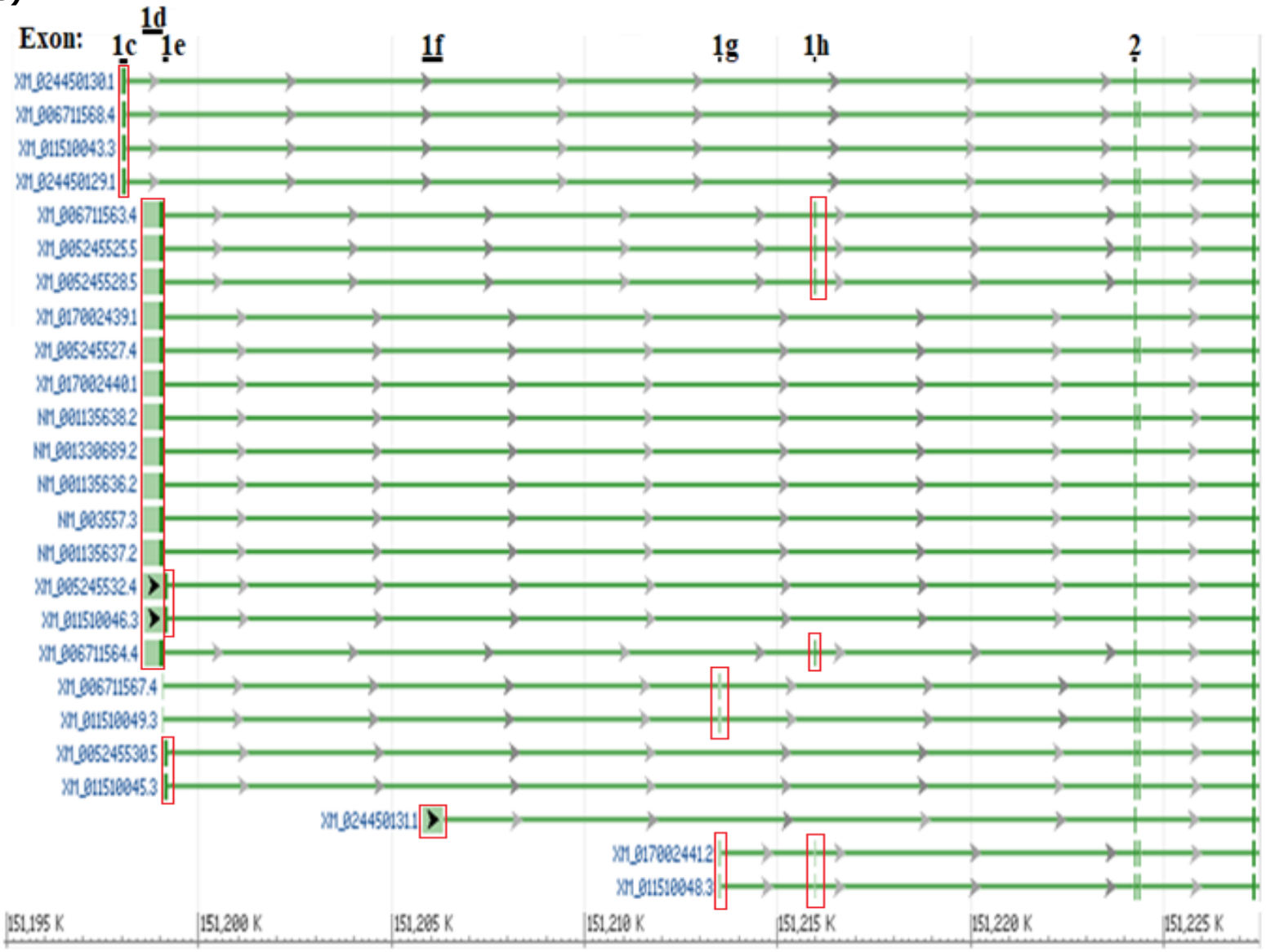
A)


1a
1b
1c
1d
[0 - 74]
11
4
151197215
151199439

B)


Find:
Tools
Tracks
Download
151,195,500
151,196 K
151,196,500
151,197 K
151,197,500
151,198 K
151,198,500
151,199 K
ANE Project (release v0.95)
NM_001135638.2
intron features, aggregate (filtered), NCBI Homo sapi...
1a
1b
1c
o sapiens Updated Annotation Release 109.20211119 on ...
XM_024450130.1
XM_006711568.4
XM_011510043.3
XM_024450129.1
XM_006711563.4
XM_005245525.5

C)


Exon:
1c
1d
1e
1f
1g
1h
2
XM_024450130.1
XM_006711568.4
XM_011510043.3
XM_024450129.1
XM_006711563.4
XM_005245525.5
XM_005245528.5
XM_017002439.1
XM_005245527.4
XM_017002440.1
NM_001135638.2
NM_001330689.2
NM_001135636.2
NM_003557.3
NM_001135637.2
XM_005245532.4
XM_011510046.3
XM_006711564.4
XM_006711567.4
XM_011510049.3
XM_005245530.5
XM_011510045.3
XM_024450131.1
XM_017002441.2
XM_011510048.3
151,195 K
151,200 K
151,205 K
151,210 K
151,215 K
151,220 K
151,225 K

D)

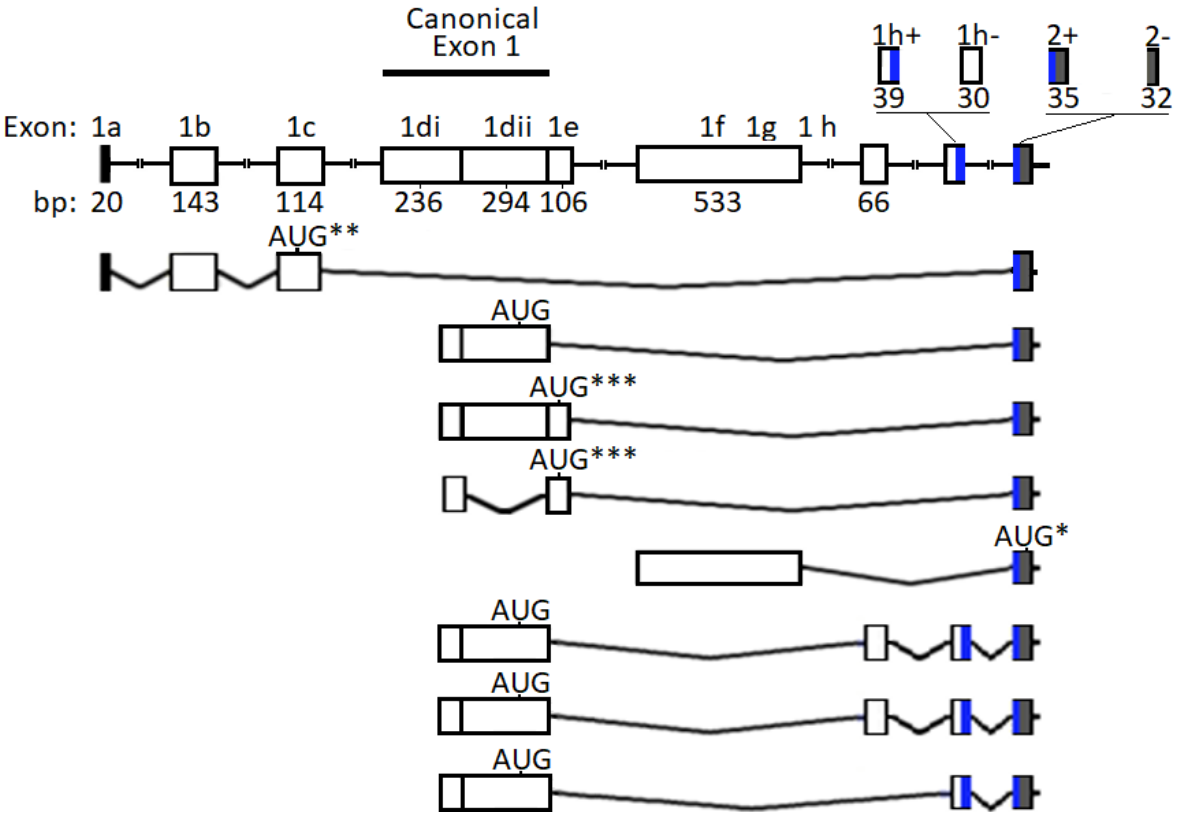


Figure 24. Alternate 5' exons of *PIP5K1A*

A) Sashimi plot of RNA-Seq reads for WT LNCaP C4-2, aligning to genomic region (base positions 151,195,500-151,199,500, Chromosome 1, GRCh38.p13, primary assembly) proximal to *PIP5K1A* Exon 1 for one of three replicates. The coverage for each alignment track is plotted as a bar graph. Arcs represent exon spanning reads, with numbers of such reads indicated.
B) Screen grab of NCBI RNA-Seq intron features proximal to *PIP5K1A* Exon 1d, showing intron spanning reads with sequences mapping three positions 5' to Exon 1d, the same positions as shown in **A**. Bars represent intron spanning (i.e. spliced) reads, with numbers on each bar representing the number of each such read. 17 and 19 reads connect to Exon 2+/Exon 2-, not shown. Image taken from [153].
C) Splicing maps of the 5' ends of annotated (those with NM_ prefix) and predicted (XM_ prefix) transcripts of *PIP5K1A* listed in NCBI. Red boxes highlight the alternate 5' exons included in such transcripts. The bottom track indicates the base position (Chromosome 1, GRCh38.p13, primary assembly).
D) Splicing maps summarising the types of splicing of *PIP5K1A* involving 5' exons either reported in the NCBI database, or

observed in this study. For simplicity, variants that differ only in use of the alternate splice donor is Exon 1h, or splice-acceptor in Exon 2, are not shown. Predicted AUG start codons for such transcripts are indicated.

Table 6. Sequences and genomic positions of putative exons of human *PIP5K1A*.

Genomic positions are according to Chromosome 1 Reference GRCh38.p13 Primary Assembly.

Exon	Sequence	Genomic Position
1a	TGACTCAGAGGACACCCCAG	151195480-151195499
1b+	ATAATTATGGAGCCCCTACTATGTACGAAGTCACTGGAGGAGCTGGAGATACAAGAGTGAACCAAACACAGTCCTTGCTCCCCAGAAGTTTTACAGTCTGCTCAGAGAGACTGGTCTTCCCACCAGGGTTCATAGCCTACGGG	151197691-151197833
1b-	ATAATTATGGAGCCCCTACTAT	151197691-151197712
1c	GTCTTCACTGATTCATGCATCAAATATTTACTGAACTCCGTCACTGTGCAGTGCCTAACTTTGTGCTTGAGGCAAACAAAACACAGATCTATTTCGAATGCCATTCCGAGGAAG	151198004-151198117
1di	GATTTTTTGCTTGGCTACCCGGAGTGAAGCGGCCGGGTTGGGCGATTAACAGGCCGTGGTTAGGAAGGACGGAGAAGGGGCGTTCGCTCCTTTGGGACTTTTCATGCCTCGTTTTTTTTTCAGATGTGGCTTGGTCTGGGCGCAAGGTCCCAGCAGCCAGCTTAAGCTTACTCTTCTGTGAAAGGGGAAAGTATCCCCTGTGGAAAGCGGTTAAACTTGTGGAGGGGGTGCGGGAC	151198552-151198787
1dii (a cryptic intron)	GTGAGTTCTTCCCCATGCCAGGCGAATGGTGTGGCCTTGAGCTGGTCCAGGAGCCGGCTCGACGTGTCTGAGGGAGGCCCCGGAGGGGGCGGGGAGGTGGCCCACAGAACGCGGGTTCTGTAAAGAGACGTTGGGAAGATTCGATTCCGAGAAGAGGAAGAACCGGATTGAAAGAGAGCCAGGCCGCTGAGGGGGAGGGGGCTGCTAAGATGGCGTCGGCCTCCTCCGGGCCGTCGTCTTCGGTCGGTTTTTCATCCTTTGATCCCGCGGTCCCTTCCTGTACCTTGTCCTCAG	151198788-151199081
1e	GTAAGCCCGGCAGGGCGTGGGTAGGGAGCTGGTGAGGAATATGCGATGGGAGGAAGAGCGAGACCTAAGAGGGTACCTGTAGCTGGGAATGTCCTACGTGTTACCG	151199082-151199187

1f	CAAAAAAAAGAAAAGGGAACAAGAAAACTTATCAG TTTATTTTTCTTTCTTCTTTTTACTGGGTTTCCTT GATCAAGGACAAAAGTGTACTTCAAGAATATACTC TTTCCTGATGGCCAGGTAGAGAGGGAATCTTTGTT TTTCTGTGCTCCATACCCAGCCCAGAGAGACTGAA TGAAACACCGATGGTGTTTCCTAGAGGAAACCCTA AGAAGCCCGTTTCCCCTATTGCACCTCACTTCTCT GCCCTTTCCCATCTAAACAGAGAATTTAATTTCGA CATATGTTTATCCTGTTTTGTTCATATCTCCTACA AGCTTTTATTCTTTTACTGTTTTCTATTTGAAGTG AACTTTTTTCCCTCAACCCCTCCATCTTTTTAAAC TGTCTCCATTTTTCTTTATTCTCCTGTGAAATTTC AGTCTGTTTATTTGTAATGGTTGTATTCAGTATGT ATTAATTTACAAATATCTAATTCCTGGAAAATCTC TCTTGTTGGAATCTGAGGAAGACATTGGCAAAAAG AGAGGAAG	151205797- 151206329
1g	GAGAGTGTTGATTTCTGAGGGGTATGGTTCCCTTT TTAAGAGAGAATCTCTTATATGTGCTGATACAG	151213480- 151213547
1h+	GTTTCAACTTTCTGTGCTGTGGGGTCAGTGGTAAG AGAG	151215962- 151216000
1h-	GTTTCAACTTTCTGTGCTGTGGGGTCAGTG	151215962- 151215991

4.4. Predicted N-terminal amino acid sequences encoded by alternate 5' exons and Kozak sequence analysis

The canonical AUG start codon of *PIP5K1A* is contained in Exon 1dii. When retained as an exon this encodes 28 amino acid residues (^{1}MASASSGPSSSVGFSSFDPAVPSCTLSS^{28}), and contributes a G nucleotide to the first codon of the subsequent exon.

Exclusion of the canonical AUG and the subsequent 27 codons is expected for the numerous predicted transcripts of *PIP5K1A* that exclude Exon 1d, either in its entirety, or partially by splicing out of 1dii. Nucleotide sequences of these transcripts have feasible alternate AUG start codons: AUG codons that are in the same frame as the PIP5KIα coding sequence, and without intervening stop codons.

The sequence context – 'Kozak sequences' – around an AUG start codon affect the probability that a scanning ribosome will recognise the AUG and initiate translation [154]. The Kozak sequences around each putative AUG start codon of the alternate *PIP5K1A* transcripts were analysed by reference to Hernández et al., [155]. The Kozak sequences were found to range from 'weak' (for Exon 1c), 'adequate' (Exon 2), to 'strong' (Exon 1e). The canonical *PIP5K1A* start codon, in Exon 1dii, was found to be 'strong' (**Table 7**).

Assuming that each of these AUG codons can initiate translation, there are several alternate N-terminal amino acid sequences for the different transcripts of *PIP5K1A*, falling into four categories:

- Transcripts, such as NCBI *PIP5K1A*_x3, x4, x8 and x9, which include Exon 1c (as shown in **Sections 4.8 and 4.9**, such transcripts likely also use Exons 1a and 1b) instead of Exon 1d. Exon 1c contains a feasible alternate AUG start codon. If translation is initiated at this AUG, the resulting protein would have a substitution of residues 1-28 of canonical PIP5KIα with 33 residues: ^{1}MHQIFTELRHCAVPNFVLEANKTQIYFECHSEE^{33}.
- Transcripts, such NCBI *PIP5K1A*_x10, x11, x12 and x13, which use Exon 1e. Exon 1e contains a feasible alternate AUG start codon. The deduced amino acid sequence for such transcripts has substitution of residues 1-28 of canonical PIP5KIα with 20 residues: ^{1}MGGRARPKRVPVAGNVLRVT^{20}.
- Transcripts, such as *PIP5K1A*_x14, which lack Exon 1d and instead include Exon 1f. The first AUG in such transcripts that is in-frame with the PIP5KIα ORF is in Exon 2, equivalent to the codon M$^{35/36}$ for canonical transcripts.

- Transcripts, such as NCBI *PIP5K1A*_x1, x2, x5 and x7, which include exon 1d, but with the additional inclusion of Exon 1h. These gain an additional 13 (1h+) or 10 (1h-) codons. Following translation of such transcripts, encoded isoforms are predicted to gain the following sequences ^{29}GFNFLCCGVSGKR^{41} (if 1h+ is included) or ^{29}GFNFLCCGVS^{38} (if 1h- is included).

Clustal Omega multiple sequence alignments were performed for the deduced N-terminal amino acid sequences putatively encoded by transcripts with alternate 5' exon use. The alignment indicated conservation of residues encoded by Exon 1h+ (^{30}FNFLCCGVSGKR^{41}) with residues putatively encoded by Exon 1c and 1e, such as in the transcripts x3/x8 and x10/x12 (**Figure 25**).

The N-termini and C-termini (i.e. the portions of each protein outside of the conserved kinase core) of the PIP5K1s are divergent between isozymes, but conserved amongst isoforms between species, leading to the suggestion that the termini each have evolutionarily conserved functions that effect different properties of each isozyme [69; 83; 156].

Conservation of amino acid sequences encoded by different alternative splice variants of the same gene could indicate convergent evolution of the different exonic sequences, such

that no matter which splice variant is generated, the encoded isoforms have similar functions. In this model, the different exons could have arisen by mutation of intronic sequence to match consensus splice sites, and not been selected against as they are neutral mutations.

Without knowing the precise function of the canonical PIP5KIα N-terminus (e.g. in specific protein-protein interactions), it is difficult to assess impacts of alternate N-terminal sequences.

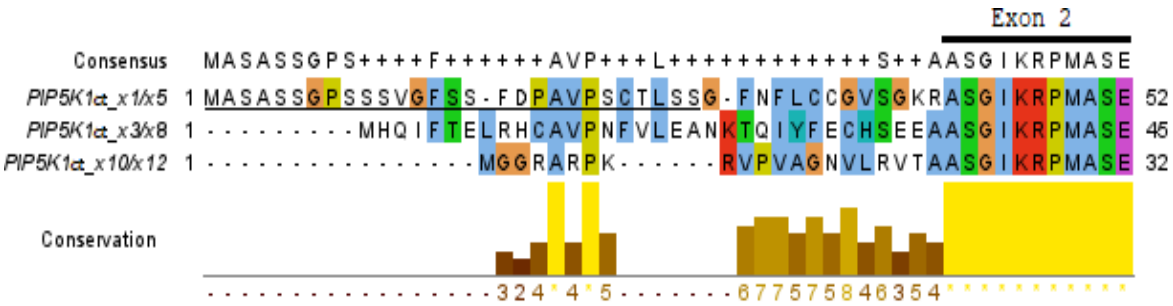


Figure 25. Sequence alignment of deduced amino acid sequences for predicted alternate transcripts of *PIPI5K1A*.
Deduced amino acid sequences encoded by predicted transcripts of *PIP5K1A* were aligned by Clustal Omega sequence alignment. Amino acid grouping colours and conservation scores performed in Jalview.

Table 7. Comparison of Kozak motifs to context of in-frame AUG start codons for *PIP5K1A* transcripts.
IUPAC single letter nucleotide codes are used (R = A/G; Y = C/U; H = A/C/U). Start codons are indicated in bold. Positions identical to the consensus Kozak motif are underlined. Classifications are made according to those proposed by Hernández et al., [155]

Context	PIP5KIα isoforms (NCBI)	Sequence	Classification
Consensus Kozak/'Optimal' motif		GCCGCCRCC**AUG**G	Optimal
'Strong' motif		NNNNNN*R*NN**AUG**G	Strong
'Adequate' motif		NNNNNN*Y*NN**AUG**G or NNNNNN*R*NN**AUG**H	Adequate
'Weak' motif		NNNNNN*Y*NN**AUG**H	Weak
Exon 1dii AUG (Canonical Met 1)	i1-i5,x1,x2,x5-x7,x15,x16	GCUGCUAAG**AUG**G	Strong
Exon 2 AUG (canonical Met 36)	x14,x17	AAGAGACCC**AUG**G	Adequate
Exon 1e in-frame AUG	x10-x13	GAAUAUGCG**AUG**G	Strong
Exon 1c in-frame AUG	x3,x4,x8,x9	CACUGAUUC**AUG**C	Weak

4.5. Evolutionary conservation of putative alternate PIP5KIα N-terminal sequences

Evolutionary Conservation is the presence of similar genes, portions of genes, chromosomal segments or encoded amino acid sequences amongst different species. Evolutionary Conservation is often analysed because conservation of a particular sequence amongst species in different branches of a phylogenetic tree typically indicates that sequence has been maintained despite the tendency of genomes to acquire mutations with successive generations, and hence indicates that sequence has been maintained by natural selection. The conservation of genomic sequence and expression of *PIP5K1A* transcripts was investigated for further evidence of the existence alternate 5' exons of *PIP5K1A*.

4.5.1. Use of Exon 1h

BLASTn searches were performed with the 5' sequences of Human predicted transcripts *PIP5KIA*_x1/x2/x5/x7 (those that contain the sequence of Exon 1h). These searches returned putative orthologues from species of many mammalian orders. Inspection of the identified transcript sequences indicated their similarity to the predicted human *PIP5K1A* transcripts resulting from Exon 1d - Exon 1h - Exon 2 splicing events. The deduced amino acid sequences encoded by these transcripts possessed

high conservation of residues encoded by Exon 1h (**Figure 26**). Species for which such transcripts were identified by BLASTn searches come from several orders of the Boreoeutherian magnorder of mammals: Primates, Rodentia, Chiroptera, Artiodactyla and Carnivora. A single representative of the Atlantogenatae was present in the list, *Trichechus manatus latirostris* (Florida manatee), an Afrotherian. However, manual searching for the six Atlantogenata species for which predicted *PIP5K1A* transcripts are available in NCBI also uncovered transcripts containing an Exon 1h like sequence in another Aftrotherian species: *Loxodonta africana* (African bush elephant): transcripts XM_023547328.1, XM_023547327.1, XM_023547324.1 and XM_023547323.1; each contains an exon approximately the same position of Exon 1h relative to Exon 2, with deduced amino acid sequence FNFLCCGVSRKR. Likewise, Exon 1h like exon sequences were found in available predicted *PIP5K1A* transcripts for the several species of Marsupial: *Dromiciops gliroides* (Monito del monte; transcript XM_044003573.1), *Gracilinanus agilis* (Agile Gracile Opossum; transcript XM_044672680.1) and *Monodelphis domestica* (Gray short-tailed opossum; transcript XM_016428845.1). A similar search was performed for the single monotreme species for which *PIP5K1A* transcripts are available in the NCBI database,

Ornithorhynchus anatinus (platypus), but no transcript containing an Exon 1h like sequence was found.

Presence of Exon 1h like sequences in *PIP5K1A* transcripts for species representing placental mammals and marsupials but not of monotremes seemingly dates the emergence Exon 1h usage to after the divergence of therian and monotreme lineages, ~187 MYA [157] but before the divergence of marsupials and placental mammals ~136 MYA [158].

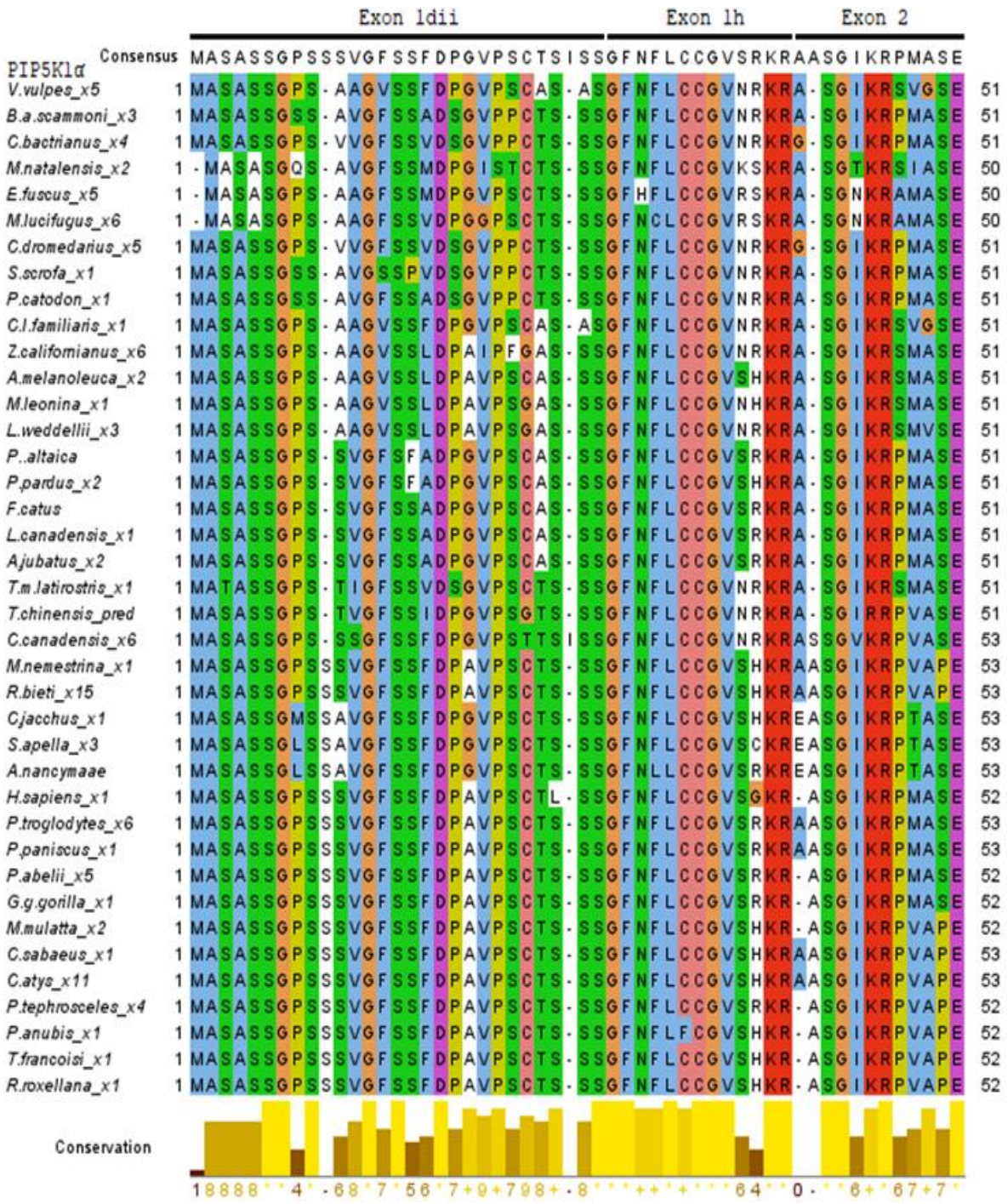


Figure 26. Conservation of Exon 1h like encoded sequences.

Sequence alignments of N-terminal deduced amino acid sequences of the putative orthologues of human *PIP5K1A*_x1, x2, x5 & x7, which contain Exon 1h, identified by BLASTn and BLASTx searches.

4.5.2. Use of Exons 1a, 1b and 1c

BLASTn searches were performed with the unique 5' end of human PIP5KIA_x3/x4/x8/x9 (which contain Exon 1c sequence) predicted transcripts, both with and without the additional sequences from Exons 1a and 1b. These BLASTn searches returned orthologous *PIP5K1A* transcripts only from *H. sapiens* and one other species: *Macaca nemestrina* (the pig-tailed macaque*;* an Old World monkey). Interestingly this *M. nemestrina PIP5K1A* transcript (*M. nemestrina PIP5K1A*_x10; XM_011769314.1) includes sequences aligning to both human *PIP5K1A* Exon 1b (alignment of a 24 nucleotide sequence) and Exon 1c (alignment spanning the length of human Exon 1c, with absence of 13 nucleotide sequence from the centre of the exon), contiguous with sequences aligning to human *PIP5K1A* Exon 1h and then Exon 2. Analysis of the *M. nemestrina* PIP5K1A gene shows these sequences occupy approximately the same positions relative to the equivalent of *PIP5K1A* Exon 1d, as they do in the human *PIP5K1A* gene sequence. This suggests that Exon 1b - Exon 1c splicing is conserved at-least between *Cercopithecoidea* and *Hominoidea* superfamilies (LCA ~23 MYA; [159]). Despite this, *M. nemestrina* appears to be the only species for which Exon 1b - Exon 1c splicing events are reported in the NCBI database of intron-spanning reads, other than *Homo sapiens*.

In-silico translation of the *M. nemestina* PIP5K1A_x10 transcript shows that the equivalent of the Exon 1c predicted AUG start codon is not in frame with the PIP5KIα ORF, due to the absence (with respect to the human sequence) of 13 nucleotides from the centre of Exon 1c. The first available AUG in the PIP5KIα reading frame in this transcript is equivalent to codon M^{72} of *M. nemestrina* PIP5KIα x1 isoform, equivalent to M^{46} of human PIP5KIα i2.

BLASTx searches with the N-terminal deduced amino acid sequences of Human PIP5KIα_x3/x4/x8/x9 identified a predicted PIP5KIα isoform from *Pan troglodytes* (the common chimpanzee; Sequence ID: PNI22897.1). The N-terminal deduced amino acid sequence had near exact match with the N-terminus of human PIP5KIα_x3/x4/x8/x9 (**Figure 27 A**).

Additional BLASTn hits were for chromosomal sequences proximal to orthologue PIP5K1A genes for two primate species: *Pongo abelii* (the Sumatran orangutan; a great ape; matches to both Exon 1b and Exon 1c detected- <90% matched nucleotides) and *Callithrix jacchus* (the common marmoset; a New World monkey; Exon 1b only detected- 84% matched nucleotides). The detection of similar nucleotide and deduced amino acid sequences in the great apes *P. abelii* and *P.troglodytes*, respectively, and of both nucleotide sequences

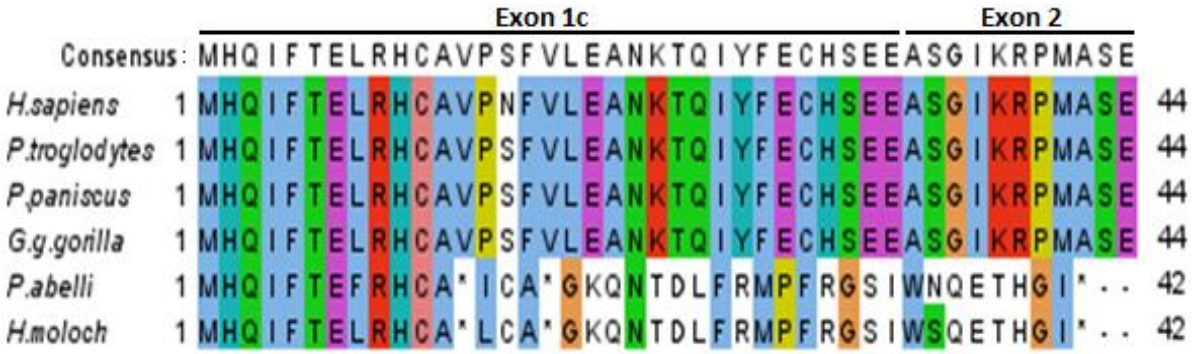
and splicing events in *P.troglodytes* and the Old World monkey *M. nemestrina*, made the lack of such findings for other apes conspicuous. Chromosome sequences proximal to the PIP5K1A canonical Exon 1 of several hominoid species were therefore inspected for all the great apes for which genomic sequences are available, as well as *Hylobates moloch* (*the* silvery gibbon) a member of the sister Hylobatidae family. Though they had not been identified by BLASTn searches, near exact matches to the human *PIP5K1A* Exon 1a, 1b and 1c sequences were found within these genomes, with similar conformity of intervening sequences to consensus splice sites (**Figure 27 B, C & D**), indicating that these could be functional exons in these species.

Deduced amino acid sequences of these species Exon 1c-like sequences indicate that the predicted AUG start codon in human Exon 1c is in frame with the PIP5KIα ORF only in species of the *Homininae* lineage (*H. sapiens*, *P. troglodytes*, *P. paniscus* and *G. g. gorilla*). This is due to an additional five nucleotides (5' GTGCA 3') 25 nucleotides downstream of the predicted Exon 1c AUG start codon in these species (**Figure 27 D**). The most parsimonious model for this is an insertion of these five nucleotides in a common ancestor of *Gorillini* and *Hominini* lineages, after divergence of *Homininae* and *Ponginae* lineages. No alternate AUG in the same frame as the PIP5KIα

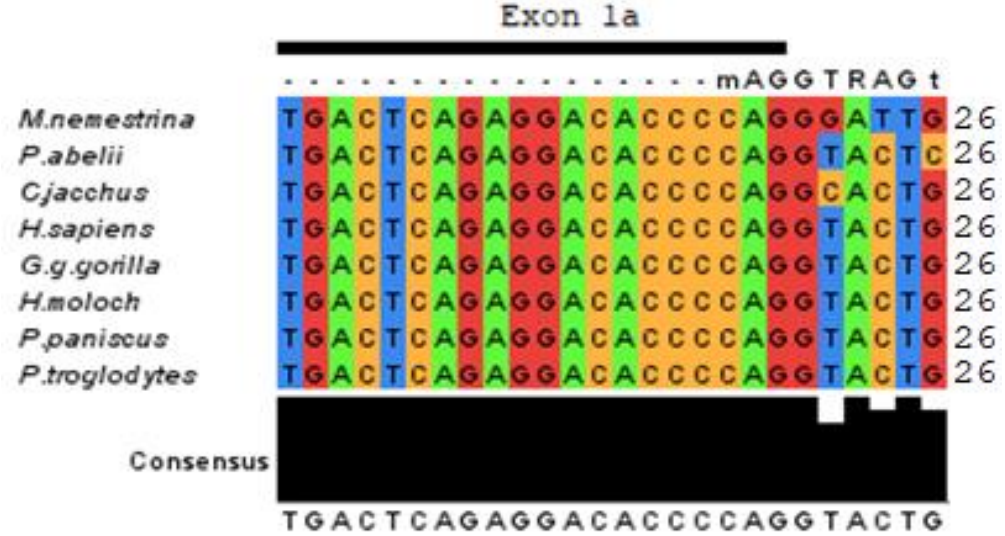
ORF exists in the theoretical Exon 1a-1b-1c sequences for *P. abelii* or *H. moloch* transcripts.

Further evidence for conservation of transcription of, and any evidence of translation of, Exon 1a-1b-1c containing PIP5K1A transcripts in *Homininae* species is necessary to determine whether the uniformity of the five-nucleotide insertion in these species is A) a lineage-specific innovation that has been selected for due to conferring novel function; or B) a result of genetic drift of a neutral mutation.

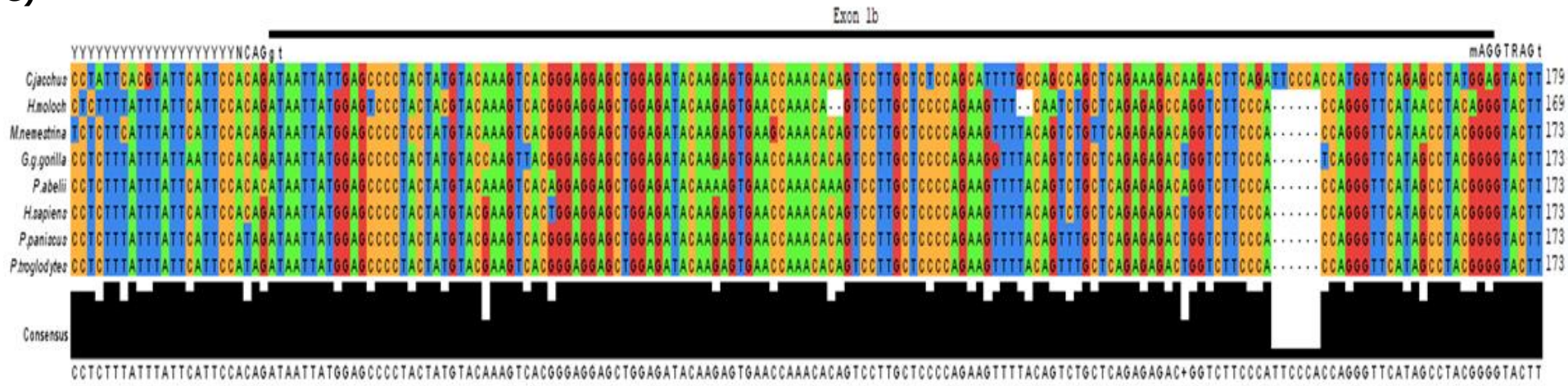
A)



B)



C)


Exon 1b
YYYYYYYYYYYYYYYYYYYYNCAGgt
mAGGTRAGt
C.jacchus
H.moloch
M.nemestrina
G.g.gorilla
P.abelii
H.sapiens
P.paniscus
P.troglodytes
Consensus
CCTCTTTATTTATTCATTCCACAGATAATTATGGAGCCCCTACTATGTACAAAGTCACGGGAGGAGCTGGAGATACAAGAGTGAACCAAACACAGTCCTTGCTCCCCAGAAGTTTTACAGTCTGCTCAGAGAGAC+GGTCTTCCCATTCCCACCAGGGTTCATAGCCTACGGGGTACTT

D)

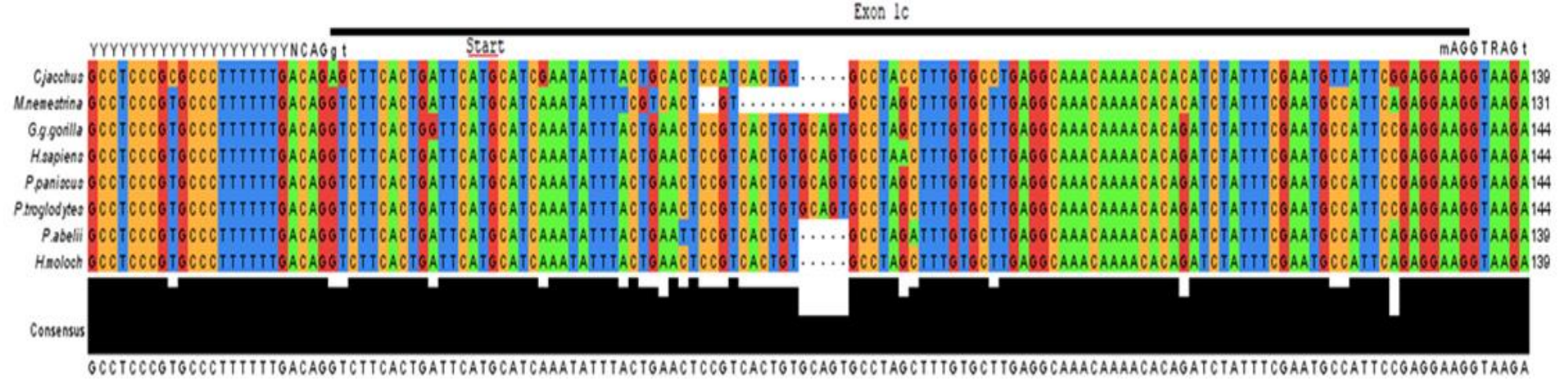

Exon 1c
YYYYYYYYYYYYYYYYYYYYNCAGgt
Start
mAGGTRAGt
C.jacchus
M.nemestrina
G.g.gorilla
H.sapiens
P.paniscus
P.troglodytes
P.abelii
H.moloch
Consensus
GCCTCCCGTGCCCTTTTTTGACAGGTCTTCACTGATTCATGCATCAAATATTTACTGAACTCCGTCACTGTGCAGTGCCTAGCTTTGTGCTTGAGGCAAACAAAACACAGATCTATTTCGAATGCCATTCCGAGGAAGGTAAGA

Figure 27. Conservation of *PIP5K1A* Exon 1a, 1b 1c like encoded sequences.

A) Multiple sequence alignments of N-terminal deduced amino acid sequences of putative orthologues of *PIP5K1A*_x3, x4, x8 & x9, which contain Exon 1c-2 splicing events for the indicated species.

B-D) Alignments of 1a, 1b and 1c like sequences identified from primate species by both BLASTn searches, and individual searches of sequences proximal to orthologues of canonical *PIP5K1A* Exon 1. Conformity of borders to consensus splice sites are indicated. Site of the putative AUG start codon is indicated for putative orthologues of Exon 1c

4.5.3. Use of Exon 1e

Finally, BLASTn and BLASTx searches were performed with the 5' sequences of human *PIP5K1A*_x10/x11/x12/x13 predicted transcripts (which contain Exon 1E), identifying putative PIP5KIα orthologues from nine non-human mammalian species of the *Carnivora* and *Artiodactyla* orders. The four *Artiodactyla* members identified were four of the seven extant members of the *Camelidae* family (*Camelus bactrianus*, the Bactrian camel*; Camelus ferus*, the wild Bactrian camel; *Vicugna pacos*, the alpaca; and *Camelus dromedarius*, the Arabian camel). The five *Carnivora* members were of diverse families in the *Caniformia* suborder (*Vulpes vulpes*, the red fox; *Aliuropoda melanoleuca*, the giant panda; *Phoca vitulina*, the common harbour seal; *Callorhinus ursinus*, the northern fur seal; and *Zalophus californianus*, the California sea lion).

The deduced amino acid sequences of these transcripts show conservation, particularly of the latter portion of sequence encoded by Exon 1e (**Figure 28**). The genome sequences of the *PIP5K1A* genes for each of these species were inspected in the NCBI genome browser, confirming the same relative position of the sequences corresponding to human *PIP5K1A* Exon 1e and immediately after the 3' end of the canonical Exon 1.

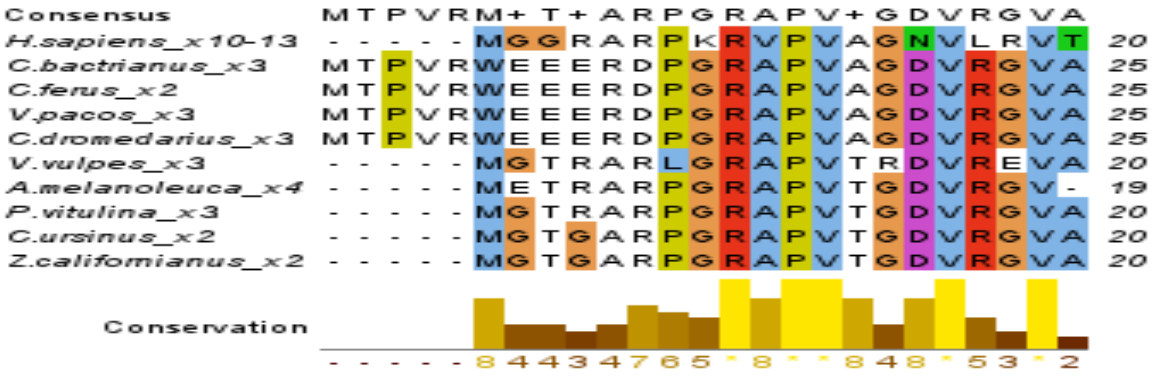


Figure 28. *PIP5K1A* Exon 1e like encoded sequences.

Multiple sequence alignments of N terminal deduced amino acid sequences of putative orthologues of human of *PIP5K1A*_x10, x11, x12 & x13 transcripts (which contain Exon 1e), identified by BLASTn and BLASTx searches.

Of note, the predicted start codon of the human and *Carnivora* transcripts is instead a Tryptophan (TGG) codon in the four *Camelidae* species, which instead have a predicted start codon 15 nucleotides to the 5'. The deduced translations of the transcripts in the *Camelidae* species thus gain five residues (^{1}MTPVR^{5}). The Kozak motifs of the predicted start codons in the transcripts are classified as 'strong' for human and the five *Caniformia* species, and 'weak' for the four *Camelidae* species [155]. The diversity of these species suggests conservation of *PIP5K1A* Exon 1e-like sequences and exon use at least among the *Boreoeutheria* magnorder.

Again, the absence of non-human primate species' transcripts from these results was conspicuous. The chromosome sequences immediately 3' to the canonical Exon 1

of PIP5K1A gene were inspected for the great ape species, finding 99% and 100% sequence identities with human *PIP5K1A* Exon 1e sequence (data not shown). Borders conformed to consensus donor-splice and acceptor-splice sequences. Exon 1e-like sequences may therefore be included in PIP5K1A transcripts of these species also, though as in our analysis of LNCaP C4-2 (**Section 4.3**), at such low frequency as to not be included in the exon-spanning reads datasets.

The chromosome sequences immediately 3' to the canonical Exon 1 of the PIP5K1A gene were likewise inspected for more distantly related species, those of Atlantogenatae, Marsupalia and Monotremata lineages. Exon 1e-like sequences were present immediately 3' of the canonical PIP5K1A Exon 1 sequence in the genomes of *Loxodonta Africana, Trichechus manatus latirostris* and *Elephantulus edwardii,* (all species within the Atlantogenata clade) including the predicted ATG start codon, Kozak motifs and splice-acceptor and donor sequences, though in the latter two species, an in-frame stop codon was present shortly after the ATG codon. Such sequences were not present for *Dromiciops gliroides* or *Gracilinanus agilis*, both marsupials, or *Ornithorhynchus anatinus*, a monotreme. This suggests that Exon 1e like sequences emerged within the PIP5K1A gene after the divergence of Marsupials and placental mammals, ~136 MYA.

4.5.4. Conclusions on evolutionary conservation of alternate 5' exons of *PIP5K1A*

In conclusion, there is evidence for existence of *PIP5K1A* transcripts containing Exon 1h or Exon 1e like sequences in mammalian species of diverse orders, with the emergence of such exons in an ancestral *PIP5K1A* gene dating to at least ≈136 MYA. The deduced amino acid sequences of such predicted transcripts show conservation across these species, hinting that if such transcripts are translated, the portions of the protein encoded by these exons may have evolutionarily conserved function, possibly functions distinct from that of the canonical isoforms of PIP5KIα. However, in depth analyses of this conservation vs expected rates of mutation were not performed, and experimental analysis of any such differential functions are difficult considering functions of the canonical N-terminal sequences remain unknown. Such experimental analyses would require, for example, high-throughput Y2H screens for protein-protein interactors for each unique N-terminal sequence, which is beyond the scope of the current project.

Analysis of the NCBI intron spanning read data (**Section 4.3**) indicates that the inclusion of Exon 1e and Exon 1h in human *PIP5K1A* transcripts is rare relative to inclusion of the Exon 1d – Exon 2 splice event. This is also the case for the other

species in which Exon 1h or Exon 1e appear in predicted transcripts (data not shown).

The conserved rarity of splice events joining to these exons supports a hypothesis that the emergence of such exons in an ancestral *PIP5K1A* gene are neutral mutations that have persisted across evolutionary time by genetic drift. i.e. rare transcripts are less likely to have an effect on cellular function, and thus less likely to effect an organisms fitness (either positively or negatively) and may thus be retained across evolutionary time because they are not selected against.

There is some evidence from transcript and chromosomal sequence data for conservation of existence of Exons 1a, 1b and 1c among primates though predicted transcript data is limited to only to two species: *H. sapiens* and *M. nemestrina*. Deduced amino acid sequences for transcripts that were predicted from the chromosomal sequences for primate species suggest conservation of the deduced amino acid sequences only among the *Hominidae* lineage. This limited evolutionary conservation, the relative infrequency of splice events involving the Exons 1a, 1b or 1c (**Section 4.3**), and the weak Kozak context of the predicted AUG start codon in Exon 1c, all suggest that the emergence of these exons is a neutral mutation, and that any resulting PIP5KIα isoform with alternate N-terminal sequence

likely have no necessary or beneficial function distinct from canonical PIP5KIα isoforms. Alternately, PIP5KIα isoforms putatively encoded by this splice event may represent a *Homininae* lineage-specific innovation, of unknown function.

Key to species:

V. Vulpes, Vulpes vulpes (red fox); *Balaenoptera acutorostrata scammoni* (Scammon's minke whale); *C. bactrianus, Camelus bactrianus* (Bactrian camel); *M. natalensis, Mastomys natalensis* (Natal multimammate mouse); *C. dromedarius, Camelus dromedarius* (Arabian camel); *S. scrofa, Sus scrofa* (wild boar); *P. catodon, Physeter catodon* (sperm whale); *C. l. familiaris, Canis lupis familiris* (dog); *Z. californianus, Zalophus californianus* (California sea lion); *A. melanoleuca, Aluirapoda melanoleuca* (giant panda); *M. leonina, Mirounga leonina* (southern elephant seal); *L. weddellii, Leptonychotes weddellii* (Weddell seal); *P. t. altaica, Panthera tigris altaica* (Siberian tiger); *P. pardus, Panthera pardus* (leopard); *F. catus, Felis catus* (cat); *L. canadensis, Lynx canadensis* (Canada lynx); *A. jubatus, Acinonyx jubatus* (cheetah); *T. m. latirostris, Trichechus manatus latirostris* (Florida manatee); *T. chinensis, Tupaia chinensis* (Chinese tree shrew); *C. canadensis, Castor canadensis* (American beaver); *M. nemestrina, Macacca nemestrina* (southern pig-tailed macaque); *R. bieti, Rhinopithecus bieti* (black-and-white snub-nosed monkey); *C. jacchus, Callithrix jacchus* (common marmoset); *S. apella, Sapajus apella* (turted capuchin); *A. nancymaae, Aotus nancymaae* (Nancy Ma's night monkey); *H sapiens, Homo sapiens* (human); *P. troglodytes, Pan troglodytes* (common chimpanzee); *P. paniscus, Pan paniscus* (bonobo); *P. abelli, Pongo abelli* (Sumatran orangutan); *G. g. gorilla, Gorilla gorilla gorilla* (western lowland gorilla); *M. mulatta, Macacca mulata* (rhesus macaque); *C. sabaeus, Chlorocebus sabaeus* (green monkey); *C. atys, Cercocebus atys* (sooty mangabey); *P. tephrosceles, Piliocolobus tephrosceles* (Ugandan red colobus); *P. anubis, Papio anubis* (olive baboon); *T. francoisi, Trachypithecus francois* (François' langur); *R. roxellana, Rhinopithecus roxellana* (golden snub-nosed monkey); *H. moloch, Hylobates moloch* (silvery gibbon); *C. ferus*, *Camelus ferus* (wild Bactrian camel); *V. pacos, Vicugna pacos* (Alpaca); *P. vitulina, Phoca vitulina* (common harbour seal); *C. ursinus, Callorhinus ursinus* (northern fur seal).

