The Role Of Phospholipase C Enzymes In Health And Disease

Ioannis Ladas

Submitted for the Degree of Doctor of Philosophy (PhD)

Centre for Endocrine and Diabetes Sciences

School of Medicine

Cardiff University

Submitted May 2015

AUTHOR'S DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate) Date ……

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD

Signed …… (candidate) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.
Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (candidate) Date ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …… (candidate) Date ……

Summary

Phospholipases C (PLCs) are a large family of enzymes that regulate the cleavage of phosphatidylinositol-4,5- bisphosphate into Inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). This project included research on PLCζ and PLCβ1 on male infertility and thyroid cancer respectively.

Human PLCζ (hPLCζ) is the key protein for egg activation following sperm-egg fusion and triggers the release of Ca^{2+} producing a series of oscillations which initiate embryogenesis. Defects in hPLCζ lead to infertility, thus the project aimed at its expression and enrichment suitable for biomedical applications.

The results included 1) Comparison of hPLCζ bacterial expression from twenty four plasmid constructs with varying protein tags. 2) Expression of active soluble hPLCζ, free from its bacterial solubility partners.

The PLCβ1 project involved a novel InDel (1076 bp deletion, ATAA junction insertion) identified in the 3rd intron of PLCβ1 (Chr 20) by genome-wide linkage analysis (GWLA, LOD score 3.01) of a large kindred with multinodular goitre (MNG) progressing to papillary thyroid cancer (PTC). Affected individuals exhibited increased PLCβ1 transcript levels in their thyroids and the InDel has a putative ERα binding site.

The project aimed to, 1) Develop a genotyping technique for high-throughput PLCβ1 InDel screening in large cohorts 2) Determine the InDel's mechanism of action in modifying thyrocyte proliferation. 3) Next generation sequencing (NGS) of the chr20 region identified by GWLA to identify potential variants responsible for MNG.

1) A QPCR genotyping method was established and the InDel identified in patients with benign thyroid disease. 2) Reporter gene assays suggested silencer elements within the InDel. 3) PLCβ1 knockdown using siRNA exhibited a trend in inhibiting proliferation of a thyroid cell-line. 4) NGS of the 20cM-Chr20 GWLA identified region found no additional disease loci confirming a role for the PLCβ1 InDel. The InDel provides a biomarker for MNG patients most likely to develop PTC.

Acknowledgments

Firstly, I would like to express my sincere gratitude to Prof. Marian Ludgate and Dr. Fiona Grennan Jones for their excellent guidance and contribution throughout the course. I feel extremely lucky to had Prof. Marian Ludgate as advisor; she is the most talented academic I have ever seen and she will always be an example in my future career.

I would also like to give many thanks to Prof. Colin Dayan, Dr. Philip James, Prof. Alan Williams and Prof. Maurice Hallett as without them, their support and guidance I wouldn't be able to complete this PhD.

A big thank-you to my friends and colleagues at both the Wales Heart Research Institute and the Centre for Endocrine and Diabetes Science for their support throughout this PhD. I would particularly like to thank Dr. Spyros Zissimopoulos for his immense support during the first year of my course and Dr. Shazli Mohd Draman, Dr. Lei Zhang, Dr. Ameen Bakhsh, Dr. Illaria Muller, Dionne Shillabeer, Charlotte Hales and Lynn Tailor for their friendship and for brightening my day every morning.

Many thanks to the collaborators; Dr. Roderick J Clifton-Bligh and Dr. Martyn Bullock for their valuable input in my work.

I would also like to thank Dr. Nigel Williams and Dr. Kiran Mantripragada for giving me the opportunity to undertake training in NGS in their department.

Special thanks to Onassis Foundation and Medical Research Council for supporting my PhD.

Finally, I would like to thank my parents for their continued support over the course of my studies.

For my Grandfather,

Father Demetrius Koutsikos

(1929-2010)

"ΑΓΝΩΣ ΔΕ ΚΑΙ ΟΣΙΩΣ ΔΙΑΤΗΡΗΣΩ ΒΙΟΝ ΤΟΝ ΕΜΟΝ ΚΑΙ ΤΕΧΝΗΝ ΤΗΝ ΕΜΗΝ."

"With purity and with holiness I will pass my life and practice my Art."

From Hippocratic Oath

Hippocrates (460 – c. 370 BC)

Abbreviations

ADAM	A Disintegrin And Metalloproteinase
AGO	Argonaute Protein
ATC	Anaplastic Thyroid Cancer
BAPTA	1,2-bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid) is a calcium-specific aminopolycarboxylic acid
BCA	Bicinchoninic Acid
Ca^{2+}	Calcium
CADPR	Cyclic ADP Ribose
CBP	Calcium Binding Protein
cDNA	Complementary DNA
CIP	Calf-intestinal alkaline phosphatase
cM	Centi-Morgan
CNV	Copy number Variations
Co_2	Carbon dioxide
Co^{2+}	Cobalt
COSMIC	Catalogue of Somatic Mutations In Cancer
Ct	Cycle Threshold
Cu^{2+}	Copper
DAG	Diacylglycerol
DEHAL	Iodotyrosine Dehalogenase
DMEM	Dulbecco's Modified Eagle's Minimal
dNTP	Deoxynucleotide
DTC	Differentiated Thyroid Cancer
DTT	Dithiothreitol
DUOX	Dual Oxidase
E.Coli	Escherichia Coli
EGTA	Thylene Glycol Tetraacetic Acid
ELISA	Enzyme-Linked ImmunoSorbent Assay
ER	Endoplasmic Reticulum

ERa	Oestrogen Receptor
FA	Follicular Adenoma
FCS	Fetal Calf Serum
FNA	Fine Needle Aspiration
FOX	Forkhead Box
FTC	Follicular Thyroid Cancer
FTRTL5	Fisher thyroid cell line
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
Gb	Giga bases
GDP	Guanosine Diphosphate
GEF	Guanine Nucleotide Exchange Factor
GFP	Green Fluorescent Protein
GPCR	G Protein Coupled Receptor
GPI	Glycophosphatidylinositol
GRR	Graphical Representation of Relationships
GST	Glutathione S-transferase
GTP	Guanosine Triphosphate
GWLA	Genome Wide Linkage Analysis
H_2O_2	Hydrogen Peroxide
HCL	Hydrogen Chloride
HCO_{3-}	Bicarbonate
HEK	Human Embryonic Kidney
HIS_6	6x Histidines
HIV	Human Immunodeficiency Virus
HNSCC	Head and Neck Squamous Cell Carcinoma
IBD	Identity By Descent
ICGC	International Cancer Genome Consortium
ICSI	Intracytoplasmic Sperm Injection
IGF	Insulin-like Growth Factor
IGV	Integrative Genomics Viewer
IMAC	Ion Affinity Chromatography
InDel	Insertion-Deletion