4.6. Identification of alternate splice events involving *PIP5K1A* Exons 12, 13, 14 and 15

One reported splice variant of *PIP5K1A*, *PIP5K1A*_v4 (NM_001135637.2) differs from the variants v1, v2, v3 and v5 in its exclusion of a 147 bp sequence, Exon 13, indicative of exon skipping type alternative splicing. Absence of the Exon 13 nucleotide sequence from *PIP5K1A* transcripts results in deduced amino acid sequence in which 49 residues from the divergent C-terminus are substituted for a single cysteine (**Table 8**).

Table 8. Variant C-terminal amino acid sequences encode.
Isoforms to which the residue numbers refer are underlined.

PIP5KIα Isoform	Exon 12 splice event	C-terminal sequence
i1, i2 and i5	Exon 12-13	**427**RFQRFMCNTVFKKIPLKPSPSKKFRSGS SFSRRAGSSGNSCITYQPSVSGEHKAQVT TKAEVEPGVHLGRPDVLPQTPPLEEISEGSP IPDPSFSPLVGETLQMLTTSTTLEKLEVAES EFTH**550**
i4	Exon 12 -14	**427**RFQRFMCNTVFKKIPCVHLGRPDVLPQT PPLEEISEGSPIPDPSFSPLVGETLQMLTTST TLEKLEVAESEFTH**500**
x15 and x16	Exon 12-15	**428**RFQRFMCNTVFKKIPLQPWKSLKLQSQ SSPIKRKASEDLEQDSAISVIPRCQPLPQQ C**485**

Both our own RNA-Seq data for LNCaP C4-2, and the NCBI aggregate intron spanning read data were analyzed to quantify alternative splicing of *PIP5K1A* Exon 13. Within our WT LNCaP C4-2 data set, there were a total of 7 ($\bar{x}$ = 2.33; n = 3) intron-spanning reads adjoining Exon 12 to Exon 14, representing 2.07 % of all reads exiting Exon 12 (**Figure 29 A**). Within the NCBI dataset of intron spanning reads, no such Exon 12 to Exon 14 intron spanning reads are reported, in contrast to 8340 intron spanning reads adjoining Exon 12 to Exon 13, and 9707 intron spanning reads adjoining Exon 13 to Exon 14 (**Figure 29 B**).

In addition to intron-spanning reads representing splice events joining Exon 12 to Exon 13 and joining Exon 12 to Exon 14, intron-spanning reads joining Exon 12 to Exon 15 were identified. Such reads were observed both within our WT LNCaP C4-2 data ($\bar{x}$ = 11; n=3; 9.76 % of all reads exiting Exon 12; **Figure 29 B; Table 9**), and within the NCBI intron spanning read data (366 reads; 4.2% of all reads exiting Exon 12; **Table 10**).

The existence of splice events joining *PIP5K1A* Exon 12 to Exon 15 are particularly interesting because of its effects on the deduced amino acid sequence of encoded PIP5KIα isoforms (i.e PIP5KIα x15 and PIP5KIα x16). Skipping of both Exon 13 and Exon 14 removes 277 bp from the coding sequence thus

removing 93 codons. As 277 is not divisible by three, a shift in reading frame of the subsequent Exons, Exon 15 and Exon 16, also occurs. The result on the deduced amino acid for such encoded isoforms is that residues 428-550 (numbered in reference to PIP5KIα i5) are substituted for an amino acid sequence encoded by an alternate reading frame of Exon 15 and Exon 16.

The residues ^{443}LQPWKSLKLQSQSSPI^{458} are encoded by the frameshifted Exon 15. The predicted stop codon of canonical *PIP5K1A* transcripts is the first codon of Exon 16 (TAA). In predicted transcripts in which Exon 12 is spliced to Exon 15, this codon is no longer in-frame to be read as a stop codon. The next in-frame stop codon (TGA) for such transcripts is 83-85 bp 3' to the start of Exon 16; and so 27 residues (^{459}KRKASEDLEQDSAISVIPRCQPLPQQC^{485}) are encoded by Exon 16.

While the Exon 12 to Exon 15 splice event causing a frameshift is predicted to change the position of the stop codon, the stop codon of the alternate reading frame remains within the final exon, and therefore seems unlikely to trigger activate NMD.

The phenomenon of exons having multiple viable translations due to alternative splicing inducing shifts in the

reading frame is called dual-coding. Dual-coding is increasingly recognised as having functional importance in eukaryotes [160-162].

PIP5K1A, *PIP5K1B* and *PIP5K1C* genes arose from gene duplication events from an ancestral gene, and display conservation of exon structure of the at least the exons encoding the central kinase core. It was therefore considered that the C-terminal sequence encoded by the alternate of reading frame of *PIP5K1A* Exons 15 and 16 may show homology to the C-termini of PIP5KIβ and PIP5KIγ.

Sequence alignments of the C-terminal amino acid sequences of the major isoforms of PIP5KIα, PIP5KIβ and PIP5KIγ were performed using Clustal Omega, including also the C-terminal sequence of PIP5KIα x15.

As expected, the sequence encoded by *PIP5K1A* Exon 12, which is present in both *PIP5K1A*_v1 and *PIP5K1A*_x15, as well as being conserved in the gene sequences of PIP5K1B and PIP5K1C, shows conservation, but the following sequences showed little conservation. Curiously the alignment indicated conservation of a small sequence of residues encoded by the frameshifted *PIP5K1A* Exon 15 and 16 with those of the canonical PIP5KIα isoforms, but also with PIP5KIβ and PIP5KIγ, particularly of a previously unreported conserved sequence

(PIP5KIα i5 ^{496}RPDVLPQTPPLEE^{508}; consensus RPD+LP+TPPLFE+) within the divergent C-termini (**Figure 29 C**). While the finding of a previously unreported conserved sequence among the C-termini of PIP5KIα, PIP5KIβ and PIP5KIγ is interesting, it was not investigated further.

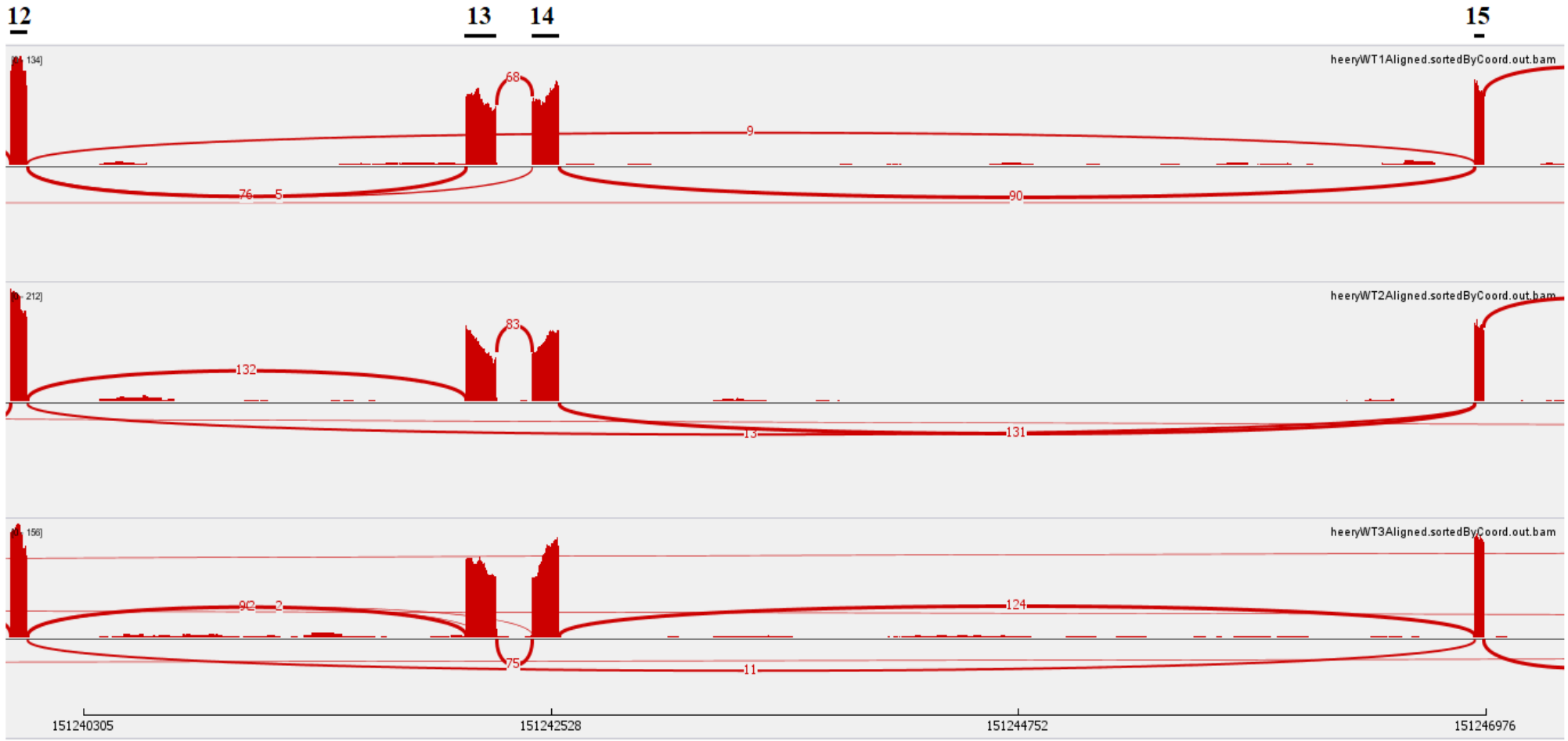
Table 9. Numbers of intron spanning reads adjoining *PIP5K1A* Exon 12 and subsequent exons in our RNA-Seq dataset for WT LNCaP C4-2.
Percentage of reads joining exons are expressed as a percentage of all reads exiting Exon 12.

	Reads joining Exon 12 to 13	Reads joining Exon 12 to 14	Reads joining Exon 12 to 15
WT 1	174	13	9
WT 2	233	0	19
WT 3	152	3	18
WT Mean	186.33	5.33	15.33
Percentage	90.02	2.58	7.41

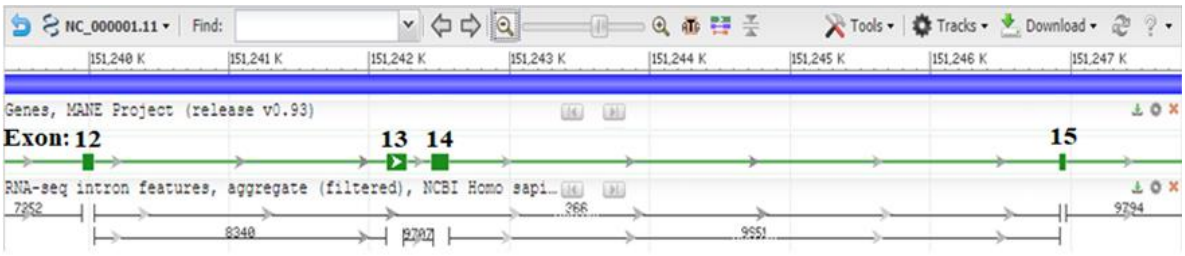
Table 10. Numbers of intron spanning reads adjoining *PIP5K1A* Exon 12 and subsequent exons in NCBI database
Percentage of reads joining exons are expressed as a percentage of all reads exiting Exon 12.

	Reads joining Exon 12- 13	Reads joining Exon 12- 14	Reads joining Exon 12- 15
Reads	8340	0	366
Percentage	95.80	0	4.20

A)


Exon:
12
13
14
15
heeryWT1Aligned.sortedByCoord.out.bam
68
9
76
5
90
heeryWT2Aligned.sortedByCoord.out.bam
83
132
13
131
heeryWT3Aligned.sortedByCoord.out.bam
92
2
124
75
11
151240305
151242528
151244752
151246976

B)



C)

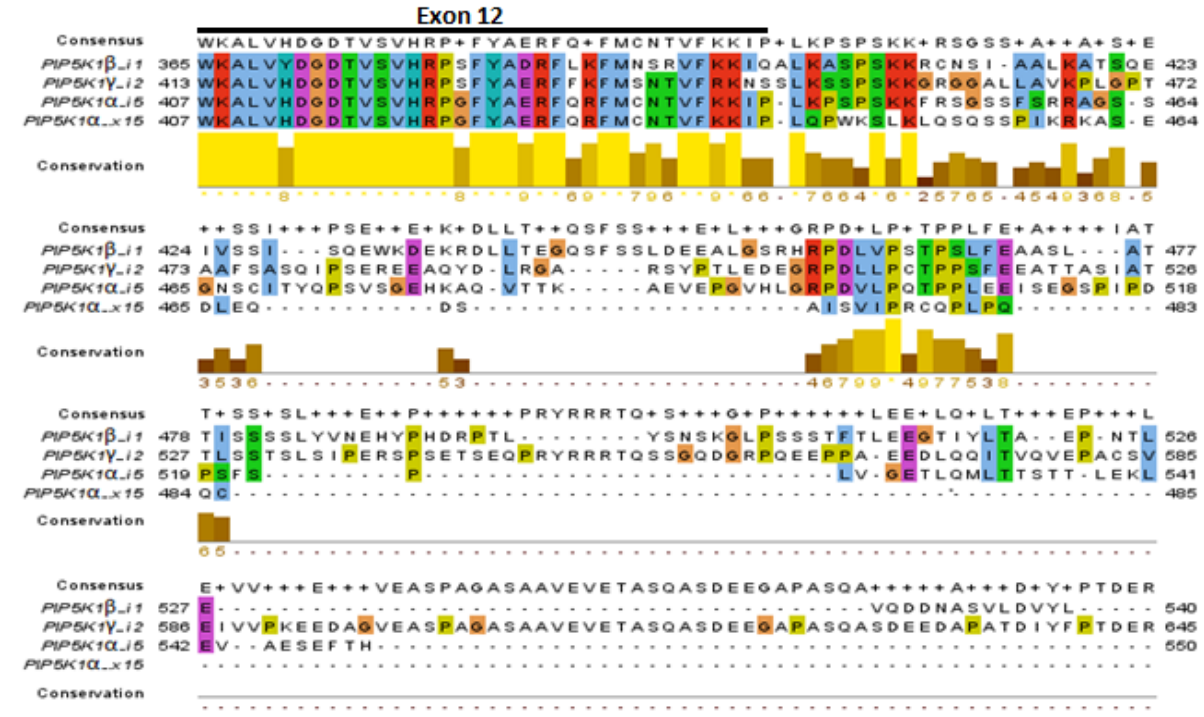


Figure 29. Existence of *PIP5K1A* Exon 12-Exon 15 splice events in LNCaP C4-2 and public RNA-Seq datasets

A) Sashimi plots of RNA-Seq reads for WT LNCaP C4-2, aligning to genomic region surrounding *PIP5K1A* Exon 12- Exon 16 for three triplicates. The coverage for each alignment track is plotted as a bar graph. Arcs represent intron spanning reads, with the number of such reads indicated. Base positions (Chromosome 1, GRCh38.p13, primary assembly) are indicated on the bottom track.

B) Screen grab of the NCBI RNA-Seq intron features in genomic region surrounding *H. sapiens PIP5K1A* Exon 12- Exon 15. Bars represent intron spanning (i.e. spliced) reads, with numbers on each bar representing the number of each such read.

C) CLUSTALW multiple sequence alignment of the C-termini of PIP5KIα (i5 and x15), PIP5KIβ i1 and PIP5K1γ i2. Conservation of residues was determined by analysis with Jalview.

4.7. Evolutionary conservation of *PIP5K1A* Exon 12 – Exon 15 splice events

BLASTn and BLASTx searches were performed with the nucleotide sequences resulting from the *PIP5K1A* Exon 12 – Exon 15 splicing event and the deduced amino acid sequence, respectively, with the intention of assessing evolutionary conservation of this splicing event. It was found that equivalents of the *PIP5K1A* Exon 12 – Exon 15 splicing event occur among Boreotherian mammals or diverse orders and that there is conservation of the encoded residues (**Figure 30**). Species in which the equivalent of the Exon 12 – Exon 15 splice event occur include members of the Primate, Pholidota, Carnivora, Artiodactyla, Perissodactyla and Chiroptera orders. No such sequences were found among members belonging to Monotremata, Marsupalia or Atlantogenata lineages.

The conservation of the equivalents to *PIP5K1A* Exon 12 - Exon 15 splicing event among many orders of Boreotherian mammals, but apparent absence of this phenomenon amongst Atlantogenatae, or the more distance Marsupalia and Monotremata, dates the emergence of this phenomenon to after the divergence of Atlantogenatae and Boreotherians ~100 MYA.

Members of the Glires clade, (orders Rodentia and Lagomorpha) were noticeably absent from this list of mammals

with orthologous transcripts containing an equivalent to *PIP5K1A* Exon 12- Exon 15 splicing event. Inspection of *PIP5K1A* gene sequence and intron spanning read data in the NCBI genome browser for three model species from these orders, *Rattus norvegicus* (Norway rat), *Mus musculus* (house mouse), and *Oryctolagus cuniculus* (European rabbit) indicated that while exons of *PIP5K1A* with high homology to Exon 12, Exon 13, Exon 14 and Exon 15 are present in these species, there are no reported reads joining the equivalents to Exon 12 – Exon 15. Divergence of the Primate and Glires lineages occurred approximately 90 MYA.

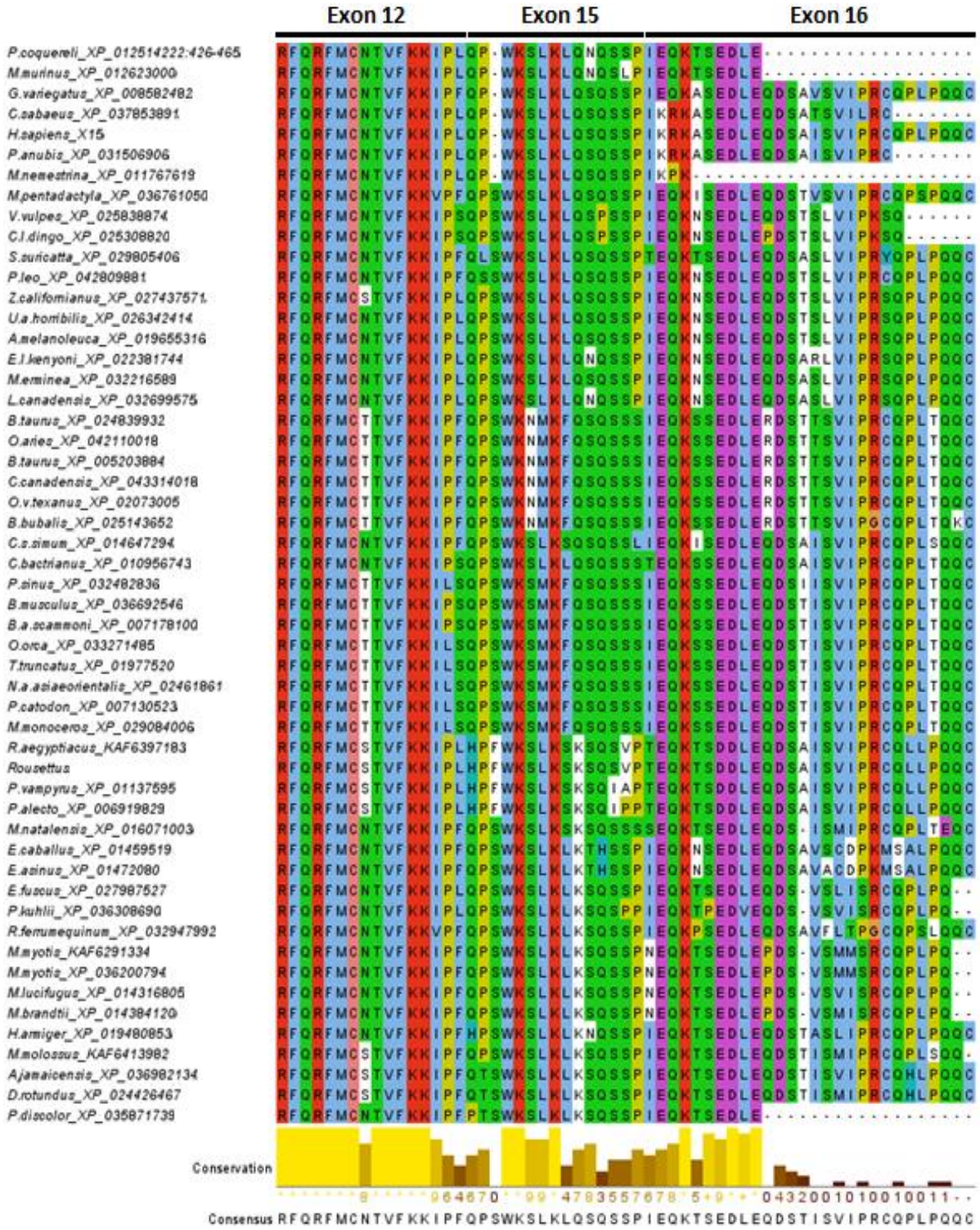


Figure 30. Evolutinonary conservation of amino acid sequences encoded by *PIP5K1A* Exons 12 – Exon 15 like splicing events.
Putative orthologues of Human PIP5KIα x15 and x16, which result from an Exon 12 - Exon 15 splicing event, were identified through BLASTn and BLASTp searches. Subsequently, CLUSTALW sequence alignment was performed for the C-terminal deduced amino acid sequence. Conservation of residues was determined by analysis with Jalview.

Key to species:

P. coquereli, Propithecus coquereli (Coquerel's sifaka); *M. murinus, Microcebus murinus* (gray mouse lemur); *G. variegatus, Galeopterus variegatus* (Sunda flying lemur); *C. sabaeus, Chlorocebus sabaeus* (green monkey); *H. sapiens, Homo sapiens* (Human)*; P. anubis, Papio Anubis* (olive baboon); *M. nemestrina, Macaca nemestrina* (southern pig-tailed macaque); M. pentadactyla, *Manis pentadactyla* (Chinese pangolin); *V. vulpes, Vulpes vulpes* (red fox); *C. l. dingo, Canis lupis dingo* (dingo); *S. suricatta, Suricata suricatta* (meerkat); *P. leo, Panthera leo* (lion); *Z. californianus, Zalophus californianus* (California sea lion); U. a. horribillis, Ursus arctos horribilis (grizzly bear); *A. melanoleuca, Ailuropda melanoleuca* (giant panda); *E. L. keyoni, Enhydra lutris kenyoni* (northern sea otter); *M. erminea, Mustela erminea* (stoat); *L. canadensis, Lontra canadensis* (North American river otter); *B. taurus, Bos taurus* (Eurasian cattle); O. aries, Ovis aries (domestic sheep); *C. canadensis, Cervus canadensis* (elk); *O. v. texanus, Odocoileus virginianus texanus* (Texas white tailed deer); *B. bubalis, Bubalus bubalis* (water buffalo); *C. s. simum, Ceratotherium simum simum* (southern white rhinoceros); *C. bactrianus, Camelus bactrianus* (Bactrian camel); *P. sinus, Phocoena sinus* (vaquita); *B. musculus, Balaenoptera musculus* (blue whale); *B. a scammoni, Balaenoptera acutorostrata scammoni* (Scammon's minke whale); *O. orca, Orcinus orca* (killer whale); *T. truncatus, Tursiops truncatus* (common bottlenose dolphin); *N. a. asiaorientalis, Neophocaena asiaeorientalis asiaeorientalis* (Yangtze finless porpoise); *P. catodon, Physeter catodon* (sperm whale); *M. monoceros, Monodon monoceros* (Narwhal); *R. aeyptiacus, Rousettus aegyptiacus* (Egyptian fruit bat); *P. vampyrus, Pteropus vampyrus* (Large flying fox); *P. alecto, Pteropus alecto* (black flying fox); *M. natalensis, Miniopterpus natalensis* (Natal long-fingered bat); *E. caballus, Equus caballus* (horse); *E. asinus, Equus asinus* (donkey); *E. fuscus, Equus fuscus* (wild horse); *P. kuhlii, Pipistrellus kuhlii* (Kuhl's pipistrelle); *R. ferrumequinium, Rhinolophus ferrumequinum* (greater horseshoe bat); *M. myotis, Myotis myotis* (greater mouse-eared bat); *M. lucifugus, Myotis lucifugus* (little brown bat); *M. brandtii, Myotis brandtii* (Brandt's bat); *H. arminger, Hipposideros armiger* (great Himalayan leaf-nosed bat); *M. molossus, Molossus molossus* (Velvety free-tailed bat); *A. jamaciensis, Artibeus jamaicensis* (Jamaican fruit bat); *D. rotundus, Desmodus rotundus* (common vamire bat); *P. discolor, Phyllostomus discolor* (pale spear-nosed bat).

4.8. PCR validation of novel upstream exons in *PIP5K1A* expressed in human prostate cancer cells

To confirm the existence of *PIP5K1A* transcripts containing the putative exons Exons 1b and 1c (Exon 1a was not included at this stage), PCR amplification was employed. For template, cDNA first strand synthesis using LNCaP C4-2 total RNA was performed. The PCRs were performed with forward primers binding to the sequence of the putative exons, and a common reverse primer for Exon 4. Control reactions with primers binding to the canonical Exon 1d were also performed. Primers sequences were as listed in **Supplementary Table 6.**

PCR products were resolved by agarose gel electrophoresis. Bands of the expected sizes were observed and these were extracted and sequenced. Variable inclusion of CAG at the 5' end of Exon 2 was predicted to cause amplification of two products, one with the CAG, one without, causing mixed sequencing results for extracted bands. Such mixed sequences were obtained for bands in lane 2, 3 and 4. The chromatograms of the mixed sequencing results were subjected to computational deconvolution with Poly Peak Parser [163], with the output sequences indicating that each band was indeed composed of two DNA species that differed by presence/absence of CAG at the *PIP5K1A* Exon 2 5' border. The

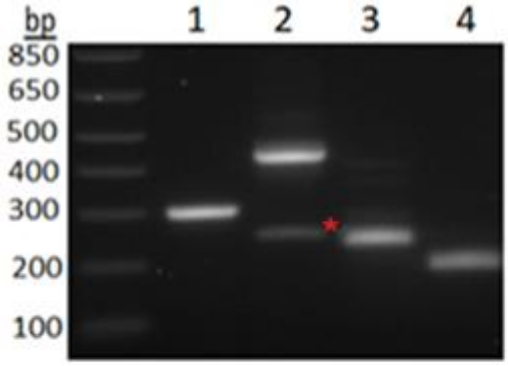
sequences obtained matched those of the putative *PIP5K1A* Exons 1b and 1c, contiguous with the Exons 2 and 4, as predicted from exon spanning read sequences in our RNA-Seq data. Sequences matching Exon 3 were not present, consistent with our RNA-Seq data in which exon spanning reads corresponding to Exon 2-Exon 3 splicing vs Exon 2-Exon 4 splicing occur at a ratio of 0.03 to 1 (reads spanning Exon 2-3: $\bar{x}$ = 2.00; reads spanning Exon 2-4: $\bar{x}$ = 78.67; n =3). Inclusion of Exon 3 would increase the size of any amplicon by 36 bp, but corresponding bands were not observed.

These results thus confirmed the existence of transcripts in the LNCaP C4-2 cell line using the Exons 1b and 1c, and lacking Exon 1d. As the reverse primer used bound within Exon 4, these data do not exclude the existence of additional variants linking exons 1b and 1c to exons distal to Exon 4, although this was addressed in subsequent experiments (**Section 4.9**).

An unexpected PCR product was observed at approximately 250 bp for the control PCR reaction using the forward primer in Exon 1d and the reverse primer in Exon 4. Interestingly, the sequence of this band matched a 5' portion of 1d contiguous with the putative Exon 1e, then Exon 2 then Exon 4. The final 294 bp sequence of Exon 1d (294 bp labelled 1dii in **Figure 31 B**), which contains the canonical ATG start codon,

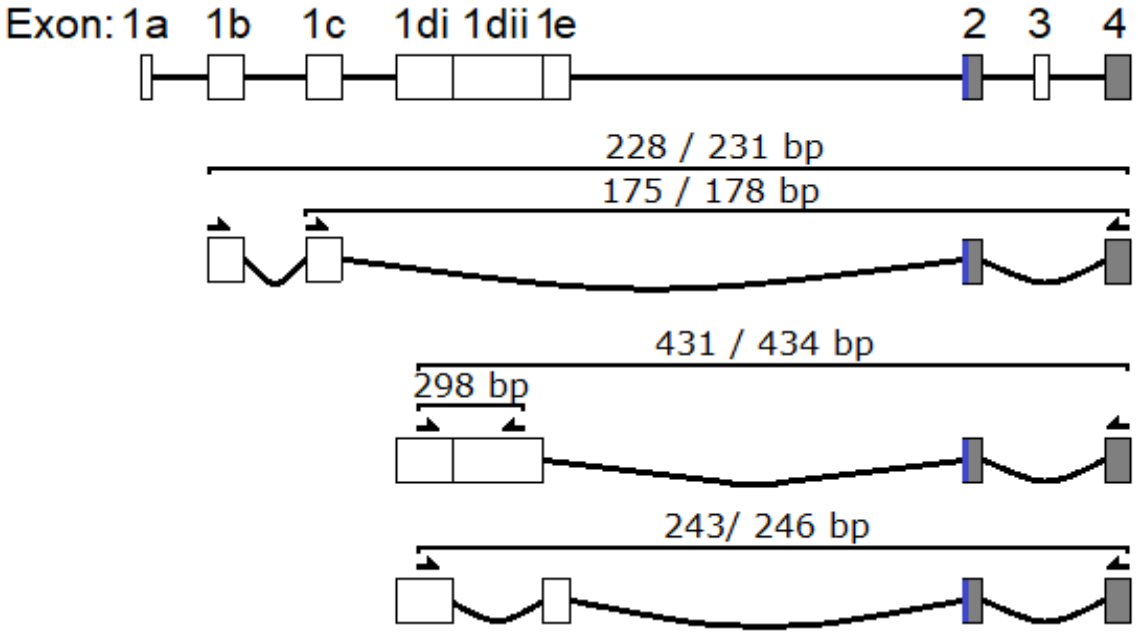
was absent from the sequencing results for this minor band, suggesting that it is variably spliced out of *PIP5K1A* transcripts, and is thus a cryptic intron. Further, analysis of this 294 bp sequence shows it possesses features of an intron (**Figure 31 C**). Therefore the canonical Exon 1d is hereafter represented as being composed of two elements, Exon 1di the 5' 236 bp, and Exon 1dii, which is the 3' 294 bp that form a cryptic intron. Skipping of 1dii and use of Exon 1e were not observed in our RNA-Seq analysis and the corresponding band in the agarose gel was faint compared to the band representing transcripts formed by the canonical splice events (Exon 1d - Exon 2 – Exon 4). This band was not observed after a first set of 30 PCR cycles (not shown), and was only apparent after a second set of 30 cycles (which used as input a one-tenth volume aliquot of the first set). These details all indicate rarity of transcripts in which 1di is retained but 1dii is absent.

A)



Lane	Forward primer targeted to	Reverse primer targeted to	Expected amplicon (bp)
1	Exon 1di	Exon 1dii	298
2	Exon 1di	Exon 4	431/434
3	Exon 1b	Exon 4	228/231
4	Exon 1c	Exon 4	175/178

B)



C)



D)

1di splice donor site	1e splice acceptor site
mAG‖GTRAGt	YYYYYYYYYYYYYYYYYYYNCAG‖gt
.|. ||.|||	||||||||||.|.||||.||| ||| ||
GAC‖GTGAGT	TCCCTTCCTGTACCTTGTCCTCAG‖GT

Figure 31. PCR validation of existence of Exon 1b and 1c in *PIP5K1A* transcripts.

A) Agarose gel electrophoresis of PCR amplification of 5' ends of canonical and predicted *PIP5K1A* transcripts from cDNA of WT LNCaP C4-2 cells using primers targeted to each 5' exon and a common reverse primer targeted to Exon 4. One 6th reaction volume was resolved by agarose gel electrophoresis. The table below is a key for the primer pairs used and expected amplicon size for the reaction loaded in each lane. The red asterisk indicates an unexpected product in lane 2 which was sequenced and found to contain sequence of Exon 1di-Exon 1e-Exon 2-Exon 4. 1 % agarose; 1 kb plus ladder (Invitrogen).

B) Exon and splicing maps of the deduced from the sequences of amplicons resolved in **A**. Primer positions and direction are indicated by half arrows. Amplicon sizes are indicated by bars above each splice map.

C) Sequences of the Exon 1di and putative Exon 1e, separated by the exon/cryptic intron 1dii. The putative splice donor site of Exon 1di and splice-acceptor site of Exon 1e are indicated by the red and blue box, respectively. A polypyrimidine sequnence towards the 3' end of 1dii is indicated by grey highlight.

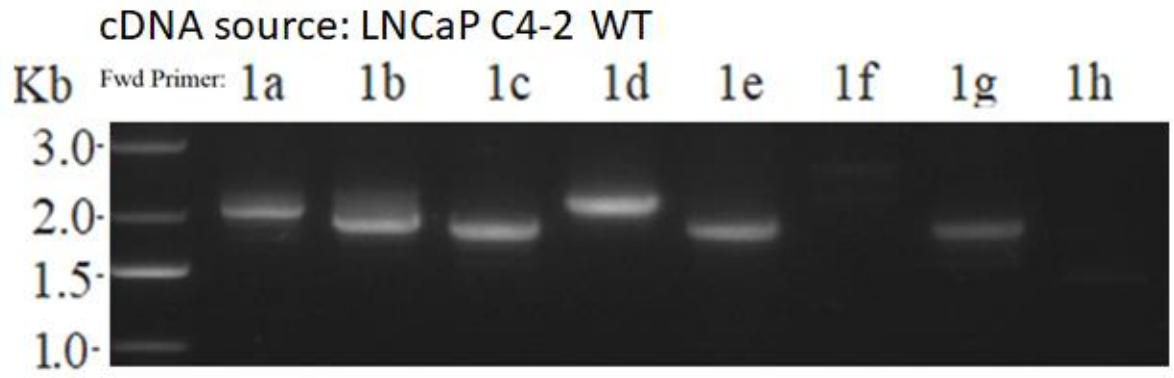
D) Comparisons of Exon 1di donor splice site and Exon 1e splice-acceptor site sequeneces (Bottom) to consensus splice sequences (Top).

4.9. Cloning alternate full-length *PIP5K1A* transcripts

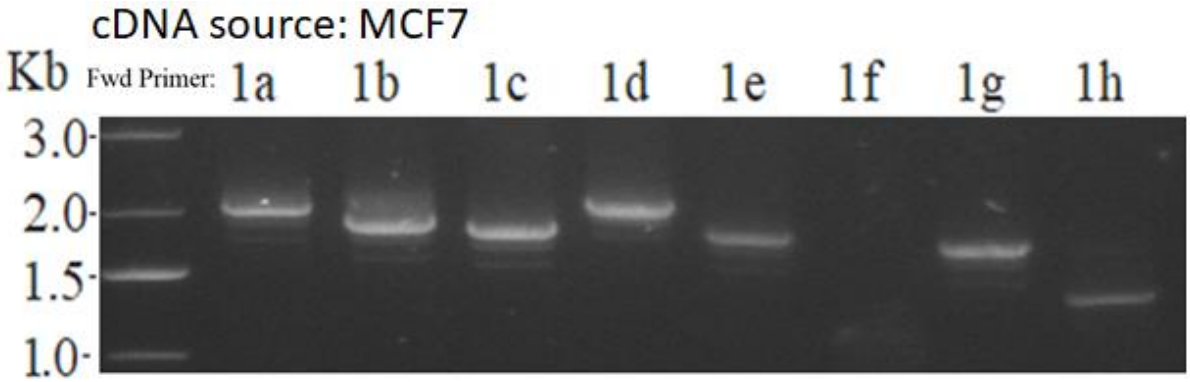
Due to the design of the experiments presented in **Section 4.8**, they could only inform as to the existence of alternate *PIP5K1A* transcripts with sequences spanning up to Exon 4, and only for those including the putative exons 1b and 1c. New primers were therefore designed targeting each of the other putative exons (1a, 1e, 1f, 1g and 1h) and a reverse primer targeting the canonical final exon, Exon 16 (**Supplementary Table 7**). As the previous experiments also only utilised cDNA synthesised from LNCaP C4-2, cDNAs were also synthesised from total RNA extracts of the human breast cancer cell line MCF7, thus allowing investigation of existence of alternate *PIP5K1A* transcripts in a second cell line.

PCRs were performed with these primers and cDNAs, and products were resolved by agarose gel electrophoresis (**Figure 32 A and B**). Bands of the expected sizes around 2 Kb were observed for all PCR reactions, except for the reactions with primer targeting Exon 1f; for which a weak band was detected for the reaction with MCF7 cDNA, but not for LNCaP C4-2.

A)


cDNA source: LNCaP C4-2 WT
Kb
Fwd Primer:
1a
1b
1c
1d
1e
1f
1g
1h
3.0
2.0
1.5
1.0

B)


cDNA source: MCF7
Kb
Fwd Primer:
1a
1b
1c
1d
1e
1f
1g
1h
3.0
2.0
1.5
1.0

C)

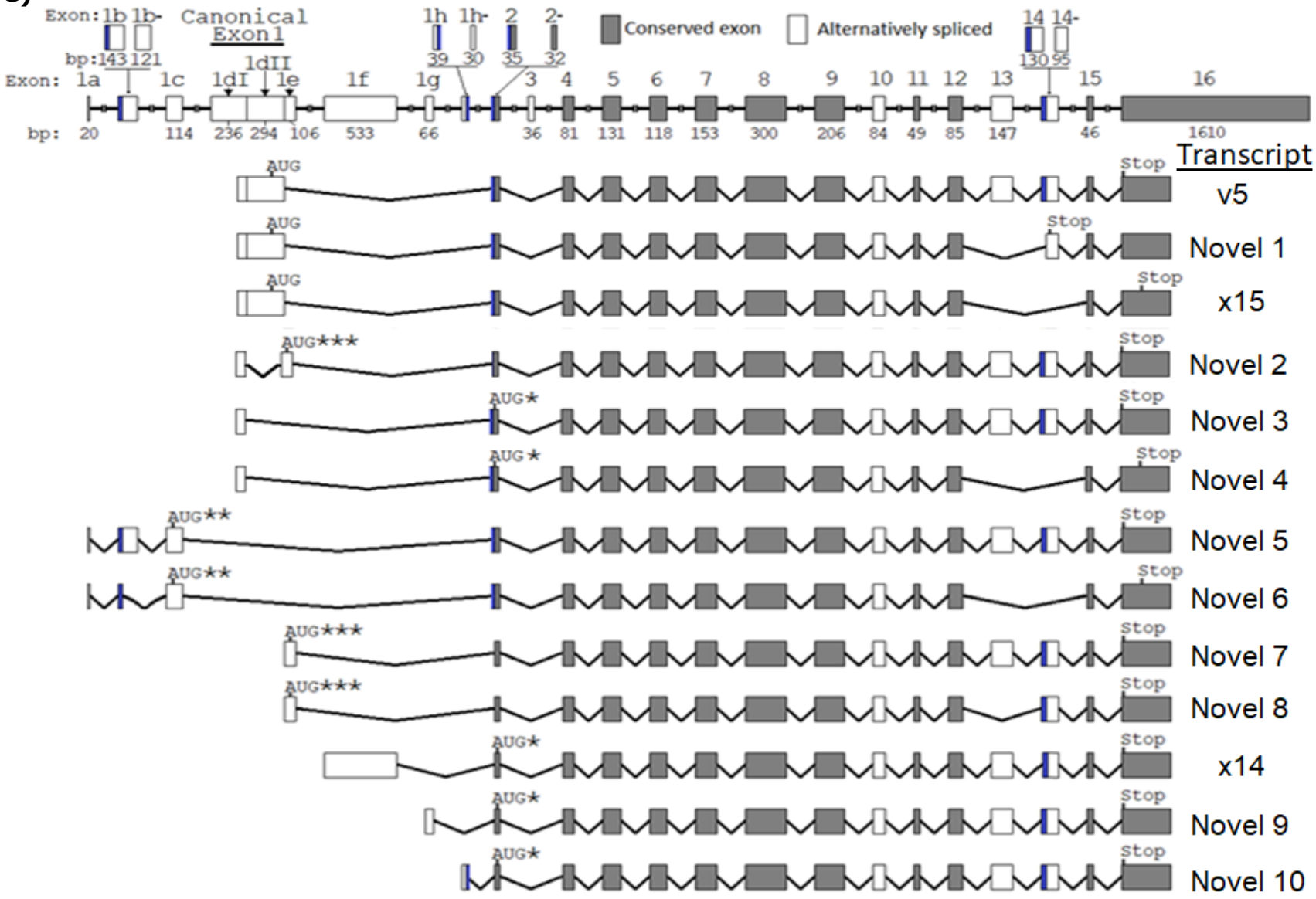

Exon: 1b 1b- Canonical Exon1 1h 1h- 2 2-
Conserved exon
Alternatively spliced
14 14-
bp: 143 121 1dII 39 30 35 32 130 95
Exon: 1a 1c 1dI 1e 1f 1g 3 4 5 6 7 8 9 10 11 12 13 15 16
bp: 20 114 236 294 106 533 66 36 81 131 118 153 300 206 84 49 85 147 46 1610
Transcript
v5
Novel 1
x15
Novel 2
Novel 3
Novel 4
Novel 5
Novel 6
Novel 7
Novel 8
x14
Novel 9
Novel 10
AUG
Stop

Figure 32. Cloning of alternate transcripts of *PIP5K1A*

A) and B) Agarose gel electrophoresis of PCR amplifications of splice variants of *PIP5K1A* using common reverse primer targeted to the conserved final exon of *PIP5K1A*, Exon 16, and various forward primers targeting each of the alternately spliced 5' exons, using cDNA from LNCaP C4-2 cells **(A)** or MCF7 cells **(B)**. 0.8 % agarose; 1 kb plus ladder (Invitrogen).

C) Splicing maps of the numerous splice variants of *PIP5K1A*. Variants identical to those previously reported (v5) or predicted (x15, and x14) are indicated. All variants newly identified in this study are designated 'Novel 1' to 'Novel 10'.

A similar pattern of bands was observed in reactions with forward primers targeting Exons 1a, 1b, 1c, 1d, 1e and 1g: an intense upper band with two weaker bands running at approximately 100 bp and 300 bp below. This pattern of bands appearing in several reactions suggests that it arises due to amplification of various transcripts each produced by different alternative splicing events of *PIP5K1A* involving exons other than those at the 5' terminus. The abundance of the different splice events involving Exon 12 in RNA-Seq data (**Section 4.6**), and the positions of the lower bands approximately 100 bp and 300 bp below the major band suggest they represent amplification of transcripts lacking either Exon 13 (147 bp) alone, and both Exon 13 and Exon 14 (147 bp and 130 bp = 277 bp), respectively.

To determine the identity of the observed bands, PCR products for each reaction were ligated into the pCR4Blunt-TOPO vector (Invitrogen), and ligation mixtures used to transform chemically-competent DH5α *E. coli*. Colonies surviving antibiotic selection were screened for *PIP5K1A* inserts by colony PCR (between 10 and 30 screened colonies per cloning reaction) with the same primer pairs that had been used in the prior reaction to generate the inserts. Inserts of various sizes around 2 Kb were detected, consistent with the prior reactions. Plasmid DNAs with putative *PIP5K1A* inserts were

isolated from screened colonies by alkaline lysis miniprep (Macherey-Nagel) and sequenced by Sanger sequencing (Source Bioscience) from both directions to obtain the full sequence of each insert.

This method of TOPO cloning with the neat PCR product mixes followed by colony PCR screening for inserts of different lengths was chosen over excision and extraction of DNA from separate bands followed by individual cloning reactions for each band. The reason for this choice was that by it, with sufficient screening by colony PCR, even relatively rare transcripts could be identified. Had the alternative method of band excision and individual cloning reactions been used, only transcripts sufficiently common to result in an easily visible and thus excisable band could be identified.

In this way an insert with sequence identical to a canonical transcript of *PIP5K1A* (NM_001330689.2; transcript variant 5) was identified.

In addition, twelve inserts representing alternate splice variants of *PIP5K1A* were identified (splicing maps shown in **Figure 32 C**). Two of these are exact matches for the NCBI predicted transcripts NCBI transcripts x15 and x14, thus confirming the existence of such transcripts.

- NCBI transcript x15 includes an Exon 12 to Exon 15 splice event, and thus validates the inclusion of this event in *PIP5K1A* transcripts. The predicted transcript x16 is almost identical to x15, with the only difference being the use of Exon 2+ (addition of a CAG sequence) in x15, but use of Exon 2- (no addition of CAG) in x16. Since the variation in use of Exon 2+ vs Exon 2- is a common feature of *PIP5K1A* transcripts, and use of Exon 2+ (as in transcripts x15) appears to be rarer than use of Exon 2-, validation of the existence of transcript x15 thus also seemingly validates existence of x16.
- NCBI transcript x14 includes Exon 1f spliced to Exon 2-. Exon 3 is excluded, and so Exon 2 is spliced to Exon 4. All subsequent exons, including Exons 13 and 14, are included. The predicted start codon for such transcripts is in Exon 2, the 36th codon of canonical *PIP5K1A* transcripts, such as transcript v5.

As for the ten insert sequences that do not exactly match any of the NCBI predicted transcripts, many are similar to predicted transcripts in their use of the alternate 5' exons instead of Exon 1d, but differ in other known splicing events such as exclusion of Exon 3, Exon 13 and Exon 14:

- ‘Novel 5’ is similar to NCBI transcript x9 (XM_024450130.1), in the inclusion of Exon 1c, with the addition of Exons 1a and 1b, which are not included in any of the NCBI predicted transcripts. The additional inclusion of Exons 1a and 1b does not change the predicted open reading frame of transcripts including Exon 1c. Possibly all transcripts that include Exon 1c also include Exon 1a and Exon 1b.
- ‘Novel 6’ is similar to ‘Novel 5’ in its inclusion of Exons 1a, 1b and 1c, but also exhibits the Exon 12-Exon 15 splicing event present in transcripts *PIP5K1A*_x15 and *PIP5K1A*_x16. Additionally only the most 5’ 143 bp of Exon 1b are present, indicating presence of a second splice donor site within Exon 1b. Exon 1b is thus depicted as two forms, 1b+ (143 bp) and 1b- (121 bp)
- ‘Novel 7’ is similar to NCBI transcript *PIP5K1A*_x11 (XM_011510045.3) in its inclusion of Exon 1e, the only difference being the absence of Exon 3 from ‘Novel 7’.
- ‘Novel 8’ is likewise similar to NCBI transcript *PIP5K1A*_x11 in its inclusion of Exon 1e, but differs in its exclusion of both Exon 3 and Exon 13. This was the only insert identified in which Exon 13 was excluded, but Exon 14 included.

- ‘Novel 9’ is similar to NCBI transcripts *PIP5K1A*_x17 and *PIP5K1A*_x18 (XM_017002441.2 and XM_011510048.3, respectively) in the inclusion of Exon 1g but differs from these in its exclusion of Exon 3, and of Exon 1h. This discovery indicates that Exon 1g can be used without being spliced to Exon 1h. Transcripts from which this insert sequence was derived may include exons 5’ to Exon 1g, such as Exon 1d, or Exons 1a, 1b and 1c, but as the forward primer used for the PCR reaction was in Exon 1g, sequence 5’ to Exon 1g was not determined. More thorough screening of clones generated following the TOPO cloning reaction with PCR products a forward primer in Exon 1d or 1a might have revealed such sequences, but this was not pursued.
- ‘Novel 10’ is similar to NCBI transcripts x17 and x18 (XM_017002441.2 and XM_011510048.3, respectively) but differs from these in its exclusion of Exon 3. Like for ‘Novel 9’ the inclusion of exons more 5’ to Exon 1h in transcripts would require more thorough screening of clones generated following PCR with forward primers more 5’ than Exon 1h, and this was not pursued.

Inserts were also detected that had other variations:

- 'Novel 1' is similar to the canonical transcript v5, but lacks Exon 13, and contains only the 3' 95 bp of Exon 14. Exon 14 therefore is represented as Exon 14+ (the canonical 130 bp Exon 14), and Exon 14-, the most 3' 95 bp.
- 'Novel 2' contains the Exon 1dii to Exon 1e splice event that was previously detected as described in **Section 4.8**. Exon 1e is subsequently spliced to Exon 2-, then to Exon 4. All subsequent exons (Exons 5-16) are included.
- 'Novel 3' contains Exon 1dii to Exon 2 splice event, which has not previously been observed. Exon 2 is subsequently spliced Exon 4 and all subsequent exons (Exons 5-16) are included.
- 'Novel 4' is similar to 'Novel 3' in the Exon 1dii to Exon 2 splicing event. 'Novel 4' additionally displays the Exon 12 – Exon 15 splicing event.

4.10. Subcellular localisation of PIP5KIα isoforms in LNCaP C4-2 cells

4.10.1. Localisation of full length PIP5KIα isoforms

The existence of C-terminal splice variants of *PIP5K1A* that include an Exon 12 to Exon 15 splice event were confirmed, as described in **Section 4.5** and **Section 4.7**. These putatively encode isoforms of PIP5KIα with substitution of residues of the C-terminus involved in sub-cellular localisation. It was therefore decided to investigate differences in sub-cellular localisation of the isoforms of PIP5KIα with alternate C-terminal sequences.

Due to the large number of splice variants of *PIP5K1A* that had been identified/ confirmed as described in **Section 4.7**, and the time required for molecular cloning experiments, sub-cellular localisation of the multiple putative isoforms of PIP5KIα with variation in the N-terminus was not investigated.

Localisation of the canonical isoforms, NCBI PIP5KIα i1, i2, and i4, and the isoform PIP5KIα x15, whose open reading frame had been previously cloned into the pCR4 blunt TOPO vector, was investigated. Mammalian expression vectors for each of these PIP5KIα variants fused to mVenus, a modified monomeric form of GFP protein, were generated by cloning into the mVenus N1 vector [164].

Construction of PIP5KIα i2–mVenus and PIP5KIα i4-mVenus plasmids are described in **Section 3.5.2**, using as template PIP5KIα coding sequences from p15aC-3D1D-5 (for isoform 2; a gift from Sanford Simon; Addgene plasmid #121050; [128]) and pLPS-*PIP5K1A* (for isoform 4; a gift from Jenny Person).

None of the CDS of PIP5KIα that had been cloned in **Section 4.9** corresponded to PIP5KIα i1. To construct a PIP5KIα i1-mVenus expression, it was therefore necessary to perform site directed mutagenesis on the PIP5KIα i2-mVenus plasmid to insert the Exon 3 sequence into the PIP5KIα i2 CDS, thus creating PIP5KIα i1 CDS. The primers used for this site directed mutagenesis are listed in **Supplementary Table 8.**

For construction of a PIP5KIα x15–mVenus plasmid, the coding sequence of PIP5KIα x15 was subcloned from the pCR4 Blunt TOPO-*PIP5K1A* x15 plasmid that was generated as described in **Section 4.7** into mVenus. The PCR for this subcloning was performed with the primers as listed in **Supplementary Table 8.** For the fusion-protein encoded by each plasmid, the amino acid sequence linking each PIP5KIα isoform to mVenus is KDPPVAT.

LNCaP C4-2 cells grown in 8 well Ibidi plates were transiently transfected using Lipofectamine 3000, and after 48 hrs fixed and analysed by confocal microscopy, as described in **Section 2.5.5**.

As shown in **Figure 33**, PIP5KIα i2-mVenus, PIP5KIα i4-mVenus, PIP5KIα i1-mVenus, PIP5KIα x15-mVenus showed a largely cell membrane associated distribution in LNCaP C4-2 cells.

Previous work from several groups reported that exogenously expressed PIP5Ks predominantly localise to the cytoplasmic membrane, in contrast to the various localisations of endogenous PIP5Ks detected by immunocytochemistry [146; 165]. Various explanations have been suggested for the apparent discrepancy between localisation of exogenous and endogenous PIP5Ks e.g. predominant localisation to the membrane might result from excessive $PI(4,5)P_2$ generation (by the exogenously expressed PIP5KI) leading to accumulation at the membrane of $PI(4,5)P_2$ effectors with which PIP5KIα interacts. This is supported by the observation that transfection with kinase deficient variants of PIP5KIγ recapitulate localisation of the endogenous enzymes [165]. However, the specificity of antibodies

used to detect endogenous PIP5Ks used in these studies is poorly characterised, and given the different distributions of the α, β and γ isozymes, it may be unwise to extrapolate such conclusions. In our hands, the Venus tagged isoforms i1, i2, i4 and X15 all showed predominant localisation to the cell membrane in C4-2 cells, with some intracellular detection in vesicle-like structures (**Figure 33**). Interestingly, expression of a kinase deficient variant, PIP5KIα i4 (K181A)-mVenus, also localised to the cell membrane and its projections, revealing that full kinase activity is not critical for its localisation to the cell membrane. The K181A substitution reportedly does not completely abrogate kinase activity of PIP5KIα, reducing activity to approximately 10% of that of the WT protein when purified from mammalian cells [75] – although the authors could not exclude the possibility that this retained activity was due to co-purification of an endogenously expressed WT PIP5KI (e.g. through dimerization interactions). In contrast to reports using PIP5KIα i1-Myc and PIP5KIα i4-FLAG constructs in various cell lines (Hela, Hek293, A549) [122; 149], we did not observed any localisation of PIP5KIα-mVenus isoforms to the nucleus, nuclear speckle or nucleolus, whereas the mVenus only control displayed both nuclear and cytoplasmic distribution.

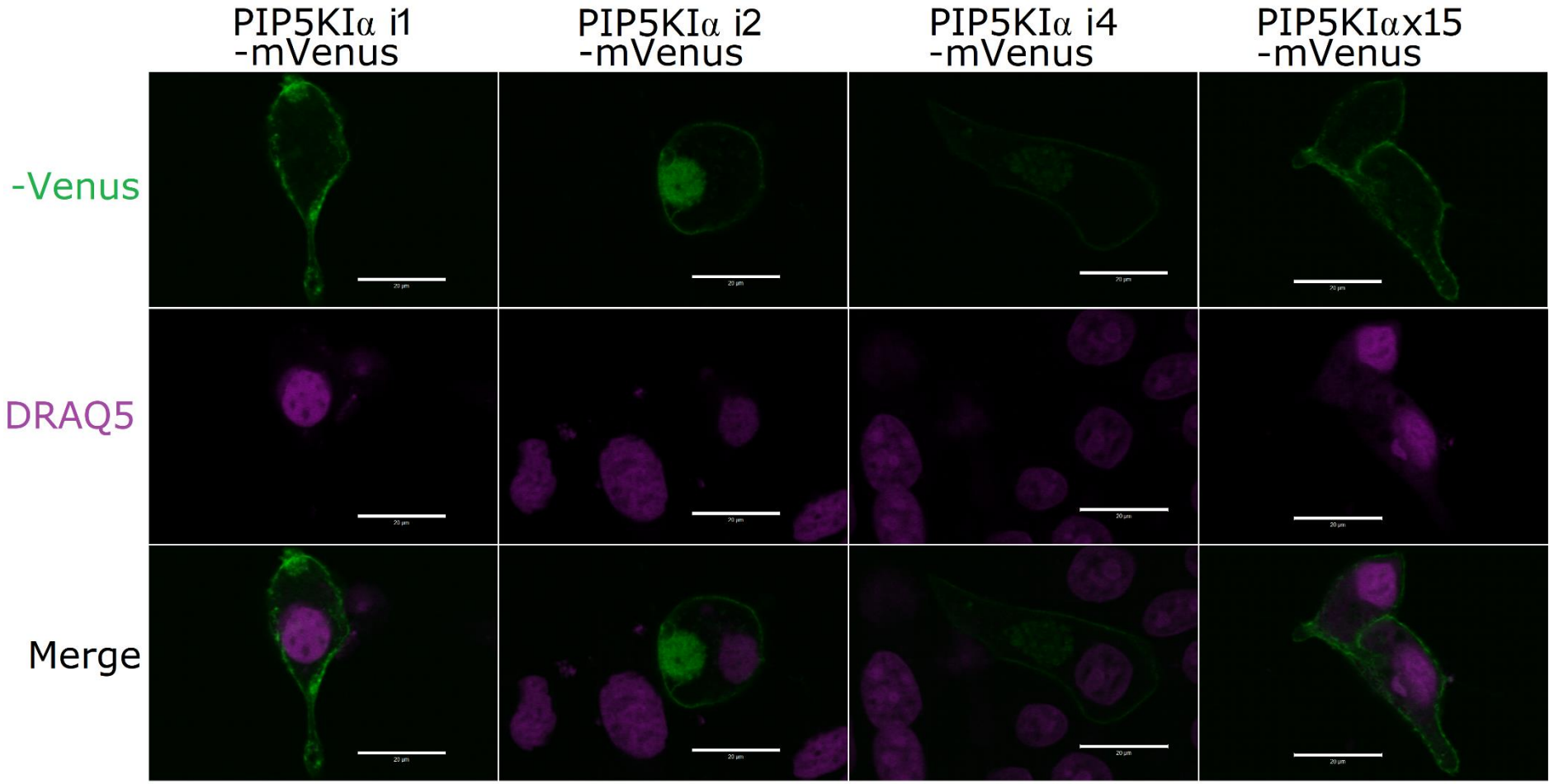

PIP5KIα i1
-mVenus
PIP5KIα i2
-mVenus
PIP5KIα i4
-mVenus
PIP5KIαx15
-mVenus
-Venus
DRAQ5
Merge

Figure 33. Subcellular localisation of variant of PIP5KIα isoforms.

Expression vectors for three canonical isoforms of PIP5KIα: PIP5KIα i1, PIP5KIα i2 and PIP5KIα i4, and a novel isoform x15 (the CDS for x15 was cloned from LNCaP C4-2 cDNA in this study) each fused to mVenus, were expressed in LNCaP C4-2 KO cells. Scale bars represent 20 μm.

mVenus-tagged PIP5KIα protein variants (green) and DNA staining (purple). Cells were the KO subline of LNCaP C4-2. Imaged using fluorescence microscopy.

4.10.2. Isoform-specific C-terminal sequences of PIP5KIα exhibit different subcellular localisations

Expression of mVenus fusions to PIP5KIα isoforms in their entirety did not result in any observable difference in localisation, each displaying mostly cell membrane localisation – which may or may not be artefactual. Mellman et al., have previously shown that an fusion protein of PIP5KIα i1 residues 440-562 with an HA tag localised to nuclear speckles [149], and the PIP5KIα C-terminus has been shown to function in nucleolar localisation [122]. It was decided to investigate localisation of the unique C-terminal sequences of each isoform.

Because the C-terminal sequence of the canonical isoforms of PIP5KIα could feasibly contain nuclear localisation signals that contribute to their reported nuclear localisation, the PIP5KIα C-terminal variants were analysed for the presence of nuclear localisation signals (NLS). The distinct C-terminal sequences of each isoform of PIP5KIα were submitted for *in-silico* predictions of NLSs using the online tool, NLSmapper [166], finding distinct predicted NLSs in the C-terminus of PI5KIα i2 and PIP5KIα x15, but not in PI5KIα i4 (**Table 20**).

PIP5KIα i4 differs from PIP5KIα i2 in its C-terminus because of the Exon 12-Exon 14 splicing events in transcripts encoding the former, leading to loss of 50 residues encoded by

Exon 13, and their substitution for a Cysteine. Likewise, PIP5KIα x15 differs in its C-terminus because of the absence of both Exons 13 and 14 from its transcripts, and the subsequent frame shift in Exons 15 and 16, leading to a substitution of residues 441-549 for an alternate 43 residue sequence. The unique C-termini sequences are presented in **Figure 34**.

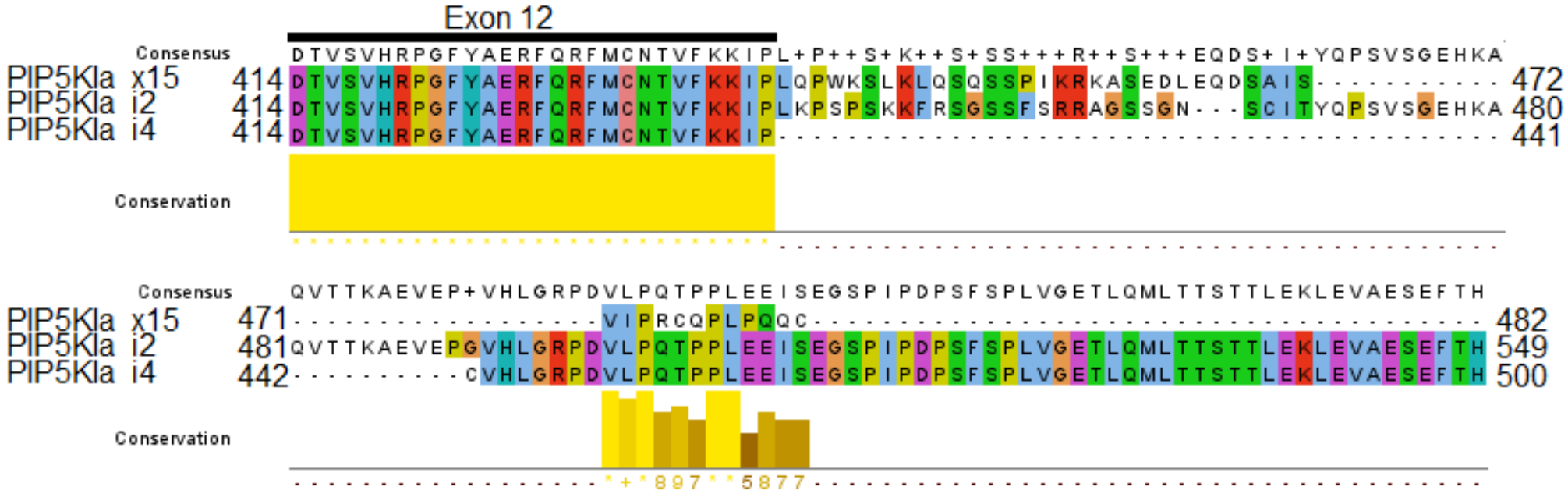


Figure 34. Multiple sequence alignments of the alternate C-terminal sequences of PIP5KIα isoforms: canonoical (PIP5KIα i2 and i4) and novel (PIP5KIα x15).
The C-terminus of PIP5KIα i2 is identical to the C-terminus of i1. The C-terminus of x15 is identical to that of x16. Residues encoded by Exon 12, and thus common to all isoforms are indicated by a black bar.

Table 20. Computationally predicted Nuclear Localisation Signals within the various C-termini of PIP5KIα isoforms

Predictions were generated with the NLSmapper tool [166]. The higher the score, the more confidence the prediction software has in the veracity of the NLS.

Submitted sequence	Predicted NLS	NLS Type	Score
427-550 PIP5KIα i2	• PSPSKKFRSGSSF	Monopartite	5.0
	• RSGSSFSRRAGSSGNSCIT YQPSVSGEHKAQVTT	Bipartite	2.1
427-500 PIP5KIα i4	None		-
427-485 PIP5KIα x15	• SPIKRKASED	Monopartite	8.0
	• RFMCNTVFKKIPLQPWKSL KLQSQSSPIKRKASED	Bipartite	3.3
	• PIKRKASEDLEQDSAISVIP RCQPLPQ	Bipartite	5.2

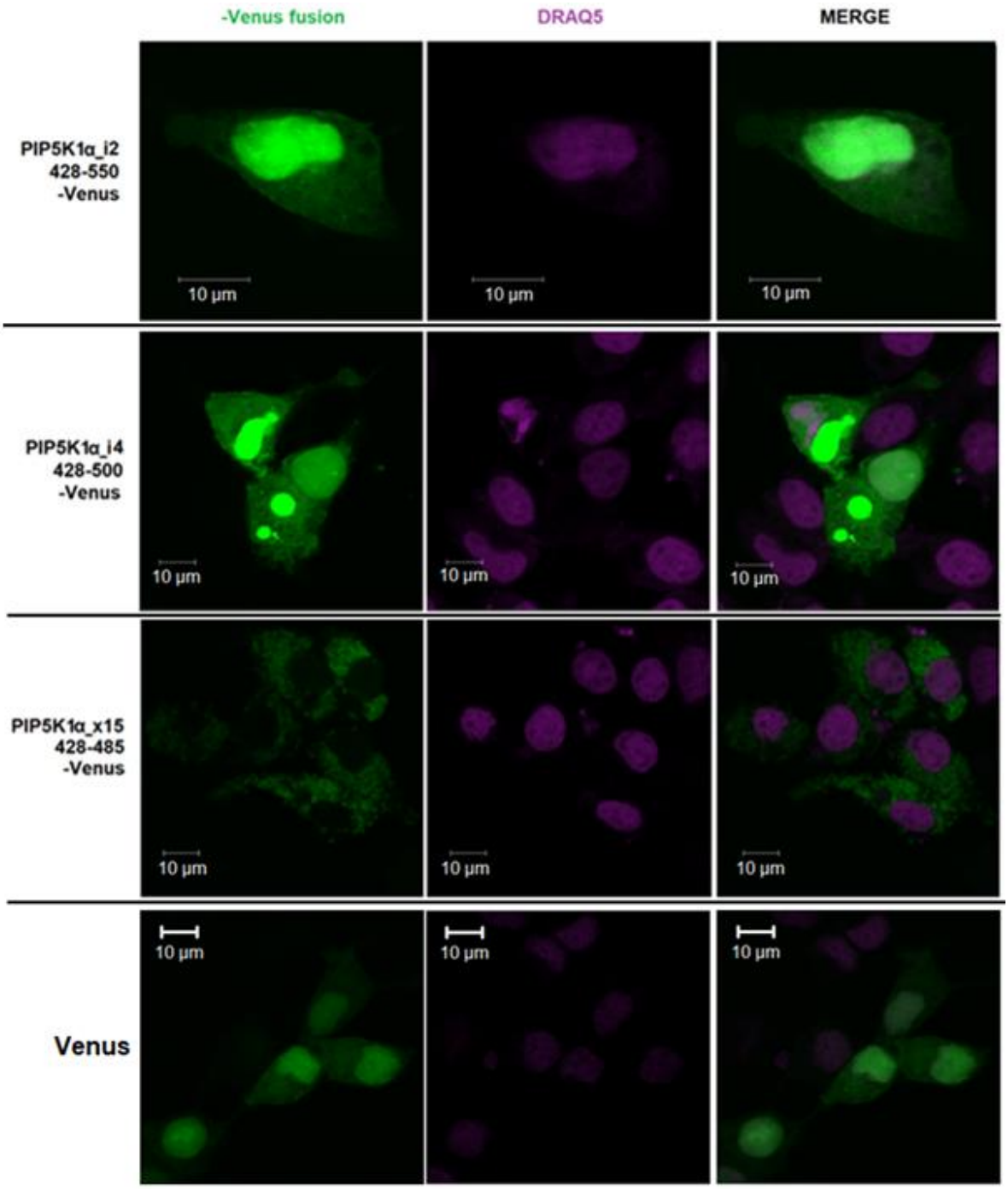
As Mellman et al. had mapped nuclear speckle targetting to residues 440-562 of PIP5KIα i1 (identical to PIP5KIα i2 427-549), this finding was used as the basis for further localisation experiments with the unique C-terminal portions of each isoform.

Coding sequences of residues 427-549 of PIP5KIα i2 (which are identical to residues 440-562 of PIP5KIα i1), residues 427-500 of PIP5KIα i4 and 427-485 of PIP5KIα x15 were each subcloned into the mVenus vector using the primers as listed in **Supplementary Table 9**.

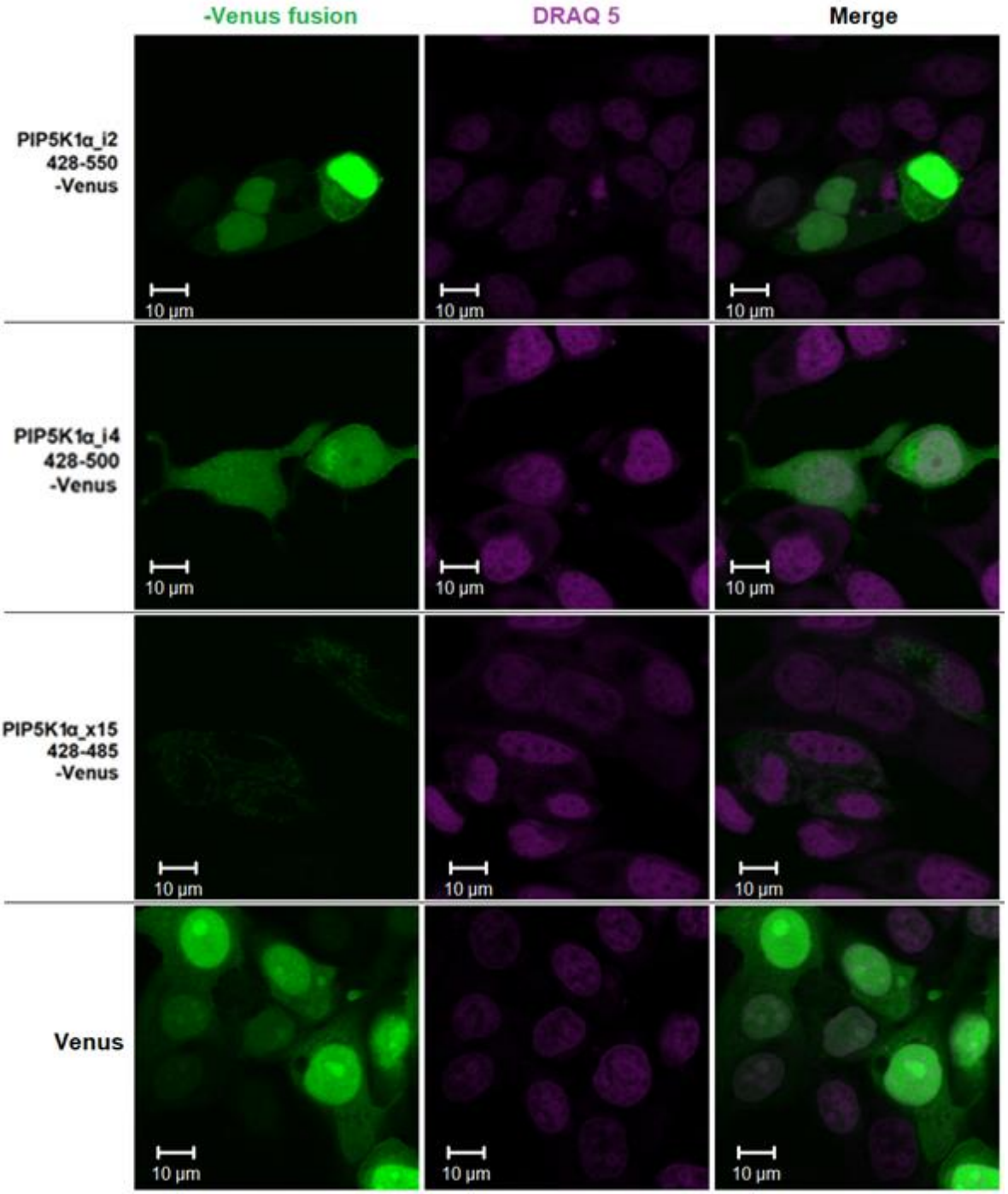
The mVenus fusions to the C-termini of the different isoforms of PIP5KIα were expressed in both LNCaP C4-2 and MCF7 cells. As illustrated in **Figure 35 A and B** the PIP5KIα i2 427-549-mVenus fusion was observed distributed throughout both cell types, including strongly within the nucleus. This distribution is indistinguishable from that of mVenus alone. Localisation specifically to nuclear speckles, as reported by Mellman et al., for HeLa cells, was not observed with the PIP5KIα i2 428-549-mVenus construct in either cell line. Possible explanations for this disparity between our results and those of Mellman et al., could be the cell type used, since Mellman et al., reported the nuclear speckle localisation in HeLa cells. Feasibly, difference in expression between the cell lines of

a key protein-binding partner, or some other factor necessary for this localisation could explain this difference between our result and Mellman et al's. Alternatively, another difference between the two experiments is the fusion partner used to detect the localisation, Mellman et al., used an HA tag and immunocytochemistry with an anti-HA antibody, whereas our experiment used direct detection via florescent protein tag. Some factor relating to the tag used could also explain the difference. There was not sufficient time to substitute the mVenus tags for HA to investigate this possibility.

A)


-Venus fusion
DRAQ5
MERGE
PIP5K1α_i2
428-550
-Venus
PIP5K1α_i4
428-500
-Venus
PIP5K1α_x15
428-485
-Venus
Venus
10 µm

B)


-Venus fusion
DRAQ 5
Merge
PIP5K1α_i2
428-550
-Venus
PIP5K1α_i4
428-500
-Venus
PIP5K1α_x15
428-485
-Venus
Venus
10 µm

C)

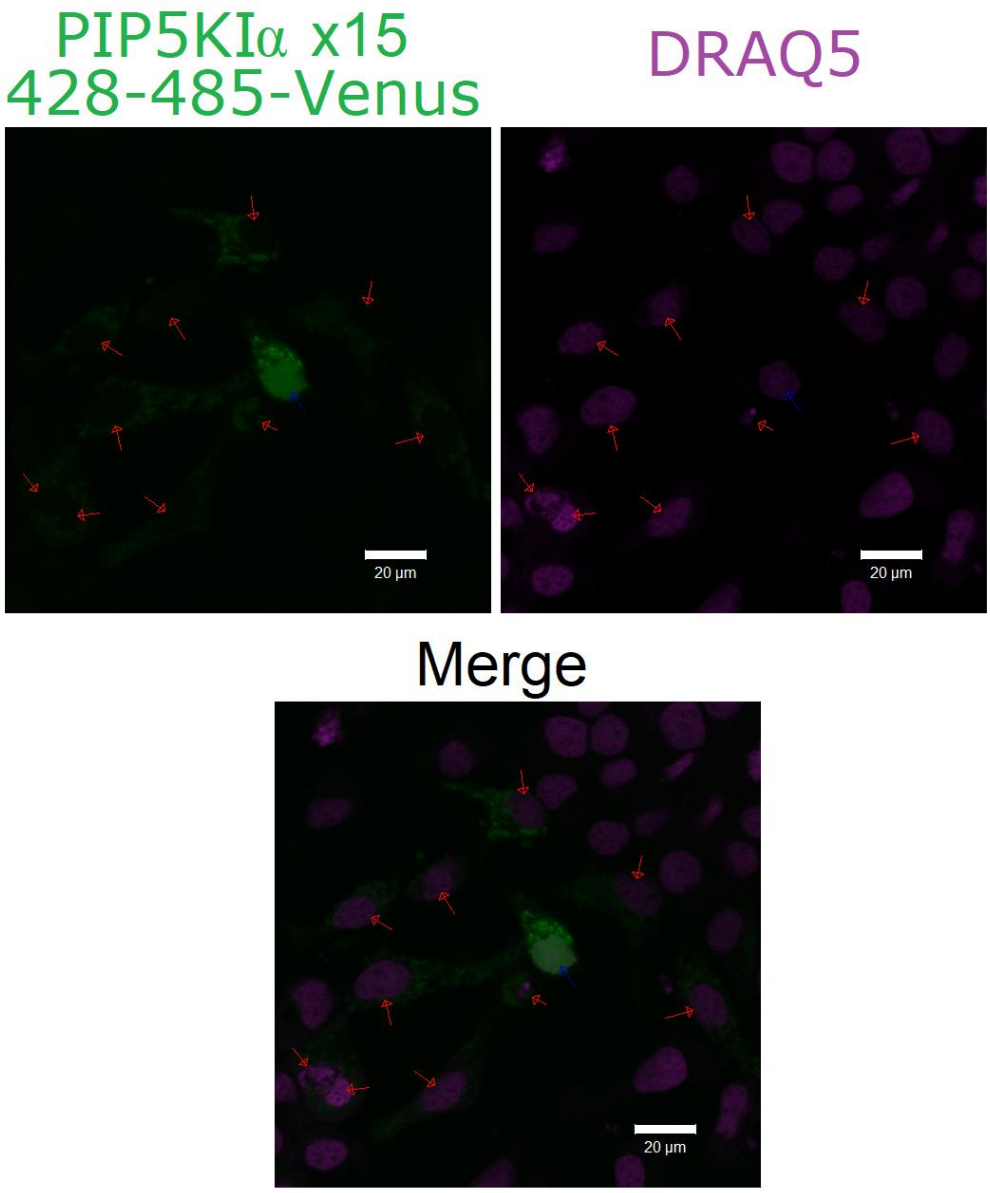


Figure 35. Subcellular localisation of variant termini of PIP5KIα isoforms.

A) and **B)** Localisation of exogenously expressed PIP5KIα C-termini-mVenus fusions in LNCaP C4-2 **(A)** and MCF7 **(B)** cells. Representative images are presented. Scale bars represent 10 μm.

C) An example of nuclear exclusion (red arrows), vs nuclear inclusion (blue arrow) of the PIP5KIα x15 428-485-mVenus fusion expressed in LNCaP C4-2. Scale bars represent 20 μm.

mVenus-tagged fusions (green) and DNA staining (purple). Imaged using fluorescence microscopy.

The PIP5KIα i4 428-500-mVenus fusion was also present throughout cells, though with a lower intensity in nuclei than was observed for the PIP5KIα i2 428—549-mVenus or mVenus alone. Again, localisation specifically to nuclear speckles, as reported by Mellman et al., was not observed in either cell line. Regions of intense fluorescence ranging from 2-12 μm in diameter were also observed outside of nuclei in ~60% of LNCaP C4-2 cells expressing this construct. PIP5KIα has been previously reported to localise to endosomes [167], though the observed regions appear too large to be endosomes (which are ~100–500 nm in diameter; [168]). Such structures were not apparent in either cell type transfected with any of the other of the constructs, and these structures were not apparent in MCF7 cells transfected with the same construct. It was not clear if these observations indicate aggregates of the protein in LNCaP C4-2, which are not observed in MCF7 cells, or whether this represents localisation to a subcellular organelle. Although this is potentially an intriguing finding, it was not investigated further.

The PIP5KIα x15 428-485-mVenus fusion was present within the cytoplasm, but absent from nuclei in the majority (77 cells; 89.5%; n = 86) of transfected cells (**Figure 35**), indicating exclusion from the nucleus. Given the ability of mVenus or mVenus-fusions to enter the nucleus, this exclusion is not

therefore likely due to the mass of the protein, but may indicate active nuclear export. Categorisation of transfected cells as displaying nuclear inclusion vs nuclear exclusion was performed systematically by the following criteria:

- "Nuclear exclusion"- cells with green signal, but weak or no green signal in regions staining with DRAQ5 (nuclei).
- "Nuclear inclusion" - cells with green signal in cell body, and similar or stronger green intensity in regions staining with DRAQ5.

Examples for both "Nuclear exclusion" and "Nuclear inclusion" categories are presented in **Figure 35 C**.

The presence of this construct in nuclei of some transfected cells (9 cells; 10.5%) could indicate cell-cycle phase dependent localisation, for example as is reported for PIP5KIα i4, which is present in nucleoli specifically during G2/S phase [122].

This absence of PIP5KIα x15 428-485-mVenus from the nuclei of the majority of transfected cells, in contrast to the presence in nuclei of PIP5KIα i2 428-550-mVenus and PIP5KIα i4 428-500-mVenus, was an unexpected finding given the *in silico* prediction of a strong NLS within the C-terminus of PIP5KIα x15. This finding suggests the presence of a nuclear export signal (NES) within the C-terminus of PIP5KIα x15.

To investigate possible presence of an NES within the C-terminus of PIP5KIα x15, two online tools, LocNES [169] and NetNES 1.1 [170] were used, predicting presence of three nuclear export signals (**Figure 36 A**):

- Predicted NES 1 - ^{428}RFQRFMCNTVFKKIPL^{443}
- Predicted NES 2 - ^{441}IPLQPWKSLKL^{451}
- Predicted NES 3 - ^{458}IKRKASEDLEQDSAISVI^{475}

To determine if any of these functions as an NES, the PIP5KIα x15 (428-485)-mVenus plasmid was mutagenized to delete different sections of PIP5KIα x15 C-terminus. Primer pairs used for this are listed in **Supplementary Table 10.**

This generated two mVenus fusion constructs: the first, PIP5KIα x15 428-443-mVenus, which encodes just Predicted NES 1; the second, PIP5KIα x15 441-485-mVenus, encodes Predicted NES 2 and Predicted NES 3, but lacks Predicted NES 1 (**Figure 36 A**).

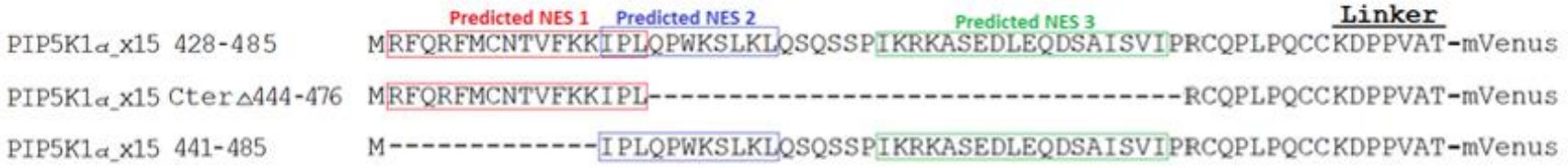
Both constructs were expressed in LNCaP C4-2 and MCF7 cells. PIP5KIα x15 428-443-mVenus displayed absence from nuclei, whereas PIP5KIα x15 441-485-mVenus did not (**Figure 36 B**), thus suggesting the existence of a nuclear export signal within residues 428-443.

Classical nuclear export signals adhere to the consensus ΦXX(X)ΦXX(X)ΦXΦ, where Φ is a hydrophobic amino acid

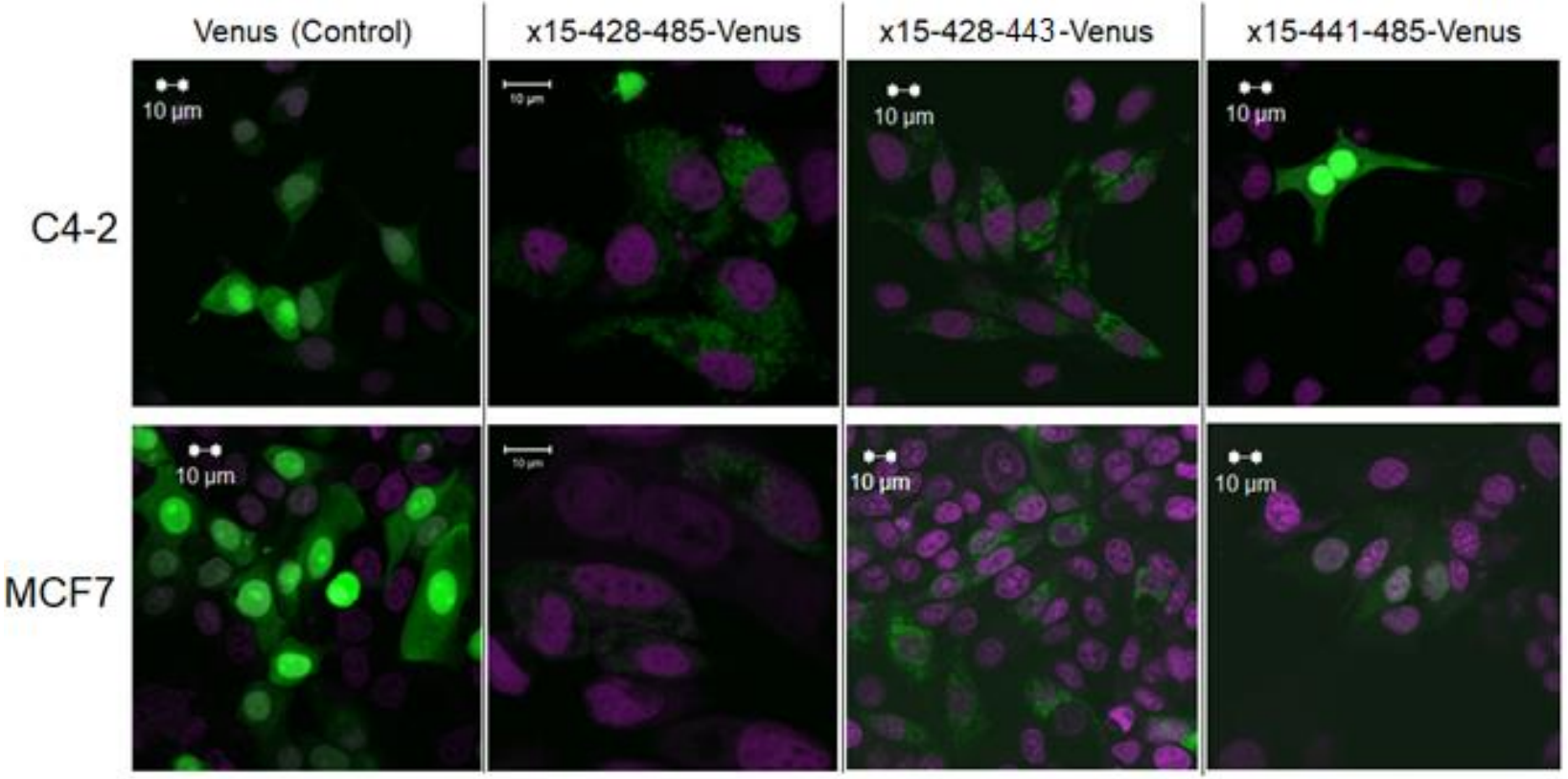
(commonly L, but also I, F, V or M) and X represents any amino acid [171]. These sequences are recognised by the exportin Crm1, facilitating export from nuclei via nuclear pores [172]. PIP5KIα x15 428-443, RFQRFMCNTVFKKIPL, contains such a sequence, with three additional hydrophobic positions within the following eight residues (**Figure 36 C**). To test the hypothesis that ^{428}RFQRFMCNTVFKKIPL^{443} is an NES, further experiments mutagenizing these hydrophobic residues to alanine (a commonly used technique when mapping NESs, e.g. [173]) were begun, but could not be completed in time for inclusion in this work.

Curiously, residues 428-443 of isoform x15 are encoded entirely by Exon 12, and are thus present in every isoform of PIP5KIα. The presence of PIP5KIα i2 (441-550)-mVenus and PIP5KIα i4 (428-550)-mVenus in nuclei, in contrast to absence from nuclei of PIP5KIα x15 (428-485)-mVenus, therefore suggests that the C-termini of PIP5KIα i2 and PIP5KIα i4 additionally contain features that antagonise nuclear export and/or that promote nuclear inclusion, whereas the C-terminus of PIP5KIα x15 does not.

A)


Predicted NES 1
Predicted NES 2
Predicted NES 3
Linker
PIP5K1α_x15 428-485 MRFQRFMCNTVFKKIPLQPWKSLKLQSQSSPIKRKASEDLEQDSAISVIPRCQPLPQCCKDPPVAT-mVenus
PIP5K1α_x15 Cter△444-476 MRFQRFMCNTVFKKIPL----------RCQPLPQCCKDPPVAT-mVenus
PIP5K1α_x15 441-485 M----------IPLQPWKSLKLQSQSSPIKRKASEDLEQDSAISVIPRCQPLPQCCKDPPVAT-mVenus

B)


Venus (Control)
x15-428-485-Venus
x15-428-443-Venus
x15-441-485-Venus
C4-2
MCF7
10 µm

C)

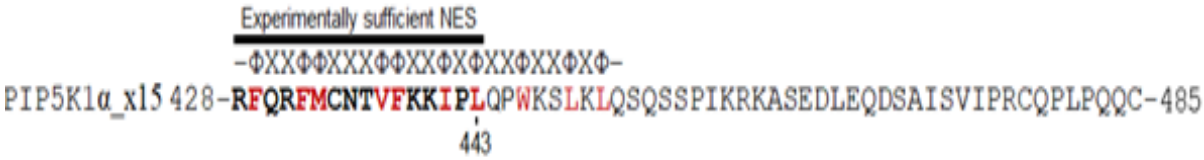


Figure 36. Mapping of a putative NES within the C-terminus of PIP5KIα x15

A) Sequences of deletion constructs made to assess and the three predicted NESs within the C-terminus of PIP5KIα x15. Predicted NES1, NES 2 and NES 3 are indicated in the red, blue and green boxes, respectively.

B) Localisation of the fusion proteins PIP5KIα x15 428-428-mVenus, PIP5KIα x15 428-443 mVenus and PIP5KIα x15 441-485 -mVenus. Scale bars represent 10 μm.

C) The C-terminus of PIP5KIα x15 contains a sequence similar to the consensus classical nuclear export signal, with multiple hydrophobic residues. PIP5KIα x15 428-443, RFQRFMCNTVFKKIPL, is sufficient in the experimental system to promote nuclear export and is therefore a putative NES. Within the following eight residues there are three additional hydrophobic residues, which could contribute to nuclear export, but were are not necessary for nuclear export in the experimental system.

4.10.3. Cloning of a Star-PAP-mCherry expression vector

A previous study proposed a direct interaction of PIP5KIα with the nucleolar protein polyA polymerase Star-PAP (the encoding gene is *TUT1*) [149]. In order to investigate nuclear speckle localisation of the PIP5KIα C-termini-mVenus constructs, and the possibility that such localisation is dependent on expression of Star-PAP, a Star-PAP-mCherry expression vector was made. mCherry is a monomeric derivative of the fluorescent protein dsRED with a Exλ of 587nm and Emλ of 610 nm, allowing for co-expression and detection of both mCherry and mVenus tagged proteins. First the mVenus coding sequence (BamHI to NotI 1407 bp fragment) within the pmVenus-N1 plasmid was swapped for mCherry coding sequence, amplified from the pcDNA3.1_AktPH-mCherry plasmid (a gift from Moritoshi Sato. Addgene plasmid #67301 [174]) with the primers as listed in **Supplementary Table 11.**

The pcDNA3.1_AktPH-mCherry plasmid does not contain a suitable MCS for the creation of fusion-protein expression plasmids, hence the generation of the new plasmid, pmCherryN1. This cloning strategy also substituted a unique AgeI site within the MCS for a MluI site, but otherwise the MCS of pmCherryN1 is identical to that of pmVenusN1. This

subcloning strategy to generate pmCherryN1 was devised by myself, but performed by Ryan Morris.

Once the pmCherryN1 had been created, I amplified the coding sequence of Star-PAP isoform 1 (NP 073741.1) from LNCaP C4-2 cDNA (primers in **Supplementary Table 11**) and inserted this into pmCherryN1 between the XhoI and MluI sites. Having generated this Star-PAP-mCherry expression plasmid, it was transfected into LNCaP C4-2 cells, both alone and in co-transfection with the various PIP5KIα C-terminal variants fused to mVenus. Detection of the Star-PAP-mCherry fusion protein was attempted, but could not be achieved. Unfortunately, there was insufficient time to investigate of PIP5KIα/Star-PAP co-localisation further.

4.11. Evolutionary conservation of CAG repeat variation in *PIP5K1A* transcripts

Having noticed the variable inclusion of a CAG nucleotide triplet in several reported transcripts of *PIP5K1A*, the genomic sequence proximal to the 5' end of *PIP5K1A* Exon 2 was assessed, leading to the conclusion that the variable inclusion of CAG resulted from presence of a tandem splice-acceptor site. Only much later in the course of this PhD project did I find the paper by Hiller *et al* [152] in which the authors investigated the occurrence of such tandem splice-acceptors in the human transcriptome. Hiller *et al*., also proposed that the 5' end of *PIP5K1A* Exon 2 contains a tandem splice-acceptor, leading to the variable inclusion of the CAG nucleotide triplet in *PIP5K1A* transcripts, by the same mechanism that I propose.

The evolutionary conservation of this phenomenon was briefly investigated by the assessment of *PIP5K1A* transcripts and genomic sequences at the 5' end of Exon 2 for several mammalian species. The approach was not exhaustive – several species of each clade for which *PIP5K1A* transcripts are listed in NCBI were assessed. This phenomenon was observed in *PIP5K1A* transcripts of Primate and Rodent species, but not for any species of any other clade.

For all primate species accessed (including all extant great apes, and representatives of lesser apes and both Old World and New World monkeys), similar variable inclusion of CAG (AAG in one observed instance; *Sapajus apella*) at Exon 2 of *PIP5K1A* transcripts was observed, with corresponding presence of NAGNAG tandem splice-acceptors at the 5' end of Exon 2 in the genomic sequence (**Figure 37, A-F**).

Pip5k1a transcripts of three rodent species, *Mus musculus, Rattus norvegicus* and *Castor canadensis* were likewise assessed. The genomic sequence at the 5' end of *Pip5k1a* Exon 2 in *Mus musculus, Rattus norvegicus* is AGCAGCAG, containing three AG dinucleotides and thus feasibly three splice-acceptor sites. However, it appears that only two of these possible splice-acceptor sites are used, with intron spanning reads connecting to most 5' AG and the middle AG, but not the most 3' (**Figure 37, G and H**).

In contrast, the equivalent position in the genome of *Castor canadensis* is AGCATCAT, thus containing only one AG dinucleotide; reported *Pip5k1a* transcripts for this species do not show any such variation at this exon junction (**Figure 37 I**), supporting the hypothesis that a tandem splice acceptor sequence at the Exon 2 5' end is responsible for the varying inclusion of CAG in the transcripts of the various other species.

The ratio of intron spanning reads joining to each splice-acceptor site in these species was assessed (**Table 24**) indicating a bias to use of the second splice-acceptor site in all Hominoidea (ape) species assessed, whereas in *Sapajus apella*, a New world monkey, and *Macaca mulatta,* an Old world monkey, there was a bias towards use of the first acceptor splice site. For the two rodent species assessed, there was bias towards the use of the first splice site.

Analysis of the ratio of use of first vs second splice acceptor site in each of the species in which this phenomenon was observed indicates that the frequency varies greatly, with bias to use of second splice acceptor among the Hominoidea, and bias to use of the first splice acceptor site among the two non-ape Primates and the two Rodentia species assessed.

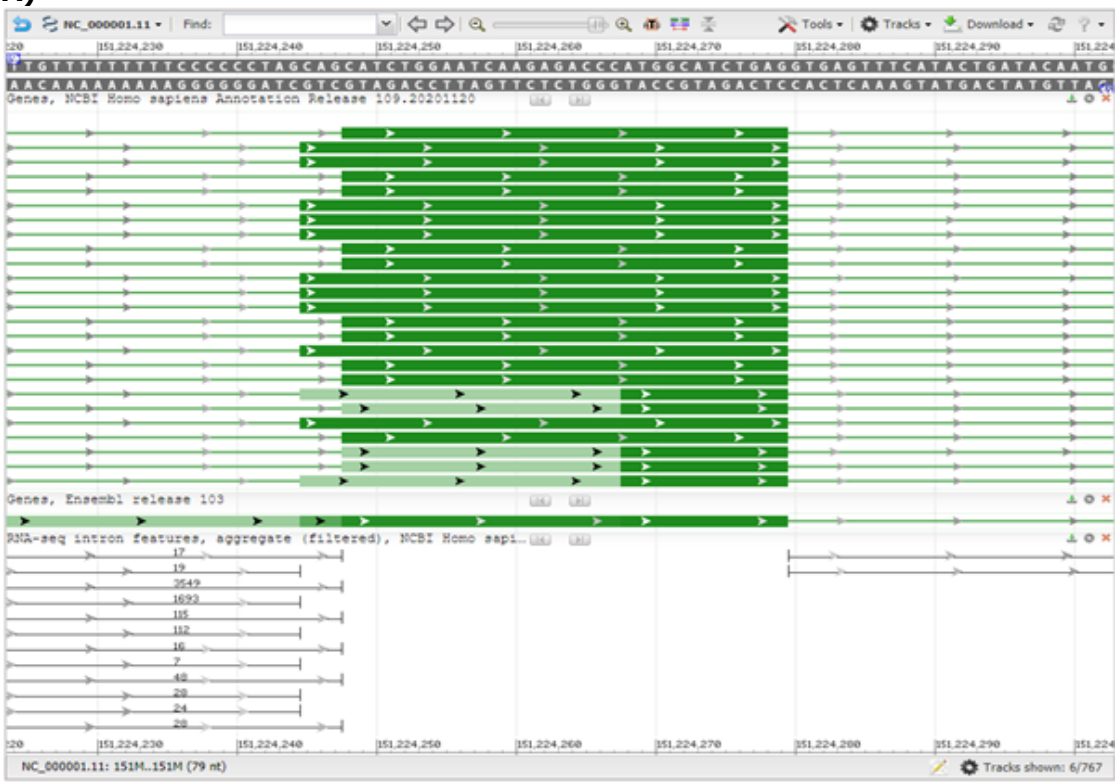
In conclusion, the phenomenon of varying inclusion in *PIP5K1A* transcripts of a CAG nucleotide triplet at the 5’ end of Exon 2 occurs in numerous Primate and Rodent species. Analysis of the genomic and transcript sequences of *PIP5K1A* for these species confirms the phenomenon results from presence of tandem splice acceptors at the 5’ end of Exon 2, since for species without repeated AG dinucleotides at the 5’ end of *PIP5K1A* Exon 2, only one type of splice event entering Exon 2 is observed. The exact sequence of the dual tandem

splice acceptor varies between species of the two orders varies, with a third CAG in transcripts of rodent species. This phenomenon was not observed in the Lagomorpha order, which is closely related to the Rodentia. Along with the lack of this phenomenon among members of every other order analysed, this indicates that the ancestral genotype included only a single splice acceptor site at *PIP5K1A* Exon 2, and that separate mutations in an ancestral Primate and an ancestral Rodent species (after the divergence of Myomorpha and Catoriomorpha) independently gave rise to this phenomenon in the two orders. Though seemingly unlikely, this is the most parsimonious explanation for the observations, requiring the fewest individual mutation events.

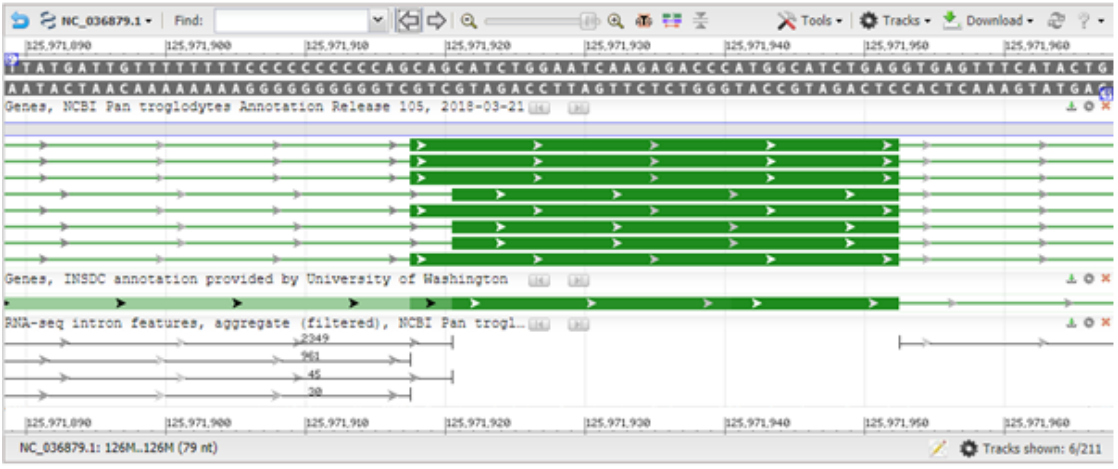
Table 24. Number of intron spanning reads joining to the equivalent of *PIP5K1A* Exon 2 for several Primate and Rodent species.