IP_3	Inositol trisphosphate
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ISPs	Ion Sphere Particles
kDa	Kilo Dalton
KPN	Importin
LDH	Lactate Dehydrogenase
LOD	Logarithm of the Odds
MBP	Maltose Binding Protein Expression
MCF7	Human Breast Adenocarcinoma Cell Line
MCT8	Monocarboxylate 8
MDS	Myelodysplastic Syndrome
MEL	Mouse Erythroleukemia
Mg^{2+}	Magnesium
min	Minutes
ml	Millilitre
MNG	Multinodular Goitre
MTC	Medullary Thyroid Cancer
Na_2PO_4	Disodium Phosphate
NaOH	Sodium Hydroxide
NEB	New England Biolabs
NES	Nuclear Export Sequence
Ni^{2+}	Nickel
NIS	Sodium Iodide Symporter
NLS	Nuclear Localization Signal
NTA	Nitrilotriacetic Acid
^oC	Calcium
OCA	Octamer-Binding Factor 1 Coactivator
PAX8	Paired Box
PCR	Polymerase Chain Reaction
PDS	Pendrin
PDTC	Poorly Differentiated Thyroid Cancer
PGM	Personal Genome Machine

pH	Pleckstrin Homology Domain
PIP_2	Phosphatidylinositol 4,5-Bisphosphate
PKC	Protein Kinase C
PLC	Phospholipase C
PP	Periprostatic Tissue
PPFP	PAX8-PPARγ Fusion Protein
PPRG	Peroxisome Proliferator-Activated Receptor. Gamma
pRb	Retinoblastoma Protein
pRL	Renilla Luciferase plasmid
PTC	Papillary Thyroid Cancer
QC	Quality Control
qPCR	Quantitative PCR
RAS	Rat Sarcoma
ret	Receptor Tyrosine-Protein Kinase
RFLP	Restriction Fragment Length Polymorphism
RISC	RNA Induced Silencing Complex
RNA	Ribonucleic Acid
RNAi	Ribonucleic Acid Interference
rpm	Revolutions per Minute
SDS	Sodium Dodecyl Sulphate
SF	Sperm Factor
SH	Subclinical Hypothyroid
SH2	src homology 2
SHIP	Study of Health in Pomerania
SiRNA	Small Interfering RNA
SNP	Single Nucleotide Polymorphism
SOLiD	Supported oligonucleotide ligation and detection
SR	Sarcoplasmic Reticulum
T3	Triiodothyronine
T4	Thyroxine
Taq	Derived from *Thermus aquaticus* Bacterium
TEV	Tobacco Etch Virus Protease

TG	Thyroglobulin
THBP	Thyroid Hormone Binding Proteins
Tm	Melting Temperature
TPC	Thyroid Papillary Carcinoma
TPO	Thyroperoxidase
TRA	Thyroid Stimulating Hormone Receptor Antibody
TRBP	Trans-Activating Response RNA-Binding Protein
TRE	Thyroid Response Elements
TRIS	Tris-Buffered Saline
tRNA	Transfer RNA
TRX	Thioredoxin
TSAb	TSHR stimulating autoantibodies
TSH	Thyroid-Stimulating Hormone
TSHR	Thyroid Stimulating Hormone Receptor
UCSC	University of California Santa Cruz
VSV-G	Vesicular Stomatitis Virus
WHO	World Health Organization
Zn^{2+}	Zinc
ZP	Zona Pellucida
µg	Microgram
µl	Microliter

List of Meetings and Abstracts Arising from this Work

Abstracts:

International Meetings:

2014: European Thyroid Association (ETA),

Santiago de Compostela, Spain

An Intronic Deletion In Phospholipase-C β1 (PLCβ1) Associated With Euthyroid Multinodular Goitre (MNG) And Risk Of Progressing To Papillary Thyroid Cancer (PTC) (Oral)

Cardiff University Meetings:

2014: Postgraduate Research Day

Next Generation Sequencing in Subjects Carrying an InDel in Phospholipase- Cβ1 to Identify Variants Associated with Euthyroid Multinodular Goitre (MNG) Progressing to Papillary Thyroid Cancer (PTC) (Poster)

2013: Postgraduate Research Day

Development and Optimization of qPCR Based Genotyping for Screening and Identification of a PLCβ1 Deletion in Euthyroid Multinodular Goitre Patients at Risk for Developing Papillary Thyroid Cancer (Poster)

Table of Contents

Contents

List of Figures

List of Tables

Chapter 1. General Introduction

1.1 Phospholipases C (PLCs)

PLCs are a large family of enzymes that have an essential role in the signalling cascade of many receptors as described below. Phosphoinositide-specific phospholipase C is comprised of six families of multi-domain eukaryotic proteins and is known to be involved in various biological processes (Bunney and Katan 2011). PLCs are known to regulate the cleavage of the polar head group from phosphatidylinositol 4,5- bisphosphate also known as PtdIns(4,5)P_2 (PIP_2) resulting in the generation of Inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). These are second messengers known to have calcium-mobilizing properties by activating a variety of proteins. More specifically, Inositol 1,4,5-trisphosphate (IP_3) diffuses into the cytosol and binds to intracellular IP_3 receptors in the endoplasmic reticulum (ER) and sarcoplasmic reticulum (SR) leading to Ca^{2+} release from the intracellular calcium stores into the cytoplasm (Katan and Williams 1997; Futatsugi *et al.* 2005). On the other hand, (DAG) activates protein kinase C (PKC) which in turn regulates transcription cell growth and survival (Hisatsune *et al.* 2005).

1.2 The structure of PLC proteins

Evidence regarding PLCs activity has been revealed in 1953 by (Hokin and Hokin 1953) with the first PLC being purified in 1981 by (Takenawa and Nagai 1981). Since then six families of mammalian PLC (β, γ, δ, ε, ζ and η) have been identified (Suh *et al.* 2008; Bunney and Katan 2011). All PLCs excluding PLCζ which is lacking the PH domain, have a conserved structure comprised of a PH domain, a series of EF hands, the X and Y domains which form the triosephosphate isomerase (TIM) α/β-barrel which is considered to be the most conserved domain among PLCs and finally a C2 domain (Figure 1.1,1.2). The subdomains in the different isozymes have also been found to play key role in the isozyme's activity, such as in PLC-γ which is known to contain two src homology 2(SH2) domains and an SH3 domain. Studies have revealed that these domains promote protein interactions with protein kinases which activate PLC –γ. Also, PLC ε contains a CDC25 domain at the N-terminus and RA domains at the C terminus which are known to regulate GTP-dependent interactions with proteins of the RAS family (Figure 1.1) (Rotin *et al.* 1992; Fukami 2002; Wen *et al.* 2006; Bunney and Katan 2011).