Species	Number of reads joining to first acceptor site	Number of reads joining to second acceptor site	Ratio of reads joining to first vs second acceptor site
Homo sapiens	1883	3773	0.50 : 1
Pan troglodytes	961	2379	0.40 : 1
Gorilla gorilla gorilla	5	41	0.12:1
Pongo abelii	307	1161	0.26:1
Hylobates moloch	27	48	0.56:1
Sapajus apella	1947	1208	1.61:1
Macaca mulatta	2749	191	14.39:1
Mus musculus	2138	2032	1.05: 1
Rattus norvegicus	5158	4199	1.23:1

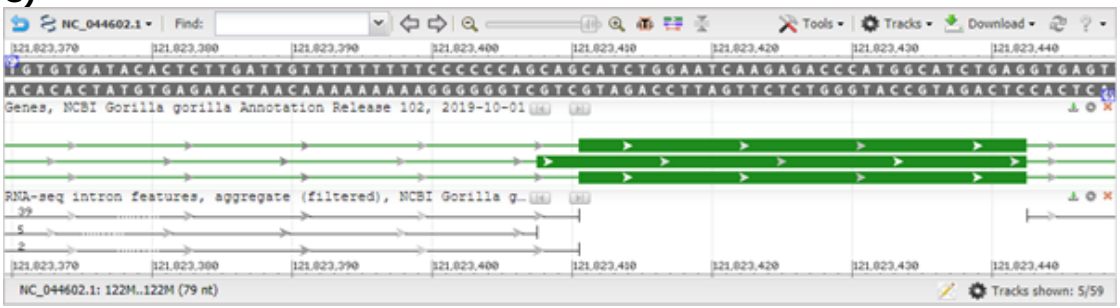
A)


NC_000001.11
Find:
Tools
Tracks
Download
Genes, NCBI Homo sapiens Annotation Release 109.20201120
Genes, Ensembl release 103
RNA-seq intron features, aggregate (filtered), NCBI Homo sapi...
17
19
3549
1693
115
112
16
7
48
28
24
28
NC_000001.11: 151M..151M (79 nt)
Tracks shown: 6/767

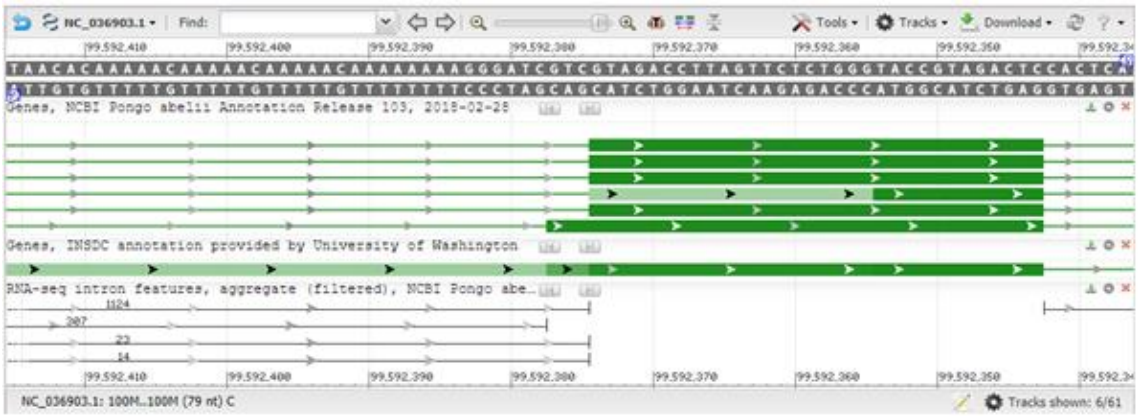
B)


NC_036879.1
Find:
Tools
Tracks
Download
Genes, NCBI Pan troglodytes Annotation Release 105, 2018-03-21
Genes, INSDC annotation provided by University of Washington
RNA-seq intron features, aggregate (filtered), NCBI Pan trogl...
2349
961
45
30
NC_036879.1: 126M..126M (79 nt)
Tracks shown: 6/211

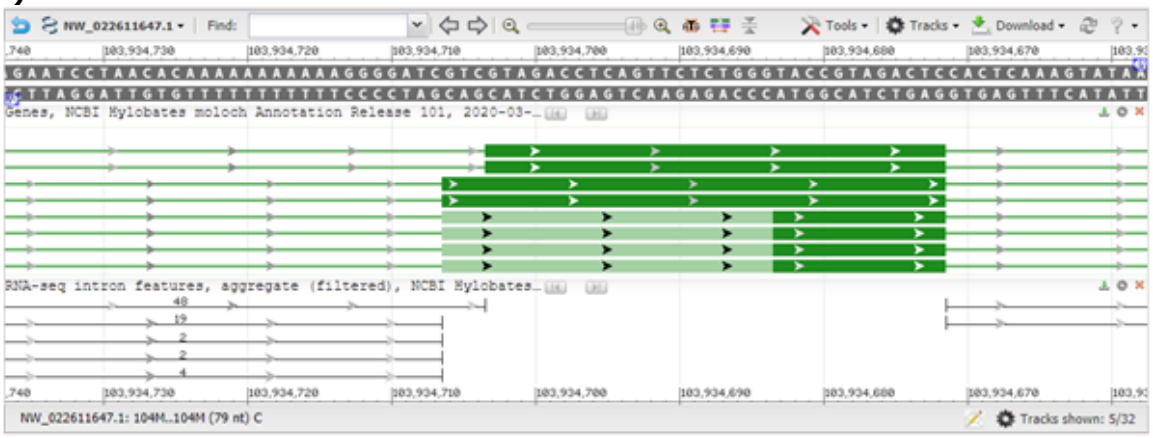
C)


NC_044602.1
Find:
Tools
Tracks
Download
Genes, NCBI Gorilla gorilla Annotation Release 102, 2019-10-01
RNA-seq intron features, aggregate (filtered), NCBI Gorilla g...
39
5
2
NC_044602.1: 122M..122M (79 nt)
Tracks shown: 5/59

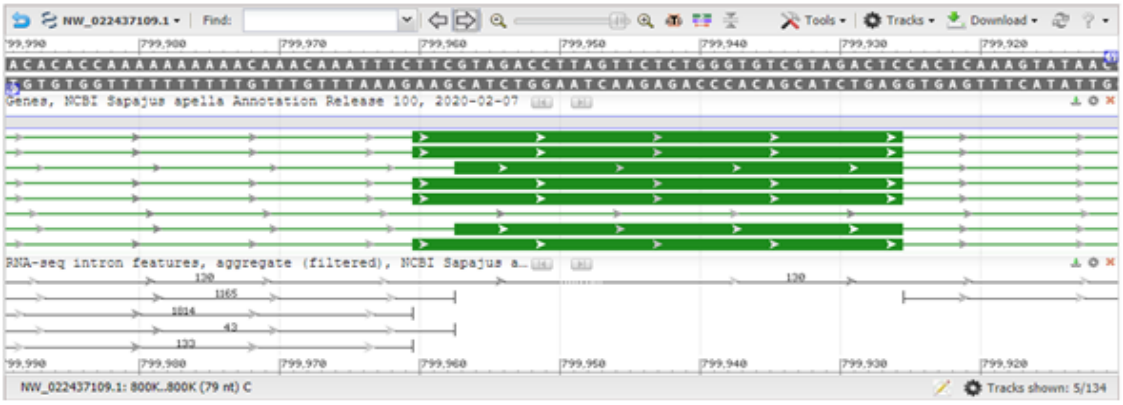
D)


NC_036903.1
Find:
Tools
Tracks
Download
Genes, NCBI Pongo abelii Annotation Release 103, 2018-02-28
Genes, INSDC annotation provided by University of Washington
RNA-seq intron features, aggregate (filtered), NCBI Pongo abe...
1124
287
23
14
NC_036903.1: 100M..100M (79 nt) C
Tracks shown: 6/61

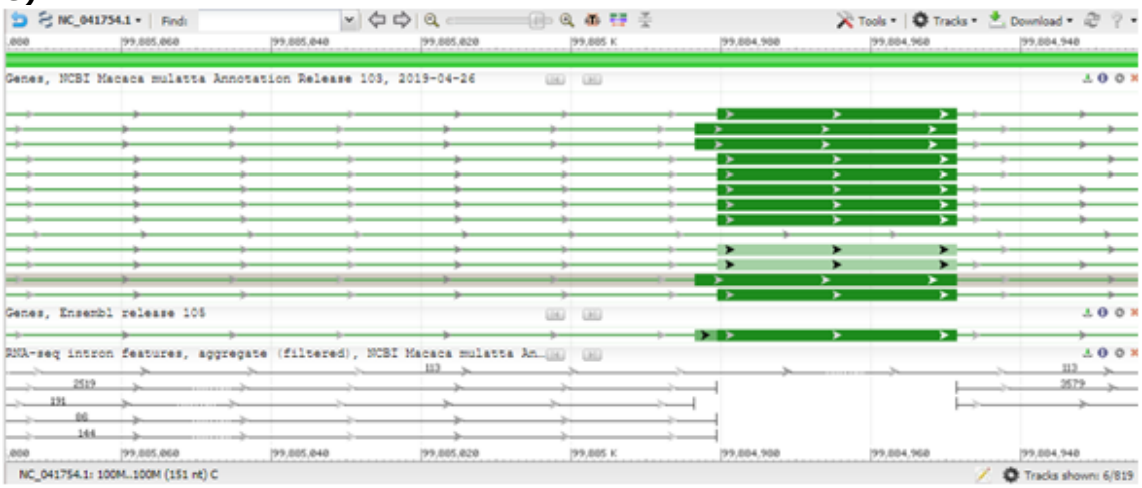
E)


NW_022611647.1
Find:
Tools
Tracks
Download
Genes, NCBI Hylobates moloch Annotation Release 101, 2020-03-...
RNA-seq intron features, aggregate (filtered), NCBI Hylobates...
48
19
2
2
4
NW_022611647.1: 104M..104M (79 nt) C
Tracks shown: 5/32

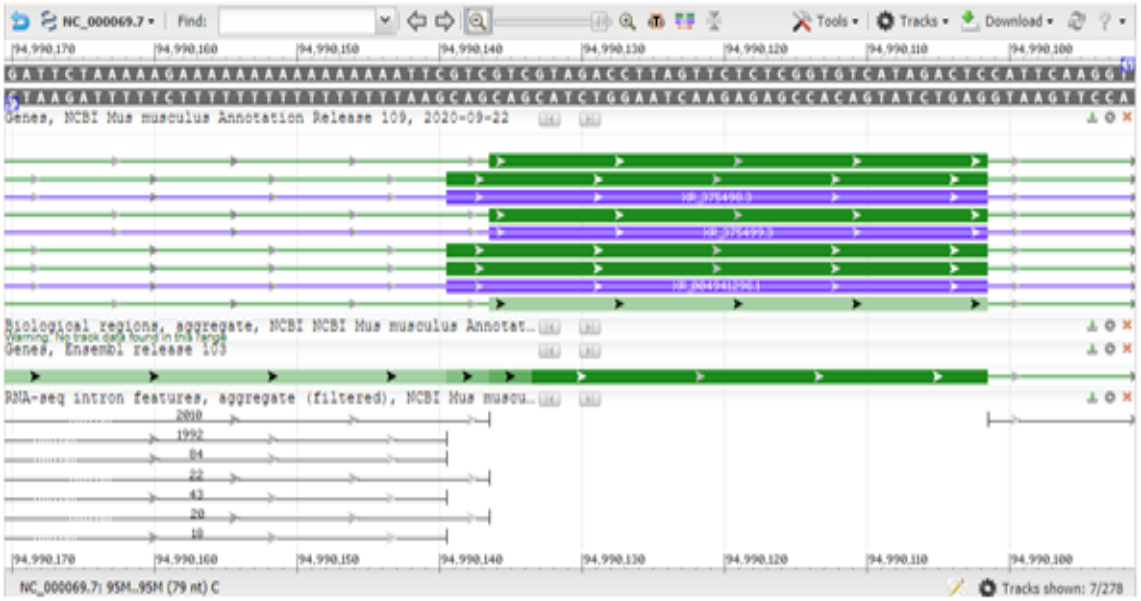
F)


NW_022437109.1
Genes, NCBI Sapajus apella Annotation Release 100, 2020-02-07
RNA-seq intron features, aggregate (filtered), NCBI Sapajus a...
NW_022437109.1: 800K..800K (79 nt) C
Tracks shown: 5/134

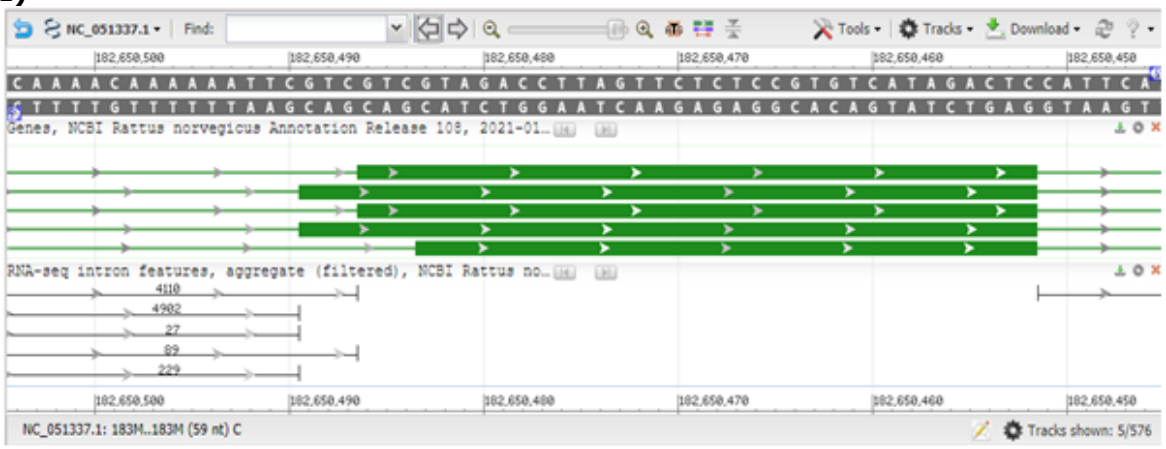
G)


NC_041754.1
Genes, NCBI Macaca mulatta Annotation Release 103, 2019-04-26
Genes, Ensembl release 105
RNA-seq intron features, aggregate (filtered), NCBI Macaca mulatta An...
NC_041754.1: 100M..100M (151 nt) C
Tracks shown: 6/819

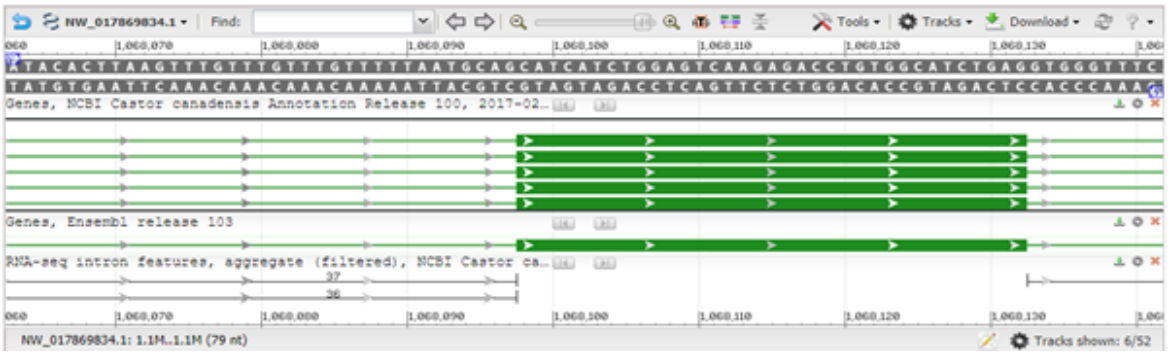
H)


NC_000069.7
Genes, NCBI Mus musculus Annotation Release 109, 2020-09-22
Biological regions, aggregate, NCBI NCBI Mus musculus Annotat...
Genes, Ensembl release 103
RNA-seq intron features, aggregate (filtered), NCBI Mus muscu...
NC_000069.7: 95M..95M (79 nt) C
Tracks shown: 7/278

I)



J)



K)

		Exon 1	Exon 2	
H.sapiens_PIP5K1A_v5	61	GUCCCUUCCUGUACCU---UGUCCUCAG---CAGCAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		120
H.sapiens_PIP5K1A_v2	61	GUCCCUUCCUGUACCU---UGUCCUCAG------CAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		117
P.troglodytes_PIP5K1A	61	GUCCCUUCCUGUACCU---CGUCCUCAG---CAGCAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		120
P.troglodytes_PIP5K1A_X3	61	GUCCCUUCCUGUACCU---CGUCCUCAG------CAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		117
G.gorilla_PIP5K1A_X2	61	GUCCCUUCCUGUACCU---CGUCCUCAG---CAGCAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		120
G.gorilla_PIP5K1A_X3	61	GUCCCUUCCUGUACCU---CGUCCUCAG------CAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		117
P.abelii_PIP5K1A_X2	61	GUCCCUUCCUGUACCU---CGUCCUCAG---CAGCAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		120
P.abelii_PIP5K1A_X3	61	GUCCCUUCCUGUACCU---CGUCCUCAG------CAUCUGGAAUCAAGAGACCCAUGGCAUCUGAG		117
S.apella_PIP5K1A_X6	61	GUACCUUCCUGUACCU---CGUCCUCAG---AAGCAUCUGGAAUCAAGAGACCCACAGCAUCUGAG		120
S.apella_PIP5K1A_X7	61	GUACCUUCCUGUACCU---CGUCCUCAG------CAUCUGGAAUCAAGAGACCCACAGCAUCUGAG		117
M.mulatta_PIP5K1A	61	GUCCCUUCCUGUACCU---CGUCCUCAG---CAGCAUCUGGAAUCAAGAGACCUGUGGCACCUGAG		120
M.mulatta_PIP5K1A_X6	61	GUCCCUUCCUGUACCU---CGUCCUCAG------CAUCUGGAAUCAAGAGACCUGUGGCACCUGAG		117
M.musculus_Pip5k1a_v1	58	GCCCCUGCUGGUACC---------GCAGCAGCAGCAUCUGGAAUCAAGAGAGCCACAGUAUCUGAG		114
M.musculus_PIP5K1A_v2	58	GCCCCUGCUGGUACC---------GCAG---CAGCAUCUGGAAUCAAGAGAGGCCACAGUAUCUGAG		111
R.norvegicus_PIP5K1A_X1	58	GUCCCUGCCGGUACC---------GCAGCAGCAGCAUCUGGAAUCAAGAGAGGCACAGUAUCUGAG		114
R.norvegicus_PIP5K1A_X2	58	GUCCCUGCCGGUACC---------GCAG---CAGCAUCUGGAAUCAAGAGAGGCACAGUAUCUGAG		111
C.canadensis_PIP5K1A_X4	58	GUCCCUUCCACUACCUCGAUCUCAUCGG---CAUCAUCUGGAGUCAAGAGACCUGUGGCAUCUGAG		120

L)

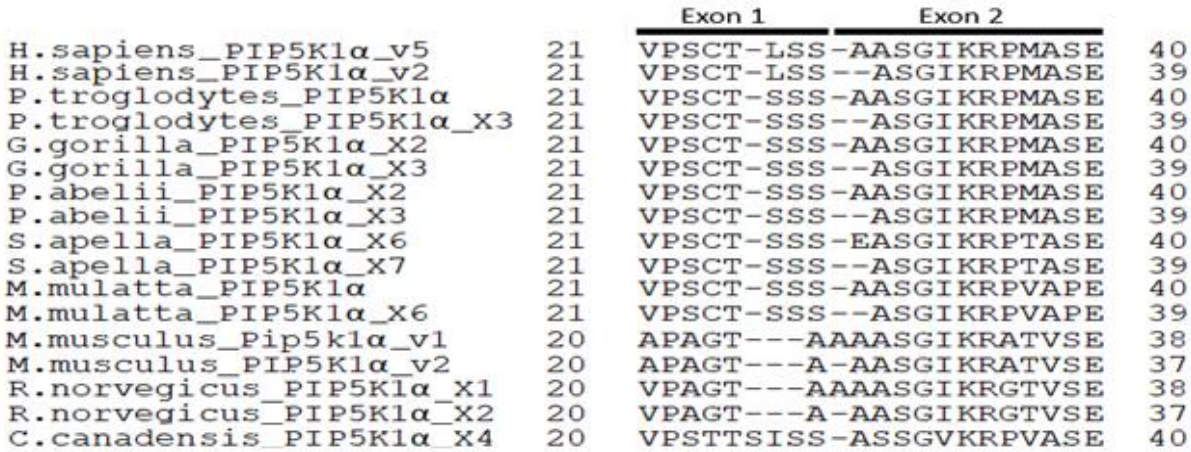


Figure 37. Alternate splice-acceptor site use at the 5' end of Exon 2 of PIP5K1A transcripts for various species.

A-I, screen grabs of NCBI Genomic regions, transcripts and products, at genomic regions proximal to PIP5K1A Exon 2, showing transcripts and Intron spanning reads for several species: ***(A)*** Homo sapiens*;* ***(B)*** Pan troglodytes*;* ***(C)*** Gorilla gorilla gorilla*;* ***(D)*** Pongo abelli*;* ***(E)*** Hylobates moloch ***(F)*** Sapajus apella*;* ***(G)*** Macaca mulatta*;* ***(H)*** Mus musculus; ***(I)*** Rattus *norvegicus;* ***(J)*** Castor canadensis. ***(K and L)*** *Sequence alignments of various PIP5K1A transcripts/ their deuced amino acid sequences at the Exon 1-Exon 2 junction for various primate and rodent species. Indicated positions are relative to the AUG start codon/first methionine.*

4.12. Summary and discussion

In summary, we have used bioinformatic analysis of RNA-Seq datasets to identify alternate exons and splice events of *PIP5K1A*, and confirmed the inclusion of these in spliced transcripts of *PIP5K1A* in two human cancer cell lines, LNCaP C4-2 and MCF7.

This work demonstrates that there is considerable variation in 5' exons in *PIP5K1A* transcripts with variable inclusion of at least 7 exons in addition to the canonical Exon 1 (**Figure 38**). In addition to use of these alternate exons, we also discovered transcripts in which a portion of the canonical Exon 1 of *PIP5K1A* functions as cryptic intron, thus further adding to the complexity of *PIP5K1A* 5' exon variation.

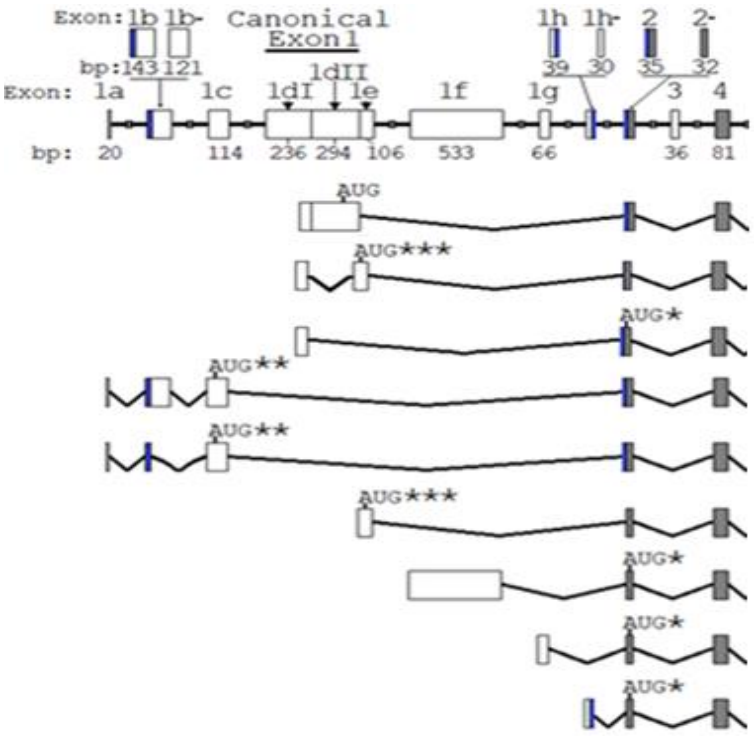


Figure 38. Complex variation in 5' Exon use in full length *PIP5K1A* transcripts demonstrated in this study.

The biological significance of this complexity in *PIP5K1A* 5’ exon use remains unknown:

Blastn and Blastp searches indicated that there is evolutionary conservation of use of these alternate exons of *PIP5K1A* among mammals, hinting at biological significance, though the time of conservation varies from ~136 MYA (Exon 1h like sequences; conserved among members of placentalia and marsupalia lineages) to ~23 MYA (Exon 1a, 1b and 1c like sequences; primate specific).

Conversely, available RNA-Seq data suggests that splice events, and thus *PIP5K1A* transcripts involving these exons are rare, seemingly limiting their possible biological importance, though this does not preclude niche or conditional functions for such transcripts and any proteins they encode.

Validation of the rarity of inclusion of these alternate exons in *PIP5K1A* transcripts, by RTqPCR, was not pursued due to time limitations, though rarity of use of Exon 1f and the skipping of the cryptic intron in the canonical Exon 1 were indicated by the end-point PCRs performed.

Functional studies of the proteins encoded by these transcripts with alternate 5’ exons were not pursued. Such studies could involve localisation experiments with the unique N-termini as pursued in **Section 4.10.2** with the unique C-

termini, or screens for protein-protein interactors with the alternate PIP5KIα N-termini.

If pursuing such work it may be prudent to also:

1) Demonstrate the translatability of such alternate *PIP5K1A* transcripts (e.g. by luciferase reporter assays with the 5' UTRs of the alternate transcripts), or the existence of the putatively encoded PIP5KIα isoforms *in vivo.*

2) Determine biological functions of the N-terminus of canonical PIP5KIα proteins, such that studies of differential function could be established.

Perhaps more interestingly, variation in 3' exon use in *PIP5K1A* transcripts was demonstrated, confirming the existence of Exon 12 to Exon 15 (i.e. Exon 13 and Exon 14 removed) splicing events.

Existence of such transcript is interesting because, by a phenomenon known as dual coding, the deduced C-terminal sequence they encode is distinct from that of canonical PIP5KIα isoforms. The C-termini of canonical isoforms of PIP5KIα are reported to promote nuclear speckle and nucleolar localisation of PIP5KIα [122; 149]. The former regulates oxidative stress and DNA damage responses (by activating Star-PAP to regulate poly(A) tail processing of select RNAs [149; 175]), while the former

facilitates interaction with chromatin silencing complexes, though the function of this interaction is currently unknown.

We thus investigated subcellular localisation of (mVenus tagged) canonical and novel (isoform x15) isoforms of PIP5KIα that vary in their C-terminal sequences. We could not recapitulate aforementioned reports of nuclear localisation of PIP5KIα, observing instead strong cell membrane localisation of all constructs. Additional experiments, such as treatment with the DNA damaging agent cisplatin, or cell cycle blockade at G1/S phase, both of which reportedly induce PIP5KIα nuclear localisation [122; 135], could be attempted with these full length PIP5KIα-mVenus fusions to investigate differences the full length isoforms' nuclear localisation.

We did however observe differences in subcellular localisation of the distinct C-termini of the PIP5KIα isoforms, finding nuclear localisation of the C-termini of canonical PIP5KIα isoforms, but exclusion from nuclei of the C-terminus of the novel PIP5KIα x15. This finding, and *in silico* predictions indicate the presence of NES, which we mapped to PIP5KIα x15 residues 427-443. These residues are also present in the C-termini of PIP5KIα i2 and i4, which were present within nuclei, suggesting that the subsequent residues in PIP5KIα i2 and i4 induce nuclear localisation (which agrees with previous reports [122; 149]), or

oppose nuclear export/cytoplasmic localisation. Experiments to further test the identity of the PIP5KIα x15 427-443 sequence as an NES, such as alanine mutagenesis scanning of hydrophobic residues, or use of Leptomycin B (which inhibits CRM1 mediated nuclear export) were planned but could not be performed due to time limitations.

This Exon 12- Exon 15 splice event was determined to be evolutionarily conserved among mammals, dating to ~100 MYA. This hints at function for such transcripts/isoforms though seemingly it is not essential, since, for example, no members of the glires clade were found to possess such transcripts.

Quantification presented in this study of the Exon 12-Exon 15 splice event present in our LNCaP C4-2 and public RNA-Seq datasets indicates this event is present in ~ 4 to 7.5% of *PIP5K1A* transcripts. *In silico* quantification with the Kallisto software places this as high as 16% of PIP5KIα transcripts (Ryan Morris and D M Herry pers comm). This is at least 3 fold higher than reads corresponding to one of the canonical isoforms, PIP5KIα i2.

Endogenous expression of the encoded isoform (PIP5KIα x15) was not investigated, though previous western blotting of LNCaP C4-2 lysates with an antibody against an epitope in the N-terminus of PIP5KIα (and thus common to the canonical

isoforms of PIP5KIα and PIP5KIα x15) did show a band of approximately the expected size for PIP5KIα x15. Because PIP5KIα i4 and PIP5KIα x15 have predicted MWs within 1.5 kDa of each other, and transcripts for both isoforms are expressed in LNCaP C4-2, designation of this band as PIP5KIα is not possible. We considered generation of antibody specific to distinct C-terminus of PIP5KIα x15, to investigate endogenous expression of this isoform, but the several companies with which we consulted each indicated the low immunogenicity of this sequence, and high possibility of cross-reactivity of any antibody generated against this sequence. We thus did not pursue anti PIP5KIα x15 isoform specific antibody further. An alternative approach to investigate endogenous expression of PIP5KIα x15 include mass spectrometry.

Validation of the relative abundance of the *PIP5K1A* Exon 12–Exon 15 vs Exon 12–Exon 13 vs Exon 12–Exon 14 splice events was pursued by use of primers specific each Exon 12-Exon 13/14/15 junction, but resulted in amplification of non specific bands in preliminary tests and this approach was set aside. Feasibly the Exon 12 - Exon 13/14/15 splicing is a regulated phenomenon, and differential expression of the corresponding splice variants/isoforms could be associated with different cell/tissue types [176] or disease states [177]. Continued development of a qPCR assay to quantify *PIP5K1A* Exon 12 -

Exon 13/14/15 splicing events is therefore desired, as such would allow for investigation of association/function of this phenomenon with/in disease.

Chapter 5

Assessment of protein-protein interactions of PIP5KIα and AR in Yeast two hybrid assays

5.1. Introduction

Previous studies and the results presented here in Chapter 3 suggest a functional relationship between androgen hormone signalling via AR and the $PI(4,5)P_2$ signalling pathways, involving the lipid kinase enzyme PIP5KIα. This suggests a crosstalk between androgen and $PI(4,5)P_2$ signalling, by mechanisms that are currently poorly understood. Although in our hands YFP-tagged PIP5KIα isoforms are localised mainly to cell membranes in LNCaP C4-2 cells (**Chapter 4**), studies from other groups have suggested PIP5KIα localises to the nucleus and interacts with nuclear proteins (pRB, p53; Star-PAP) under some circumstances [122; 149; 178]. In addition, it has been proposed that PIP5KIα and AR may undergo direct or indirect interaction in cells. Evidence for this came from reports that PIP5KIα co-immunoprecipitates with AR proteins in cell-free extracts from prostate cancer cell lines [94; 95], suggesting the possibility of either direct interaction of two proteins, or inclusion of both within a protein complex. The possibility of direct PIP5KIα interactions with AR is intriguing because it suggests nuclear functional interactions could contribute to prostate cancer development. For example PIP5KIα reportedly directly interacts with and stabilises WT p53 and mutant p53 in the nucleus [135]; stabilisation of mutant p53 contributes to its oncogenic activities [179; 180]. In addition to co-

immunoprecipitation of AR and PIP5KIα, co-immunoprecipitation of PIP5KIα and another steroid receptor, ERα, has also recently been reported [181]. Intriguingly, evidence for cross talk between $PI(4,5)P_2$/$PI(3,4,5)P_3$ and NR signalling comes from the observation that nuclear receptor Steroidogenic Factor-1 (SF-1; NR5A1) reportedly binds the acyl tails of PI4P, $PI(4,5)P_2$ and $PI(3,4,5)P_3$ which is one mechanism by which the typically membrane integrated PIPs may be retained in nuclear speckles, which are membraneless compartments within the nucleus [59; 182]. SF-1 reportedly forms heterodimers with AR [183], and previous work within our group has found ligand (mibolerone) dependent heterodimerisation of AR with the family member of SF-1, Liver receptor homolog-1 (LRH-1; NR5A2) [184]. PIP5KIα, AR and SF-1 are each reported to localise to nuclear speckles [149; 185; 186], meaning complex formation is a feasibility.

Interestingly, examination of the sequences of PIP5KIα, PIP5KIβ and PIP5KIγ shows that each contains several potential LXXLL-like motifs and a potential CoRNR box motif which might mediate putative interaction between the PI5KIs and AR or other NRs. In PIP5KIα these are: ^{109}MQDFY^{113}, ^{149}FRELF^{153}, ^{204}LQKLL^{208}, ^{208}LPGYY^{212}, ^{226}FYGLY^{230} and ^{244}MNNLL^{248}; and the potential coRNR-like motif: ^{85}LKGAIQLGI^{93} (residue numbering as *per H. sapiens* PIP5KIα i1; **Figure 39**). Equivalent residues

to H. Sapiens PIP5KIα [149]FRELF[153] in *D. Rerio* PIP5KIα are [127]FREMF[131], which are necessary for interaction with Dishevelled *in-vivo* [66], and may therefore be positioned at a surface of the protein suitable for protein-protein interactions. FxxLF is a known AR LBD binding motif [187; 188]. These could feasibly mediate a direct interaction between PIP5KIα and AR. The equivalent positions of several of these ([85]LKGAIQLGI[93], [109]MQDFY[113], [149]FRELF[153], [204]LQKLL[208]) are within helices, i.e. like known functional LxxLL and coRNR motifs, in the published crystal structure of *D. rerio* PIP5KIα [66].

Given the importance of both PIP5KIα and AR in prostate cancer, the possible importance of any such interaction, and the Heery lab's long record of investigating protein-protein interactions of steroid nuclear receptors [37; 101; 189], we set out to investigate possible direct interactions of PIP5KIα with AR by yeast two-hybrid (Y2H). Given that protein-protein interactions of PIP5KIα with Star-PAP and K-Ras isoform 2B have previously been reported, it was decided to attempt to validate these interactions in the Y2H system. If validated, these would serve as positive controls for the functionality of Y2H PIP5KIα constructs, and enable an analysis of potential isoform specific interaction of PIP5KIα with these reported partners.

Finally, due to the poor efficacy of subcloning in large Y2H vectors using standard methods, I also set out to assess whether modifications to cloning protocols could improve cloning success with these plasmids. The results of these approaches are also described within this chapter.

A)

```
PIP5KIβ(i2)      -------------------------MSSAAENGEAAPGKQNEEKTYKKTQVKNQILSRL   34
PIP5KIα(i1)      MA--SASSGPSSSVGFSSFDPAVPSCTLSSAASGIKRPMASEVLE---A--RQDSYISLV   53
PIP5KIγ(i1)      MELEVPDEAESAE------AGAVPSEAAWAAESGAAAGLAQK--K---A--APTEVLSMT   47
                                          :* .*      .:  :         . :*

PIP5KIβ(i2)      PKIPCTFIQAWTNTEMITLLACYYFRSVSTASSAIKGAIQLGIGYTVGNLTSKPERDVLM   94
PIP5KIα(i1)      PYASGMPIKKIGHRS---VDSSGETTYKKTTSSALKGAIQLGITHTVGSLSTKPERDVLM  110
PIP5KIγ(i1)      AQPGPGHGKKLGHRG---VDASGETTYKKTTSSTLKGAIQLGIGYTVGHLSSKPERDVLM  104
                         :   :     : :.      .*:**:******** :*** *::********

PIP5KIβ(i2)      QDFYVVESVFLPSEGSNLTPAHHYPDFRFKTYAPLAFRYFRELFGIKPDDYLYSICSEPL  154
PIP5KIα(i1)      QDFYVVESIFFPSEGSNLTPAHHYNDFRFKTYAPVAFRYFRELFGIRPDDYLYSLCSEPL  170
PIP5KIγ(i1)      QDFYVVESIFFPSEGSNLTPAHHFQDFRFKTYAPVAFRYFRELFGIRPDDYLYSLCNEPL  164
                 ********:*:************: *********:**********:*******:*.***

PIP5KIβ(i2)      IELSNPGASGSLFFVTSDDEFIIKTVQHKEAEFLQKLLPGYYMNLNQNPRTLLPKFYGLY  214
PIP5KIα(i1)      IELCSSGASGSLFYVSSDDEFIIKTVQHKEAEFLQKLLPGYYMNLNQNPRTLLPKFYGLY  230
PIP5KIγ(i1)      IELSNPGASGSLFYVTSDDEFIIKTVMHKEAEFLQKLLPGYYMNLNQNPRTLLPKFYGLY  224
                 ***.. *******:*:********** ********************************

PIP5KIβ(i2)      CMQSGGINIRIVVMNNVLPRSMRMHFTYDLKGSTYKRRASRKEREKSNPTFKDLDFLQDM  274
PIP5KIα(i1)      CVQAGGKNIRIVVMNNLLPRSVKMHIKYDLKGSTYKRRASQKEREKPLPTFKDLDFLQDI  290
PIP5KIγ(i1)      CVQSGGKNIRVVVMNNILPRVVKMHLKFDLKGSTYKRRASKKEKEKSFPTYKDLDFMQDM  284
                 *:*:** ***:******:*** ::**:.:************:**:**  **:*****:**:
```

B)

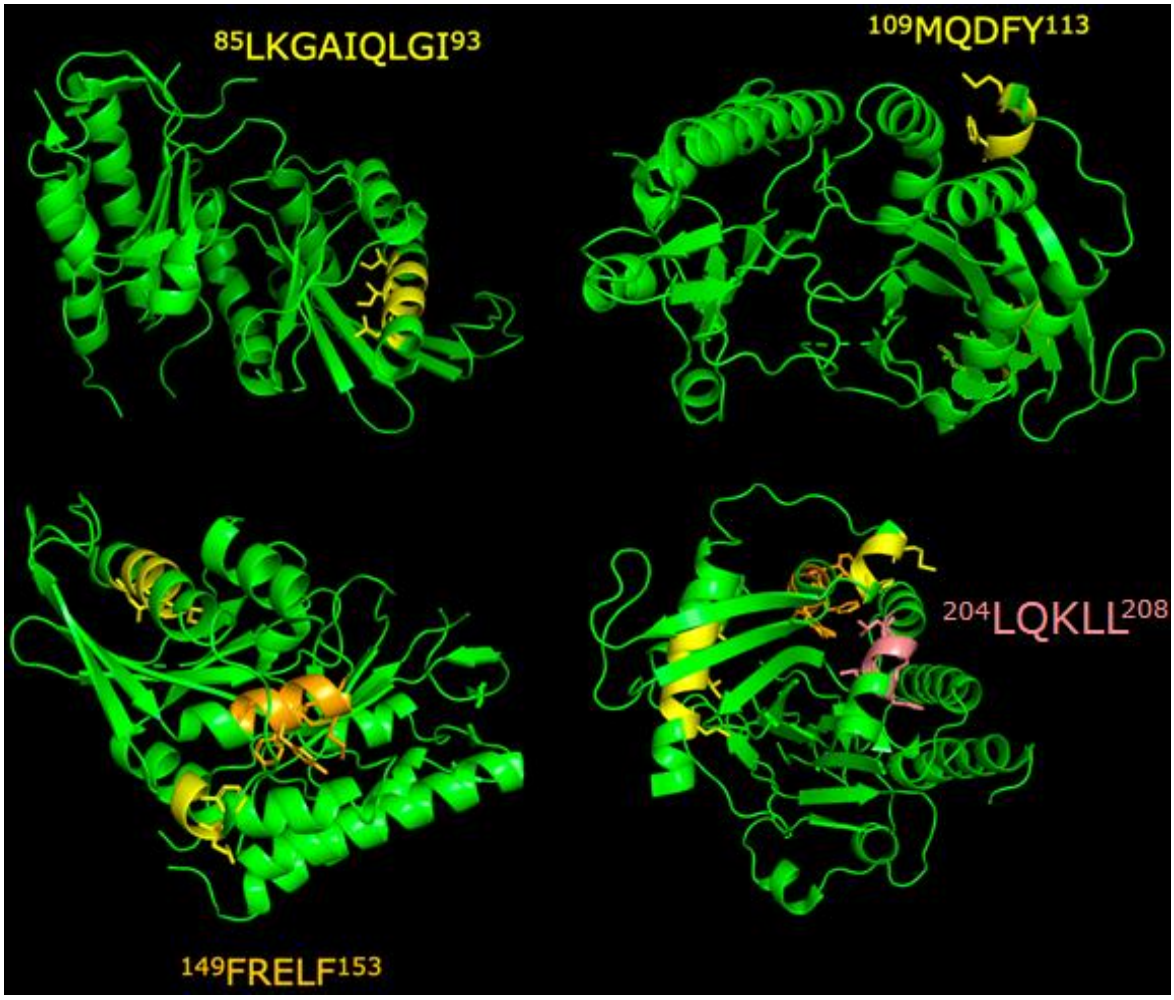


Figure 39. Sequence alignment indicating conserved potential LxxLL like motifs and a CoRNR box like motif in the three isozymes of PIP5KI.

A) Multiple sequence alignments of *H. sapiens* PIP5KIα, PIP5KIβ and PIP5KIγ, indicating conserved sequences similar to LxxLL-like motifs: ^{109}MQDFY^{113}, ^{149}FRELF^{153}, ^{204}LQKLL^{208}, ^{208}LPGYY^{212}, ^{226}FYGLY^{230} and ^{244}MNNLL^{248}; and one potential coRNR-like motif: ^{85}LKGAIQLGI^{93} (residue numbers as *per H. sapiens* PIP5KIα i1). These are well conserved among the PIP5KIs (red boxes).

B) Predicted 3D structure of *H. sapiens* PIP5KIα, generated by Prof. Barrie Kellam (University of Nottingham) by reference to published crystal structure of *Danio Rerio* PIP5KIα [66], visualised in PYMOL. Although PIP5KIα contains several potential LXXLL or CoRNR box like sequences, only four of these are present in helices: ^{85}LKGAIQLGI^{93}, ^{109}MQDFY^{113}, ^{149}FRELF^{153}, ^{204}LQKLL^{208}.

5.2. Initial experiments assessing PIP5KIα interaction with steroid receptor LBDs by Y2H ONPG assay

5.2.1. Construction of yeast expression vectors of LexA-PIP5KIα and NLS-VP16-PIP5KIα

Initial yeast two hybrid experiments were conducted with PIP5KIα i4, (1-500) since the coding sequence for this isoform was immediately available in a donor plasmid, pLPS *PIP5K1A* eGFP; a gift from Dr Jenny Perrson, Umeå University. This subcloning was performed similarly to the process as described in **Section 3.3**, ligating a PCR amplified PIP5KIα i4 CDS into the multiple cloning sites of pBTM116mod and pASV3mod (**Figure 40**), between XhoI and BamHI sites. Primers used for the amplification of PIP5KIα i4 CDS are listed in **Supplementary Table 13**.

Following transformation of the ligation products into competent DH5α, colonies were screened for the *PIP5K1A* CDS insert with the same primer pair. For cloning of PIP5KIα i4 CDS into pBTM116mod, three positive colonies were obtained after a single cloning experiment. However six separate cloning attempts were required to insert the same XhoI/BamHI fragment into pASV3mod, before finally attaining a single positive clone. Nucleotide sequencing pBTM116 PI5KIα i4 and pASV3 PIP5KIα i4 plasmids confirmed the validity of LexA-

PIP5KIα and VP16-PIP5KIα fusion constructs. Additional sequencing with a primer internal to *PIP5K1A* CDS was also performed. This *PIP5K1A* internal primer targets close enough to the 3’ end of the *PIP5K1A* CDS that sequencing reads were sufficient to span the 3’ junction of the insert and the plasmid MCS, and thus confirm insertion of only a single copy of PIP5KIα CDS into each plasmid (concatamer insertions would lead to partially mixed chromatograms with such a primer). The sequences of amino acid linkers of resulting fusion proteins when cloning into pBTM116mod and pASV3mod using the XhoI site as the 5’ restriction site are EFPGLE and GAILE, respectively.

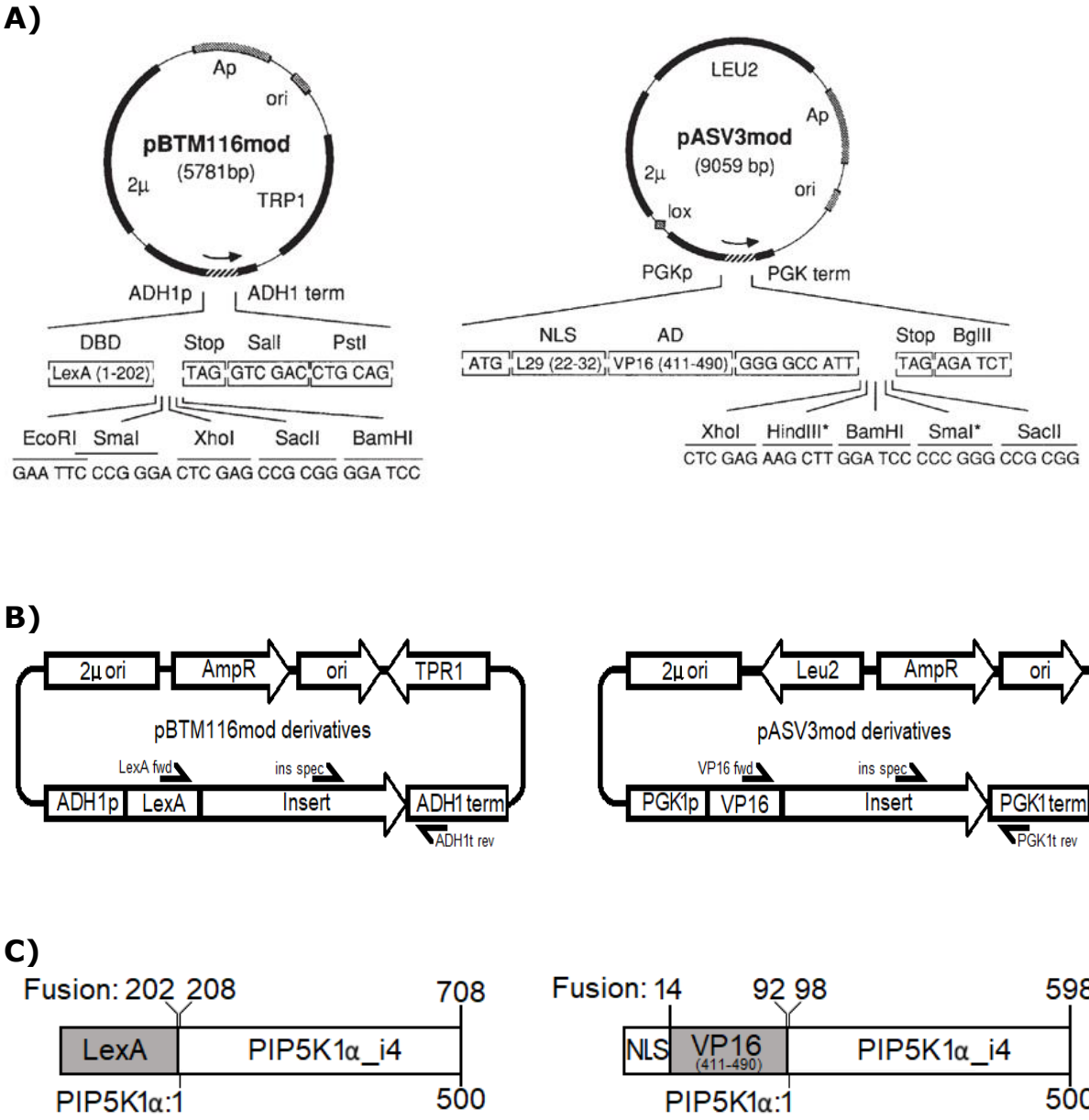


Figure 40. Molecular cloning to generate pBTM116 PIP5KIα i4 and pASV3 PIP5KIα i4 plasmids.

A) Existing plasmid maps of pBMT116mod and pASV3mod. Reproduced from [101].

B) Maps of the PIP5KIα fusion expression plasmids for Y2H, including the positons of primers used for sequencing.

C) Schematics of the LexA- and NLS-VP16-PIP5KIα fusion proteins.

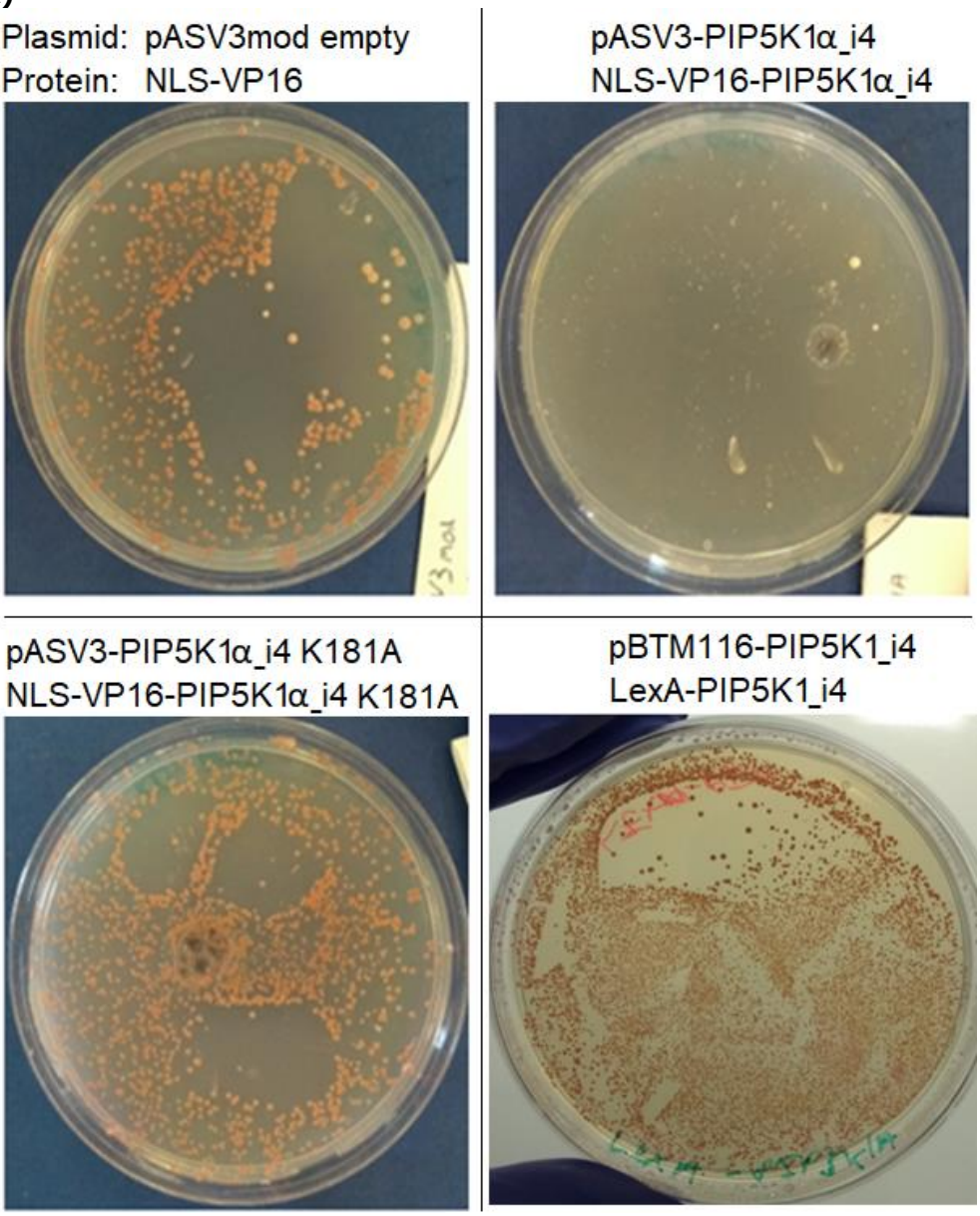
5.2.2. Expression of NLS-VP16-PIP5KIα induces growth defect in S. cerevisiae

For the purpose of conducting Y2H experiments, transformation of the yeast strain L40 was performed by Lithium Acetate mediated transformation, with the various plasmids encoding LexA/ NLS-VP16 fusion proteins to nuclear receptor LBDs/ PIP5KIα i4.

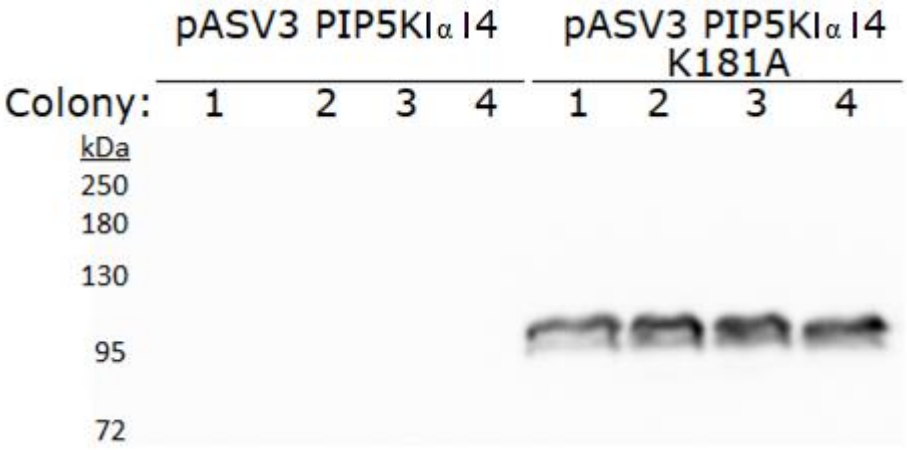
It became apparent that any transformations that included the pASV3-PIP5KIα i4 plasmid resulted in severe growth defect to the transformed yeast. When transformed with empty pASV3mod/pBTM116mod or their derivatives encoding NLS-VP16/LexA -fusion proteins, transformed colonies typically become visible 48 hours post transformation, and have a red colouration – the red colour results from *ADE2* knockout leading to accumulation of polymerised 5-aminoimidazole ribotide when grown on adenine deficient media. However, as shown in **Figure 41 A**, transformations including the pASV3-PIP5KIα i4 did not result in the appearance of visible colonies until 76 hours post transformation, and these colonies were fewer in number, smaller, and had a white colouration, all suggesting a severe growth defect in these transformants. Such growth defects may be caused by toxicity due to transcriptional "squelching" by the strong VP16 acidic activation domain [190; 191] acting in combination in someway with PIP5KIα.

Intriguingly, transformations including the plasmid pBTM116-PIP5KIα i4 (which encodes LexA-PIP5KIα i4) or pASV3-PIP5KIα i4 K181A (the latter generated by site directed mutagenesis of pASV3-PIP5KIα i4, using primers as listed in **Supplementary table 5**) which encodes NLS-VP16 PIP5KIα i4 K181A, a kinase deficient variant, did not result in any observed growth defect (**Figure 41 A**). Western blots showed robust expression of the VP16-PIP5KIα i4 K181A protein, whereas the WT version was not observed (**Figure 41 B**), suggesting low-level expression due to yeast cell toxicity. These findings indicate that the observed growth defect is dependent upon expression of a catalytically active PIP5KIα fused to a nuclear localised VP16 AAD.

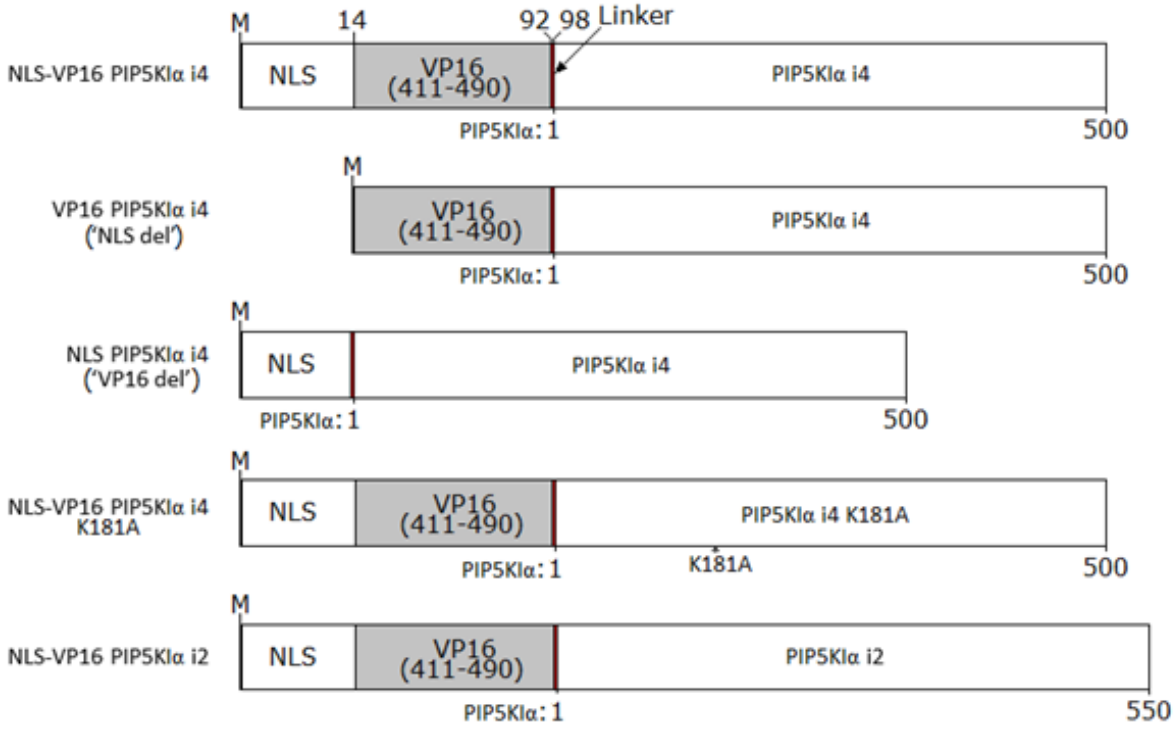
A)


Plasmid: pASV3mod empty
Protein: NLS-VP16
pASV3-PIP5K1α_i4
NLS-VP16-PIP5K1α_i4
pASV3-PIP5K1α_i4 K181A
NLS-VP16-PIP5K1α_i4 K181A
pBTM116-PIP5K1_i4
LexA-PIP5K1_i4

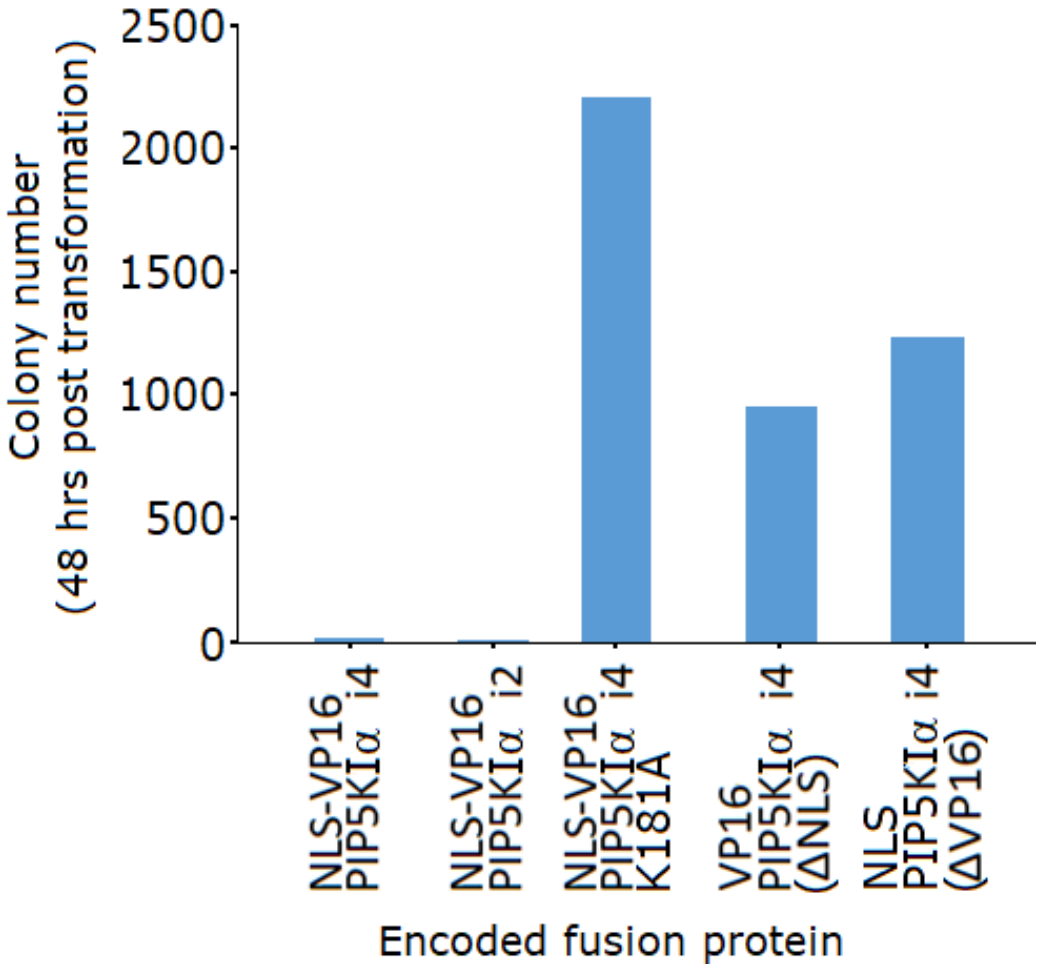
B)


pASV3 PIP5KIα I4
pASV3 PIP5KIα I4 K181A
Colony: 1 2 3 4 1 2 3 4
kDa
250
180
130
95
72

C)

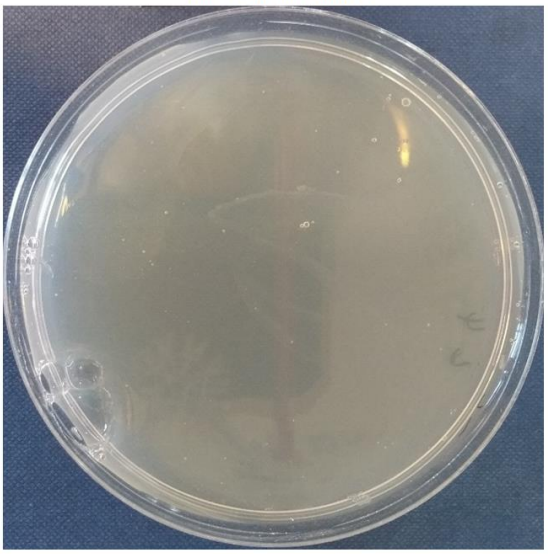

NLS-VP16 PIP5KIα i4
M
14
92 98 Linker
NLS
VP16
(411-490)
PIP5KIα i4
PIP5KIα: 1
500
VP16 PIP5KIα i4
('NLS del')
NLS PIP5KIα i4
('VP16 del')
NLS-VP16 PIP5KIα i4
K181A
PIP5KIα i4 K181A
K181A
NLS-VP16 PIP5KIα i2
PIP5KIα i2
550

D)

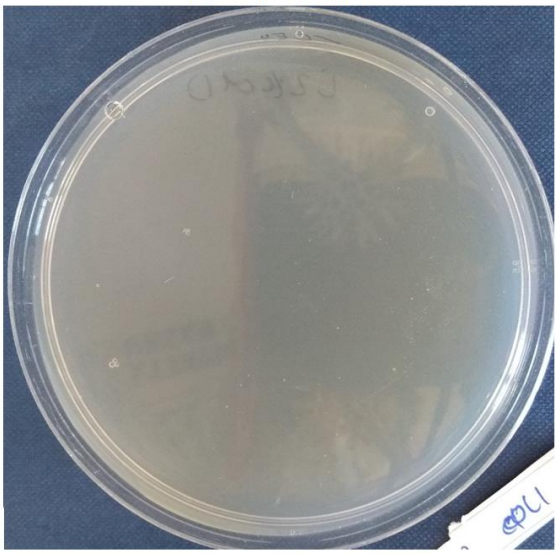

Colony number
(48 hrs post transformation)
2500
2000
1500
1000
500
0
NLS-VP16 PIP5KIα i4
NLS-VP16 PIP5KIα i2
NLS-VP16 PIP5KIα i4 K181A
VP16 PIP5KIα i4 (ΔNLS)
NLS PIP5KIα i4 (ΔVP16)
Encoded fusion protein

E)

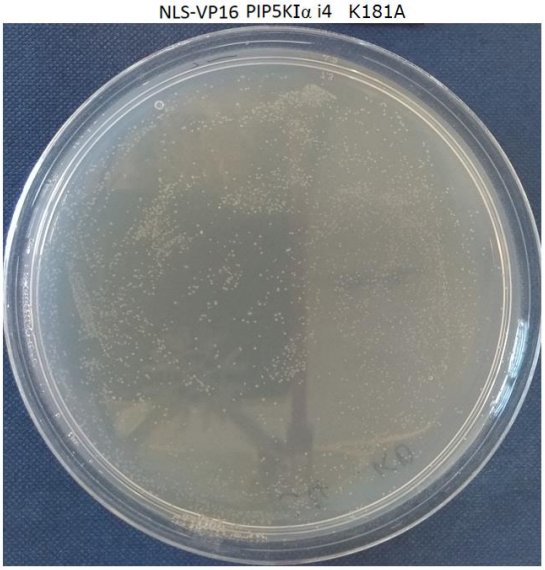
NLS-VP16 PIP5KIα i4



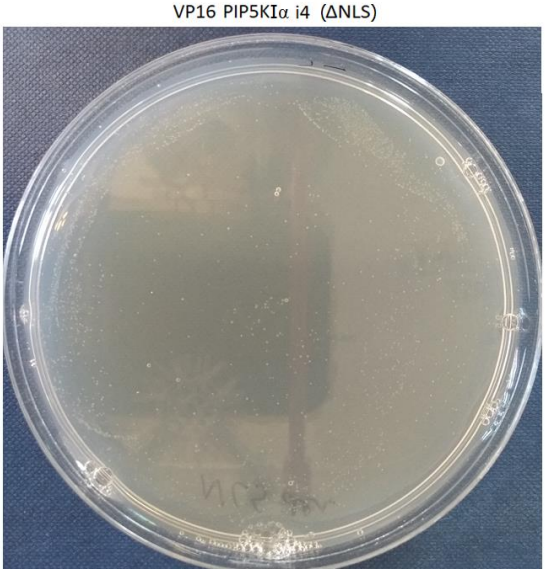
NLS-VP16 PIP5KIα i2



NLS-VP16 PIP5KIα i4 K181A



VP16 PIP5KIα i4 (ΔNLS)



NLS PIP5KIα i4 (ΔVP16)

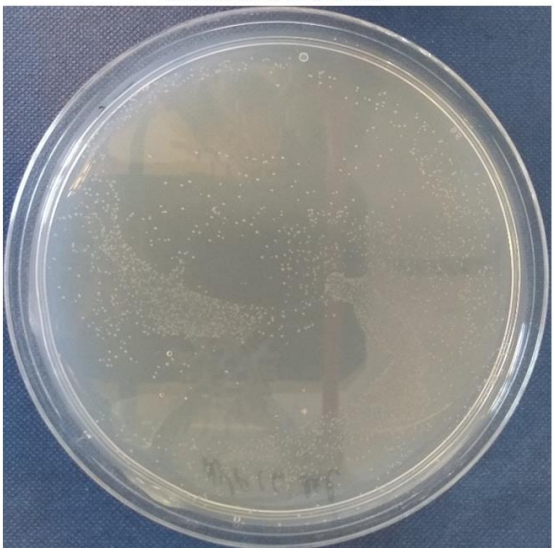


Figure 41. Expression of NLS-VP16-PIP5KIα is toxic to yeast.

A) Photographs of YSM (-Leu) agar plates spread with transformation mixtures of L40 yeast transformed with expression plasmids for various PIP5KIα fusion proteins for Y2H. Images taken 60 hours post transformation. Note the smaller size, white colouration and fewer colonies for transformation with plasmid encoding NLS-VP16-PIP5KIα i4 compared to those for NLS-VP16 alone, LexA-PIP5KIα i4, or NLS-VP16-PIP5KIα i4 K181A (kinase deficient).

B) Western blotting with an anti-VP16 antibody for lysates of yeast colonies transformed with pASV3mod PIP5KIα i4 or pASV3mod PIP5KIα i4 K181A. Detection of VP16-fusion protein for the pASV3-PIP5KIα i4 K181A, but not pASV3-PIP5KIα i4 transformants indicates attempted escape of the growth defect phenotype by somehow reducing NLS-VP16-PIP5KIα protein level, and that this is not necessary for the kinase deficient NLS-VP16-PIP5KIα K181A.

C) Schematics of various PIP5KIα fusion proteins generated for expression from pASV3mod.

D) Colony counts of at 48 hours for transformation of L40 yeast with expression plasmids for the PIP5KIα fusion proteins depicted in **C**.

E) Photographs of YSM (-Leu) agar plates spread with transformation mixtures of L40 yeast transformed with expression plasmids for various PIP5KIα fusion proteins, at 48 hours post transformation.

To further investigate the possible cause of observed growth defect of NLS-VP16-PIP5KIα i4 expression in yeast, several additional constructs were made. To determine if the effect is specific for PIP5KIα isoform 4, the CDS of a second isoform, PIP5KIα i2, was also cloned into pASV3mod (details of cloning presented in **Section 5.3.3.**). To investigate the contribution of each element in NLS-VP16-PIP5KIα i4 fusion protein to the observed growth defect, deletion constructs removing either the NLS or VP16 coding sequences were made by site directed mutagenesis. The primers used to generate these are listed in **Supplementary Table 14**. One plasmid encodes an NLS-PIP5KIα i4 fusion, with the majority of the VP16 AAD deleted (residues 483-490 of VP16 are retained). The second encodes a VP16-PIP5KIα i4 fusion, in which the 13 residues of the L29 NLS are deleted; the ATG start codon and Kozak motif are maintained. Schematics depicting the various fusion proteins are presented in **Figure 41 C**.

Transformations of L40 yeast were performed with 500 ng of each of these plasmids (plating on YSM –Leu agar plates) in order to observe effects on colony formation. Photographs of plates and colony counts for 48 hours post transformation are presented in **Figure 41 D and E.**

Colony counts of transformants with each of these expression plasmids demonstrate that both NLS-VP16-PIP5KIα i4 and NLS-VP16-PIP5KIα i2 are >100 fold lower than for the NLS-VP16-PIP5KIα i4 K181A expression plasmid, demonstrating that the growth defect is not specific to PIP5KIα i4.

Deletions of either the NLS or VP16 AAD portions of the PIP5KIα i4 fusion protein resulted in increased (>50 fold) colonies compared to the NLS-VP16-PIP5KIα i4.

In summary, though these experiments are non-exhaustive, they indicate that the greatest detrimental effect to colony growth is observed for expression plasmids for NLS-VP16-PIP5KIα i4 or NLS-VP16-PIP5KIα i2, with no effect observed for NLS-VP16-PIP5KIα i4 K181A or LexA-PIP5KIα expression plasmids. Deletion of either the NLS or VP16 components may each partially or completely relieve the toxic effect.

Based upon these findings, a putative model for the mechanism for this toxicity is through nuclear localised NLS-VP16-PIP5KIα binding to native yeast protein or proteins through interaction with PIP5KIα. Such native yeast proteins could include a yeast homologue of nuclear proteins known to interact with PIP5KIα in mammalian cells, such as pRB or p53

(though yeast do not possess a homologue of p53). This interaction, or downstream effects of this interaction is seemingly dependent on production of $PI(4,5)P_2$. Inclusion of VP16 in the fusion also appears to be necessary, since expression of a LexA-PIP5KIα fusion did not cause any growth defect, and human PIP5KIα has previously been expressed in *S. cerevisiae* without any reported toxicity [84]. Expression of VP16 fused to DNA binding domains cause squelching by sequestering limiting transcription factors or transcriptional activators, thus curtailing basal or activated transcription [190]. PIP5KIα is not presumed to bind DNA, but may interact with native yeast proteins that do, and thus a VP16-PIP5KIα fusion could bring about a similar effect, seemingly dependent on $PI(4,5)P_2$ generation.

Regardless of the exact mechanism, the apparent dependence of this phenomenon on catalytically active PIP5KIα suggested a method to screen for inhibitors of PIP5KIα activity. Such screens have previously been conducted for inhibitors of human class I PI3K in yeast [192; 193]. Work was conducted in this direction, constructing a plasmid for (cyanamide) inducible (rather than constitutive expression) of NLS-VP16-PIP5KIα, but the progress is too preliminary for inclusion in this thesis.

5.2.3. Yeast two-hybrid assay to assess interaction of PIP5KIα i2 with AR-LBD and ERα-LBD

After successful generation of the PIP5KIα LexA/ NLS VP16 vectors, Y2H assays were performed to assess interactions of LexA-PIP5KIα i4 with NLS-VP16-AR-LBD (aa 626-920) or NLS-VP16-ERα LBD (human ERα, residues 282-595). The inverse experiments, to access NLS-VP16-PIP5KIα interactions with LexA fusions to the NR-LBDs were also performed, but due to the toxic effect of transformation with pASV3-PIP5KIα i4, experiments were also performed using the pASV3-PIP5KIα i4 K181A plasmid. β-galactosidase activity, the reporter for interaction of the bait and prey proteins, was measured by quantitative ONPG assay.

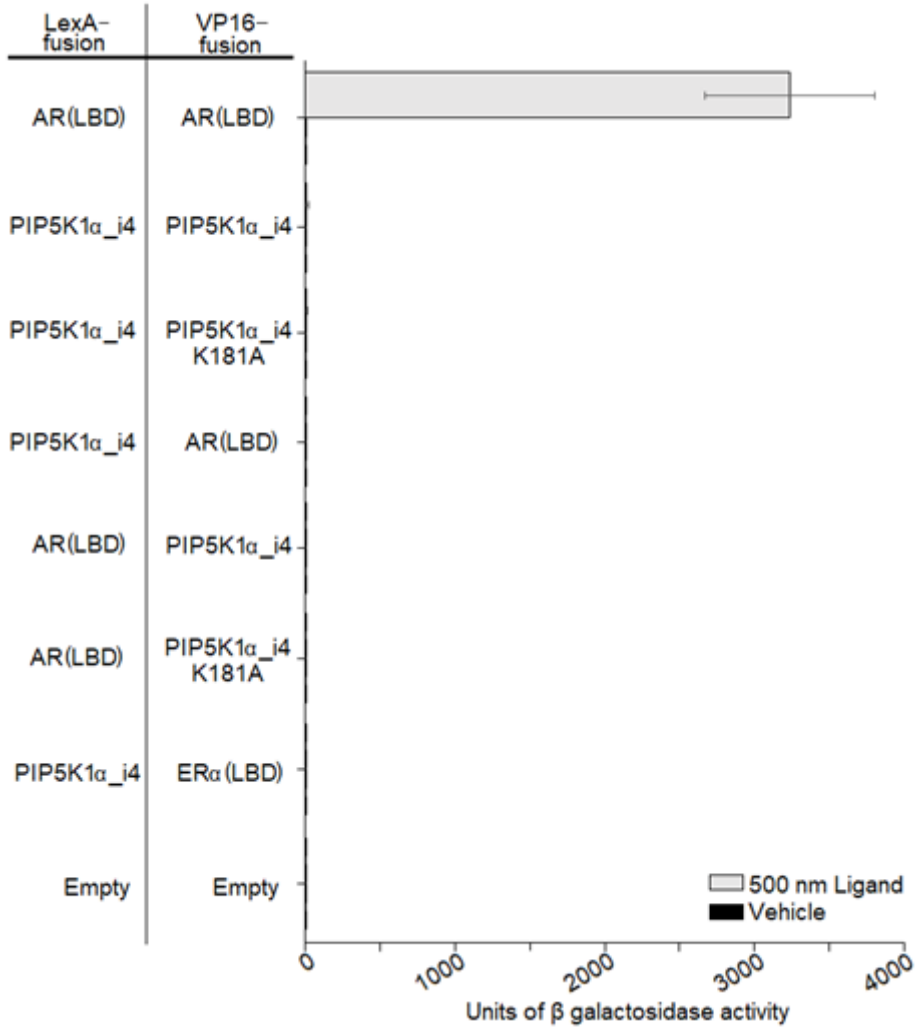
As shown in **Figure 42**, neither AR-LBD nor ERα-LBD showed either ligand–dependent or ligand independent interaction with PIP5KIα i4. Both NLS-VP16-AR LBD and LexA-PIP5KIα expression was confirmed by western blotting with extracts for the transformed yeast colonies (**Figure 42 B** and **C**), and so lack of detectable interaction (β-galactosidase reporter gene activation) cannot be explained by a lack of expression of either fusion protein. NLS-VP16-ERα LBD expression was not tested, and so lack of its expression cannot be excluded as the cause for lack of observable interaction in

this assay, though this is unlikely due to successful expression of NLS-VP16-ERα LBD from that same plasmid for in many previous experiments within this lab [184].

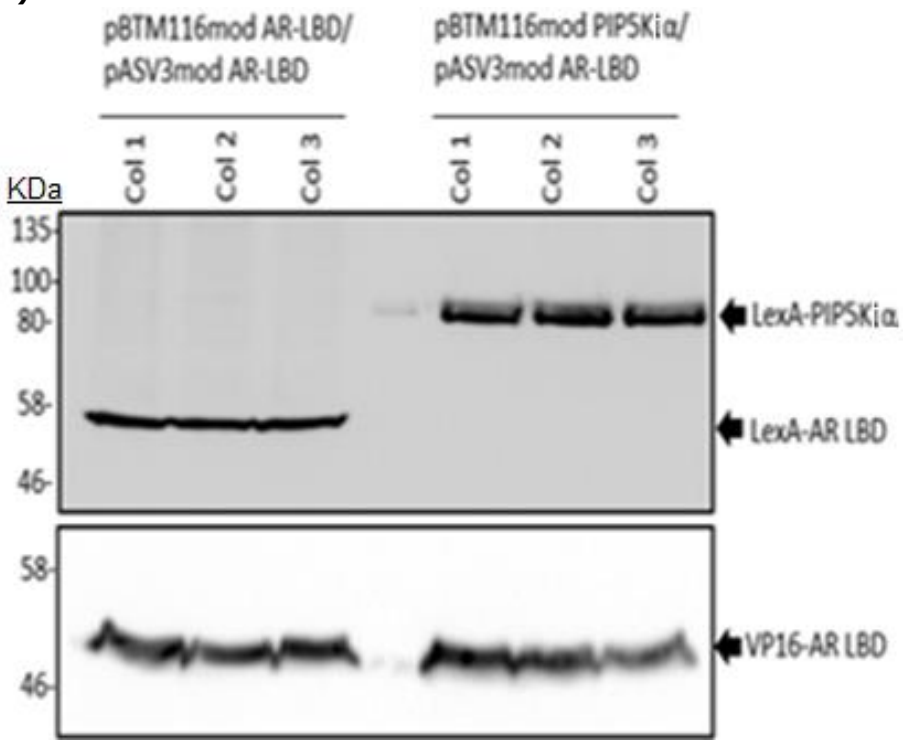
Assays with yeast colonies co-expressing LexA-AR LBD and NLS-VP16-AR LBD were performed with ligand (500 nM mibolerone) or vehicle (ethanol) to provide a positive and negative control for ligand dependent AR LBD interactions. Since PIP5KIα is reported to homodimerise [66; 194], Y2H assays with yeast colonies co-expressing LexA-PIP5KIα i4 and NLS-VP16 PIP5KIα i4 K181A were performed to provide a positive control, but β-galactosidase activity was not detectable. The lack of detectable β-galactosidase activity in these assays could be due to some fault with either of the fusion proteins, such as a PIP5KIα component inducing localisation at the cytoplasmic membrane; or very weak/ transient interactions that result in low levels of reporter gene activation below detection by the ONPG assay.

To address these possible issues, further experiments were designed using a modified form of X-gal β-galactosidase assay, which is reportedly more sensitive than ONPG based assays [102] (**Section 5.5**).

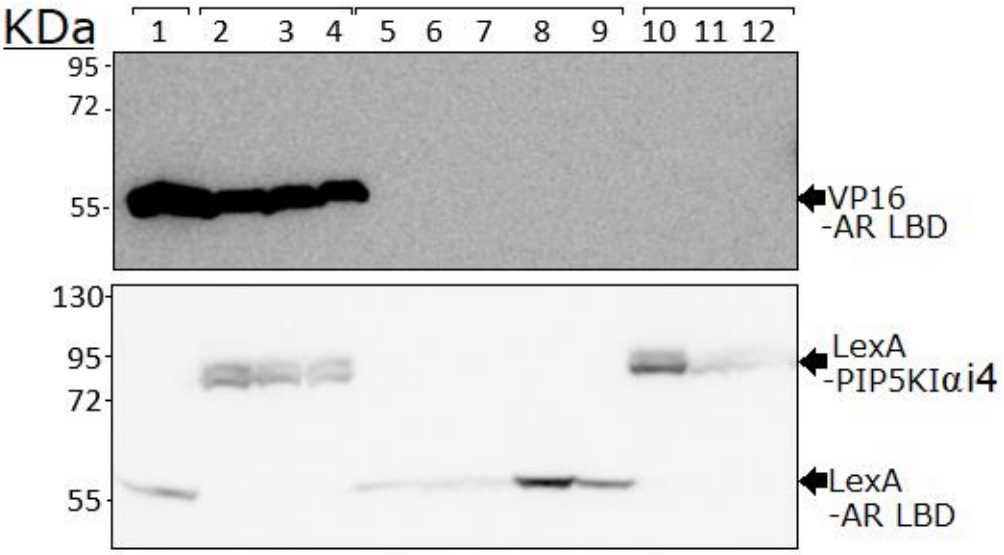
A)


LexA–fusion
VP16–fusion
AR(LBD)
AR(LBD)
PIP5K1α_i4
PIP5K1α_i4
PIP5K1α_i4
PIP5K1α_i4 K181A
PIP5K1α_i4
AR(LBD)
AR(LBD)
PIP5K1α_i4
AR(LBD)
PIP5K1α_i4 K181A
PIP5K1α_i4
ERα (LBD)
Empty
Empty
500 nm Ligand
Vehicle
0
1000
2000
3000
4000
Units of β galactosidase activity

B)


pBTM116mod AR-LBD/
pASV3mod AR-LBD
pBTM116mod PIP5Kiα/
pASV3mod AR-LBD
Col 1
Col 2
Col 3
Col 1
Col 2
Col 3
KDa
135
100
80
58
46
58
46
LexA-PIP5Kiα
LexA-AR LBD
VP16-AR LBD

C)



Lane	pBTM116mod plasmid	pASV3mod plasmid
1	LexA-AR LBD	VP16-AR LBD
2-4	LexA-PIP5KIα i4	VP16-PIP5KIα i4
5-9	LexA-AR LBD	VP16-PIP5KIα i4
10-12	LexA-PIP5KIα i4	VP16-PIP5KIα i4

Figure 42. ONPG assay to assess interactions of AR LBD and ERα LBD with PIP5KIα

A) Yeast two-hybrid β-galactosidase assay using cell-free extracts derived from yeast co-transformed with an expression vector for LexA-PIP5KIα i4 and two VP16-NR-LBDs. The receptor cognate ligands were mibolerone and 17β-estradiol for AR-LBD and ERα-LBD, respectively, at 500 nm. The results shown represent the mean β-galactosidase activity of three clones, ± standard error.

B) Western blot analyses with an anti-LexA and anti-VP16 antibody to detect the respective fusion proteins in cell-free extracts derived from yeast co-transformants

C) A second set of western blot analyses with the anti-LexA and anti-VP16 antibodies to detect respective fusion proteins in cell-free extracts derived from yeast co-transformants. The inset table indicates for each lane what plasmids were co-transformed to obtain colonies from which extracts were prepared.

5.3. Improved protocols for cloning with Y2H vectors

Cloning to generate new expression vectors for Y2H experiments, particularly inserting into the larger pASV3mod plasmid, had previously proved difficult, greatly impeding my own research and that of previous lab members. Due to this difficulty, novel approaches to molecular cloning were trialled in order to generate Y2H vectors containing the coding sequences for the AR NTD, PIP5KIα i2 and PIP5KIα x15.

5.3.1. Full sequencing of the pASV3mod plasmid

While a plasmid map and partial restriction map of the pASV3mod plasmid was available [101], the full sequence and restriction map of this plasmid were not known. These were desired so that any unnecessary elements that may hinder molecular cloning attempts could be removed. Only a portion of the sequence around the MCS of pASV3mod was known. Sequencing primers targeting the known sequence were designed and used to obtain sequence (700-1500 bp per sequencing read) beyond the already known sequence. This process was repeated sequentially 8 times with new sequencing primers until the entire pASV3mod sequence was obtained (**Figure 43**). Sequencing primers and the element of pASV3mod targeted by each are listed in **Supplementary Table 15.**

Note the discrepancy between the previously reported plasmid size of 9059 bp, and the obtained sequence, 9040 bp. A possible explanation for this discrepancy is that the plasmid was constructed from various elements without exact sequence information, so the exact lengths may not have been known. Alternatively, multiple small errors in the base calling may have occurred during the recent sequencing.

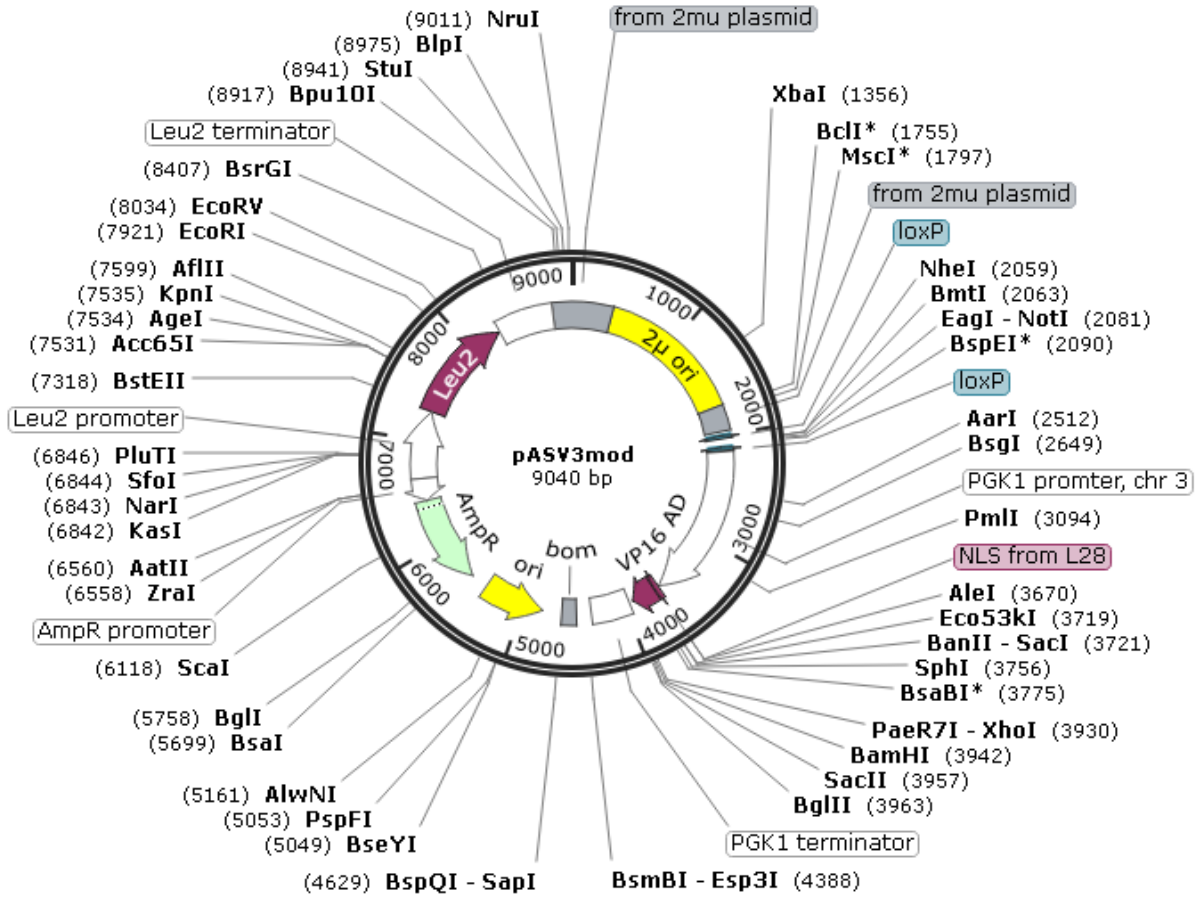


Figure 43. Complete restriction map of pASV3mod

Previously the entire sequence of pASV3mod was not known, and only a limited restriction map was available [101]. The plasmid was sequenced entirely by designing primers (**Table 1**) to known sequence (starting in the MCS) and progressively sequencing then designing new primers and sequencing until the entire plasmid sequence was obtained.

5.3.2. Cloning AR-NTD into pASV3mod by In Vivo Assembly

To investigate possible interaction of PIP5KIα with the NTD of AR, sub-cloning of the AR NTD (residues 1-538; the CDS of the first exon of *AR*) and insertion in the pASV3mod plasmid was attempted by the conventional restriction site cloning (primers in **Supplementary Table 16**). Primers were used to amplify the CDS of *AR* Exon 1 from the donor plasmid pSVARo [195]. AR Exon 1 is difficult to amplify due to high GC content and the presence CAG and GGN repeats [196]. PCR conditions were therefore optimised, finding good amplification using Phusion polymerase (NEB) with the supplied HF buffer (NEB) and DMSO inclusion at 3% final volume, an annealing temperature of 60 °C held for 1 minute, and an extension temperature of 72 °C held for 3 minutes.

Cloning of the AR NTD into the pBTM116 plasmids was not desired, because of the likelihood of that a LexA-AR NTD fusion would possess transcriptional activation activity (due to the AF1 in AR NTD) leading to high levels of β-galactosidase expression even in the absence of interaction with a VP16-fusion partner. Likewise, the coding sequence of the DNA binding domain (AR Exon 2 and Exon 3) was not included in the NLS-VP16 fusion. This was because of the possibility that constitutive transcriptional activity of a VP16-AR NTD & DBD fusion protein

would lead to squelching (exogenous transcription factor mediated inhibition of gene expression and cell growth arrest) in yeast transformants as has been observed for VP16 fusions to ERα DBD [190]. Six independent cloning experiments were performed to generate a pASV3 AR NTD plasmid, but after each transformation of competent DH5α with the ligation products, no colonies were obtained upon selection.

Given this failure, an alternative cloning strategy was therefore designed, utilising a method of homologous recombination cloning called *In Vivo Assembly* (IVA)[197] (**Figure 44 A**). This method relies on separate PCR amplifications of an insert and a vector, generating amplicons that possess 15 to 50 base pairs of homology at their ends, which can undergo homologous recombination when transformed into a suitable *E.coli* strain (such as DH5α). Note that while most *E.coli* strains used for subcloning are *recA* deficient, to avoid recombination of sequences with homology of >150bp, these strains contain a second recombinase activity that can catalyse recombination of short homologous sequences as exploited in this technology, as reviewed by Watson and Garcia [197].

Primers were designed to amplify the CDS of Exon 1 of AR (**Supplementary Table 16**), from the donor plasmid pSVARo. These primers each included 20 bp of homology to the

MCS of pASV3mod to facilitate recombination. PCR of an 8987 bp amplicon of pASV3mod was also attempted but repeatedly failed to result in a strong band upon agarose gel resolution. No colonies resulted after transformations with the mixture of DpnI treated PCR products of AR NTD and pASV3mod.

It was reasoned that the failure to achieve transformants was due to the poor amplification of the vector, likely due to its large size. To address this, the vector was amplified in two separate shorter fragments (4393 bp and 4646 bp) with 52 bp of homology to each other at one end. The other ends of each of these PCR products had 20 bp of homology to the ends of the AR NTD PCR product (**Figure 44 B**).

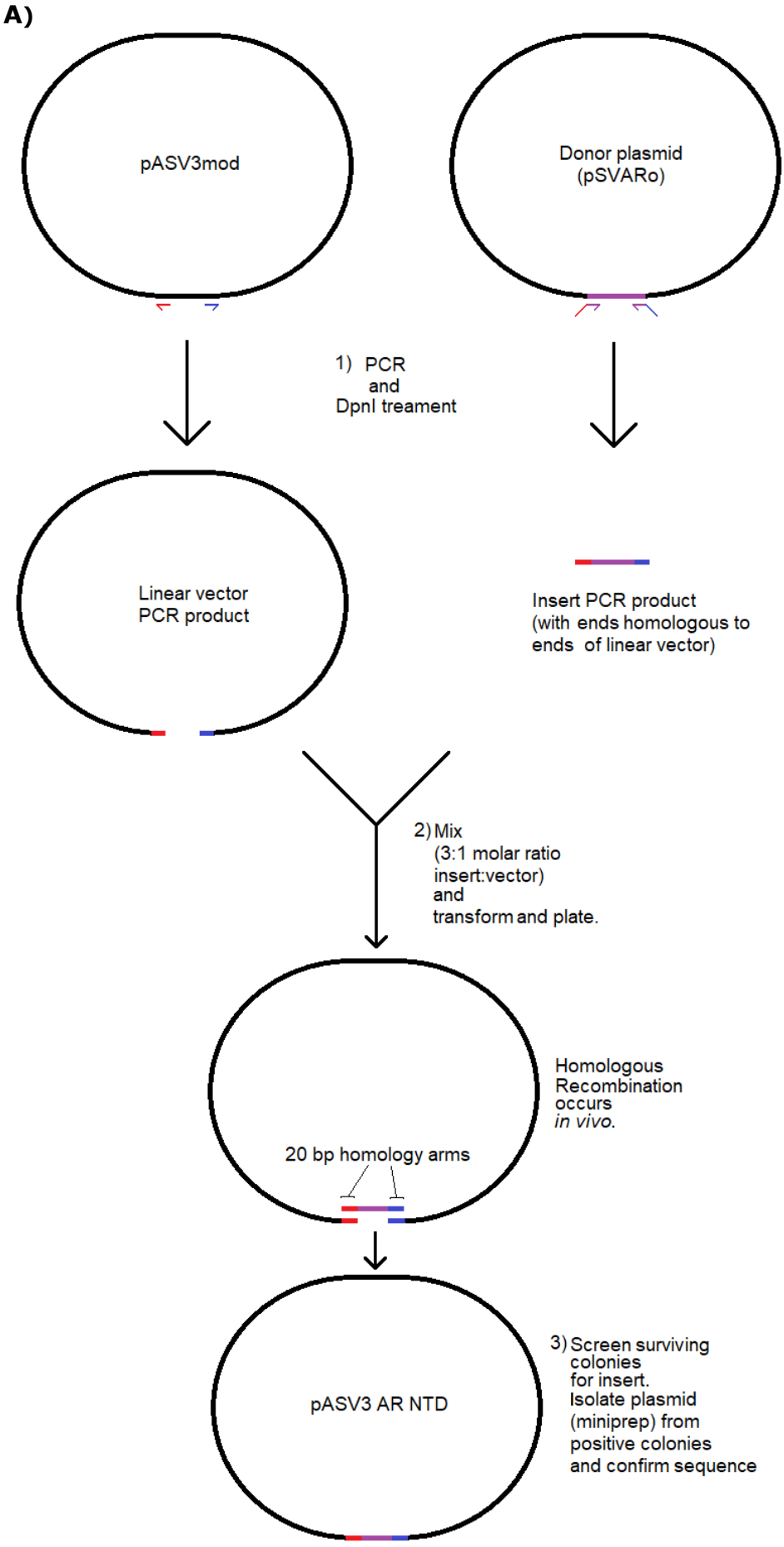
After PCR amplification, products were DpnI treated, mixed (5 μl: 5μl: 5 μl) and used to transform competent DH5α. Upon transformation and plating on selective medium, 15 colonies were obtained. These were screened for inclusion of the AR NTD CDS by colony PCR using a primer pair with an expected amplicon of 195 bp that is specific to the 3' junction of AR NTD CDS to pASV3 and in the intended plasmid. 11 of the 15 colonies (73 %) yielded an amplicon of the expected size (**Figure 44 C**). Of these, Colony 15 was selected to be expanded. The plasmid was isolated by alkaline lysis miniprep. Restriction digests with EcoRV and KpnI confirmed yielded

bands of expected size for the intend pASV3 AR NTD plasmid (**Figure 44 D**).

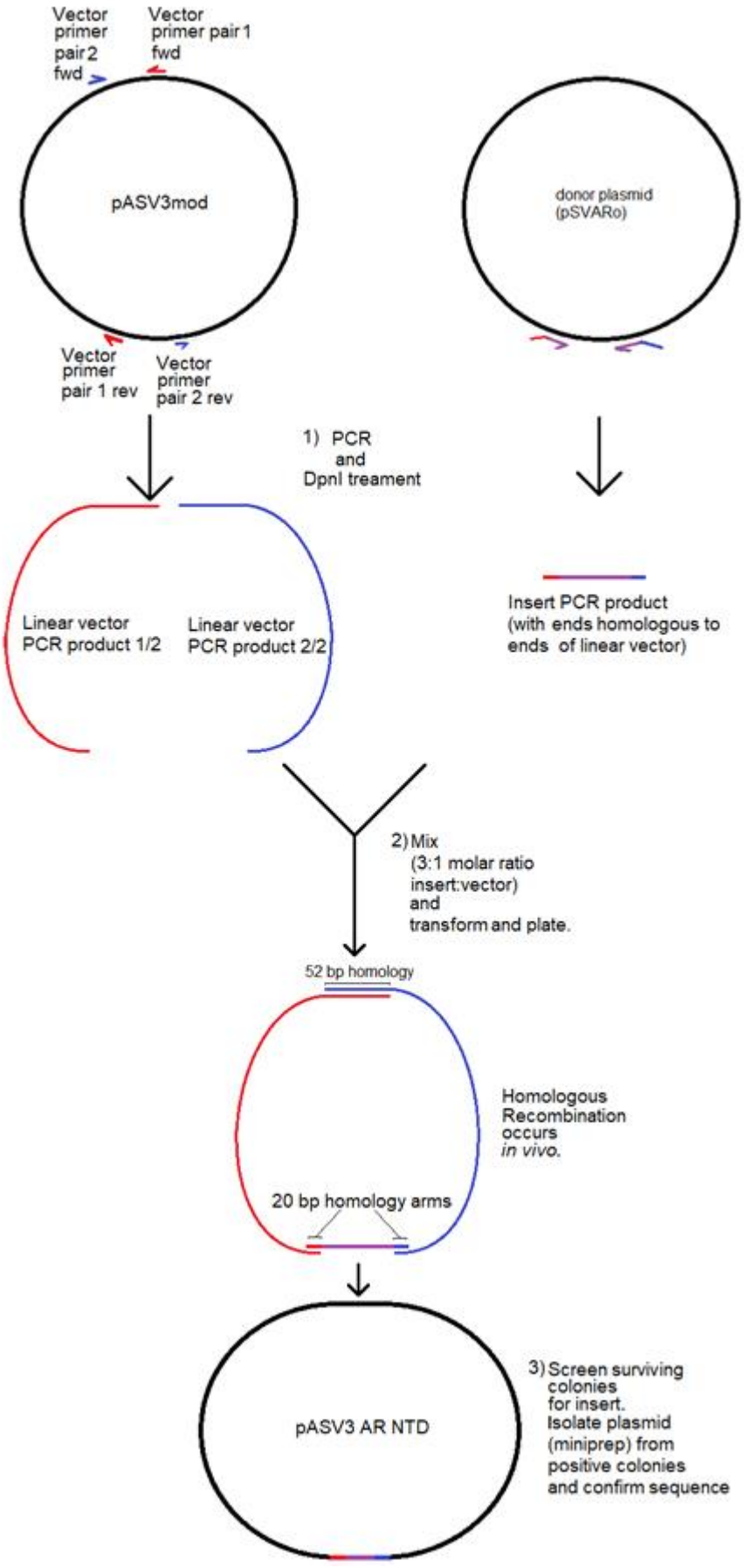
Sequencing of this plasmid confirmed correct insertion of the AR NTD CDS into the pASV3 mod, with no introduced errors and correct sequence at the junctions of vector and insert. The poly Q and poly G repeat lengths were determined to be 20 and 16 respectively, which are both within the normal range for Human AR [196].

IVA cloning was also used successfully to generate a pASV3 PIP5KIβ plasmid. PIP5KIβ CDS (human PIP5KIβ; Uniprot identifier: O14986-2; residues 1-549) was amplified from pDONR223-PIP5K1B (a gift from William Hahn & David Root - Addgene plasmid # 23848 [198]). The primers used each included 22 bp of homology to the MCS of pASV3mod to facilitate recombination.

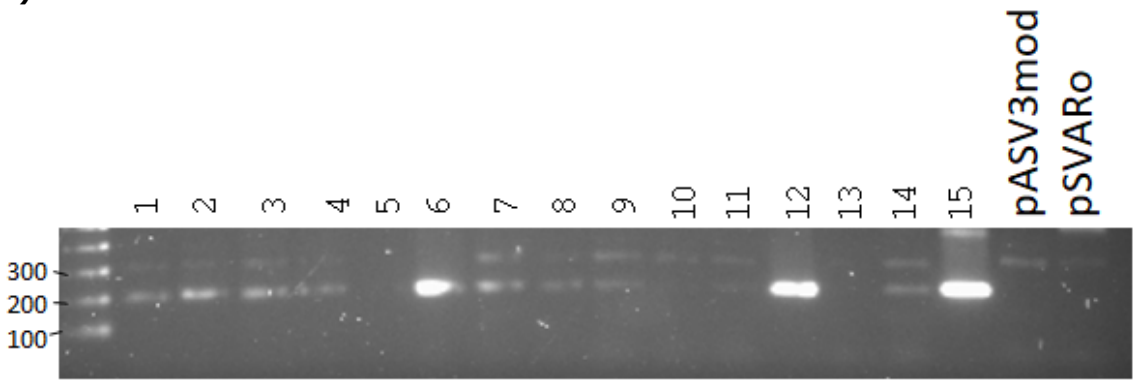
Three separate experiments to generate a pASV3 PIP5KIγ plasmid were performed. PIP5KIγ CDS (Human PIP5KIγ; Uniprot identifier: O60331-1, residues 1-668) was amplified from GFP-PIPK1 gamma 90 (a gift from Pietro De Camilli [199]). While the amplifications of PIP5KIγ CDS and pASV3 mod fragments were successful, following transformation, none of the colonies screened by colony PCR (>50) contained a PIP5K1C insert.