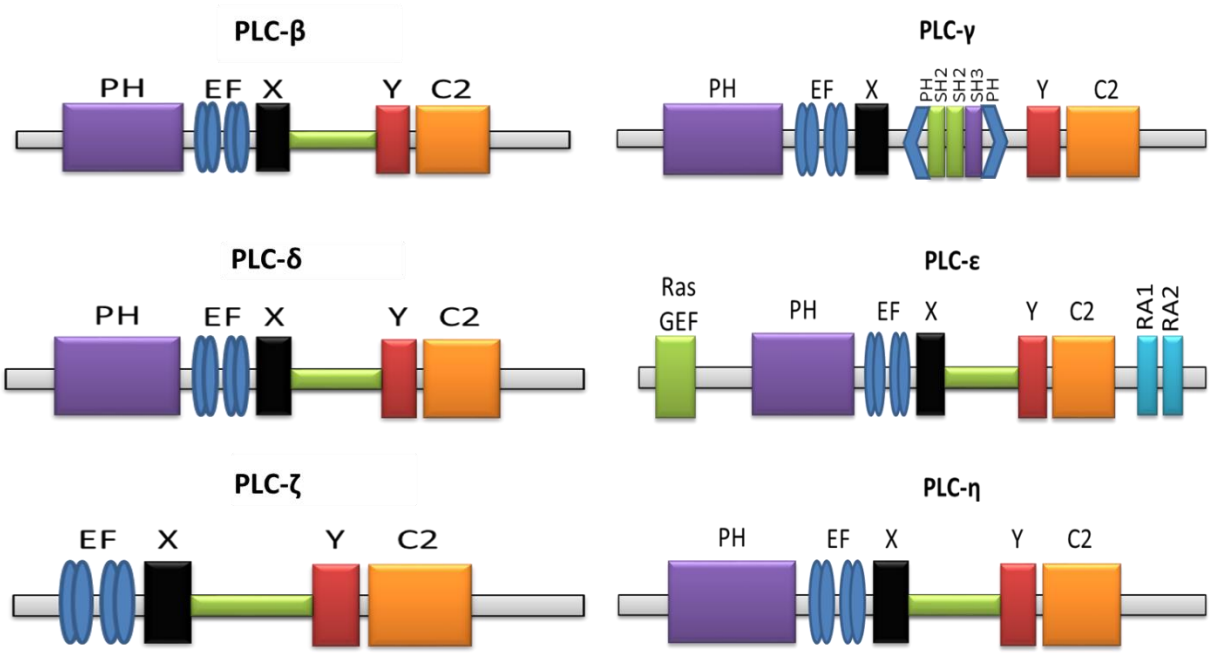


Figure 1.1 The structure of the different PLC isoenzymes. Phospholipases consist of 6 is families with distinct structure and biodistribution. PH: Pleckstrin Homology domain, SH: Src homology domain, RA: Ras association 1 domain, Ras:Rat sarcoma, GEF: Guanine nucleotide exchange factor.

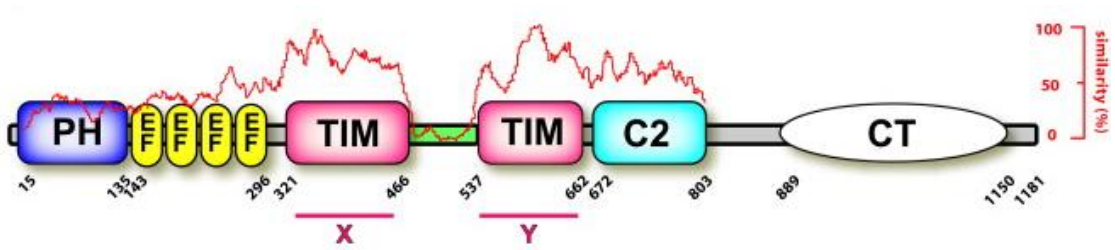


Figure 1.2 The level of conservation of the various domains across PLCs is illustrated by the red line, which give the percentage similarity. Adapted from (Hicks *et al.* 2008).

The aim of this project was the investigation of the potential use of human PLCζ in therapeutic applications such as male infertility and the role of a human PLCβ1 InDel in MNG patients likely to develop papillary thyroid cancer.

1.3 Phospholipase C ζ (PLCζ)

PLCζ it is a sperm specific protein with a molecular weight of approximately 70kDa which is the smallest among mammalian PLCs (Saunders *et al.* 2002). It is structurally similar to PLCδ1 with the distinct difference that it lacks a PH domain (Saunders *et al*. 2002). It is considered to be the ideal candidate for 'egg activation factor' as it causes multiple Ca^{2+} oscillations having a similar pattern to those occurring with in vitro fertilization (Saunders *et al*. 2002; Kouchi *et al.* 2004). Also, it has approximately 100 fold higher calcium sensitivity than PLCδ, requiring a Ca^{2+}concentration of only a few nanomolar in order to be activated. Studies have shown that the cytosolic Ca^{2+} concentration is about 50-80nM indicating that is enough concentration to activate PLCζ but not δ which requires micromolar concentration (Nomikos *et al.* 2005).

1.4 PLCζ structure

1.4.1 EF hands

The EF hand motif is a well-studied domain as it is one of the most common in animal cells with over 1000 having been identified so far within 66 protein subfamilies (Nelson *et al.* 2002). The name EF-hand derived from the calcium-binding motif identified in parvalbumin and was given by Kretsinger and Nockolds in 1973 (Figure 1.3). EF-hand motif proteins have been classified into two different types, the Ca^{2+} sensors and the Ca^{2+} signal modulators, where the Ca^{2+} signal is transduced and the shape and duration of Ca^{2+} signals are modulated respectively (Nelson *et al*. 2002). The motif was found in the vast majority of the proteins where it was identified to bind Ca^{2+} and in few cases magnesium (Lewit-Bentley and Réty 2000) (Figure 1.3).