A)
pASV3mod
Donor plasmid
(pSVARo)
1) PCR
and
DpnI treament
Linear vector
PCR product
Insert PCR product
(with ends homologous to
ends of linear vector)
2) Mix
(3:1 molar ratio
insert:vector)
and
transform and plate.
Homologous
Recombination
occurs
in vivo.
20 bp homology arms
pASV3 AR NTD
3) Screen surviving
colonies
for insert.
Isolate plasmid
(miniprep) from
positive colonies
and confirm sequence

B)

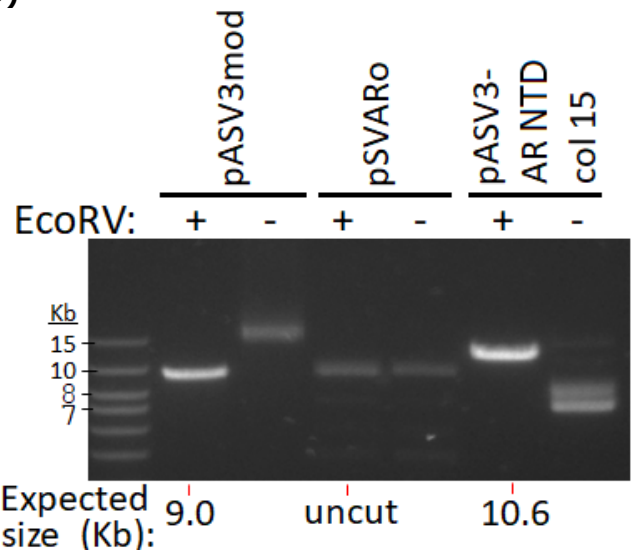

Vector primer pair 2 fwd
Vector primer pair 1 fwd
pASV3mod
donor plasmid (pSVARo)
Vector primer pair 1 rev
Vector primer pair 2 rev
1) PCR and DpnI treament
Linear vector PCR product 1/2
Linear vector PCR product 2/2
Insert PCR product (with ends homologous to ends of linear vector)
2) Mix (3:1 molar ratio insert:vector) and transform and plate.
52 bp homology
Homologous Recombination occurs *in vivo*.
20 bp homology arms
pASV3 AR NTD
3) Screen surviving colonies for insert. Isolate plasmid (miniprep) from positive colonies and confirm sequence

C)



Expected amplicon = 195 bp
Product specific to pASV3-AR NTD plasmid

D)



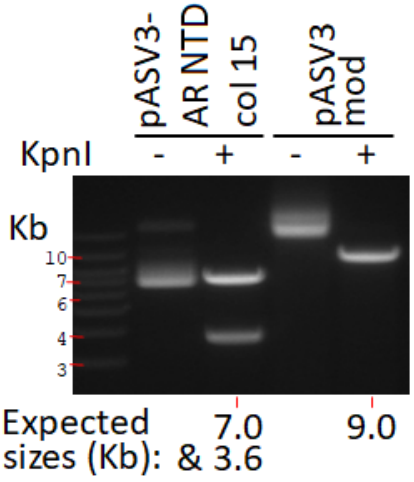


Figure 44. In Vivo Assembly cloning of AR NTD into pASV3mod

A) Work flow of IVA cloning strategy. 1) Vector and insert are separately PCR amplified so that the amplicons contain homologous ends (20 bp in this case; indicated by red and blue bars), then DpnI treated to digest methylated template DNA. 2)DpnI treated products are mixed and transformed into competent *E. coli*. Recombination occurs *in vivo;* colonies form from cells that successfully performed recombination. 3) Colonies are screened for the insert by colony PCR, then further processed.

B) Workflow of a modified IVA cloning strategy for when amplification of larger vectors, such as pASV3mod is needed. The modification is to amplify the vector in two sections (with homology at one end due to off-set primer sites) to improve amplification. Three fragments therefore undergo recombination *in vivo*, rather than two.

C) Agarose gel electrophoresis of PCR screening of colonies resulting from IVA cloning of AR NTD into pASV3mod. PCR primers specifically targeted the expected AR NTD CDS–pASV3 plasmid. 11 of 15 (73%) colonies show amplification for the expected product. Control reactions with colonies transformed with pASV3mod or pSVARo did not yield amplicons of the expected size. 1.5 % agarose; 1 kb plus ladder (Invitrogen).

D) Agarose gel electrophoresis of restriction digests of plasmid amplified and extracted by mini-prep from a single colony (Colony 15) from the pASV3-AR NTD cloning experiment. Restriction digests with EcoRV (a single cutter in pASV3mod and pASV3 AR NTD) and KpnI (a single cutter in pASV3mod, but double cutter in pASV3 AR NTD) confirmed the identity as pASV3 AR NTD. 0.6 % agarose; 1 kb plus ladder (Invitrogen).

5.3.3. Cloning PIP5KIα i2 CDS into pASV3mod and pBTM116mod by modified restriction site cloning protocols

As described, standard restriction site based protocols for cloning into the Y2H vectors, particularly pASV3mod, proved difficult and time consuming, with multiple attempts often all failing to attain a desired plasmid. While the IVA approach showed promise in that it was used successfully to generate pASV3 AR NTD and pASV3 PIP5KIβ plasmids, it was not without fault, since multiple attempts to generate a pASV3 PIP5KIγ plasmid failed.

A new approach was therefore devised to clone PIP5KIα i2 CDS into both pASV3mod and pBTM116mod. In this approach, the CDS to be inserted into the final destination vector is first TOPO cloned into an intermediate vector, in this case pCR4 Blunt TOPO, from which it is digested before ligation into the ultimate destination vector.

Other modifications made to the protocols were as follows:

- Use of chloramphenicol amplification when growing cultures of *E.coli* containing the pASV3mod plasmid or its derivatives for alkaline lysis miniprep. Typically growing an overnight culture of *E.coli* containing a pASV3mod derived plasmid in 10 ml LB + ampicillin results in a low yield of plasmid after miniprep. To address this, cultures

were instead grown overnight (approximately 16 hrs) in 20 ml of LB + ampicillin in 50 ml Falcon tubes, at which point chloramphenicol was added to a final concentration of 170 µg/ml and incubation continued for a further 8 hours before harvesting. This protocol improves the yield of plasmid from <100 ng per µl to >250 ng per µl, which is beneficial for the necessary sequencing reactions and intended yeast transformations.

- Ligation reactions with T4 ligase were reduced in final volume to 10 µl and carried out in a thermal cycler, cycling between 10 °C for 30 secs and 30 °C for 30 secs, for 16 hours. This reputedly increases cloning efficiency [99] (supposedly by increasing T4 ligase activity by overcoming the discrepancy between the optimum temperature (~37 °C) for T4 ligase and the low melting temperature of short restriction enzyme generated DNA overhangs).

PIP5KIα i2 CDS was amplified (primers in **Supplementary Table 17**) which each contain either an XhoI site or BamHI site. The PCR product was cloned into the pCR4 Blunt TOPO vector and the TOPO reaction mixture transformed. A resulting colony was amplified and the plasmid isolated by alkaline lysis miniprep. 1 µg of the plasmid was restriction digested with XhoI and BamHI and the digestion products

resolved by agarose gel electrophoresis. The 1.6 kb PIP5KIα i2 CDS band was excised and the DNA purified by DNA gel extraction kit (Monarch®). Similarly, 1 µg of pASV3mod plasmid was digested with XhoI and BamHI, resolved by agarose gel electrophoresis, excised and purified.

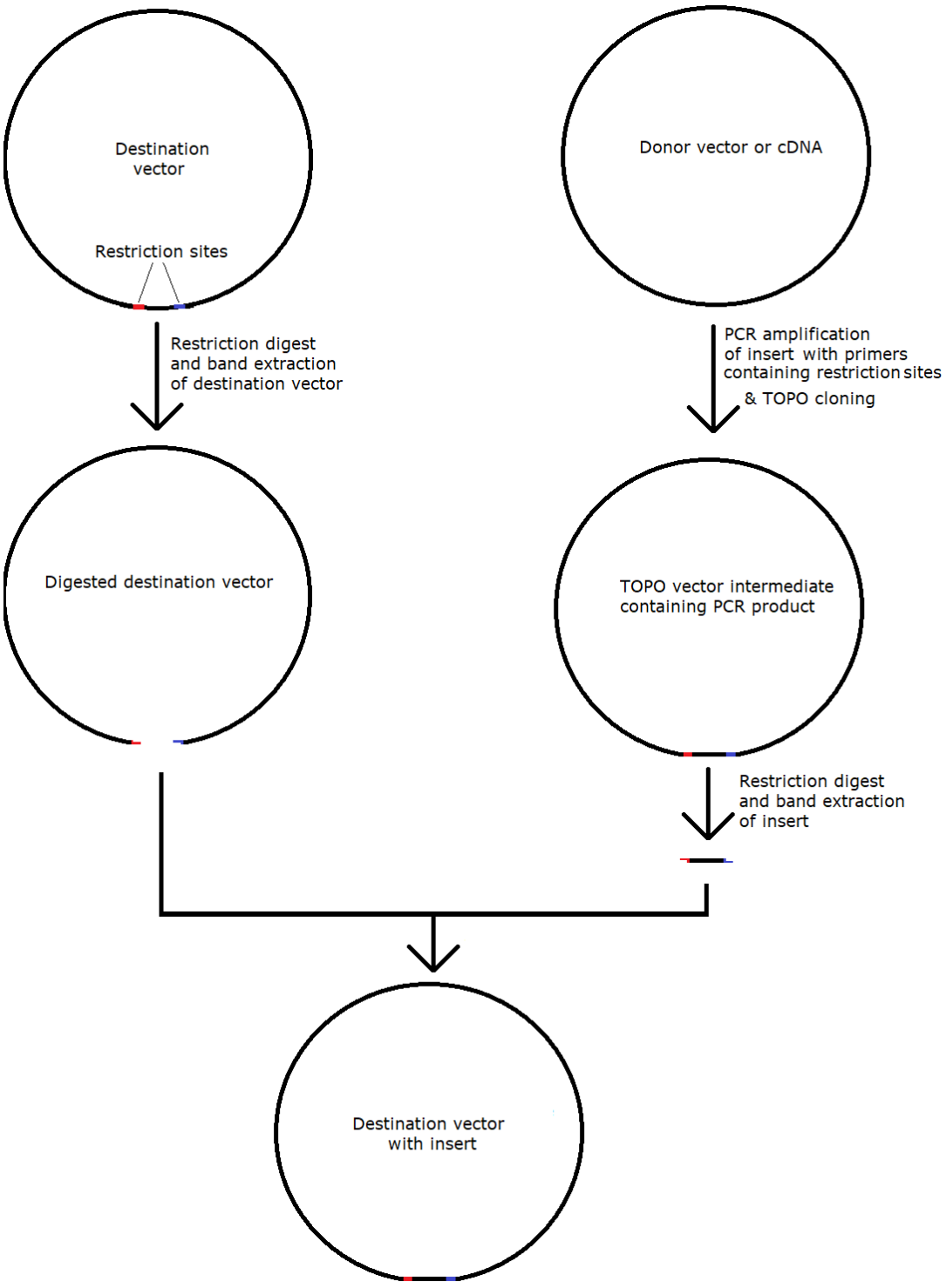
A ligation reaction between 188 ng digested PIP5KIα i2 CDS fragment and 90 ng digested pASV3mod was set up (molar ratio of ~ 3:1) and the ligation products used to transform chemically competent DH5α, which were plated on an LB ampicillin plate. 91 colonies were attained. Zero colonies were attained for a control reaction receiving no transforming DNA. 16 of the 91 colonies were screened by colony PCR with primers targeting *PIP5K1A* Exon 11 and Exon 13 (specifically amplifying PIP5KIα i2 CDS) with an expected product size of 295 bp. Of the 16 screened colonies, 15 (94 %) were positive. Three such positive colonies were amplified for alkaline lysis miniprep and the harvested plasmid subjected to restriction digest with the XhoI and BamHI, confirming the identity of these plasmids as pASV3mod PIP5KIα i2 (**Figure 45 B**). One of these isolated plasmids was submitted to Sanger sequencing, confirming the sequence of the insert as identical to the listed CDS of PIP5KIα i2, with no introduced errors and in the correct reading frame for generation of the NLS-VP16 fusion protein.

These protocols were subsequently used to clone PIP5KIα i2 CDS into pBTM116 mod, and to clone the CDS of PIP5KIα x15 (residues 1-485), Star-PAP (isoform 1; residues 1-874) and K-Ras isoform 2b (K-Ras 2b; 1-188) into both pASV3mod and pBTM116mod. A typical ligation experiment using the protocols, followed by transformation would yield 100 – 200 colonies, of which ~95 % of colonies screened by colony PCR were positive for the insert. Note, XhoI and BamHI could not be used to clone Star-PAP CDS into either pBTM116mod or pASV3mod because the Star-PAP CDS contains both of these sites. The primers for amplification of Star-PAP CDS contained SalI and BglII restriction sites, which were used to release this insert from the intermediate TOPO vector. SalI and BglII are respectively isocaudomers of XhoI and BamHI: they recognise different restriction sequences but generate compatible cohesive ends. The linker sequence for NLS-VP16 to Star-PAP is GAILD (instead of GAILE), and for LexA to Star-PAP is EFPGLD (instead of EFPGLE).

The high cloning efficiency of this approach marks a significant improvement to molecular cloning for the Y2H vectors, particularly pASV3mod. Because TOPO cloning has high efficiency, the addition of this step is straightforward. The additional time required to TOPO clone, then miniprep and release and purify the desired fragment (three days) is

compensated by the high efficiency of cloning into the final destination vector, which saves the time otherwise wasted on multiple cloning experiments.

A)


Destination vector
Restriction sites
Donor vector or cDNA
Restriction digest and band extraction of destination vector
PCR amplification of insert with primers containing restriction sites & TOPO cloning
Digested destination vector
TOPO vector intermediate containing PCR product
Restriction digest and band extraction of insert
Destination vector with insert

B)

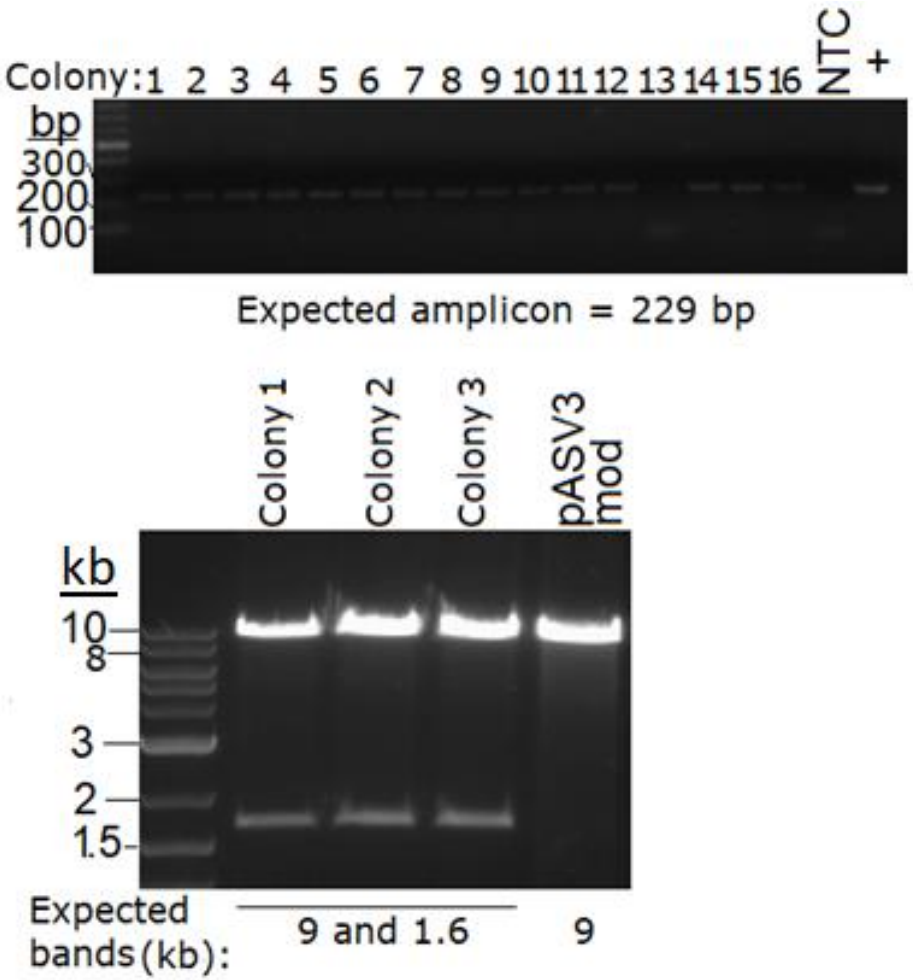


Figure 45. Improved protocols for cloning with Y2H vectors

A) Workflow cloning into pASV3mod or pBTM116mod, via a TOPO cloning vector intermediate for the insert. The principle is almost identical to standard subcloning procedures, with the exception that PCR amplified insert is first TOPO cloned into a TOPO vector (pCR4 BLUNT TOPO in this case), prior to release from that intermediate by restriction digest. The released fragment is then cloned into the final destination vector by standard T4 ligase ligation.

B) Agarose gel electrophoresis of experiments screening colonies for insertion of *PIP5K1A*_i2 into pASV3mod by colony PCR (Top) and restriction digest (Bottom). 15 of 16 colonies (94%) contained the target sequence, which is specific to PIP5KIα i2 CDS. Three colonies were mini-prepped and restriction digested with XhoI and BamHI. Bands at ~9 kb and 1.6 kb confirm that the plasmids are pASV3 PIP5KIα i2. Top: 1.5 % agarose; 1 kb plus ladder (NEB). Bottom: 0.6 % agarose; 1 kb plus ladder (NEB).

5.4. Assessing PIP5KIα interaction with AR-LBD and AR-NTD by Y2H X-gal assay

To address the possible issue that low sensitivity of the ONPG based β-galatosidase assay could prevent detection of weak interactions between PIP5KIα and itself or AR LBD, a semi-quantitative pellet X-gal β-galatosidase assay (PXG) and modified colony filter assay were performed, based upon the methods of Möckli and Auerbach[102].

Because the typical colony filter lift assay is performed by transferring a colony grown on solid media onto filter paper, it is not suitable for assessing steroid ligand dependent interactions, which require overnight growth in liquid media containing the appropriate ligand. The colony filter lift assay is reportedly the most sensitive (though only qualitative) because of the long incubation time of yeast cell lysate in X-gal solution. The assay was therefore modified to use not a streaked yeast colony, but a fixed cell number (assessed by OD 600 measurement; 1 AU) of overnight yeast culture (cultured with or without ligand), that was pelleted and resuspended in a fixed volume of water (20 μl) before spotting onto filter paper. The subsequent steps (lysis by freeze thaw in liquid Nitrogen, then overnight incubation in X-gal solution) are identical to the method used by Möckli and Auerbach. These assays were also

used to assess interactions between NLS-VP16-AR NTD and LexA-PIP5KIα.

Ligand dependent homodimerisation of AR LBD was detectable by these assays, as was ligand dependent interaction between AR LBD and AR NTD. Interestingly, co-transformation of LexA-AR LBD with NLS-VP16 alone resulted in weak β-galatosidase activity in presence of ligand (but not in absence of ligand), suggesting that the LexA-AR LBD fusion is capable of weakly activating transcription of the reporter gene in a ligand dependent manner to a level not detectable by the ONPG β-galatosidase assay. The observed β-galactosidase activity likely results from the LexA portion of the fusion binding to the reporter gene promoter, with the AF2 domain present in AR LBD portion driving transcriptional activation. This supports the report that the pellet X-gal assay is more sensitive than ONPG based assays [102].

Colour formation that could indicate interaction between LexA-AR LBD and NLS-VP16-PIP5KIα i4 K181A was observed in the presence of ligand, but this was not significantly greater (and was in fact significantly lower; two-tailed t test, t= 4.47, p= 0.01) than the observed colour formation for the ligand dependent LexA-AR LBD/ NLS-VP16-empty interactions. Thus the weak observed colour change for the former reactions is likely driven only by the weak transcriptional activity of LexA-

AR LBD: there is no detectable interaction between LexA-AR LBD and NLS-VP16-PIP5KIα I4 K181A.

Colour formation was not observed for reactions assessing homodimerisation of PIP5KIα i4 (or PIP5KIα i4 K181A), even in the most sensitive assay, the overnight qualitative colony filter assay (**Figure 46 A-C**). This indicates that either PIP5KIα i4 does not homodimerise, which is in contravention of previous literature reports [66; 194], or there is some fault with one or both of the fusion proteins. Faults could include failure of one or both fusion proteins entering the nucleus, or misfolding of either fusion protein. Expression of the LexA-PIP5KIα i4 and NLS-VP16-PIP5KIα i4 K181A proteins were previously determined for similarly co-transformed colonies (**Figure 42 A**), indicating any such failure is not in expression of these fusion proteins.

Strong colour formation was observed for assays between LexA-AR LBD and NLS-VP16-AR NTD in ligand dependent fashion. The colour formation is greater than for LexA-AR LBD with NLS-VP16 (+ ligand), and no colour formation was observed for NLS-VP16-AR NTD with LexA (+/- ligand) indicating that the observed colour change for LexA-AR LBD/ NLS-VP16-AR NTD assays results from AR LBD/ AR NTD interactions. This is in agreement with reports of ligand dependent interactions between AR LBD and AR NTD, mediated

by helix 12 of the AR LBD, and ^{23}FQNLF^{27} and ^{33}WHTLF^{437} in the AR NTD [31; 187].

Colour formation was not observed for interaction between LexA-PIP5KIα i4 and NLS-VP16-AR NTD. The reciprocal experiment was not performed because of the likelihood that a LexA-AR NTD fusion would activate the reporter gene, and possibly also lead to squelching.

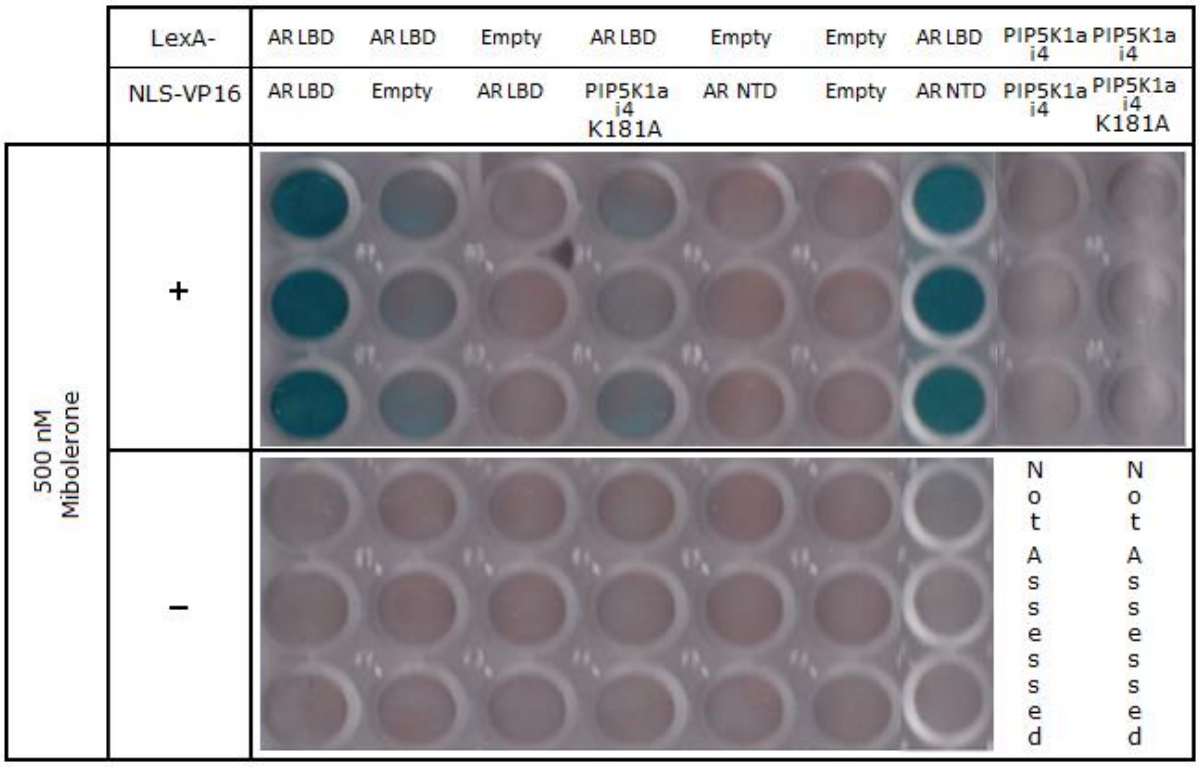
Feasibly the reported co-immunoprecipitation of AR with PIP5KIα could be isoform specific, and so Y2H pellet x-gal assays were also performed to assess interaction of three isoforms of PIP5KIα, isoform 2, isoform 4 and isoform x15 (each fused with LexA), with either NLS-VP16-AR LBD (+/- ligand) or NLS-VP16-AR NTD. No interaction was detected (**Figure 46 D**).

A final set of Y2H pellet x-gal assays was performed to assess reported dimerisation of PIP5KIα with Star-PAP and K-Ras isoform 2b, which reportedly interact [134; 149]. Again, interaction was not detectable for any tested pair. This experiment was intended to provide a positive control for PIP5KIα interactions in this Y2H system. Its failure indicates that revision of the experimental approach is required.

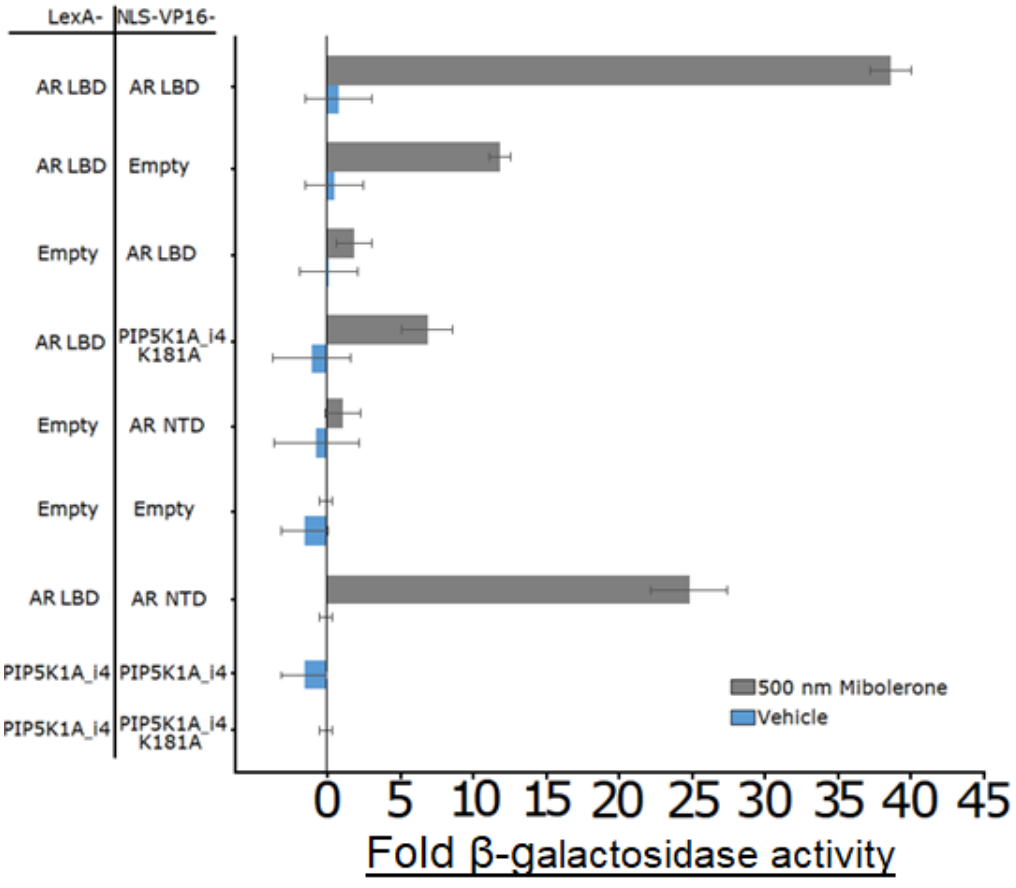
Experiments assaying interaction between LexA-PIP5KIα fusions and NLS-VP16-Star-PAP were not possible because all such transformations yielded zero yeast colonies. Confirmation

of expression of these fusion proteins by western blotting was not performed due to insufficient time.

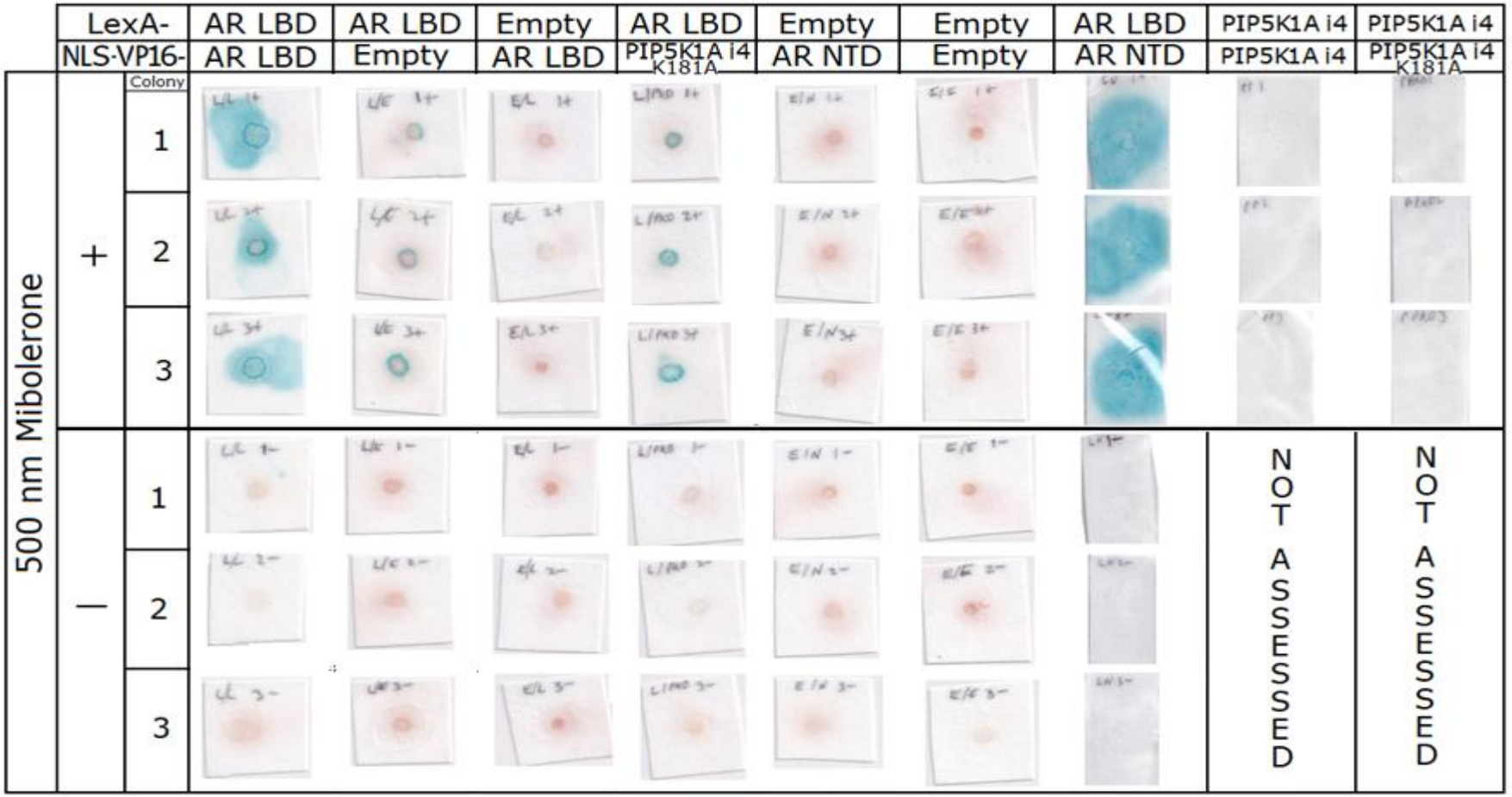
A)


LexA-
AR LBD
AR LBD
Empty
AR LBD
Empty
Empty
AR LBD
PIP5K1a i4
PIP5K1a i4
NLS-VP16
AR LBD
Empty
AR LBD
PIP5K1a i4 K181A
AR NTD
Empty
AR NTD
PIP5K1a i4
PIP5K1a i4 K181A
500 nM Mibolerone
+
–
Not Assessed
Not Assessed

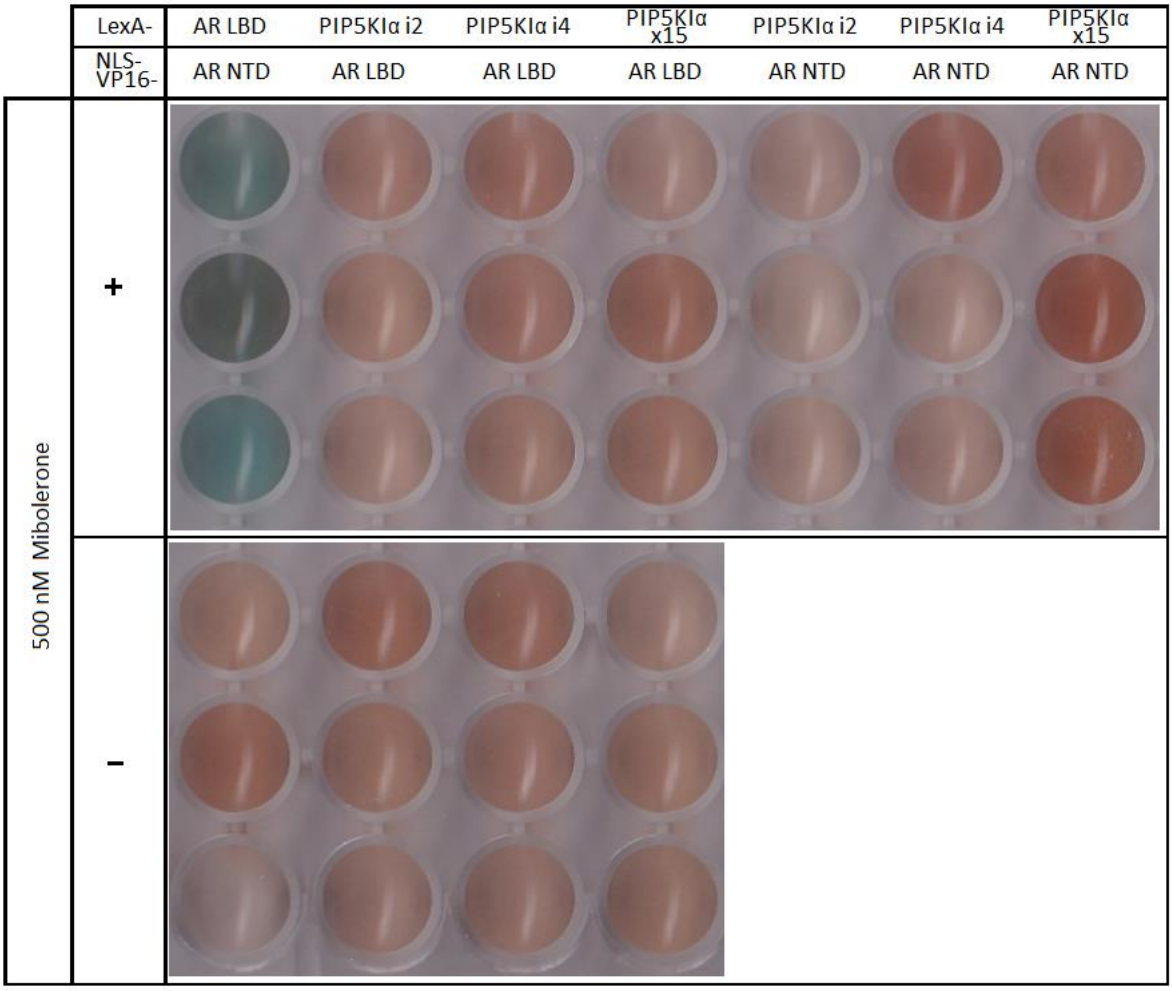
B)


LexA-
NLS-VP16-
AR LBD
AR LBD
AR LBD
Empty
Empty
AR LBD
AR LBD
PIP5K1A_i4 K181A
Empty
AR NTD
Empty
Empty
AR LBD
AR NTD
PIP5K1A_i4
PIP5K1A_i4
PIP5K1A_i4
PIP5K1A_i4 K181A
500 nm Mibolerone
Vehicle
0 5 10 15 20 25 30 35 40 45
Fold β-galactosidase activity

C)


LexA-
NLS-VP16-
AR LBD
AR LBD
Empty
AR LBD
Empty
Empty
AR LBD
PIP5K1A i4
PIP5K1A i4
AR LBD
Empty
AR LBD
PIP5K1A i4
AR NTD
Empty
AR NTD
PIP5K1A i4
K181A
500 nm Mibolerone
Colony
+
—
1
2
3
NOT ASSESSED
NOT ASSESSED

D)


LexA-
AR LBD
PIP5KIα i2
PIP5KIα i4
PIP5KIα x15
PIP5KIα i2
PIP5KIα i4
PIP5KIα x15
NLS-VP16-
AR NTD
AR LBD
AR LBD
AR LBD
AR NTD
AR NTD
AR NTD
500 nM Mibolerone
+
–

E)

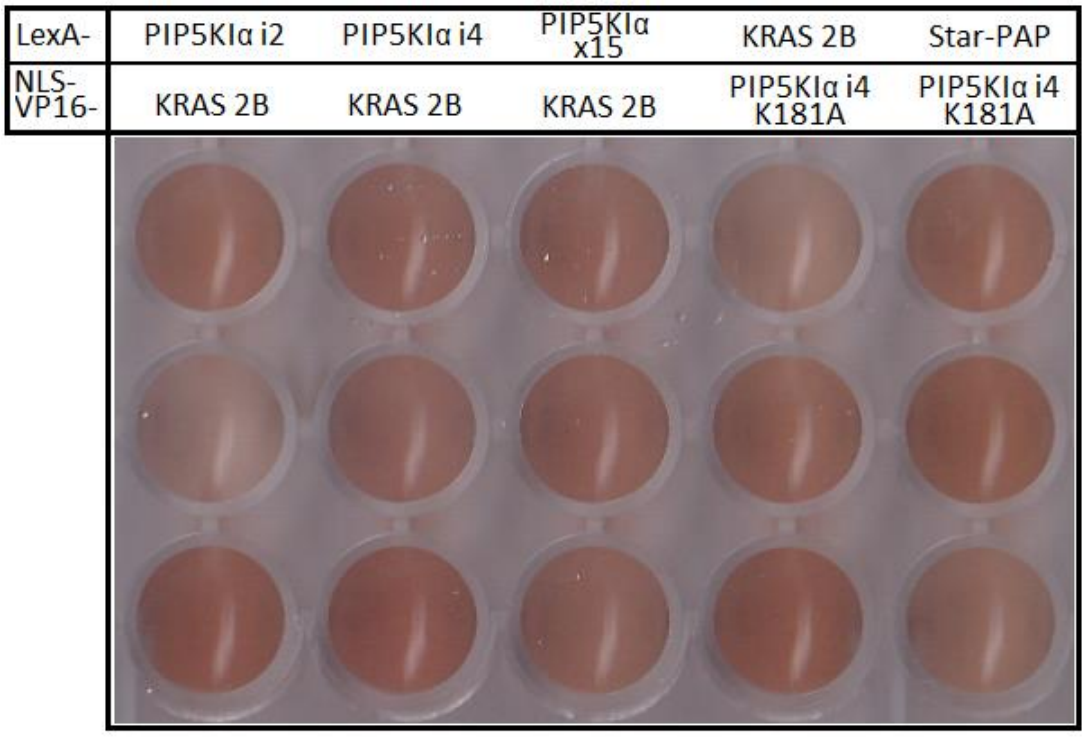

LexA-
PIP5KIα i2
PIP5KIα i4
PIP5KIα x15
KRAS 2B
Star-PAP
NLS-VP16-
KRAS 2B
KRAS 2B
KRAS 2B
PIP5KIα i4 K181A
PIP5KIα i4 K181A

Figure 46. Assessing PIP5KIα i4 interaction with AR-LBD and AR-NTD by Y2H X-gal assay.

A) Scanned image of a PXG assay performed in a 96-well microtiter plate after 30 min of incubation. Three independent transformants were assayed per protein interaction pair. Blue colour formation results from β-galactosidase activity, converting colourless X-gal to galactose and 5-bromo-4-chloro-3-hydroxyindole, the latter of which spontaneously dimerises and is oxidized to form 5,5'-dibromo-4,4'-dichloro-indigo, which has a dark blue colour. Note the weak colour formation in wells for lysates with LexA-AR LBD fusion and either NLS-VP16 or NLS-VP16-PIP5KIα.

B) Fold β-gal activity in the PXG assay, relative to readings for the reactions with lysates of colonies transformed with empty pBTM116 mod and empty pASV3mod. Error bars represent the standard deviation from three replicates. Activity in wells with LexA-AR LBD / NLS-VP16-PIP5KIα i4 K181A (+ ligand) was not significantly greater than for LexA-AR LBD / NLS-VP16 (+ ligand), indicating no interaction between AR LBD and PIP5KIα i4 K181A. Importantly, no colour formation was observed LexA-PIP5KIα i4 / NLS-VP16-PIP5KIα i4 K181A or NLS-VP16-PIP5KIα i4 K181A, meaning there is no positive control for PIP5KIα interactions.

C) Modified colony filter lifts assays were performed by spotting of pellets of overnight culture, lysis in liquid N_2, then overnight incubation in X-gal buffer. This assay is reportedly more sensitive than the pellet X-gal assay, but is not quantitative. Note that there is strong colour formation for assays testing LexA-AR LBD with either NLS-VP16-AR LBD or NLS-VP16-AR NTD with ligand. Weak ligand dependent colour formation also seen

D) Scanned image of a PXG assay performed in a 96-well microtiter plate after 30 min of incubation. Three independent transformants were assayed per protein interaction pair. Blue colour formation results from β-galactosidase activity. This experiment was designed to assess interactions of PIP5KIα isoforms with AR LBD or AR NTD.

E) Scanned image of a PXG assay performed in a 96-well microtiter plate after 30 min of incubation. Three independent transformants were assayed per protein interaction pair. Blue colour formation results from β-galactosidase activity. This experiment was designed to assess interactions of PIP5KIα isoforms with K-Ras 2b or Star-PAP.

5.5. Summary and discussion

Although co-immunoprecipitation of ERα or AR with PIP5KIα has been reported [95; 181], these studies did not show supporting evidence for direct interactions e.g. using two hybrid, GST pulldown assays, co-localization in vivo, or by mapping putative interaction interfaces or site-directed mutagenesis. In this study, we used Y2H to explore if direct interactions of AR LBDs and PIP5KIα occur in cells. Y2H has been phenomenally useful in demonstrating and probing interactions of proteins in solution and enabled the identification/validation of many important NR/ cofactor complexes. While AR LBD functionality was confirmed in this work by ligand-dependent interaction with the AR NTD, we did not detect any interactions of AR LBD with PIP5KIα fusion constructs.

The toxicity of the WT NLS-VP16-PIP5KIα construct in contrast to the K181A mutant also suggests that the PIP5KIα enzymatic activity may be conserved in yeast and thus the protein is likely to be functional. However, to exclude the possibility that lack of interaction of AR and PIP5KIα in Y2H assays is due to poorly folded PIP5KIα, positive controls for PIP5KIα protein-protein interactions are required.

PIP5KIα and other PIP5KIs have been reported to undergo homodimeric interactions ([66; 194]) which may involve orientation

of the proteins at the plasma membrane by binding to PIPs. Hu *et al*, reported the structure of the zebrafish PIP5KIα catalytic domain and identified regions involved in dimerization and interaction with DVL (dishevelled), a WNT signalling component thought to activate PIP5KIα and PI(4,5)P_2 production *in-vivo*. In this study we assessed whether co-expression of LexA (DNA binding domain)-PIP5KIα i4 (NCBI) and NLS-VP16 (activation domain)-PIP5KIα i4, could activate reporter activity in a Y2H assays, as evidence for dimerization interactions. Using quantitative or qualitative assays of reporter activation, no evidence of homodimerisation of PIP5KIα was detected in our hands, which may be due to a number of factors and limitations of the assay. Detection of Y2H interactions requires assembly of the fusion proteins on the promoter of the reporter gene in the nucleus. Although the AR fusion constructs clearly are transcriptionally active, we cannot verify that PIP5KIα constructs are present in the yeast nucleus. It is possible that the high affinity for localisation to cell membranes overrides nuclear localisation in our assay. In light of this finding, an alternative and possibly more suitable approach might be to explore membrane-based two hybrid systems, however there was not sufficient time to initiate those studies.

Secondly, it is possible that the localisation of PIP5Ks proteins to the nucleus may be isoform-specific, or alternatively

may be signal dependent e.g. as a consequence of post translational modifications. It would therefore be important to establish the subcellular localisation of human PIP5KIα fusion proteins expressed in yeast, e.g. through use of fluorescent tags. A recent study of the yeast cells revealed that PI(4,5)P_2 localises to the cytoplasmic leaflet of the yeast cell membrane [200]. PI(4,5)P_2 is generated from PI4P at the membrane by the MSS4p the only homolog of PIP5Ks in yeast cells. This suggests that functional human NLS-VP16-PIP5KIα may also localise to the yeast cell membrane. Interestingly, another yeast protein, BCP1, has been identified as a molecular chaperone that facilitates export of MSS4p from the yeast nucleus. This protein is related to the BCCIP family in humans (BRCA1 and CDNK1A interacting protein) that ubiquitinates p53 and is implicated in regulation of translation. Future studies will need to unravel the complex protein-protein interaction networks of PIP5ks to fully understand their roles in cells.

Although we found no evidence for AR LBD-PIP5KIα or PIP5KIα homodimeric interactions in yeast cells, the failure of the positive control for PIP5KIα interactions, meant that the possibility of PIP5KIα interaction with either NR LBD could not be completely excluded. Feasibly there could be some fault with either of the LexA-PIP5KIα /NLS-VP16-PIP5KIα-fusion proteins, such as sequestration at membranes (cytoplasmic or nuclear);

or very weak/transient interactions that result in low levels of β-galactosidase below detection. A final set of Y2H experiments were performed to test interactions of several isoforms of PIP5KIα with reported PIP5KIα interactors, namely Star-PAP and K-Ras isoform 2B [134; 149] also failed to detect of β-galactosidase activity, supporting the idea that there is a fault in using PIP5KIα in Y2H experiments. PIP5KIα has previously been used in Y2H, but only a C-terminal portion (residues 427-549 of PIP5KIα i2; [149]). Had this assay worked, it would have been interesting to determine PIP5KIα isoform specific interactions, particularly for Star-PAP/PIP5KIα i2 (previously demonstrated to occur through PIP5KIα i2 C-terminus) vs Star-PAP/PIP5KIα x15 (which has a different C-terminus).

For any further experiments utilising PIP5KIs in Y2H, it would be prudent to investigate removal of the membrane targeting activation loop, or else use only sequences of PIP5KIα unlikely to localise to membranes.

Finally, to overcome inherent difficulties of working with the Y2H vectors, particularly pASV3mod, it was shown here that *in vivo* assembly cloning or restriction site cloning via an intermediate TOPO vector are useful alternative approaches. The later greatly enhances subcloning efficiency using this vector (resulting in hundreds of colonies to be obtained and

~95% cloning efficiency for pASV3mod). This is a great improvement over previous protocols, which will facilitate future Y2H studies with these plasmids, overcoming the bottleneck that the cloning previously presented.

Concluding remarks

The Type I phosphatidylinositol 4-phosphate 5-kinases, including PIP5KIα, are responsible for the generation of the majority of cellular $PI(4,5)P_2$, an important signalling molecule required for cellular processes such as membrane dynamics, vesicular transport, actin remodelling and cell motility. $PI(4,5)P_2$ is the central substrate of the PI3K family of enzymes that generate $PI(3,4,5)P_3$, to activate the AKT/mTOR signalling pathway, and thus PIP5KIs lie upstream of one of the most frequently altered signalling pathways in cancer. An important aspect of activating these pathways in prostate tumours is deletion of one or both copies of *PTEN*, a gene encoding a phosphatase that converts $PI(3,4,5)P_3$ to $PI(4,5)P_2$, which combined with mutations in *AR*, *KRAS* and *TP53* may drive cancer progression to CRPC. The importance of PIP5KIα to growth and metastasis of breast [201] and prostate carcinoma [94; 95] has recently been shown in animal models. Moreover, evidence for direct or indirect interactions of PIP5KIα with AR, p53 and K-Ras proteins have been reported [95; 134; 135].

In prostate tumours, PIP5KIα appears to contribute to disease progression, since levels of PIP5KIα and $PI(4,5)P_2$ correlate with disease grade [94] and are greatly elevated in lymph node and bone metastases of PCa patients [95]. Proof of principle studies indicate that pharmacological targeting of PIP5KIα can curtail tumour growth and invasiveness in

xenograft mouse models [95; 201], although the precise mechanisms involved are unclear. Contribution of PIP5KIα to these processes appears to involve crosstalk with AR function: expression of both AR and AR-V7 correlate with expression of PIP5KIα in PCa tissue specimens, exogenous expression of AR-V7 in PCa cell lines leads to an increase in expression of PIP5KIα ,and pharmacological targeting or siRNA knockdown of PIP5KIα also reduces expression of AR and AR-V7. Further to this, Sarwar et al. demonstrated co-immunoprecipitation of PIP5KIα and AR/AR-V7 in nuclei of PCa cells, suggesting the existence of a protein complex which the authors posit to regulate transcriptional activity of AR and AR-V7, contributing to PCa disease progression and castration resistance [95]. Thus, while a role for PIP5KIα in the PCa and its relationship with AR has begun to be uncovered, the molecular mechanisms behind these phenomena remain unclear.

Another potential aspect of *PIP5K1A* function in cancer emerging in the literature comes from reports of the detection of circular *PIP5K1A* RNAs in colon [202], ovarian[203], bone [204] and lung tumours [205]. These are proposed to act as microRNA sponges, selective for particular miRNAs in different tissue types. Further work will be required to validate these observations to discriminate between actions of the circular transcripts and the encoded proteins.

This study undertook an investigation of PIP5KIα function using CRISPR edited cells to expand understanding of its molecular actions in a metastatic prostate cancer cell line, LNCaP C4-2. I continued prior work from the group, genotyping and characterising expression of PIP5KIα in CRISPR Cas9 nickase edited clones of the LNCaP C4-2 cell line, which is a standard model of androgen resistant metastatic prostate cancer. This resulted in the selection of a clone in which expression of PIP5KIα was largely ablated (C4-2 PIP5K1A-KO; KO). A second clone targeted at the canonical Exon 1 start codon was found to express an N-terminally truncated PIP5KIα protein (C4-2 PIP5K1A-ΔN; ΔN).

RNA-Seq analysis comparing the transcriptomes of the LNCaP C4-2 unedited cells with those of the KO and ΔN sublines identified a substantial number of DEGs in KO cells, involving genes implicated in cell cycle regulation, K-Ras signalling, p53 pathways, metastasis, steroid hormone signalling and epithelial mesenchymal transition. To our knowledge this is the first report to profile gene expression changes due to loss of PIP5KIα function, and this highlights the importance of PIP5KIα in prostate cancer cells. Our findings of alterations in K-Ras and P53 signalling are interesting given recent reports of physical interaction of PIP5KIα with K-Ras and P53 [134; 135].