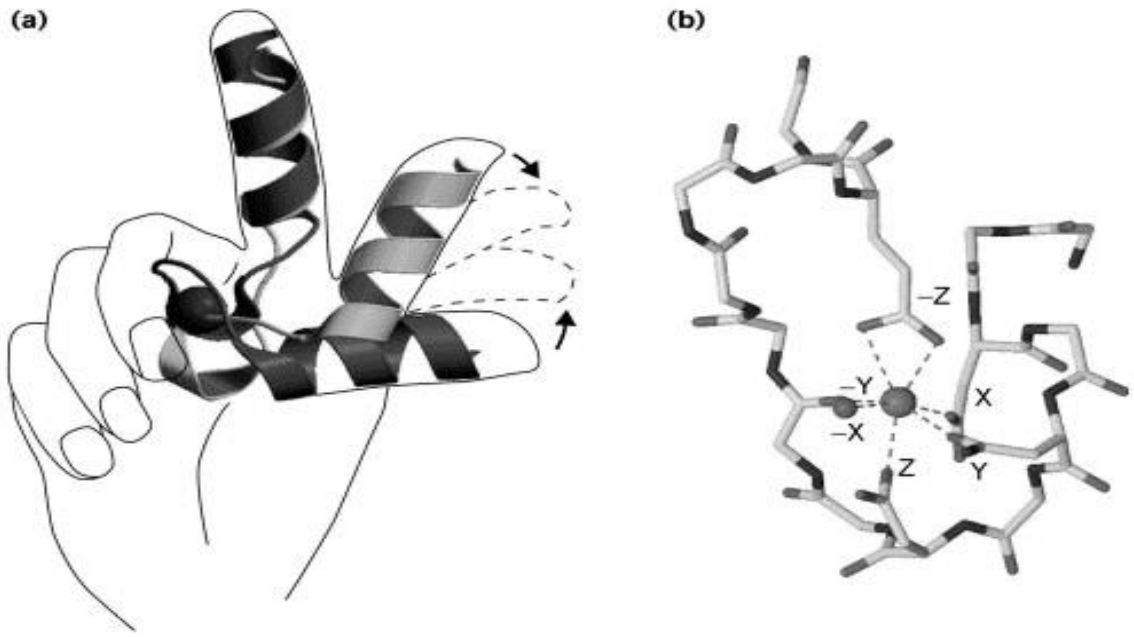


Figure 1.3 The typical structure of EF hand motif. A) EF hands motif is characterised by its helix loop helix motif. B) The sequence pattern is X.Y.Z.-Y.-X..-Z, with the letters representing the ligands in metal coordination and the dots the intervening residues. Adapted from (Lewit-Bentley and Réty 2000).

The typical structure of the EF-hand motif is comprised usually from 40 residues forming a helix-loop-helix motif where two a-helices are bridged by a Ca^{2+} chelation loop (Taylor *et al.* 1991; Lewit-Bentley and Réty 2000; Gifford *et al.* 2007). The sequence pattern is X.Y.Z.-Y.-X..-Z, with the letters representing the ligands in metal coordination and the dots the intervening residues (Lewit-Bentley and Réty 2000) (Figure 1.3). In order for the interaction of the EF-hands and the Ca^{2+} to be maximised, the EF loops are rich with negatively charged amino acids such as glutamic acid and aspartic acid with the latter being the more common of the two (Gifford *et al*. 2007). The EF- hand motif has been found to be present in pairs forming a four-helix bundle with amphipathic helices following packaging which lead to the formation of a hydrophobic core. This structure is stabilised further from the β-sheets which are formed between the EF-loop pairs (Gifford *et al*. 2007).

Mammalian PLCζ consists of 4 domains, 2 pairs of EF hand domains, the X and Y catalytic domains, an intervening region known as the XY linker and the C2 domain (Figure 1.4). The EF hands consist of 4 EF hand motifs, which exist in pairs in order to stabilize each other. EF-hands have been shown by a number of studies to play an important role in the proteins' calcium sensitivity and activity. Mutational studies revealed that the EF hand domains are

critical for Ca^{2+} oscillation- inducing activity (Kouchi *et al.* 2005; Nomikos *et al*. 2005). The PLC activity of PLCζ was remarkably reduced when the EF-hand domains were deleted one by one from the N-terminus, with the data indicating that EF-1 and EF-2 are responsible for the PLC activity and EF-3 contributing to the high Ca^{2+} sensitivity (Kouchi *et al*. 2005; Nomikos *et al*. 2005). EF hands were found to be important for the interaction of the X and Y domains with the PIP2 substrate and therefore for PLC enzyme activity. The removal of EF hands reduced the minimum number of interacting sites from 4 to 1 and deletion of the EF hands abolished the PLC activity in vivo (Nomikos *et al*. 2005). Studies performed in PLCδ1 have indicated that the loop regions of EF-hands 3 and 4 do not have metal binding sites. In contrast, EF hands 1 and 2 contain calcium ligands suggesting that calcium binding will occur at least to the first EF hand loop (Essen *et al.* 1997). Also it has been suggested that EF hands may have a role in substrate targeting. Replacement of human PLCζ EF hands with mouse led to the decrease of Ca^{2+} sensitivity combined with a 6 fold decrease in substrate affinity (Nomikos *et al.* 2014)

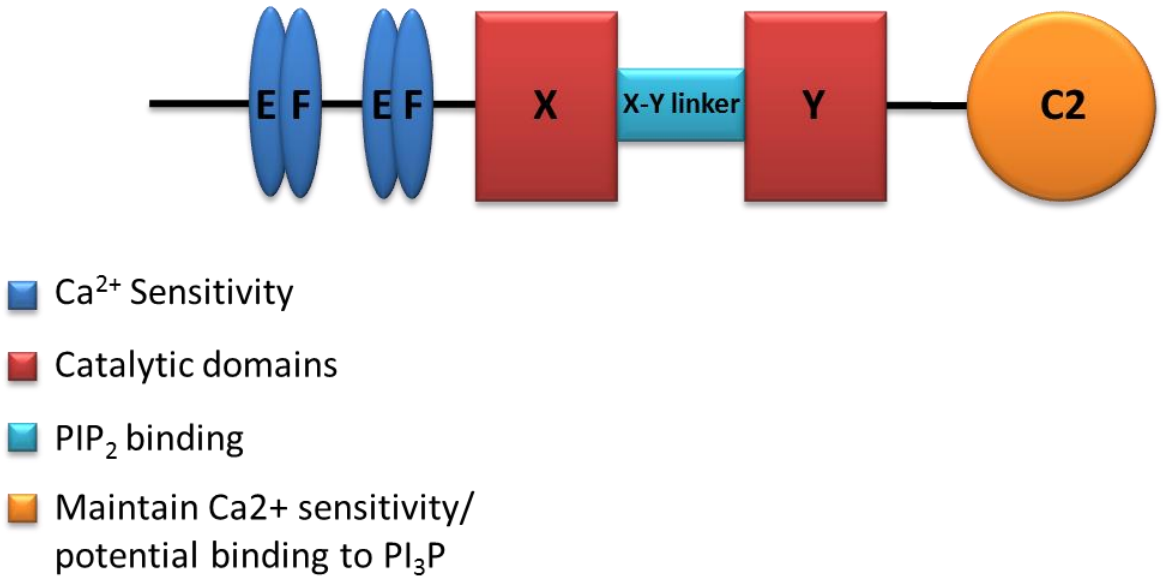


Figure 1.4 The structure of human PLCζ. PLCζ differs from the other PLCs, as it does not contain a PH domain. Recent studies have revealed the role of its domains in egg activation. The EF –hand domain is shown to be responsible for the Ca^{2+} sensitivity, The X and Y domains are the catalytic domains, the X-Y linker has a critical role in PIP_2 binding and the C2 domain has been shown to assist in maintaining Ca^{2+} sensitivity. Studies have also indicated possible binding of C2 domain to PI_3P.

1.4.2 XY catalytic domain and linker

The XY catalytic domain is one of the most conserved regions, with about 64% similarity with PLCδ1 and is responsible for PLC enzymatic activity (Swann *et al.* 2006). The X and Y domains are intervened by a region known as the XY linker. This region is considered to be the least conserved among the species and is comprised mainly of basic residues (Nomikos *et al.* 2007). Recent studies have revealed that XY linker due to its positive charge, interacts with the negatively charged membrane surface, resulting in the binding of PLCζ to PIP_2 (Nomikos *et al.* 2011b). Interestingly, a negative charge in the XY linker region has been observed in PLCβ, δ and ε which regulates PIP_2 hydrolysis regulating an auto-inhibition of PLC activity, which can occur only following a conformation change which keeps away the X-Y linker allowing the binding of PLC to PIP_2 (Hicks *et al*. 2008). Studies have shown that substitution of three of the lysines to alanines (K374A, K375A and K377A) in PLCζ X-Y linker, reduced *in vivo* Ca^{2+} oscillations and *in vitro* activity as, the positive charge was significantly reduced disrupting the PIP_2 binding efficiency (Nomikos *et al.* 2013a). Replacement of human X-Y linker with the mouse region had a significant effect on calcium oscillations confirming its vital role in enzymatic activity (Nomikos *et al*. 2014). Also a nuclear localisation signal (NLS) has been predicted within the X-Y linker and is suggested to have a key role in the shuttling of the protein between the cytoplasm and the nucleus following egg activation, in mouse oocytes (Nomikos *et al*. 2013a; Kashir *et al.* 2015).

1.4.3 The C2 domain

The C2 domain is a conserved region among PLCs comprising approximately 120 amino acids and having an important role in membrane anchoring (Amdani *et al.* 2013). In PLCζ, it is suggested to have an important role in *in vivo* activity, as removal of C2 abolished Ca^{2+} oscillations in mouse eggs (Nomikos *et al*. 2014). The C2 domain has also been suggested to have a potential role in activation of PLCζ. It has been found that interacts with PI_3P reducing PLCζ *in vitro* activity (Kouchi *et al*. 2005) suggesting that this interaction could be responsible for its inactive status in sperm prior sperm – egg fusion. Also, it has been suggested to have a role in PLCζ localisation. A computational prediction has suggested that C2 may interact with the EF hand hydrophobic residues forming a loop which in turn exposes the NLS region of the X-Y linker (Kuroda *et al.* 2006)(Figure 1.5).

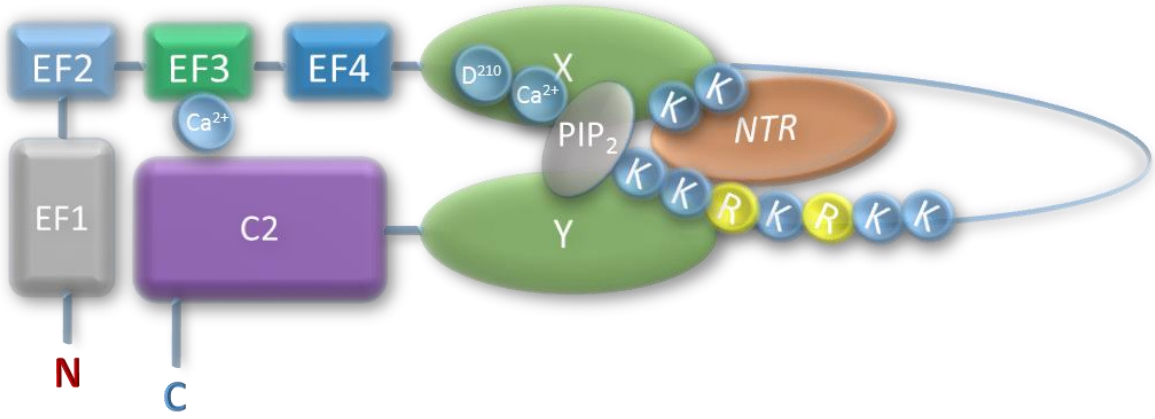


Figure 1.5 Illustration of the predicted interaction of C2 with the hydrophobic residues of EF hands forming a loop exposing the NLS region of XY the linker. Adapted from (Kuroda *et al.* 2006).

1.5 Human PLCζ

1.5.1 Signalling

Calcium (Ca^{2+}) signalling is known to have a vital role in a wide range of cellular mechanisms. However, studies have found that all calcium signalling systems share a common characteristic, i.e. they cause calcium pulses. This calcium release has been shown to be regulated by a number of messengers such as Inositol- 1, 4, 5-trisphosphate ($InsP_3$), cyclic ADP ribose (cADPR) and others. A number of studies by (Burgess *et al.* 1983) in hepatocytes, (Streb et al. 1983) in permealized pancreatic acinar cells and others using a range of different cells revealed that ER is the major Ca^{2+} store which triggers Ca^{2+} release due to IP_3 activity (Taylor and Tovey 2010). During fertilization, the calcium oscillations were indicated to be regulated by IP_3 which is generated by the activity of PLCζ (Berridge *et al.* 2003) (Figure 1.6).