There was a strong degree of consistency amongst the up- and down-regulated DEG cohorts in the KO and ΔN cells, although the number of DEGs was higher in the ΔN cells. This may be a secondary effect due to the hyperactivation of the AKT pathway, as noted by high levels of phospho-AKT. We noted that amongst the most significantly altered DEGs were androgen regulated genes such as *KLK3* and *KLK4* (androgen activated) and *UGT2B15, UGT2B17* (androgen repressed), consistent with crosstalk between PI(4,5)P_2 and androgen signalling. These DEGs were validated by RTqPCR. Importantly, rescue experiments in which an mVenus-tagged version of PIP5KIα was exogenously expressed in KO cells showed at least partial reversal of the changes in *KLK3* and *UGT2B17* expression back to levels seen in unedited C4-2 cells. In future studies it would be of interest to explore expression of these genes in response to androgens and antiandrogens, given they are known to be differentially regulated by AR. This supports the hypothesis that loss of PIP5KIα results in significant changes in the expression of AR-regulated genes, but we did not observe substantial changes in *AR* expression. Moreover, the data indicate a potential role for PIP5KIα in regulating expression of enzymes (UGT2Bs) that control the bioavailability of androgens in prostate tumours, which could be another mechanism by which elevated PIP5KIα expression contributes to PCa

progression. Future research in this direction should include analysis of UGT2B protein levels in the CRISPR edited cells, as well as of the cells' androgen sensitivity and their intracellular levels of glucuronidated androgen metabolites. Further to this, the molecular mechanisms leading to differential expression of these androgen responsive genes in the CRISPR edited cells should be investigated.

Several studies have proposed nuclear roles for PIP5KIα [122; 135; 149; 178], including a mechanism by which PIP5KIα activity regulates expression of certain transcripts by physical interaction with the polyA polymerase, Star-PAP [149]. It would be interesting to determine if androgen responsive genes are subject to similar mechanisms of regulation by PIP5KIα in PCa, and if so, how this intersects with AR transcriptional regulation of these genes, and what role such mechanisms play in the pathogenesis of PCa. Nuclear PIP5KIα has also been proposed to regulate p53 through direct interaction and association of p53 with $PI(4,5)P_2$. This was proposed to stabilise mutant p53 by promoting interaction with heatshock chaperones HSPB1 and HSPB5 [135]. Finally, nuclear $PI(4,5)P_2$ has been proposed to allosterically regulate the interactions of TAF3, ING2 and UHRF1 with histones/ chromatin through binding to PHD finger domains or other motifs involving polybasic residues [206; 207]. Similarly, phosphoinositides have been proposed to bind to

Nuclear Receptors such as LRH1/NR5A2 and SF1/NR5A1, or HDAC proteins (reviewed in [208]), thus future studies may uncover further detail of the function of these complexes.

Remarkably, we noted a strong upregulation of AKT phosphorylation in the ΔN subline, suggesting a mechanism linked to deletion of the N-terminal 36 or 46 amino acids of PIP5KIα. Though the divergence of the N-termini and C-termini of PIP5KIα, PIP5KIβ and PIP5KIγ has previously been suggested as a possible cause for the divergent functions of the three isozymes [156] no previous study has specifically investigated the function the divergent region of the N-termini of PIP5KIα, PIP5KIβ or PIP5KIγ. We hypothesised that the elevated level of phospho-AKT may result from altered localisation of N-terminally truncated PIP5KIα, but N-terminally truncated variants of PIP5KIα fused to mVenus (a derivative of yellow fluorescent protein) showed no difference in localisation compared to full-length PIP5KIα-mVenus fusions. Likewise, fusions of the first 36 amino acids of PIP5KIα did not display different localisation from mVenus alone. Alternatively, the N-terminus of PIP5KIα could have a role in regulating catalytic activity of PIP5KIα, or in recruiting other proteins that gather at sites of PI(4,5)P_2 synthesis, such as AKT; at least one C-terminal variant of PIP5KIγ (PIP5KIγ i2) associates in such

complexes which effect AKT activation [209]. Further experiments to assess the effect of exogenous expression of N-terminally truncated PIP5KIα on AKT phosphorylation were considered, but given the poor transfection efficiency of LNCaP C4-2, such experiments were not pursued beyond preliminary stages. Such experiments could be considered in future using lentiviral delivery systems to overcome the poor efficiency of LNCaP C4-2 transfection. Although mechanisms of the elevated AKT phosphorylation that we observed were not explored further in this study, they are currently being investigated by another researcher in our lab, Ryan Morris.

Curious as to the identity of the PIP5KIα protein expressed in the ΔN subline, I investigated splice variants of *PIP5K1A* expressed in LNCaP C4-2 cells. Initial bioinformatic analyses showed that the number of potential splice variants of *PIP5K1A* is far greater than the five variants that have previously been reported [145].

Through these analyses I identified seven 5' exons of *PIP5K1A* which are used instead of/in combination with the previously identified first exon, and I also found that the 3' portion of this canonical first exon is a cryptic intron. Molecular cloning confirmed the existence of *PIP5K1A* transcripts containing these alternate exons in cDNA from LNCaP C4-2 and MCF7, a breast cancer cell line. Analysis of RNA-Seq data

indicates that splice events involving these exons are very rare compared to those involving the canonical Exon 1, and it remains to be determined if such transcripts have functional significance e.g. in production of circular RNAs or in isoform specific targeting to cellular compartments.

Analysis of the evolutionary conservation of such sequences was performed, finding conservation of splice events involving these exons to varying degrees in diverse orders of mammals, suggesting that such transcripts of *PIP5K1A* have been evolutionarily selected. Whether these transcripts are translated, and if so whether the encoded isoforms are functionally distinct from the canonical isoforms are open questions. Since the N-termini of these isoforms is distinct from that of the canonical isoforms of PIP5KIα, any function of the N-terminus would likely be altered in these isoforms.

In addition to these splice variants with altered 5' sequences, I detected novel splice variants arising from altered 3' Exon usage, again confirming their existence in LNCaP C4-2 and MCF7 cells by molecular cloning. The deduced translation of this splice variant showed that it encodes an isoform of PIP5KIα (PIP5KIα x15) with a greatly altered C-terminus, due to the phenomenon of dual coding exons. As the C-terminus of one canonical isoform of PIP5KIα had previously been shown to function in localisation of PIP5KIα to nuclear speckles [149], I

investigated the localisation of the C-terminus of PIP5KIα x15 by expression of mVenus fusions of the unique termini of the PIP5KIα isoforms. Interestingly, I found that the C-terminus of PIP5KIα x15 does not localise to nuclei in transfected cells, in contrast the termini of canonical variants of PIP5KIα. Further experiments mapped the sequence responsible for this behaviour to residues that resemble a CRM1 dependent nuclear exclusion sequence. Analysis of RNA-Seq data sets indicated that the splice event characteristic of the PIP5KIα x15 isoform is present in approximately 4-8% of reads involving these exons, being more abundant than reads characteristic of the previously reported PIP5KIα i4 isoform. Analysis of the evolutionary conservation of the splice event characteristic of the *PIP5K1A* x15 splice variant showed that it is also conserved amongst mammals of diverse orders. Together this indicates that the *PIP5K1A* x15 splice variant is a relatively abundant transcript, and the encoded isoform may have an evolutionarily conserved function that is distinct from previously reported isoforms of PIP5KIα.

Future work on the splice variants of *PIP5K1A* should begin with development of assays (such as RT qPCR or Nanopore sequencing) to quantify the expression levels of each transcript. Such an assay could then be used to quantify expression levels of *PIP5KIA* splice variants in various prostate

and PCa cell lines or tissue specimens to assess any association between their expression and the various stages of PCa.

An investigation into the functions of isoforms encoded by these splice variants was begun for the PIP5KIα x15 isoform. This work should continue to delineate functional differences between the various PIP5KIα isoforms. A screen of protein-protein interactors with the C-terminus of PIP5KIα i2 by Y2H previously uncovered its interaction with Star-PAP [149]. Our lab has the capability to conduct similar screens for interactors with the unique C-terminus of PIP5KIα x15, as well as that of PIP5KIα i4, which has also not been previously studied. Like the numerous C-terminal variants of PIP5KIγ, the various C-terminally variant isoforms of PIP5KIα may have unique binding partners and functions.

Work could also continue on the numerous N-terminal variants of PIP5KIα putatively encoded by the 5' splice variants, seeking to determine any function of the N-terminus of canonical PIP5KIα isoforms and how this is altered in the isoforms with distinct N-termini.

Finally, I used Y2H to explore interactions of PIP5KIα proteins with domains from the AR, K-RAS and Star-PAP, which have been reported to interact with PIP5KIα in previous studies [95; 134; 149]. While interactions of the AR ligand-binding domain with AR N-terminal domain were successfully detected in

response to ligand, no interactions with PIP5KIα constructs were detectable. I suspect this is due to the localisation of PIP5KIα to the yeast cell membrane, making use of full length PIP5KIα unsuitable for the Y2H system used in this study. Future work in this area should repeat experiments but remove the membrane targeting activation loop from the PIP5KIα constructs, or else use a Y2H system capable of detecting protein-protein interactions that occur at membranes, such as the split-ubiquitin system [210].

In summary this thesis has characterised the expression and function of PIP5KIα in the prostate cancer CRPC cell line LNCaP C4-2, including two novel CRISPR edited subclones having loss or mutation of PIP5KIα. These cells will be used to explore molecular functions of PIP5KIα in future. It has been shown for the first time that loss of PIP5KIα expression results in substantial transcriptome changes that include AR/ androgen regulated genes. Alternative splice variants of *PIP5K1A* that expand the repertoire of PIP5KIα proteins have been identified, and deletion analysis performed to assess how N- or C-terminal domains of PIP5KIα impact subcellular localisation. Proposed direct protein-protein interactions of PIP5KIα were explored by Y2H, and the limitations of this approach to study interactions of PIP5KIα were identified.

Appendices

Supplementary data

qPCR primer validation.

Primers targeting the *UGT2B* genes, *UGT2B10, UGT2B11, UGT2B15, UGT2B17* and *GUSB* were designed using Primer BLAST [211] (**Supplementary Table 18**). Validation of these primers for use in qPCR is described here.

Because of the large number *UGT2B* genes and pseudogenes, designing primers specific to transcripts of any one *UGT2B* gene is difficult. It was therefore necessary to target primers to the first or final exons of the target *UGT2B* genes, as these display greater diversity than the interior exons. This approach is in contrast to conventional approaches for design of qPCR primers in which each primer for a particular gene is targeted to a different exon of that gene, or to an exon-exon boundary. These are typically followed to ensure amplification occurs only from cDNA and not genomic DNA; the large distance between the primer binding sites in gDNA precludes amplification, but absence of introns from mRNA, and thus cDNA, allows amplification from cDNA. To control for the possibility of amplification from contaminating genomic DNA, degradation of genomic DNA was performed during cDNA synthesis, using the DNase ('gDNA wipeout') supplied with the

cDNA synthesis kit (Quiagen). Mock cDNA synthesis reactions were also performed, which underwent DNase treatment but to which no reverse transcriptase was added- these were used as input for PCR (no RT control), to assess amplification from contaminating gDNA.

Validation of these primers was performed as per the recommendations of Taylor *et al*., [212]. Briefly, a problem associated with working with qPCR is that amplification with a primer pair may not represent the true concentration of target nucleic acid in a given sample. This is of particular concern when working with samples from a new experiment where it is uncertain if protein and chemical contaminants residual from the RNA extraction vary between the individual biological replicates. Such contamination can affect efficiency of a PCR, and thus varying levels of contamination between samples can lead to CT values that are artefactual and do not represent the true abundance of target cDNA.

To address this issue, it is recommended that an optimum annealing temperature (TA) be determined for each primer pair, and a dilution of samples at which PCR is not inhibited be ascertained.

The optimum annealing temperature is defined as the minimal annealing temperature at which amplification of only

the intended target occurs. It is recommended that this be assessed by conducting thermal gradient qPCRs for each primer pair for a range of TAs:

1) These experiments should be performed with a sample consisting of a pool of each of the individual biological replicates, thus representing the average of the quantity of targets, and average of contaminants.

2) The thermal a range of TAs should be around the predicted Ta for each primer pair. The Ct value should decrease with TA, until the Cts converge – TAs within the temperature range at which the Ct values converge, and for which no secondary products form, are the optimal (most efficient) TAs for the primer pair. Products should be run on an agarose gel to confirm a single band of the expected size for the intended amplicon. Melt curve analysis should also be performed for SYBR based assays, also to confirm a single peak, and thus single product.

Without access to a qPCR instrument capable of performing a thermal gradient, gradient PCR (in a thermal cycler capable of gradient PCR but not quantification) was instead performed for each primer pair, followed by qualitative assessment of product intensity by agarose gel electrophoresis (**Supplementary Figure 1 A**). This also allowed determination of a minimal TA for each primer pair at which no amplification

of secondary products was detectable. A control reaction was also performed for each primer pair, using as input a 1:20 dilution of pooled RNA sample that had undergone DNase treatment but no reverse transcription (no RT control).

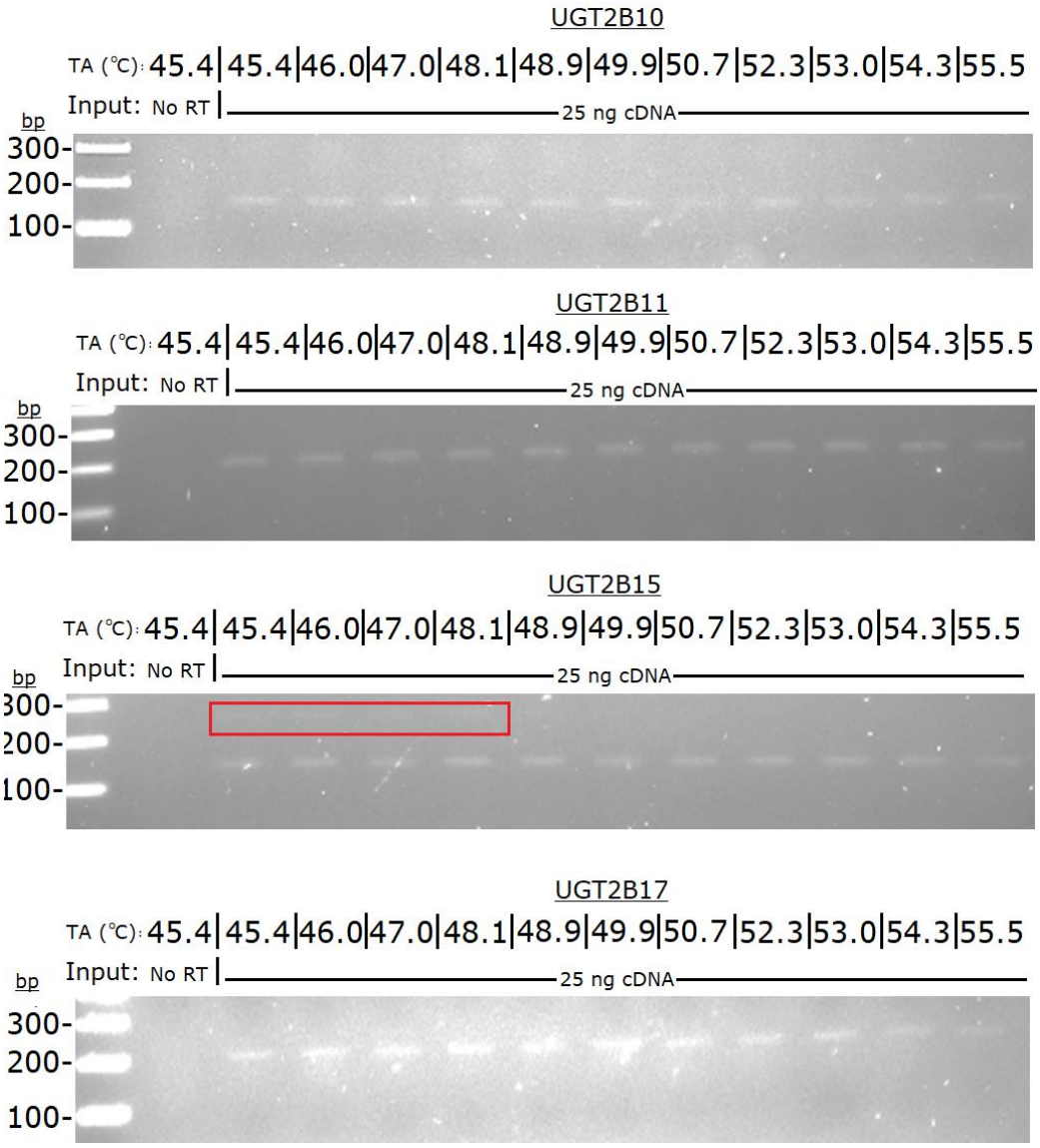
This assessment indicated that for each of the *UGT2B10*, *UGT2B11* and *UGT2B17* primer pairs, the minimum TA for which no secondary products were apparent was 45.4 °C, the lowest temperature tested. For the *UGT2B15* primer pair, a secondary product was very faintly visible just below 300 bp for TAs 48.1 °C and below. Because the qPCRs with each of these primers needed to be performed simultaneously, and because the intensity of amplicons for *UGT2B10, UGT2B11* and *UGT2B17* primers at 45.4 °C and 48.9 °C appeared to be equally (i.e. the range of TAs of maximal efficiency of primer binding includes 48.9 °C), a TA of 49 °C was chosen for all primer pairs. Validation of the KLK3 and GUSB primer pairs was performed separately by Ryan Morris; 49 °C was determined to be an adequate TA for these primer pairs also.

No amplification was detected for the reactions receiving the no RT control reaction as input, indicating successful degradation of gDNA.

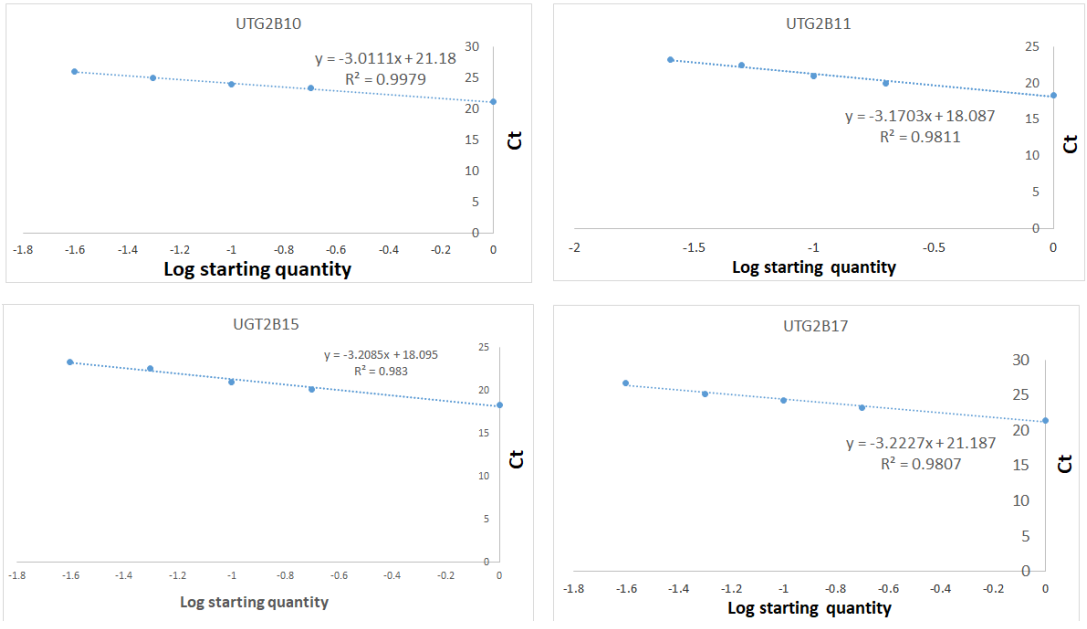
Having determined a TA for qPCR with these primer pairs, a standard curve was performed to determine reaction

efficiency. For this, qPCR was performed with the selected TA and a dilution series of a pooled sample consisting each of the individual biological replicates. Log starting quantity (undiluted sample) was plotted against Ct value for that dilution, and the reaction efficiency determined: 100 x R^2 value of the gradient (**Supplementary Figure 1 B**). Reaction efficiencies for the primer pairs were determined to be 99.79% for *UGT2B10*, 98.11% for *UGT2B11*, 98.3 % for *UGT2B15* and 98.07 % for *UGT2B17*. Reaction efficiencies between 90% and 110% are recommended [212].

A)


UGT2B10
TA (℃): 45.4|45.4|46.0|47.0|48.1|48.9|49.9|50.7|52.3|53.0|54.3|55.5
Input: No RT | 25 ng cDNA
bp
300-
200-
100-
UGT2B11
TA (℃): 45.4|45.4|46.0|47.0|48.1|48.9|49.9|50.7|52.3|53.0|54.3|55.5
Input: No RT | 25 ng cDNA
bp
300-
200-
100-
UGT2B15
TA (℃): 45.4|45.4|46.0|47.0|48.1|48.9|49.9|50.7|52.3|53.0|54.3|55.5
Input: No RT | 25 ng cDNA
bp
300-
200-
100-
UGT2B17
TA (℃): 45.4|45.4|46.0|47.0|48.1|48.9|49.9|50.7|52.3|53.0|54.3|55.5
Input: No RT | 25 ng cDNA
bp
300-
200-
100-

B)


UTG2B10
y = -3.0111x + 21.18
R² = 0.9979
Ct
Log starting quantity
UTG2B11
y = -3.1703x + 18.087
R² = 0.9811
Ct
Log starting quantity
UTG2B15
y = -3.2085x + 18.095
R² = 0.983
Ct
Log starting quantity
UTG2B17
y = -3.2227x + 21.187
R² = 0.9807
Ct
Log starting quantity

Supplementary Figure 1. Validation of qPCR primers for UGT2B genes.

A) Agarose gel electrophoresis resolution of amplicons formed in reactions with gradient PCR to validate minimum temperatures at which there is amplification of single products with the primers targeting *UGT2B10*, *UGT2B11*, *UGT2B15* and *UGT2B17.* Input cDNA for PCRs was pooled LNCaP C4-2 WT, ΔN and KO subline cDNA. 1.8 % agarose; 1 kb plus ladder (Invitrogen).

B) Determination of reaction efficiencies, as determined by performance of qPCR with serial dilution of pooled LNCaP C4-2 WT, ΔN and KO subline cDNA.

Discovery of an alternative splice variant of *TUT1*

As described in **Section 4.10.3**, Star-PAP isoform 1 CDS (encoded by the *TUT1* gene, transcript NM_022830.3) was cloned using the listed in **Supplementary Table 12**. These respectively flank the predicted ATG start codon, and predicted stop codon of NM_022830.3, using LNCaP C4-2 cDNA for template. The expected amplicon for this PCR was 2648 bp.

Interestingly, when the PCR was performed and the products resolved, a second band was observed at approximately 1500 bp (**Supplementary Figure 2 A**). This second band was of greater intensity than the expected band, indicating that the second band was unlikely to have resulted from off-target binding of the primers, and more likely to have instead resulted from amplification of a splice variant of *TUT1*. The NCBI database of splice variants of *TUT1* was consulted [213],

but no transcript, either annotated or predicted, that could explain such an amplicon was listed.

Two splice variants of *TUT1* are listed: NM_022830.3 and NM_001367906.1, both containing nine exons mapping to Chromosome 11, 62,575,052 – 62,591,523. The two variants differ in their use of the seventh exon, which appears to contain a second slice acceptor site: NM_001367906.1 transcripts result from the use of this second splice-acceptor site in *TUT1* Exon 7, and thus lose 59 bp from the CDS. As a consequence, the reading frame of the subsequent exons (Exon 8 and Exon 9) is shifted in NM_001367906 transcripts, leading to use of a different stop codon (also within Exon 9) and a substitution of the C-terminal 451 residues in the encoded protein.

Based upon intron spanning reads listed in NCBI, NM_022830.3 is the major transcript, with 4478 reads joining *TUT1* Exon 6 to the first splice junction of Exon 7, and no intron spanning reads joining Exon 6 to the second splice junction of Exon 7. Likewise, no intron spanning reads joining Exon 6 to the second splice junction of Exon 7 are present in our RNA-Seq data for WT LNCaP C4-2 (**Supplementary Figure 2 D**), but there are a total of 813 reads joining Exon 6 to the first splice junction of Exon 7 ($\bar{x}$ = 271 reads, n =3) in the same data set.

To identify what transcript had resulted in the unexpected ≈1500 bp amplicon, the band was cloned by TOPO cloning and a resulting colony sequenced. The sequencing result demonstrated that this amplicon was indeed amplified from a splice variant of *TUT1*, with a splice event exiting Exon 8 joining to a second splice-acceptor in Exon 9, 1115 bp 3' to the canonical Exon 9 splice-acceptor.

Though there is no corresponding predicted transcript in NCBI, this splice event is reported in the NCBI intron spanning read data, with 351 such reads. Reads corresponding to this splice event also appear in our RNA-Seq data for WT LNCaP C4-2, with a total of 61 reads (17, 29 and 15 reads, $\bar{x}$ = 20.33, n = 3) in contrast to a total of 372 ($\bar{x}$ = 124, n = 3) reads exiting Exon 8 joining to the first splice-acceptor of Exon 9 (**Supplementary Figure 2 D**). An additional type of read exiting Exon 8 is observed, joining to

This splice event skipping 1115 bp of *TUT1* Exon 9 causes a shift in reading frame, so that the stop codon used in NM_022830.3 is not in-frame. Interestingly, there is no stop codon in this reading frame within the subsequent sequence of Exon 9 as the boundaries of Exon 9 are defined in NCBI. Instead, the next possible in-frame stop codon is 422 bp 3' to the novel splice-acceptor in Exon 9. Assuming that *TUT1*

transcripts continue past the defined end of Exon 9, so that this in-frame stop codon can be used, the amino acid sequence encoded by such transcripts was deduced to be:

```
  1   MAAVDSDVES LPRGGFRCCL CHVTTANRPS LDAHLGGRKH
 41   RHLVELRAAR KAQGLRSVFV SGFPRDVDSA QLSEYFLAFG
 81   PVASVVMDKD KGVFAIVEMG DVGAREAVLS QSQHSLGGHR
121   LRVRPREQKE FQSPASKSPK GAAPDSHQLA KALAEAADVG
161   AQMIKLVGLR ELSEAERQLR SLVVALMQEV FTEFFPGCVV
201   HPFGSSINSF DVHGCDLDLF LDLGDLEEPQ PVPKAPESPS
241   LDSALASPLD PQALACTPAS PPDSQPPASP QDSEALDFET
281   PSSSLAPQTP DSALASETLA SPQSLPPASP LLEDREEGDL
321   GKASELAETP KEEKAEGAAM LELVGSILRG CVPGVYRVQT
361   VPSARRPVVK FCHRPSGLHG DVSLSNRLAL HNSRFLSLCS
401   ELDGRVRPLV YTLRCWAQGR GLSGSGPLLS NYALTLLVIY
441   FLQTRDPPVL PTVSQLTQKA GEGEQVEVDG WDCSFPRDAS
481   RLEPSINVEP LIRTLEGSAG AAPRVAPGTR AEFLNGRAVG
521   GSSSTEKSQV CLHILPVASD LCAQGRRFAF SVLHSLL
```

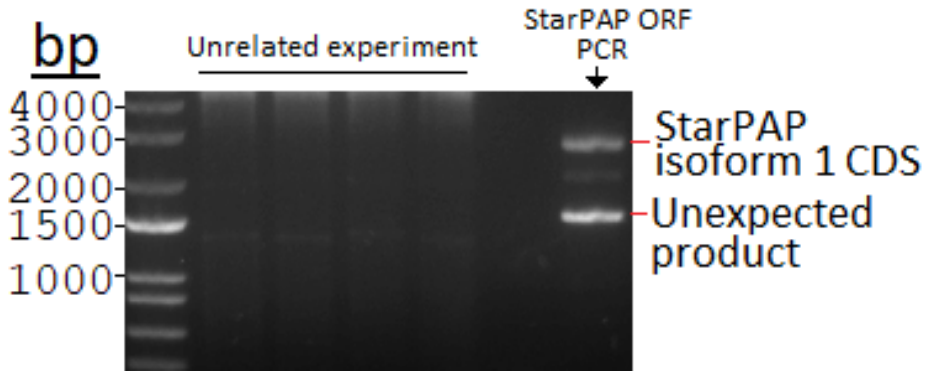
A BLASTn searches was performed with the nucleotide sequence of these putative transcripts, finding three predicted transcripts resulting from the equivalent splice event, for three primate species: XM_031016476.1 (from *G. g. gorilla*); XM_030810963.1 (from *N. leucogenys*) and XM_032153092.1 (from *H. moloch*). Deduced amino acid sequences for these transcripts showed conservation with the deduced amino acid sequence of the predicted alternate human transcript (data not shown).

To confirm existence of transcripts containing this slice event and continuing to the predicted stop codon 3' to the Exon 9 boundary, a new reverse primer was designed (**Supplementary Table 12**) which flanks the predicted stop

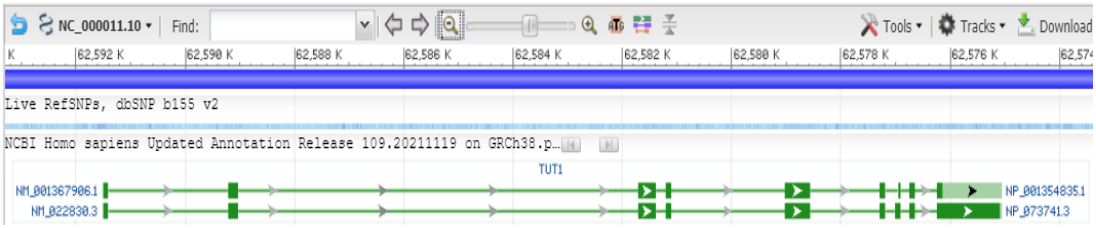
codon of the novel *TUT1* transcript. This was used in conjunction with a forward primer flanking the ATG start codon of Star-PAP isoform 1 CDS. PCRs with these primers were used with cDNA from LNCaP C4-2 or HEK293 cells as input. The expected amplicon for these PCRs was 1915 bp. Upon resolution of the products of these PCRs, a band of the expected size was observed for the reactions with cDNA of each cell type, confirming the existence of this alternate transcript of *TUT1* (**Supplementary Figure 2 E**).

The C-terminal amino acids of Star-PAP isoform 1 encoded by Exon 9 (residues 491-874) constitute nearly half of the protein, including a portion of the fingers domain (which forms a cleft with the palm domain to hold UTP or ATP substrate), and the KA-1 domain which is proposed to facilitate binding to small nuclear RNA [214]. The predicted protein encoded by the newly discovered splice variant of *TUT1* is therefore unlikely to possess poly(A) polymerase or uridyltransferase activity. Whether this alternate isoform of Star-PAP is generated from the novel alternate transcript of *TUT1*, and any distinct functions it may possess were not investigated further.

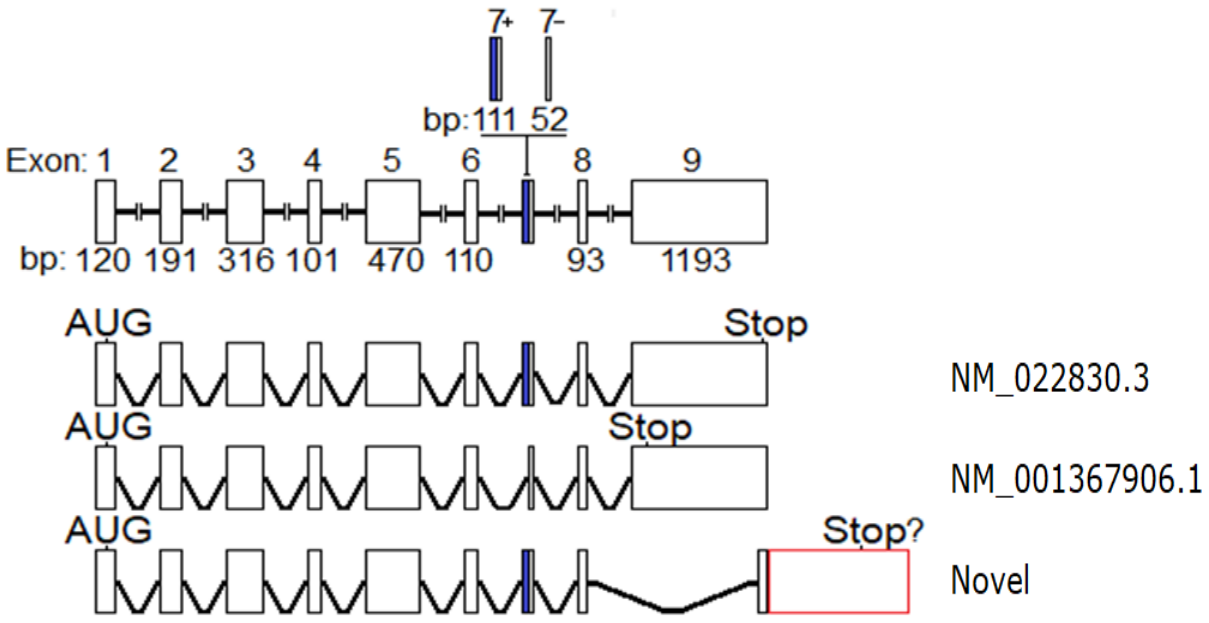
A)


bp
Unrelated experiment
StarPAP ORF PCR
4000
3000
2000
1500
1000
StarPAP isoform 1 CDS
Unexpected product

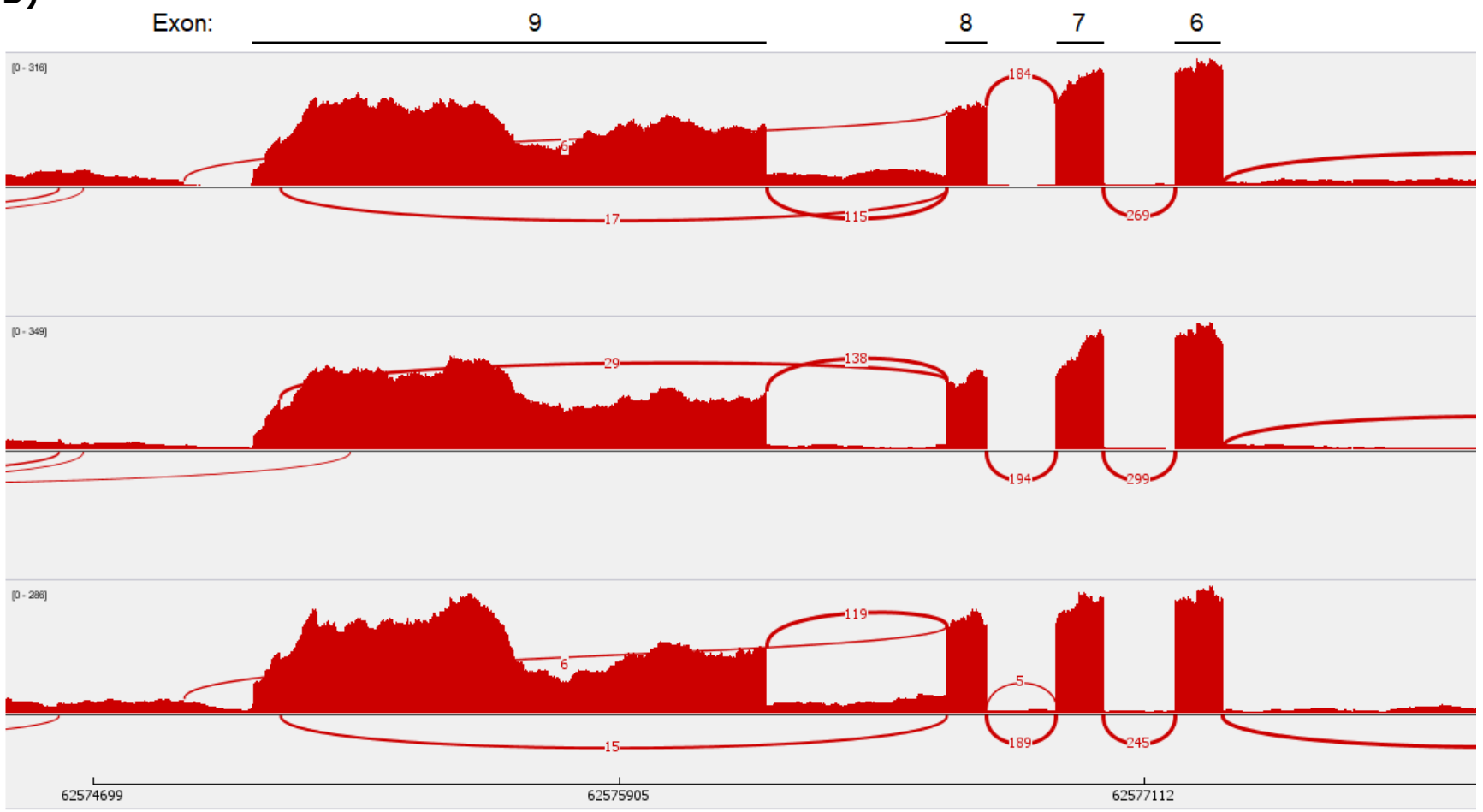
B)


NC_000011.10
Find:
Tools
Tracks
Download
Live RefSNPs, dbSNP b155 v2
NCBI Homo sapiens Updated Annotation Release 109.20211119 on GRCh38.p...
TUT1
NM_001367906.1
NM_022830.3
NP_001354835.1
NP_073741.3

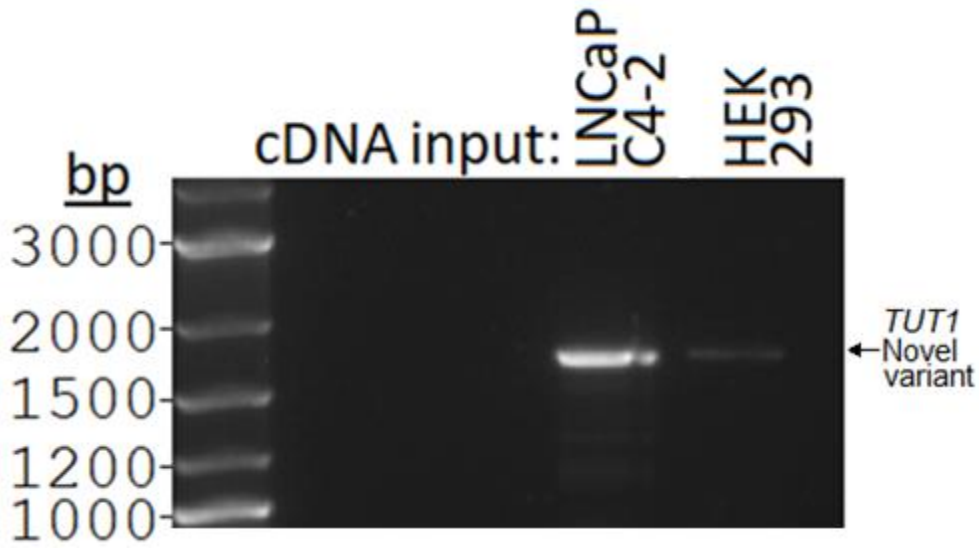
C)


7+
7-
bp: 111 52
Exon: 1 2 3 4 5 6 8 9
bp: 120 191 316 101 470 110 93 1193
AUG
Stop
NM_022830.3
AUG
Stop
NM_001367906.1
AUG
Stop?
Novel

D)


Exon:
9
8
7
6
[0 - 316]
184
6
17
115
269
[0 - 349]
29
138
194
299
[0 - 286]
119
6
5
15
189
245
62574699
62575905
62577112

E)



Supplementary Figure 2. Identification of a novel splice variant of *TUT1*.

A) Agarose gel electrophoresis of PCR amplification of the Star-PAP isoform 1 CDS from *TUT1* transcript variant 1 using LNCaP C4-2 cDNA as input. The expected product was 2648 bp. The expected product was amplified, but another product of approximately 1500 bp was also observed. The high intensity of this unexpected band made it unlikely to have resulted from off-target amplification, prompting investigation of its identity. 0.8 % agarose; 1 kb plus ladder (Invitrogen).

B) NCBI lists two transcripts for *TUT1*, *TUT1*_v1 (NM_022830.3) and *TUT1*_v2 (NM_001354835.1). Neither of these could explain the approximately 1500 bp product observed in **(A)**.

C) Splicing map of the two NCBI listed transcripts of *TUT1*, and a third splice variant discovered in this project. The red box indicates sequence of the novel transcripts that extends beyond the 3' end of Exon 9 as it is defined by NCBI.

D) Sashimi plot of terminal exons of *TUT1*. The majority of reads (115, 138 and 119 reads) exiting Exon 8 join to the canonical splice-acceptor of Exon 9. A minority of reads exiting Exon 8 (17, 29 and 15 reads) instead join to a second splice-acceptor site within Exon 9.

E) Agarose gel electrophoresis of PCR amplification from cDNA from LNCaP C4-2 and HEK293 cells to validate existence of *TUT1* transcripts containing the splice event skipping 1115 bp Exon 9, and continuing past the defined end of Exon 9. 0.8 % agarose; 1 kb plus ladder (NEB).

Tables of oligonucleotide sequences

Supplementary Table 1. Spacer sequences and corresponding DNA oligonucleotide pairs for sgRNAs targeting the canonical first exon of *PIP5K1A*.

Name	Type	Sequence (5'- -3')
sgRNA A	Spacer sequence	GGCCUGGCUCUCUUUCAAUC
sgRNA B	Spacer sequence	GGGGGAGGGGGCUGCUAAGA
A top and A bottom	DNA oligos for cloning into pX461	CACCGGGCCTGGCTCTCTTTCAATC AAACGATTGAAAGAGAGCCAGGCCC
B Top and B bottom	DNA oligos for cloning into pX461	CACCGGGGGGAGGGGGCTGCTAAGA AAACTCTTAGCAGCCCCCTCCCCCC

Supplementary Table 2. Primer pairs used for genotyping of edited clones targeted at *PIP5KIA* Exon 1.

Target	expected amplicon for WT (bp)	Primer pair sequences (5'- -3')
Exon 1 (1dii) Intron 1-2	317	CCTTGAGCTGGTCCAGGAG and TCTTCCTCCCATCGCATATT
Exon 1 (1di) Intron 1-2	578	AGTATCCCCTGTGGAAAGCG and CCTTCAATCTCCATCCGTCCT

Supplementary Table 3. Spacer sequences and corresponding DNA oligonucleotide pairs for sgRNAs targeting Exon 6–intron boundary of *PIP5K1A*.

Name	Type	Sequence (5'- -3')
sgRNA C	Spacer sequence	ATGCTTACCAAGTAATCAT
sgRNA D	Spacer sequence	CACTGTGACTCCAGTATCAG
C top and C bottom	DNA oligonucleotides for cloning into pX461	CACCGATGCTTACCAAGTAATCAT and AAACATGATTACTTGGTAAGCATC
D Top and D bottom	DNA oligonucleotides for cloning into pX461	CACCGCACTGTGACTCCAGTATCAG and AAACCTGATACTGGAGTCACAGTGC

Supplementary Table 4. Primer pairs used for genotyping of edited clones targeted at *PIP5KIA* Exon 6-intron boundary.

Target	Expected amplicon for WT (bp)	Primer pair sequences (5'- -3')
Intron 5-6 Intron 6-7	583	AGGGGTTCAGGTCGGACAAT and CAGAGCGAGACTGTGGAGAAAA
Intron 5-6 Intron 7-8	863	TACAATCAGGGCGATGTGGG and TTTACCCTGCTGAAGACAGCC

Supplementary Table 5. Primer pairs used for generation of pCMV, and for cloning PIP5KIα CDS into pCMV and mVenus N1,

Purpose	Notes	Primer Pair Sequence (5'- -3')
Site directed mutagenesis to delete mVenus tag from mVenus N1 plasmid	-	CCCGGGATCCACCGGTTAAAGCGGCCGC GACTCTAG and ACCGGTGGATCCCGGGCCC
Cloning PIP5KIα i2 and i4 CDS into pCMV and mVenusN1	XhoI and BamHI sites underlined	AAAACTCGAGCACCATGGCGTCGGCCTC CTCC and AAAAGGATCCTTATGGGTGAACTCTGAC TC
Cloning Δ36 PIP5KIα CDS into pCMV and mVenusN1	XhoI and BamHI sites underlined	AAAACTCGAGCACCATGGCATCTGAGGT GCCT and AAAAGGATCCTTATGGGTGAACTCTGAC TC
Cloning Δ46 PIP5KIα CDS into pCMV and mVenusN1	XhoI and BamHI sites underlined	AAAACTCGAGCACCATGCCCATCAAGAA AATAGGC and AAAAGGATCCTTATGGGTGAACTCTGAC TC

Site directed mutagenesis to generate PIP5KIα K181A	GATGAGTTCATTATTGCGACAGTCCAACATAAAG and CTTTATGTTGGACTGTCGCAATAATGAACTCAT

Supplementary Table 6. PCR primers targeting 5' exons of *PIP5K1A*, including novel exons 1b and 1c.

Primers were designed (with the exception of one control reaction) to exclude amplification from genomic DNA because of the long distance between the primer sites in the genome.

Exon targeted	Primer sequence (5'- -3')	Direction of primer
1b	CCACCAGGGTTCATAGCCTAC	Forward
1c	ACTGAACTCCGTCACTGTGC	Forward
1d (1di)	AGTATCCCCTGTGGAAAGCG	Forward
1d (1dii)	AGTATCCCCTGTGGAAAGCG	Reverse
4	AGGAATCAACACTTCTATGGCCT	Reverse

Supplementary Table 7. PCR primers used to amplify transcripts of *PIP5K1A* including those with the putative novel exons 1a-1c and 1e-1h. A common reverse primer targeting the final exon, Exon 16, was also used.

Primers were designed to exclude amplification from genomic DNA because of the long distance between the primer sites in the genome.

Targeted exon	Primer sequence (5'- -3')	Direction
1a	TGACTCAGAGGACACCCCAG	Forward
1b	CCACCAGGGTTCATAGCCTAC	Forward
1c	ACTGAACTCCGTCACTGTGC	Forward
1e	GGGAGCTGGTGAGGAATATGC	Forward
1d	AGTATCCCCTGTGGAAAGCG	Forward
1f	ACTCTTTCCTGATGGCCAGG	Forward
1g	GTGTTGATTTCTGAGGGGTATGGTTC	Forward
1h	GTTTCAACTTTCTGTGCTGTGGG	Forward
16	GAGGCACTAAGGCCCATTCA	Reverse

Supplementary Table 8. Primer sequences used for cloning full length CDS of various PIP5KIα isoforms

Purpose	Primer pair sequences (5'- -3')
Amplification of PIP5KIα i2 and i4 CDS for cloning into mVenus	AAAACTCGAGCACCATGGCGTCGGCCTCCTCC and AAAAGGATCCTTATGGGTGAACTCTGACTC
Site directed mutagenesis, insertion of Exon 3 CDS into PIP5KIα i2-mVenus plasmid, generating PIP5KIα i1-mVenus plasmid	GTCTTGGAAGCTAGACAGGATTCTTACATCT CATTGGTGCCTTATGCCTCTGGCATGCCC and CAATGAGATGTAAGAATCCTGTCTAGCTTCCAAG ACCTCAGATGCCATGGGTCTCTTGATTCCAG
Amplification of PIP5KIα x15 CDS for cloning into mVenus	AAAACTCGAGCACCATGGCGTCGGCCTCCTCC and AAAAGGATCCTTGCATTGCTGGGGCAAGGG

Supplementary Table 9. Primers used for subcloning of PIP5KIα C-terminal variants into the mVenus plasmid.

Residue 427 of PIP5KIα i2, i4 and x15 is an arginine, and thus a methionine codon and CACC partial Kozak sequence were included in the forward primer to facilitate initiation of translation.

Amplicon	Restriction sites	Primer Pairs (5'- -3')
PIP5KIα i2 427-549 and PIP5KIα i4 427-500	XhoI and BamHI	AAAACTCGAGCACCATGCGGTTCCAGCGCTTCATGT and AAAAGGATCCTTATGGGTGAACTCTGACTC
PIP5KIα x15 427-485	XhoI and BamHI	AAAACTCGAGCACCATGCGGTTCCAGCGCTTCATGT and AAAAGGATCCTTGCATTGCTGGGGCAAGGG

Supplementary Table 10. Primers used for site directed mutagenesis to delete different sections containing predicted NESs from PIP5KIα x15 428-485-mVenus.

Purpose	Primer sequences (5'- -3')
Site directed mutagenesis to generate PIP5KIα x15 428-443-mVenus.	GCAACACAGTATTTAAGAAGATTCCCTTACCAAG ATGTCAGCCCTTGCC and TAAGGGAATCTTCTTAAATACTGTGTTGC
Site directed mutagenesis to generate PIP5KIα x15 441-485-mVenus.	ACTCAGATCTCGAGCACCATGCAACCTTGGAAAAG CTTGAAGTTG and CATGGTGCTCGAGATCTGAGTCCGG

Supplementary Table 11. Primers used to for subcloning to generate pmCherryN1 and Star-PAP-mCherry plasmids.

Amplicon	Restriction site	Primer pair sequence (5'- -3')
mCherry CDS	BamHI and NotI	AAAAGGATCCCACCATGGTGAGCAAGGGCGAG and AAAAGCGGCCGCTTACTTGTACAGCTCGTCCATGC
Star-PAP 1 CDS	XhoI and MluI	AAAAAGTCGACCACCATGGCGGCGGTGGATTC and AAAAAACGCGTAACTTGAGATGTCGAATTGCTTGAGGG

Supplementary Table 12. Primer pairs used to amplify CDS of canonical and putative novel isoform of Star-PAP.

Amplicon	Restriction sites	Primer pair sequences (5'- -3')
Star-PAP i1 CDS	XhoI and MluI	AAAAAGTCGACCACCATGGCGGCGGTGGATTC and AAAAAACGCGTAACTTGAGATGTCGAATTGCTTGAGGG
Novel Star-PAP isoform CDS	XhoI	AAAAAGTCGACCACCATGGCGGCGGTGGATTC and TTAAAGGAGAGAATGAAGAACGCTGAAC

Supplementary Table 13. Primers used for molecular cloning of PIP5KIα i4 CDS into pBTM116mod and pASV3mod and sequencing.

Purpose	Primer sequences (5'- -3')
Amplification of PIP5KIα i4 CDS for cloning. XhoI and BamHI sites underlined	AAAACTCGAGATGGCGTCGGCCTCCTCC and AAAAGGATCCTTAATGGGTGAACTCTGACTC
Sequencing 5' end of inserts in pBTM116	CGAAGTTACCGTTAAGCGCC
Sequencing 3' end of inserts in pBTM116	CCTGAGAAAGCAACCTGACCT
Sequencing 5' end of inserts in pASV3	TGAGCAGATGTTTACCGATG
Sequencing 3' end of inserts in pASV3	CTGGCAATTCCTTACCTTCC
Sequence internal section of *PIP5K1A*	ATGGGTGGCATCCCTGCC

Supplementary Table 14. Primer pairs used for site directed deletion of VP16 or NLS of pASV3 PIP5KIα i4.

Construct	Site directed mutagenesis primer sequences (5'- -3')
VP16 PIP5KIα i4 (NLS deleted)	TACAACAAATATAAAACAAGATCCACCATGGCCCCC CCGACCGATGTC and CATGGTGGATCTTGTTTTATATTTGTTGTA
NLS PIP5KIα i4 (VP16 deleted)	CACAGAAAGCACCCAGGTGGTAGATCGACGCTTGGA ATTGACGAGTACGGTGGG and CGTCGATCTACCACCTGGGTGCTTTCTGTG

Supplementary Table 15. Sequencing primers used to obtain the entire sequence pASV3mod

Primer sequence (5'- -3')	Targeted element	Strand
GGAAGGTAAGGAATTGCCAG	PGK1 Terminator	+
TTCTATGTTCGGGTTCAGCGT	PGK1 Terminator	+
AAACTTGGTCTGACAGTTACCA	AmpR Terminator	+
TTAACTGTGGGGAATACTCAGG	Leu2 Promoter	+
GCGATAGTTCCTCACTCTTTCCT	PGK1 Promoter	-
TAACGGATGGGGAGGGATCAA	PGK1 Promoter	-
TGCGAATACCGCTTCCACAA	2mu	-
TCGCCACTATCTTGTCTGCT	Leu2 CDS	+

Supplementary Table 16. Primer sequences used for attempted restriction site cloning of AR NTD, and IVA cloning of AR NTD, PIP5K1B and PIP5K1C into pASV3mod.

Purpose	Primer pair sequences (5'--3')
Amplification of AR NTD (residues 1-538) for cloning into pASV3mod by restriction site cloning. XhoI and BamHI sites underlined. Multiple cloning attempts failed.	AAAACTCGAGATGGAAGTGCAGTTAGGG and AAAAGGATCCTTACGCATGTCCCCGTAAG GTC
Amplification of AR NTD (residues 1-538) for cloning into pASV3mod by IVA. Homology arms to pASV3mod underlined.	GAATTGACGAGTACGGTGGGATGGAAGTG CAGTTAGGGCTGGGA and GCACCACCACCAGTAGAGACCTAACGCATG TCCCCGTAAGGTCCG
Amplification of 8987 bp of pASVmod. Note: Amplification was poor.	CCCACCGTACTCGTCAATTCCAAG and GTCTCTACTGGTGGTGGTGCTTC
Amplification of two fragments of pASV3mod: 4393 bp and 4646 bp. Amplification successful.	GTCTCTACTGGTGGTGGTGCTTC and CTTAACTTCTTCGGCGACAGCAT (4393 bp fragment); and CCCACCGTACTCGTCAATTCCAAG and TGGTTCCAACAGTACCACCG (4646 bp fragment).
PCR screening of colonies for inclusion of pASV3 AR NTD; 195 bp amplicon.	GAAAGCGACTTCACCGCAC and CTGGCAATTCCTTACCTTCC
Amplification of PIP5K1B for cloning into pASV3mod by IVA. 22 base pair of homology to pASV3mod MCS are underlined.	TGGAATTGACGAGTACGGTGGGATGTCTTC TGCTGCTGAAAATGGAGAG and AAGCACCACCACCAGTAGAGACTTAATGCT CTGATGTTGCCATCTTGAAC
Amplification of PIP5K1C for cloning into pASV3mod by IVA. 22 base pair of homology to pASV3mod MCS are underlined. Note, amplification was successful but IVA repeatedly failed.	TGGAATTGACGAGTACGGTGGGATGGAGC TGGAGGTACCGG and AAGCACCACCACCAGTAGAGACTTATGTGT CGCTCTCGCCGT

Supplementary Table 17. Primers used for cloning of various inserts into the pASV3mod and pBTM116mod plasmids. Restriction sites are underlined.

Target	Sites	Primer pair sequences (5'- -3')
PIP5KIα i2 CDS	XhoI BamHI	AAACTCGAGATGGCGTCGGCCTCCTCC and AAAGGATCCTTAATGGGTGAACTCTGAC TC
PIP5K1A Exon 11 and 13 (for colony PCR)		CACTCTTGGAAAGCCCTGGT and CCTTGTGTTCCCCAGAGACC
PIP5KIα x15 CDS	XhoI BamHI	AAAACTCGAGATGGCGTCGGCCTCCTCC and AAAAGGATCCTTAGCATTGCTGGGGCAA GGG
Star-PAP CDS	SalI BglII	AAAAGTCGACATGGCGGCGGTGGATTC and AAAAAGATCTTCACTTGAGATGTCGAATT GCTTGAG
KRAS 2B CDS	XhoI BamHI	AAAACTCGAGATGACTGAATATAAACTT GTGGTAGTTGG and AAAAGGATCCTTACATAATTACACACTTT GTCTTTGACTTC

Supplementary Table 18. Primer pairs targeting the UGT2B genes, KLK3 and GUSB for qPCR

Target gene	Target Exon	Sequence (5'- -3')	Product length (bp)	Predicted TA (°C)
UGT2B10	Exon 1 (CDS) Exon 1 (CDS)	TGACATCGTTTTTGCAGATGCTTA and CAGGTACGTAGGAAGGAGGGAA	152	52
UGT2B11	Exon 6 (3' UTR) Exon 6 (3' UTR)	AGCTCATATTGAAATTTGTTGCAC and CAGAATCCTGTTCAAAATGAAGCC	190	49
UGT2B15	Exon 1 (CDS) Exon 1 (CDS)	GTGTTGGGAATATTATGACTACAGTAAC and CAGCCAGTAGCTCACCACAG	141	50
UGT2B17	Exon 6 (3' UTR) Exon 6 (3' UTR)	GATTTCCTTTTTCCTGCGACA and TTTGTGTGGGGTAACTTTTGGA	200	50
KLK3	Exon 2 Exon 2-3 junction	AGTGCGAGAAGCATTCCCAA and GCAAGATCACGCTTTTGTTCCT	135	53
GUSB	Exon 9 Exon 10	CTCATTTGGAATTTTGCCGATT and CCGAGTGAAGATCCCCTTTTTA	81	49

References

1. Toivanen, R., and Shen, M.M. (2017). Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. Development 144, 1382-1398.
2. Robert, M., Gibbs, B.F., Jacobson, E., and Gagnon, C. (1997). Characterization of prostate-specific antigen proteolytic activity on its major physiological substrate, the sperm motility inhibitor precursor/semenogelin I. Biochemistry 36, 3811-3819.
3. Gupta, N., Sudhakar, D.V., Gangwar, P.K., Sankhwar, S.N., Gupta, N.J., Chakraborty, B., Thangaraj, K., Gupta, G., and Rajender, S. (2017). Mutations in the prostate specific antigen (PSA/KLK3) correlate with male infertility. Scientific reports 7, 1-9.
4. Wilson, J.D. (2011). The critical role of androgens in prostate development. Endocrinology and Metabolism Clinics 40, 577-590.
5. Marker, P.C., Donjacour, A.A., Dahiya, R., and Cunha, G.R. (2003). Hormonal, cellular, and molecular control of prostatic development. Developmental biology 253, 165-174.
6. Yeh, S., Tsai, M.-Y., Xu, Q., Mu, X.-M., Lardy, H., Huang, K.-E., Lin, H., Yeh, S.-D., Altuwaijri, S., and Zhou, X. (2002). Generation and characterization of androgen receptor knockout (ARKO) mice: an in vivo model for the study of androgen functions in selective tissues. Proceedings of the National Academy of Sciences 99, 13498-13503.
7. English, H.F., Kyprianou, N., and Isaacs, J.T. (1989). Relationship between DNA fragmentation and apoptosis in the programmed cell death in the rat prostate following castration. The Prostate 15, 233-250.
8. Sugimura, Y., Cunha, G., and Donjacour, A. (1986). Morphological and histological study of castration-induced degeneration and androgen-induced regeneration in the mouse prostate. Biology of reproduction 34, 973-983.
9. Prins, G.S., Birch, L., and Greene, G.L. (1991). Androgen receptor localization in different cell types of the adult rat prostate. Endocrinology 129, 3187-3199.
10. Zhou, Y., Bolton, E.C., and Jones, J.O. (2015). Androgens and androgen receptor signaling in prostate tumorigenesis. Journal of molecular endocrinology 54, R15-R29.
11. Anon. Cancer Research UK [online]. Available at: *https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer* [Accessed on 1.10.2022].
12. McKenney, J.K. (2018). Mesenchymal tumors of the prostate. Modern Pathology 31, 133-142.
13. Fine, S.W. (2018). Neuroendocrine tumors of the prostate. Modern Pathology 31, 122-132.
14. Choi, N., Zhang, B., Zhang, L., Ittmann, M., and Xin, L. (2012). Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. Cancer cell 21, 253-265.
15. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. cell 144, 646-674.
16. Datta, D., Aftabuddin, M., Gupta, D.K., Raha, S., and Sen, P. (2016). Human prostate cancer hallmarks map. Scientific reports 6, 1-14.

17. Quon, H., and Loblaw, D. (2010). Androgen deprivation therapy for prostate cancer—review of indications in 2010. Current Oncology 17, 38-44.
18. Loblaw, D.A., Pickles, T., Cheung, P.C., Lukka, H., Faria, S., and Klotz, L. (2009). Hormone use after radiotherapy failure: a survey of Canadian uro-oncology specialists. Canadian Urological Association Journal 3, 460.
19. Kirby, M., Hirst, C., and Crawford, E. (2011). Characterising the castration-resistant prostate cancer population: a systematic review. International journal of clinical practice 65, 1180-1192.
20. Saad, F., and Hotte, S.J. (2010). Guidelines for the management of castrate-resistant prostate cancer. Canadian Urological Association Journal 4, 380.
21. Body, A., Pranavan, G., Tan, T.H., and Slobodian, P. (2018). Medical management of metastatic prostate cancer. Australian prescriber 41, 154.
22. Wade, C.A., and Kyprianou, N. (2018). Profiling prostate cancer therapeutic resistance. International Journal of Molecular Sciences 19, 904.
23. Callewaert, L., Van Tilborgh, N., and Claessens, F. (2006). Interplay between two hormone-independent activation domains in the androgen receptor. Cancer research 66, 543-553.
24. Haile, S., and Sadar, M.D. (2011). Androgen receptor and its splice variants in prostate cancer. Cellular and molecular life sciences 68, 3971-3981.
25. Chen, S.-y., Wang, J., Yu, G.-q., Liu, W., and Pearce, D. (1997). Androgen and Glucocorticoid Receptor Heterodimer Formation A POSSIBLE MECHANISM FOR MUTUAL INHIBITION OF TRANSCRIPTIONAL ACTIVITY. Journal of Biological Chemistry 272, 14087-14092.
26. Lee, Y.-F., Shyr, C.-R., Thin, T.H., Lin, W.-J., and Chang, C. (1999). Convergence of two repressors through heterodimer formation of androgen receptor and testicular orphan receptor-4: a unique signaling pathway in the steroid receptor superfamily. Proceedings of the National Academy of Sciences 96, 14724-14729.
27. Verrijdt, G., Haelens, A., and Claessens, F. (2003). Selective DNA recognition by the androgen receptor as a mechanism for hormone-specific regulation of gene expression. Molecular genetics and metabolism 78, 175-185.
28. Bolton, E.C., So, A.Y., Chaivorapol, C., Haqq, C.M., Li, H., and Yamamoto, K.R. (2007). Cell-and gene-specific regulation of primary target genes by the androgen receptor. Genes & development 21, 2005-2017.
29. Tan, M.E., Li, J., Xu, H.E., Melcher, K., and Yong, E.-l. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacologica Sinica 36, 3.
30. Lavery, D.N., and Mcewan, I.J. (2005). Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. Biochemical Journal 391, 449-464.
31. He, B., Kemppainen, J.A., Voegel, J.J., Gronemeyer, H., and Wilson, E.M. (1999). Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH2-terminal domain. Journal of Biological Chemistry 274, 37219-37225.
32. Jenster, G., Trapman, J., and Brinkmann, A. (1993). Nuclear import of the human androgen receptor. Biochemical Journal 293, 761-768.

33. Rosner, W., Hryb, D.J., Khan, M.S., Nakhla, A.M., and Romas, N.A. (1991). Sex hormone-binding globulin: anatomy and physiology of a new regulatory system. The Journal of steroid biochemistry and molecular biology 40, 813-820.
34. "Artoria2e5". Hypothalamic–pituitary–gonadal axis. Wikipedia [online]. Available from: https://en.wikipedia.org/wiki/Hypothalamic%E2%80%93pituitary%E2%80%93gonadal_axis#/media/File:Hypothalamic%E2%80%93pituitary%E2%80%93gonadal_axis.svg [Accessed on 1.10.2022].
35. Anon. Creative Commons Notice. Creative Commons [Online]. Available at: https://creativecommons.org/licenses/by/3.0/legalcode [Accessed on 1.10.2022].
36. Cano, L.Q., Lavery, D.N., and Bevan, C.L. (2013). Mini-review: Foldosome regulation of androgen receptor action in prostate cancer. Molecular and cellular endocrinology 369, 52-62.
37. Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature 387, 733.
38. Estébanez-Perpiñá, E., Moore, J.M., Mar, E., Delgado-Rodrigues, E., Nguyen, P., Baxter, J.D., Buehrer, B.M., Webb, P., Fletterick, R.J., and Guy, R.K. (2005). The molecular mechanisms of coactivator utilization in ligand-dependent transactivation by the androgen receptor. Journal of Biological Chemistry 280, 8060-8068.
39. Hur, E., Pfaff, S.J., Payne, E.S., Grøn, H., Buehrer, B.M., and Fletterick, R.J. (2004). Recognition and accommodation at the androgen receptor coactivator binding interface. PLoS biology 2, e274.
40. Jenster, G., van der Korput, H.A., Trapman, J., and Brinkmann, A.O. (1995). Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. Journal of Biological Chemistry 270, 7341-7346.
41. Simental, J.A., Sar, M., Lane, M.V., French, F.S., and Wilson, E.M. (1991). Transcriptional activation and nuclear targeting signals of the human androgen receptor. Journal of Biological Chemistry 266, 510-518.
42. Cao, B., Qi, Y., Zhang, G., Xu, D., Zhan, Y., Alvarez, X., Guo, Z., Fu, X., Plymate, S.R., and Sartor, O. (2014). Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy. Oncotarget 5, 1646.
43. van de Wijngaart, D.J., Dubbink, H.J., van Royen, M.E., Trapman, J., and Jenster, G. (2012). Androgen receptor coregulators: recruitment via the coactivator binding groove. Molecular and cellular endocrinology 352, 57-69.
44. Bevan, C.L., Hoare, S., Claessens, F., Heery, D.M., and Parker, M.G. (1999). The AF1 and AF2 domains of the androgen receptor interact with distinct regions of SRC1. Molecular and cellular biology 19, 8383-8392.
45. Berrevoets, C.A., Doesburg, P., Steketee, K., Trapman, J., and Brinkmann, A.O. (1998). Functional interactions of the AF-2 activation domain core region of the human androgen receptor with the amino-terminal domain and with the transcriptional coactivator TIF2 (transcriptional intermediary factor 2). Molecular Endocrinology 12, 1172-1183.
46. Heemers, H.V., and Tindall, D.J. (2007). Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. Endocrine reviews 28, 778-808.

47. Marshall, T.W., Link, K.A., Petre-Draviam, C.E., and Knudsen, K.E. (2003). Differential requirement of SWI/SNF for androgen receptor activity. Journal of Biological Chemistry 278, 30605-30613.
48. Wang, Q., Sharma, D., Ren, Y., and Fondell, J.D. (2002). A coregulatory role for the TRAP-mediator complex in androgen receptor-mediated gene expression. Journal of Biological Chemistry 277, 42852-42858.
49. McEwan, I.J., and Gustafsson, J.-Å. (1997). Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. Proceedings of the National Academy of Sciences 94, 8485-8490.
50. Lee, D.K., Li, M., and Chang, C. (2003). The second largest subunit of RNA polymerase II interacts with and enhances transactivation of androgen receptor. Biochemical and biophysical research communications 302, 162-169.
51. Burd, C.J., Morey, L.M., and Knudsen, K.E. (2006). Androgen receptor corepressors and prostate cancer. Endocrine-Related Cancer 13, 979-994.
52. Leung, J.K., and Sadar, M.D. (2017). Non-genomic actions of the androgen receptor in prostate cancer. Frontiers in endocrinology 8, 2.
53. Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., and Kuefer, R. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. science 310, 644-648.
54. Massie, C.E., Lynch, A., Ramos-Montoya, A., Boren, J., Stark, R., Fazli, L., Warren, A., Scott, H., Madhu, B., and Sharma, N. (2011). The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. The EMBO journal 30, 2719-2733.
55. Blunsom, N.J., and Cockcroft, S. (2020). Phosphatidylinositol synthesis at the endoplasmic reticulum. Biochimica Et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids 1865, 158471.
56. Balla, T. (2013). Phosphoinositides: tiny lipids with giant impact on cell regulation. Physiological reviews 93, 1019-1137.
57. Burke, J.E. (2018). Structural basis for regulation of phosphoinositide kinases and their involvement in human disease. Molecular cell 71, 653-673.
58. Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature 443, 651-657.
59. Fiume, R., Faenza, I., Sheth, B., Poli, A., Vidalle, M., Mazzetti, C., Abdul, S., Campagnoli, F., Fabbrini, M., and Kimber, S. (2019). Nuclear phosphoinositides: their regulation and roles in nuclear functions. International journal of molecular sciences 20, 2991.
60. Choy, C.H., Han, B.K., and Botelho, R.J. (2017). Phosphoinositide diversity, distribution, and effector function: stepping out of the box. BioEssays 39, 1700121.
61. Lemmon, M.A. (2007). Pleckstrin homology (PH) domains and phosphoinositides. In Biochemical Society Symposia. (London; Portland on behalf of The Biochemical Society; 1999), p 81.
62. Brown, J.R., and Auger, K.R. (2011). Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. BMC evolutionary biology 11, 1-14.
63. Liu, G.Y., and Sabatini, D.M. (2020). mTOR at the nexus of nutrition, growth, ageing and disease. Nature reviews Molecular cell biology 21, 183-203.

64. Boura, E., and Nencka, R. (2015). Phosphatidylinositol 4-kinases: Function, structure, and inhibition. Experimental cell research 337, 136-145.
65. González, B., Schell, M.J., Letcher, A.J., Veprintsev, D.B., Irvine, R.F., and Williams, R.L. (2004). Structure of a human inositol 1, 4, 5-trisphosphate 3-kinase: substrate binding reveals why it is not a phosphoinositide 3-kinase. Molecular cell 15, 689-701.
66. Hu, J., Yuan, Q., Kang, X., Qin, Y., Li, L., Ha, Y., and Wu, D. (2015). Resolution of structure of PIP5K1A reveals molecular mechanism for its regulation by dimerization and dishevelled. Nature communications 6, 8205.
67. Zeng, X., Uyar, A., Sui, D., Donyapour, N., Wu, D., Dickson, A., and Hu, J. (2018). Structural insights into lethal contractural syndrome type 3 (LCCS3) caused by a missense mutation of PIP5Kγ. Biochemical Journal 475, 2257-2269.
68. Heck, J.N., Mellman, D.L., Ling, K., Sun, Y., Wagoner, M.P., Schill, N.J., and Anderson, R.A. (2007). A conspicuous connection: structure defines function for the phosphatidylinositol-phosphate kinase family. Critical reviews in biochemistry and molecular biology 42, 15-39.
69. Schill, N.J., and Anderson, R.A. (2009). Two novel phosphatidylinositol-4-phosphate 5-kinase type Iγ splice variants expressed in human cells display distinctive cellular targeting. Biochemical Journal 422, 473-482.
70. Schill, N.J., Hedman, A.C., Choi, S., and Anderson, R.A. (2014). Isoform 5 of PIPKIγ regulates the endosomal trafficking and degradation of E-cadherin. Journal of cell science 127, 2189-2203.
71. Sbrissa, D., Ikonomov, O.C., and Shisheva, A. (2002). Phosphatidylinositol 3-phosphate-interacting domains in PIKfyve: binding specificity and role in PIKfyve endomembrane localization. Journal of Biological Chemistry 277, 6073-6079.
72. Shisheva, A. (2008). PIKfyve: Partners, significance, debates and paradoxes. Cell biology international 32, 591-604.
73. Muftuoglu, Y., Xue, Y., Gao, X., Wu, D., and Ha, Y. (2016). Mechanism of substrate specificity of phosphatidylinositol phosphate kinases. Proceedings of the National Academy of Sciences 113, 8711-8716.
74. Ishihara, H., Shibasaki, Y., Kizuki, N., Wada, T., Yazaki, Y., Asano, T., and Oka, Y. (1998). Type I phosphatidylinositol-4-phosphate 5-kinases: cloning of the third isoform and deletion/substitution analysis of members of this novel lipid kinase family. Journal of Biological Chemistry 273, 8741-8748.
75. Coppolino, M.G., Dierckman, R., Loijens, J., Collins, R.F., Pouladi, M., Jongstra-Bilen, J., Schreiber, A.D., Trimble, W.S., Anderson, R., and Grinstein, S. (2002). Inhibition of phosphatidylinositol-4-phosphate 5-kinase Iα impairs localized actin remodeling and suppresses phagocytosis. Journal of Biological Chemistry 277, 43849-43857.
76. Narkis, G., Ofir, R., Landau, D., Manor, E., Volokita, M., Hershkowitz, R., Elbedour, K., and Birk, O.S. (2007). Lethal contractural syndrome type 3 (LCCS3) is caused by a mutation in PIP5K1C, which encodes PIPKIγ of the phophatidylinsitol pathway. The American Journal of Human Genetics 81, 530-539.
77. Schramp, M., Hedman, A., Li, W., Tan, X., and Anderson, R. (2012). PIP kinases from the cell membrane to the nucleus. Phosphoinositides I: Enzymes of Synthesis and Degradation, 25-59.
78. Rao, V.D., Misra, S., Boronenkov, I.V., Anderson, R.A., and Hurley, J.H. (1998). Structure of type IIβ phosphatidylinositol phosphate

kinase: a protein kinase fold flattened for interfacial phosphorylation. Cell 94, 829-839.
79. Amos, S.-B.T., Kalli, A.C., Shi, J., and Sansom, M.S. (2019). Membrane recognition and binding by the phosphatidylinositol phosphate kinase PIP5K1A: a multiscale simulation study. Structure 27, 1336-1346. e1332.
80. CIRUELA, A., HINCHLIFFE, K.A., Divecha, N., and IRVINE, R.F. (2000). Nuclear targeting of the β isoform of Type II phosphatidylinositol phosphate kinase (phosphatidylinositol 5-phosphate 4-kinase) by its α-helix 7. Biochemical Journal 346, 587-591.
81. Hinchliffe, K.A., Ciruela, A., Letcher, A.J., Divecha, N., and Irvine, R.F. (1999). Regulation of type IIα phosphatidylinositol phosphate kinase localisation by the protein kinase CK2. Current biology 9, 983-S981.
82. Anon. Cell Signalling Technology [online]; Available at: https://www.phosphosite.org/staticAboutPhosphosite [Accessed on 1.10.22].
83. Anderson, R.A., Boronenkov, I.V., Doughman, S.D., Kunz, J., and Loijens, J.C. (1999). Phosphatidylinositol phosphate kinases, a multifaceted family of signaling enzymes. Journal of Biological Chemistry 274, 9907-9910.
84. Kunz, J., Wilson, M.P., Kisseleva, M., Hurley, J.H., Majerus, P.W., and Anderson, R.A. (2000). The activation loop of phosphatidylinositol phosphate kinases determines signaling specificity. Molecular cell 5, 1-11.
85. Liu, A., Sui, D., Wu, D., and Hu, J. (2016). The activation loop of PIP5K functions as a membrane sensor essential for lipid substrate processing. Science advances 2, e1600925.
86. Anon. Cell Signalling Technology; [online]. Available at https://www.phosphosite.org/proteinAction.action?id=11798&showAllSites=true [Accessed on 1.10.2022].
87. Samuels, Y., and Waldman, T. (2010). Oncogenic mutations of PIK3CA in human cancers. In Phosphoinositide 3-kinase in Health and Disease. (Springer), pp 21-41.
88. Shayesteh, L., Lu, Y., Kuo, W.-L., Baldocchi, R., Godfrey, T., Collins, C., Pinkel, D., Powell, B., Mills, G.B., and Gray, J.W. (1999). PIK3CA is implicated as an oncogene in ovarian cancer. Nature genetics 21, 99.
89. Philp, A.J., Campbell, I.G., Leet, C., Vincan, E., Rockman, S.P., Whitehead, R.H., Thomas, R.J., and Phillips, W.A. (2001). The phosphatidylinositol 3′-kinase p85α gene is an oncogene in human ovarian and colon tumors. Cancer research 61, 7426-7429.
90. Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., and McCombie, R. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. science 275, 1943-1947.
91. Jamaspishvili, T., Berman, D.M., Ross, A.E., Scher, H.I., De Marzo, A.M., Squire, J.A., and Lotan, T.L. (2018). Clinical implications of PTEN loss in prostate cancer. Nature Reviews Urology.
92. Lien, E.C., Dibble, C.C., and Toker, A. (2017). PI3K signaling in cancer: beyond AKT. Current opinion in cell biology 45, 62-71.
93. Crumbaker, M., Khoja, L., and Joshua, A.M. (2017). AR signaling and the PI3K pathway in prostate cancer. Cancers 9, 34.
94. Semenas, J., Hedblom, A., Miftakhova, R.R., Sarwar, M., Larsson, R., Shcherbina, L., Johansson, M.E., Härkönen, P., Sterner, O., and Persson, J.L. (2014). The role of PI3K/AKT-related PIP5K1α and the discovery of its selective inhibitor for treatment of advanced

prostate cancer. Proceedings of the National Academy of Sciences 111, E3689-E3698.
95. Sarwar, M., Semenas, J., Miftakhova, R., Simoulis, A., Robinson, B., Wingren, A.G., Mongan, N.P., Heery, D.M., Johnsson, H., and Abrahamsson, P.-A. (2016). Targeted suppression of AR-V7 using PIP5K1α inhibitor overcomes enzalutamide resistance in prostate cancer cells. Oncotarget 7, 63065.
96. Strätker, K., Haidar, S., Amesty, Á., El-Awaad, E., Götz, C., Estévez-Braun, A., and Jose, J. (2020). Development of an in vitro screening assay for PIP5K1α lipid kinase and identification of potent inhibitors. The FEBS Journal 287, 3042-3064.
97. Anon. Thermo Scientific [online]; Available at: https://dna.uga.edu/wp-content/uploads/sites/51/2019/02/Note-on-the-260_280-and-260_230-Ratios.pdf [Accessed on 1.10.2022].
98. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual.(Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
99. Lund, A.H., Duch, M., and Skou Pedersen, F. (1996). Increased cloning efficiency by temperature-cycle ligation. Nucleic acids research 24, 800-801.
100. Bartel, P.L., and Fields, S. (1995). Analyzing Protein-Protein Interactions Using Two-Hybrid System. Methods Enzymol 254, 241-263.
101. Le Douarin, B., Heery, D.M., Gaudon, C., vom Baur, E., and Losson, R. (2001). Yeast Two-Hybrid Screening for Proteins that Interact with Nuclear Hormone Receptors. In Steroid Receptor Methods: Protocols and Assays, B.A. Lieberman, ed. (Totowa, NJ, Humana Press), pp 227-248.
102. Möckli, N., and Auerbach, D. (2004). Quantitative β-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. Biotechniques 36, 872-876.
103. Anon. Thermo Fischer Scientific [online]; Available at: https://www.thermofisher.com/uk/en/home/references/protocols/cell-culture/transfection-protocol/lncap-cells-protocol.html [Accessed on 01/02/2022].
104. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 29, 15-21.
105. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930.
106. Ge, S.X., Son, E.W., and Yao, R. (2018). iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC bioinformatics 19, 1-24.
107. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. bioinformatics 26, 139-140.
108. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome biology 15, 1-21.
109. Goedhart, J., and Luijsterburg, M.S. (2020). VolcaNoseR is a web app for creating, exploring, labeling and sharing volcano plots. Scientific reports 10, 1-5.
110. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.

Proceedings of the National Academy of Sciences 102, 15545-15550.
111. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. Cell systems 1, 417-425.
112. David, M., and Leslie, S. (2021). Prostate Specific Antigen. StatPearls.
113. Turgeon, D., Carrier, J.-S.b., Lévesque, E.r., Hum, D.W., and Bélanger, A. (2001). Relative enzymatic activity, protein stability, and tissue distribution of human steroid-metabolizing UGT2B subfamily members. Endocrinology 142, 778-787.
114. Meech, R., Hu, D.G., McKinnon, R.A., Mubarokah, S.N., Haines, A.Z., Nair, P.C., Rowland, A., and Mackenzie, P.I. (2019). The UDP-glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. Physiological reviews 99, 1153-1222.
115. Horoszewicz, J.S., Leong, S.S., Kawinski, E., Karr, J.P., Rosenthal, H., Chu, T.M., Mirand, E.A., and Murphy, G.P. (1983). LNCaP Model of Human Prostatic Carcinoma. Cancer Research 43, 1809-1818.
116. Wu, H.-C., Hsieh, J.-T., Gleave, M.E., Brown, N.M., Pathak, S., and Chung, L.W.K. (1994). Derivation of androgen-independent human LNCaP prostatic cancer cell sublines: Role of bone stromal cells. International Journal of Cancer 57, 406-412.
117. Pfitzenmaier, J., Quinn, J.E., Odman, A.M., Zhang, J., Keller, E.T., Vessella, R.L., and Corey, E. (2003). Characterization of C4–2 prostate cancer bone metastases and their response to castration. Journal of Bone and Mineral Research 18, 1882-1888.
118. Hernandez-Quiles, M. (2016). TARGETING ANDROGEN RECEPTOR AND PIP5K1A IN PROSTATE CANCER CELLS USING CRISPR-CAS9 GENOME EDITING TOOL. University of Nottingham.
119. Gearing M.; addgene Blog [online]. Available at: https://blog.addgene.org/crispr-101-cas9-nickase-design-and-homology-directed-repair; [Accessed on 1.10.2022].
120. Riet, J., Ramos, L., Lewis, R.V., and Marins, L. (2017). Improving the PCR protocol to amplify a repetitive DNA sequence. Genetics and Molecular Research 16, 1.
121. Anon. Compute pI/Mw. Expasy [online]. Available at: https://web.expasy.org/compute_pi/ [Accessed 1.10.22].
122. Chakrabarti, R., Sanyal, S., Ghosh, A., Bhar, K., Das, C., and Siddhanta, A. (2015). Phosphatidylinositol-4-phosphate 5-kinase 1α modulates ribosomal RNA gene silencing through its interaction with histone H3 lysine 9 trimethylation and heterochromatin protein HP1-α. Journal of Biological Chemistry 290, 20893-20903.
123. Fruman, D.A., Meyers, R.E., and Cantley, L.C. (1998). Phosphoinositide kinases. Annual review of biochemistry 67, 481-507.
124. Hilgarth, R.S., and Sarge, K.D. (2005). Detection of sumoylated proteins. In Ubiquitin-Proteasome Protocols. (Springer), pp 329-337.
125. Lee, C.-R., Park, Y.-H., Min, H., Kim, Y.-R., and Seok, Y.-J. (2019). Determination of protein phosphorylation by polyacrylamide gel electrophoresis. Journal of microbiology 57, 93-100.
126. Anon. Cell Signalling Technology [online]. Available at: https://www.cellsignal.co.uk/products/primary-antibodies/pip5k1a-antibody/9693; [Accessed on 1.10.2022].
127. Anon. addgene [online]. Available at:https://www.addgene.org/27793/ [Accessed on 1.10.2022].

128. Anon. addgene [online]. Available at: https://www.addgene.org/121050/ [Accessed on 1.10.22].
129. Babushok, D.V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian, H.H. (2007). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome Research 17, 1129-1138.
130. Hart, J.R., and Vogt, P.K. (2011). Phosphorylation of AKT: a mutational analysis. Oncotarget 2, 467.
131. Xu, N., Lao, Y., Zhang, Y., and Gillespie, D.A. (2012). Akt: a double-edged sword in cell proliferation and genome stability. Journal of oncology 2012.
132. Wang, T., Whitchurch, J.B., Collins, H.M., Green, T., Semenas, J., Roberts, C.J., Morris, R.D., Hubert, M., Chen, S., and El-Schich, Z. PIP5K1α is required for promoting tumor progression in castrationresistant prostate cancer. Frontiers in Cell and Developmental Biology, 439.
133. Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nature biotechnology 20, 87.
134. Adhikari, H., and Counter, C.M. (2018). Interrogating the protein interactomes of RAS isoforms identifies PIP5K1A as a KRAS-specific vulnerability. Nature communications 9, 1-17.
135. Choi, S., Chen, M., Cryns, V.L., and Anderson, R.A. (2019). A nuclear phosphoinositide kinase complex regulates p53. Nature cell biology 21, 462-475.
136. Kohli, M., Ho, Y., Hillman, D.W., Van Etten, J.L., Henzler, C., Yang, R., Sperger, J.M., Li, Y., Tseng, E., and Hon, T. (2017). Androgen receptor variant AR-V9 is coexpressed with AR-V7 in prostate cancer metastases and predicts abiraterone resistance. Clinical Cancer Research 23, 4704-4715.
137. Lawrence, M.G., Stephens, C.R., Need, E.F., Lai, J., Buchanan, G., and Clements, J.A. (2012). Long terminal repeats act as androgen-responsive enhancers for the PSA-kallikrein locus. Endocrinology 153, 3199-3210.
138. Lai, J., An, J., Nelson, C.C., Lehman, M.L., Batra, J., and Clements, J.A. (2014). Analysis of androgen and anti-androgen regulation of KLK-related peptidase 2, 3, and 4 alternative transcripts in prostate cancer. Biological chemistry 395, 1127-1132.
139. Lévesque, É., Turgeon, D., Carrier, J.-S., Montminy, V., Beaulieu, M., and Bélanger, A. (2001). Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. Biochemistry 40, 3869-3881.
140. Beaulieu, M., Lávesque, E., Hum, D.W., and Bálanger, A. (1996). Isolation and characterization of a novel cDNA encoding a human UDP-glucuronosyltransferase active on C19 steroids. Journal of Biological Chemistry 271, 22855-22862.
141. Levesque, E., Beaulieu, M., GreeN, M.D., Tephly, T.R., Belanger, A., and Hum, D.W. (1997). Isolation and characterization of UGT2B15 (Y85): a UDP-glucuronosyltransferase encoded by a polymorphic gene. Pharmacogenetics 7, 317-325.
142. Lévesque, E., Labriet, A., Hovington, H., Allain, É.P., Melo-Garcia, L., Rouleau, M., Brisson, H., Turcotte, V., Caron, P., and Villeneuve, L. (2020). Alternative promoters control UGT2B17-dependent androgen catabolism in prostate cancer and its influence on progression. British journal of cancer 122, 1068-1076.

143. Li, H., Xie, N., Chen, R., Verreault, M., Fazli, L., Gleave, M.E., Barbier, O., and Dong, X. (2016). UGT2B17 expedites progression of castration-resistant prostate cancers by promoting ligand-independent AR signaling. Cancer research 76, 6701-6711.
144. Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics 40, 1413-1415.
145. Loijens, J.C., and Anderson, R.A. (1996). Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family. Journal of Biological Chemistry 271, 32937-32943.
146. Van den Bout, I., and Divecha, N. (2009). PIP5K-driven PtdIns (4, 5) P2 synthesis: regulation and cellular functions. Journal of cell science 122, 3837-3850.
147. NCBI Homo sapiens Annotation release 109.20201120, Gene ID: 8394.
148. Wang, L., Li, G., and Sugita, S. (2005). A central kinase domain of type I phosphatidylinositol phosphate kinases is sufficient to prime exocytosis: isoform specificity and its underlying mechanism. Journal of Biological Chemistry 280, 16522-16527.
149. Mellman, D.L., Gonzales, M.L., Song, C., Barlow, C.A., Wang, P., Kendziorski, C., and Anderson, R.A. (2008). A PtdIns4, 5P 2-regulated nuclear poly (A) polymerase controls expression of select mRNAs. Nature 451, 1013-1017.
150. Xie, Y., Zhu, L., and Zhao, G. (2000). Assignment1 of type I phosphatidylinositol-4-phosphate 5-kinase (PIP5K1A) to human chromosome bands 1q22→ q24 by in situ hybridization. Cytogenetic and Genome Research 88, 197-199.
151. Anon. NetPhos 3.1. DTU Health Tech [online]. Available at: https://services.healthtech.dtu.dk/service.php?NetPhos-3.1 [Accessed on 1.10.2022].
152. Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R., and Platzer, M. (2004). Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nature genetics 36, 1255-1257.
153. Anon. PIP5K1A Gene ID 8394. NCBI [online]. Available at: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=8394 [Accessed on 1.10.2022].
154. Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283-292.
155. Hernández, G., Osnaya, V.G., and Pérez-Martínez, X. (2019). Conservation and variability of the AUG initiation codon context in eukaryotes. Trends in biochemical sciences 44, 1009-1021.
156. Choi, S., Thapa, N., Tan, X., Hedman, A.C., and Anderson, R.A. (2015). PIP kinases define PI4, 5P2 signaling specificity by association with effectors. Biochimica Et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids 1851, 711-723.
157. Zhou, Y., Shearwin-Whyatt, L., Li, J., Song, Z., Hayakawa, T., Stevens, D., Fenelon, J.C., Peel, E., Cheng, Y., and Pajpach, F. (2021). Platypus and echidna genomes reveal mammalian biology and evolution. Nature 592, 756-762.
158. Hallström, B.M., Kullberg, M., Nilsson, M.A., and Janke, A. (2007). Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Molecular biology and evolution 24, 2059-2068.

159. Disotell, T.R., and Tosi, A.J. (2007). The monkey's perspective. Genome biology 8, 1-4.
160. Kovacs, E., Tompa, P., Liliom, K., and Kalmar, L. (2010). Dual coding in alternative reading frames correlates with intrinsic protein disorder. Proceedings of the National Academy of Sciences 107, 5429-5434.
161. Nord, A., Hornbeck, P., Carey, K., and Wheeler, T. (2018). Splice-Aware multiple sequence alignment of protein isoforms. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. pp 200-210.
162. Liang, H., and Landweber, L.F. (2006). A genome-wide study of dual coding regions in human alternatively spliced genes. Genome research 16, 190-196.
163. Hill J.T. PolyPeakParser [online]. Available at: http://yosttools.genetics.utah.edu/PolyPeakParser/ [Accessed on 1.10.2022].
164. Kremers, G.-J., Goedhart, J., van Munster, E.B., and Gadella, T.W. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. Biochemistry 45, 6570-6580.
165. Giudici, M.-L., Lee, K., Lim, R., and Irvine, R.F. (2006). The intracellular localisation and mobility of Type Iγ phosphatidylinositol 4P 5-kinase splice variants. FEBS letters 580, 6933-6937.
166. Kosugi S. cNLS Mapper [online]. Available at: https://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi [Accessed on 1.10.2022].
167. Fairn, G.D., Ogata, K., Botelho, R.J., Stahl, P.D., Anderson, R.A., De Camilli, P., Meyer, T., Wodak, S., and Grinstein, S. (2009). An electrostatic switch displaces phosphatidylinositol phosphate kinases from the membrane during phagocytosis. Journal of Cell Biology 187, 701-714.
168. Klumperman, J., and Raposo, G. (2014). The complex ultrastructure of the endolysosomal system. Cold Spring Harbor perspectives in biology 6, a016857.
169. Xu D. LocNES. UT Southwestern Medical Centre [online]. Availavle at: http://prodata.swmed.edu/LocNES/LocNES.php [Accessed on 1.10.2022].
170. Anon. NetNes1.1. DTU Health Tech [online]. Available at: https://services.healthtech.dtu.dk/service.php?NetNES-1.1 [Accessed on 1.10.21].
171. Tynell, J., Melén, K., and Julkunen, I. (2014). Mutations within the conserved NS1 nuclear export signal lead to inhibition of influenza A virus replication. Virology journal 11, 1-12.
172. Sorokin, A., Kim, E., and Ovchinnikov, L. (2007). Nucleocytoplasmic transport of proteins. Biochemistry (Moscow) 72, 1439-1457.
173. Zhao, L., Xia, H., Huang, J., Zheng, Y., Liu, C., Su, J., and Ping, J. (2020). Features of nuclear export signals of NS2 protein of influenza D virus. Viruses 12, 1100.
174. Anon. addgene [online]. Available at: http://n2t.net/addgene:67301 [Accessed on 1.10.2022].
175. Li, W., Laishram, R.S., Ji, Z., Barlow, C.A., Tian, B., and Anderson, R.A. (2012). Star-PAP control of BIK expression and apoptosis is regulated by nuclear PIPKIα and PKCδ signaling. Molecular cell 45, 25-37.

176. Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. Nature reviews Molecular cell biology 18, 437-451.
177. Watson, P.M., and Watson, D.K. (2010). Alternative splicing in prostate and breast cancer. The Open Cancer Journal 3.
178. Divecha, N., Roefs, M., Los, A., Halstead, J., Bannister, A., and D'Santos, C. (2002). Type I PIPkinases Interact with and Are Regulated by the Retinoblastoma Susceptibility Gene Product—pRB. Current biology 12, 582-587.
179. Alexandrova, E.M., Yallowitz, A., Li, D., Xu, S., Schulz, R., Proia, D., Lozano, G., Dobbelstein, M., and Moll, U. (2015). Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment. Nature 523, 352-356.
180. Muller, P.A., and Vousden, K.H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer cell 25, 304-317.
181. Semenas, J., Wang, T., Sajid Syed Khaja, A., Firoj Mahmud, A., Simoulis, A., Grundström, T., Fällman, M., and Persson, J.L. (2021). Targeted inhibition of ERα signaling and PIP5K1α/Akt pathways in castration-resistant prostate cancer. Molecular Oncology 15, 968-986.
182. Blind, R.D., Suzawa, M., and Ingraham, H.A. (2012). Direct modification and activation of a nuclear receptor–PIP2 complex by the inositol lipid kinase IPMK. Science signaling 5, ra44-ra44.
183. Jorgensen, J.S., and Nilson, J.H. (2001). AR suppresses transcription of the LHβ subunit by interacting with steroidogenic factor-1. Molecular Endocrinology 15, 1505-1516.
184. Fulton, J. (2012). Selective interactions of nuclear receptors and cofactors: novel targets for drug discovery. University of Nottingham.
185. Chen, W.-Y., Lee, W.-C., Hsu, N.-C., Huang, F., and Chung, B.-c. (2004). SUMO modification of repression domains modulates function of nuclear receptor 5A1 (steroidogenic factor-1). Journal of Biological Chemistry 279, 38730-38735.
186. Karvonen, U., Jänne, O.A., and Palvimo, J.J. (2006). Androgen receptor regulates nuclear trafficking and nuclear domain residency of corepressor HDAC7 in a ligand-dependent fashion. Experimental cell research 312, 3165-3183.
187. He, B., Kemppainen, J.A., and Wilson, E.M. (2000). FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. Journal of Biological Chemistry 275, 22986-22994.
188. Dubbink, H.J., Hersmus, R., Verma, C.S., van der Korput, H.A., Berrevoets, C.A., van Tol, J., Ziel-van der Made, A.C., Brinkmann, A.O., Pike, A.C., and Trapman, J. (2004). Distinct recognition modes of FXXLF and LXXLL motifs by the androgen receptor. Molecular endocrinology 18, 2132-2150.
189. Aksoy, O., Pencik, J., Hartenbach, M., Moazzami, A.A., Schlederer, M., Balber, T., Varady, A., Philippe, C., Baltzer, P.A., and Mazumder, B. (2021). Thyroid and androgen receptor signaling are antagonized by μ-Crystallin in prostate cancer. International journal of cancer 148, 731-747.
190. Gilbert, D.M., Heery, D.M., Losson, R., Chambon, P., and Lemoine, Y. (1993). Estradiol-inducible squelching and cell growth arrest by a chimeric VP16-estrogen receptor expressed in Saccharomyces cerevisiae: suppression by an allele of PDR1. Molecular and Cellular Biology 13, 462-472.

191. Sheppard, H., Matsuda, S., Harries, J., Kindle, K., and Heery, D. (2003). Transcriptional activation by estrogen receptor (ERalpha) and steroid receptor coactivator (SRC1) involves distinct mechanisms in yeast and mammalian cells. Journal of molecular endocrinology 30, 411-422.
192. Fernández-Acero, T., Rodríguez-Escudero, I., Vicente, F., Monteiro, M.C., Tormo, J.R., Cantizani, J., Molina, M., and Cid, V.J. (2012). A yeast-based in vivo bioassay to screen for class I phosphatidylinositol 3-kinase specific inhibitors. Journal of Biomolecular Screening 17, 1018-1029.
193. Rodríguez-Escudero, I., Roelants, F.M., Thorner, J., Nombela, C., Molina, M., and Cid, V.J. (2005). Reconstitution of the mammalian PI3K/PTEN/Akt pathway in yeast. Biochemical Journal 390, 613-623.
194. Lacalle, R.A., de Karam, J.C., Martínez-Muñoz, L., Artetxe, I., Peregil, R.M., Sot, J., Rojas, A.M., Goñi, F.M., Mellado, M., and Mañes, S. (2015). Type I phosphatidylinositol 4-phosphate 5-kinase homo-and heterodimerization determines its membrane localization and activity. The FASEB Journal 29, 2371-2385.
195. Brinkmann, A., Faber, P., Van Rooij, H., Kuiper, G., Ris, C., Klaassen, P., Van der Korput, J., Voorhorst, M., Van Laar, J., and Mulder, E. (1989). The human androgen receptor: domain structure, genomic organization and regulation of expression. Journal of steroid biochemistry 34, 307-310.
196. Hsing, A.W., Gao, Y.-T., Wu, G., Wang, X., Deng, J., Chen, Y.-L., Sesterhenn, I.A., Mostofi, F.K., Benichou, J., and Chang, C. (2000). Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer research 60, 5111-5116.
197. Watson, J.F., and García-Nafría, J. (2019). In vivo DNA assembly using common laboratory bacteria: a re-emerging tool to simplify molecular cloning. Journal of Biological Chemistry 294, 15271-15281.
198. Anon. addgene [online]. Available at: http://n2t.net/addgene:23848 [Accessed on 1.10.2022].
199. Anon. addgene [online]. Available at: http://n2t.net/addgene:22299 [Accessed on 1.10.2022].
200. Kurokawa, Y., Konishi, R., Tomioku, K., Tanabe, K., and Fujita, A. (2021). The distribution of phosphatidylinositol 4, 5-bisphosphate in the budding yeast plasma membrane. Histochemistry and Cell Biology 156, 109-121.
201. Sarwar, M., Syed Khaja, A.S., Aleskandarany, M., Karlsson, R., Althobiti, M., Ødum, N., Mongan, N.P., Dizeyi, N., Johnson, H., and Green, A.R. (2019). The role of PIP5K1α/pAKT and targeted inhibition of growth of subtypes of breast cancer using PIP5K1α inhibitor. Oncogene 38, 375-389.
202. Zhang, Q., Zhang, C., Ma, J.-X., Ren, H., Sun, Y., and Xu, J.-Z. (2019). Circular RNA PIP5K1A promotes colon cancer development through inhibiting miR-1273a. World journal of gastroenterology 25, 5300.
203. Sun, Y., Li, X., Chen, A., Shi, W., Wang, L., Yi, R., and Qiu, J. (2019). circPIP5K1A serves as a competitive endogenous RNA contributing to ovarian cancer progression via regulation of miR-661/IGFBP5 signaling. Journal of cellular biochemistry 120, 19406-19414.
204. Shi, P., Li, Y., and Guo, Q. (2021). Circular RNA circPIP5K1A contributes to cancer stemness of osteosarcoma by miR-515-5p/YAP axis. Journal of Translational Medicine 19, 1-12.

205. Chi, Y., Luo, Q., Song, Y., Yang, F., Wang, Y., Jin, M., and Zhang, D. (2019). Circular RNA circPIP5K1A promotes non-small cell lung cancer proliferation and metastasis through miR-600/HIF-1α regulation. Journal of cellular biochemistry 120, 19019-19030.
206. Stijf-Bultsma, Y., Sommer, L., Tauber, M., Baalbaki, M., Giardoglou, P., Jones, D.R., Gelato, K.A., van Pelt, J., Shah, Z., and Rahnamoun, H. (2015). The basal transcription complex component TAF3 transduces changes in nuclear phosphoinositides into transcriptional output. Molecular cell 58, 453-467.
207. Gelato, K.A., Tauber, M., Ong, M.S., Winter, S., Hiragami-Hamada, K., Sindlinger, J., Lemak, A., Bultsma, Y., Houliston, S., and Schwarzer, D. (2014). Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. Molecular cell 54, 905-919.
208. Jacobsen, R.G., Gavgani, F.M., Edson, A.J., Goris, M., Altankhuyag, A., and Lewis, A.E. (2019). Polyphosphoinositides in the nucleus: roadmap of their effectors and mechanisms of interaction. Advances in biological regulation 72, 7-21.
209. Thapa, N., Choi, S., Tan, X., Wise, T., and Anderson, R.A. (2015). Phosphatidylinositol phosphate 5-kinase Iγ and phosphoinositide 3-kinase/Akt signaling couple to promote oncogenic growth. Journal of Biological Chemistry 290, 18843-18854.
210. Fetchko, M., and Stagljar, I. (2004). Application of the split-ubiquitin membrane yeast two-hybrid system to investigate membrane protein interactions. Methods 32, 349-362.
211. Anon. PrimerBlast. NCBI [online]. Available at: https://www.ncbi.nlm.nih.gov/tools/primer-blast/ [Accessed on 1.10.22].
212. Taylor, S.C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., and Fenrich, J. (2019). The ultimate qPCR experiment: producing publication quality, reproducible data the first time. Trends in Biotechnology 37, 761-774.
213. Anon. TUT1 gene ID 64852. NCBI [online]. Available at: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=64852 [Accessed on 1.10.2022].
214. Yamashita, S., Takagi, Y., Nagaike, T., and Tomita, K. (2017). Crystal structures of U6 snRNA-specific terminal uridylyltransferase. Nature communications 8, 1-11.