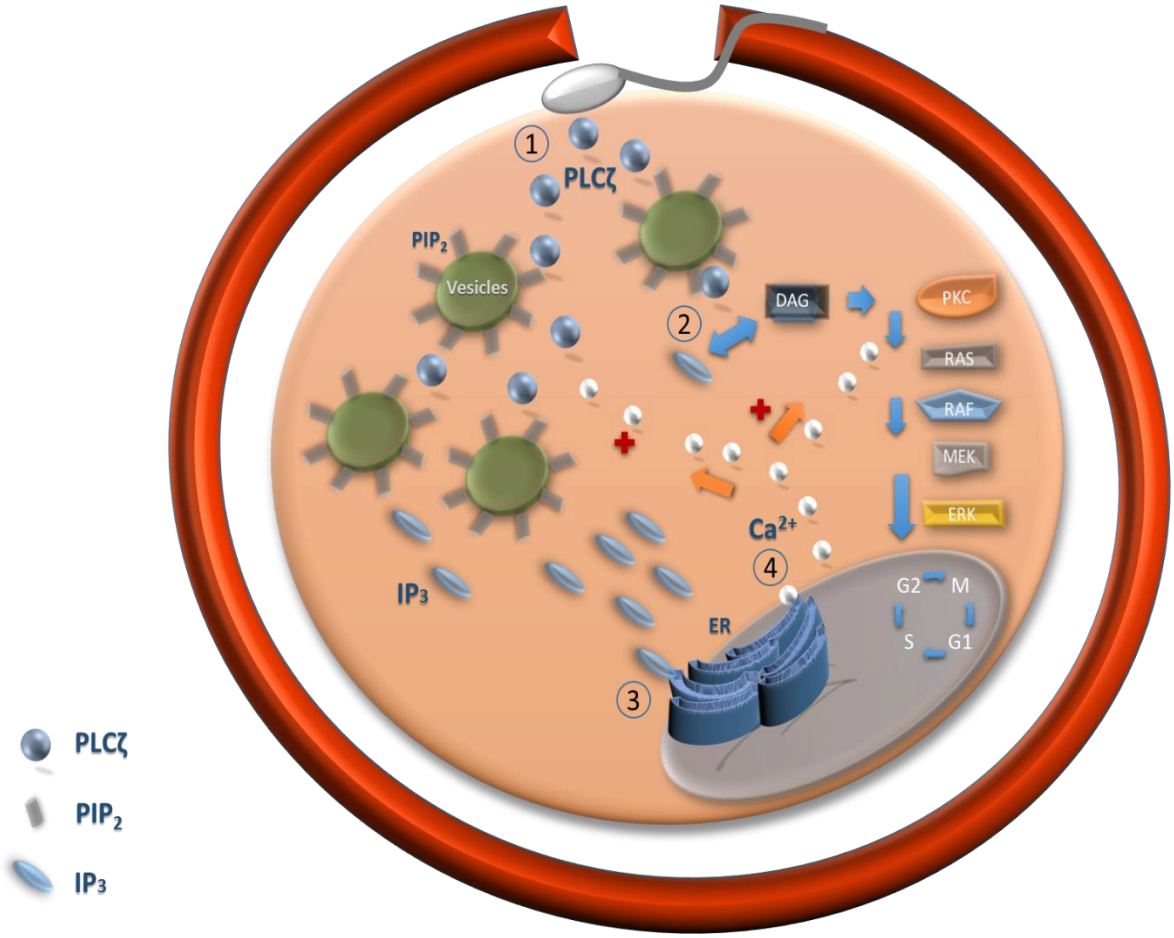


Figure 1.6 The cell signalling of PLCζ. The fusion of sperm with the egg is followed by a series of events. 1) PLC is released hydrolysing PIP_2 at internal vesicles. 2) PIP_2 hydrolysis leads in the formation of IP_3 and DAG. 3) IP_3 binds to the IP_3 receptor at the ER, 4) releasing intracellular calcium stores forming a positive feedback loop with PLCζ activity.

1.5.2 IP_3 receptor

The IP_3 receptors are Ca^{2+} channels which are comprised of 4 closely associated subunits, each of approximately 2700 amino acid residues. Each subunit has a single IP_3 binding site where, the binding of IP_3 leads to the opening of the channel and release of intracellular Ca^{2+} (Taylor *et al.* 2004; Taylor and Tovey 2010). Studies have suggested that binding to less than four IP_3-binding sites can lead to opening of the channel (Boehning and Joseph 2000); release of intracellular Ca^{2+} stores requires the binding of IP_3 to at least one of the four IP_3- binding sites of the receptor (Taylor and Tovey 2010). The binding of the IP_3 causes a large

conformational change in the receptor making the stimulatory site accessible whilst the inhibitory site is hidden. In addition intracellular Ca^{2+} store has an essential role in IP_3R sensitivity, with high stores increasing sensitivity leading to opening of the pore and loss of the Ca^{2+} store. Furthermore, the loss of Ca^{2+} desensitises the receptor to IP_3 resulting in closing of the pore until the stores are eventually refilled (Berridge 2007; Taylor and Tovey 2010).

1.6 Egg activation theories

There were a number of hypotheses proposed (Figure 1.7), until (Saunders *et al*. 2002) suggested that PLCζ is the sperm factor responsible for oocyte calcium oscillations and therefore initiation of embryo development.

1.6.1 Calcium bomb

The first hypothesis was known as the ‘’calcium bomb’’ according to which the egg, calcium stores from sperm cytosol are inserted into the egg resulting in its activation (Jaffe 1983). Studies performed by (Steinhardt *et al.* 1974; Fulton and Whittingham 1978) revealed that microinjection of Ca^{2+} or of the divalent transporting ionophore A23187 were capable of promoting Ca^{2+} release from intracellular stores in the egg resulting in its activation. However, the major argument against the bomb theory was the fact that spermatozoa volume is only a marginal proportion compared to the egg volume with less than 1%, therefore it was suggested that the amount of calcium located in the sperm head would not be sufficient to initiate Ca^{2+} release (Nomikos *et al*. 2011b). In addition, studies by (Lawrence *et al.* 1997) indicated a delay known as the ‘latent period’ between sperm fusion and Ca^{2+} release, which supported the soluble sperm factor theory.

1.6.2 Calcium conduit

The second hypothesis, was the ‘’Calcium conduit’’ which suggested the flow of extracellular calcium into the egg, through the sperm fusion with the egg (Jaffe 1983). Although, if the theory was correct, elevated levels of Ca^{2+} at the location of sperm–egg fusion would be expected (Jaffe 1991). Investigation with indo-1 dextran dye (Jones *et al.* 1998) indicated that there is no such increase and in the same study illustrated that reducing extracellular Ca^{2+} to 13nM

using chelating agents such as EGTA, did not prevent the first Ca^{2+} spike indicating that an extracellular source for the Ca^{2+} trigger is unlikely (Jones *et al*. 1998).

1.6.3 Membrane receptor hypothesis

The third hypothesis was the membrane receptor hypothesis, according to which the interaction of the sperm with a receptor at the egg surface would eventually lead to increase of the $InsP_3$ and egg activation. Initially, PLCβ was proposed to have the oscillatory role through a GTP binding protein, since in hamster eggs inhibition of G proteins lead to silencing of Ca^{2+} release. Although, inhibition of the Gaq proteins failed to eliminate Ca^{2+} oscillations in mouse eggs (Miyazaki *et al.* 1993; Williams *et al.* 1998).

1.6.4 Sperm factor

This theory was introduced when injection of cytosolic sperm extracts into hamster eggs caused Ca^{2+} oscillations as observed in fertilized mammalian eggs (Swann 1990). In addition, studies (Rice *et al.* 2000) in which they have used sperm extracts revealed that the sperm factor was calcium dependent and its activity was solely due to IP_3 production. To support that hypothesis, U73122 PLC inhibitor was used to assess the ability of PLCs to produce Ca^{2+} oscillations, which led to the conclusion that the soluble sperm factor is a PLC (Wu *et al.* 2001). Initially, activation was thought to occur via activation of PLCβ and later with PLCγ; although experiments have shown that both models were misleading (Miyazaki 1988; Runft *et al.* 1999).

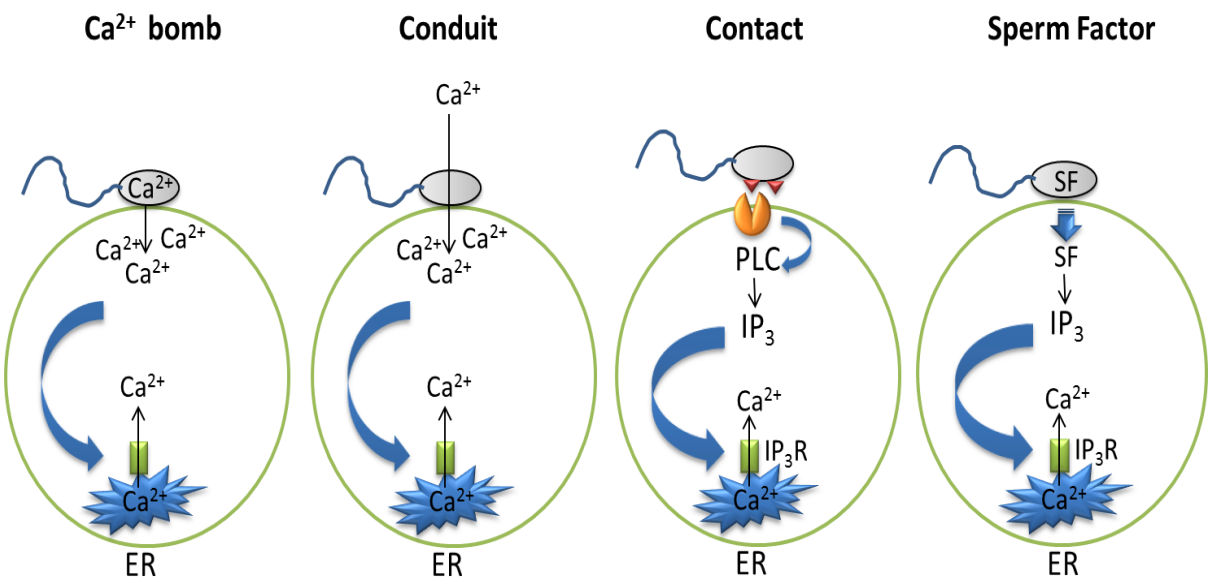


Figure 1.7 A summary of the egg activation theories. The evolution of the different hypotheses which eventually led to the sperm factor hypothesis and PLCζ. ER: Endoplasmic Reticulum, IP_3: Inositol trisphosphate, PLC: Phospholipase C, SF: Sperm Factor.

1.7 Calcium oscillations

The discovery of hPLCζ, (Saunders *et al*. 2002) confirmed the sperm factor hypothesis as this PLC isoform proved to be the best candidate. It is considered to be ideal candidate for egg activation as it causes multiple oscillations that mimic physiological fertilization (Saunders *et al*. 2002; Kouchi *et al*. 2005). PLCζ has approximately 100 fold higher calcium sensitivity than hPLCδ1, requiring only few nanomolar of Ca^{2+} in order to be activated. Studies have shown that the cytosolic Ca^{2+} concentration that is required for PLCζ activity is about 50-80nM in contrast with PLCδ1 which is approximately 100 times less sensitive to Ca^{2+} concentrations (Nomikos *et al*. 2005) (Figure 1.8).

Calcium release is a universal requirement for mammalian egg activation. Calcium release was found as either a single transient or in series of oscillations (Figure 1.9). Studies in mouse have revealed that PLCζ is causing oscillations in approximately a minute following sperm-egg fusion. In vivo studies have shown that calcium transient is mouse egg follows two stages, with initially a slower transient rate followed by a rapid oscillation rate. Oscillations continue

for approximately 3 hours following the first transient the frequency drops to approximately 1 every 10 minutes.

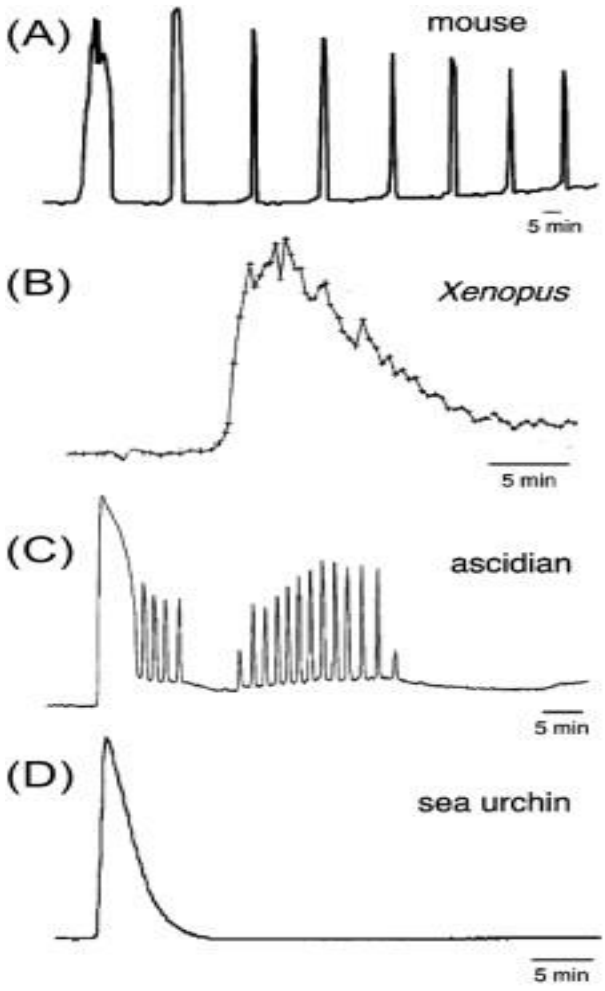


Figure 1.8 Fertilisation calcium transient across different species and the variety of different calcium transient patterns required for egg activation. Adapted from (Miyazaki and Ito 2006).

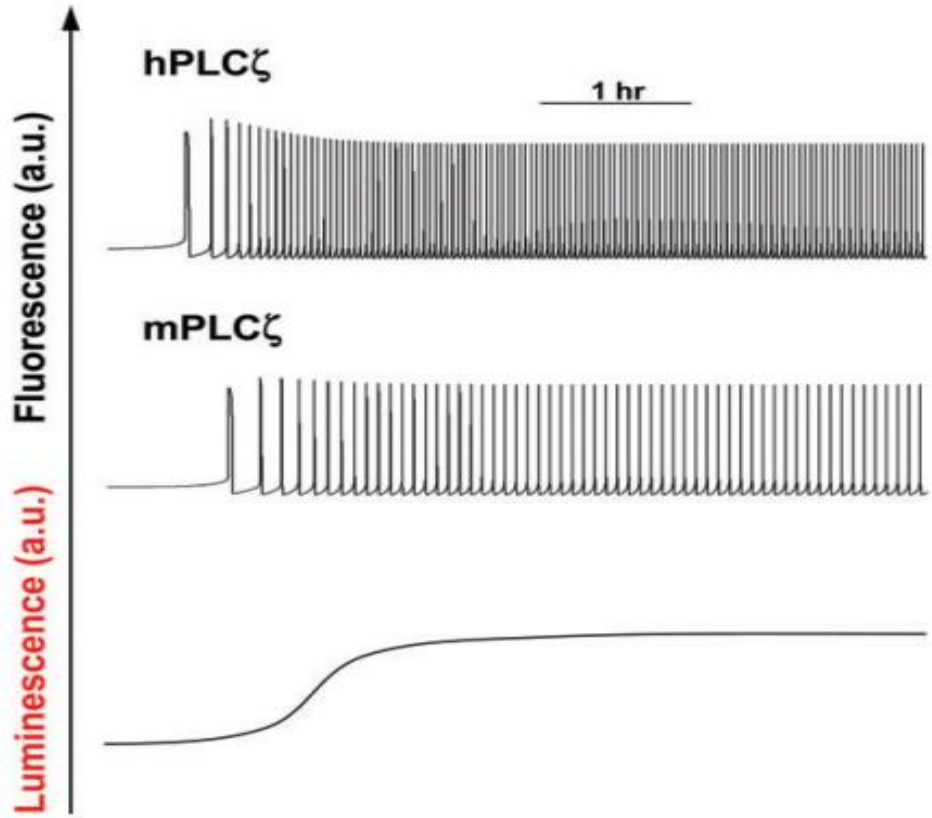


Figure 1.9 The calcium oscillations of human and mouse PLCζ in mouse egg following mathematical modelling simulation based on their in vivo enzymatic characteristics. Adapted from (Nomikos *et al*. 2013a).

1.8 Spermiogenesis

The maturation process of spermatids into mature spermatozoa is known as spermiogenesis. It is a long process which requires approximately 14 days until the sperm travels along the ductus deferens and stored for the ejaculation as seen in (Figure 1.10)(Knobil and Neill 1998). The process is initiated by formation of the acrosome, in the Golgi apparatus; vesicles migrate at the pre-acrosomal location and then the axoneme develops. A large acrosome is generated, a cup-like structure which is applied to the elongating nucleus. The acrosome functions as a modified lysosome to enable the spermatozoon penetrate the zona pellucida and fertilization to be achieved (Dale *et al.* 2010). During the acrosome stage the nucleus condenses, such that the sperm head is formed, condensation also initiates mitochondrial migration and midpiece formation at the opposite pole of the pre-acrosome. This is accompanied by inactivation of mRNA transcription and replacement of histones by protamines. As the

spermatids elongate, a flagellum is formed a structure which is responsible for sperm movement as seen in (Figure 1.11) (Knobil and Neill 1998; Falcone and Hurd 2007)

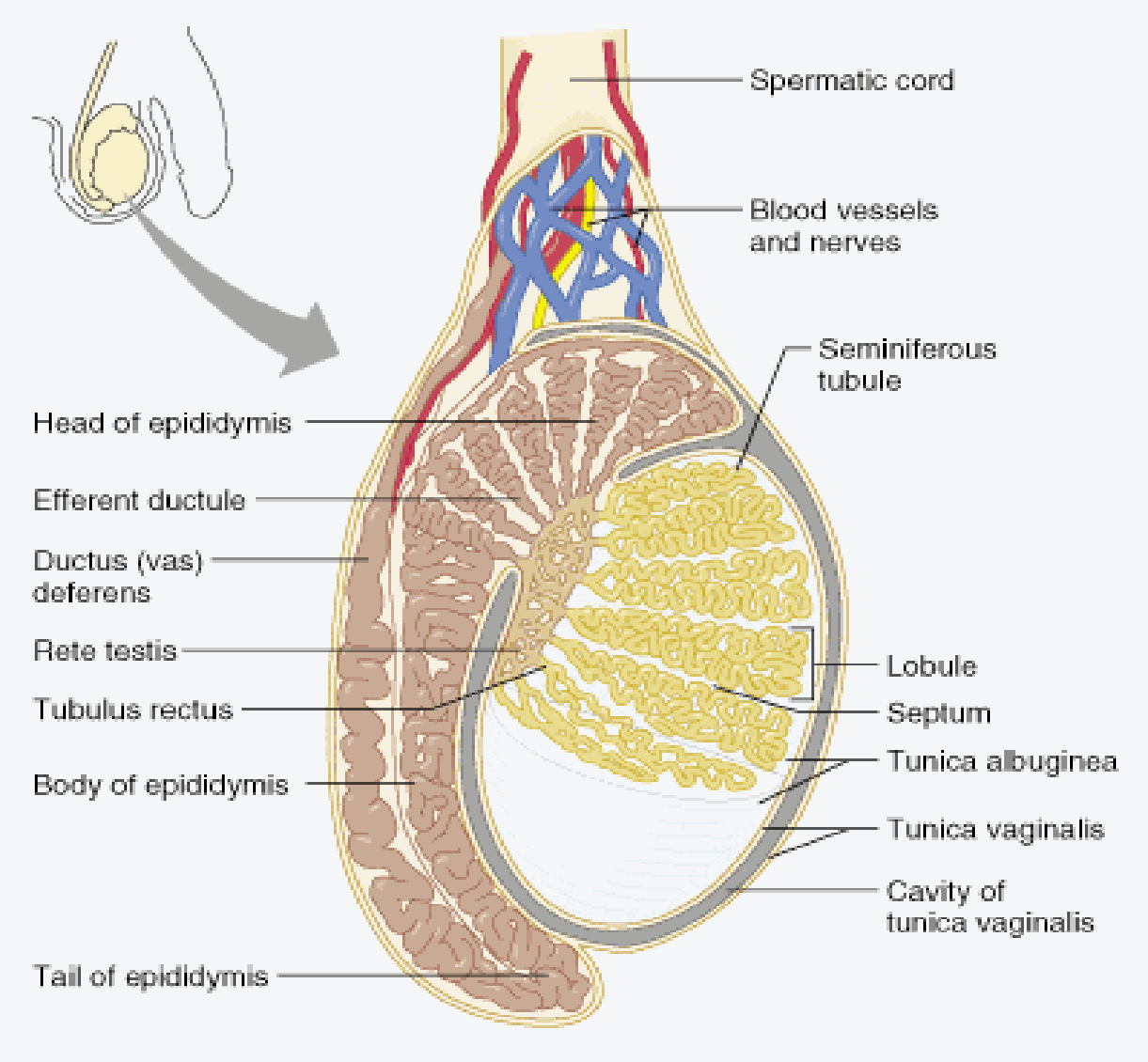


Figure 1.10 A map of Spermiogensis. Spermatogenesis takes place in seminiferous tubules where Sertoli cells are located; the epididymis is a vital structure which is involved in sperm maturation and storage. Taken from www.anatomytopics.wordpress.com

1.9 Spermiation

The change in sperm morphology is followed by spermiation, a process where mature spermatids are transferred from Sertoli cells into the lumen and towards the rete testis. Cytoplasmic portions often remain on the spermatozoon after spermiation which indicates the level of sperm maturity(Ross and Dobler 1975).

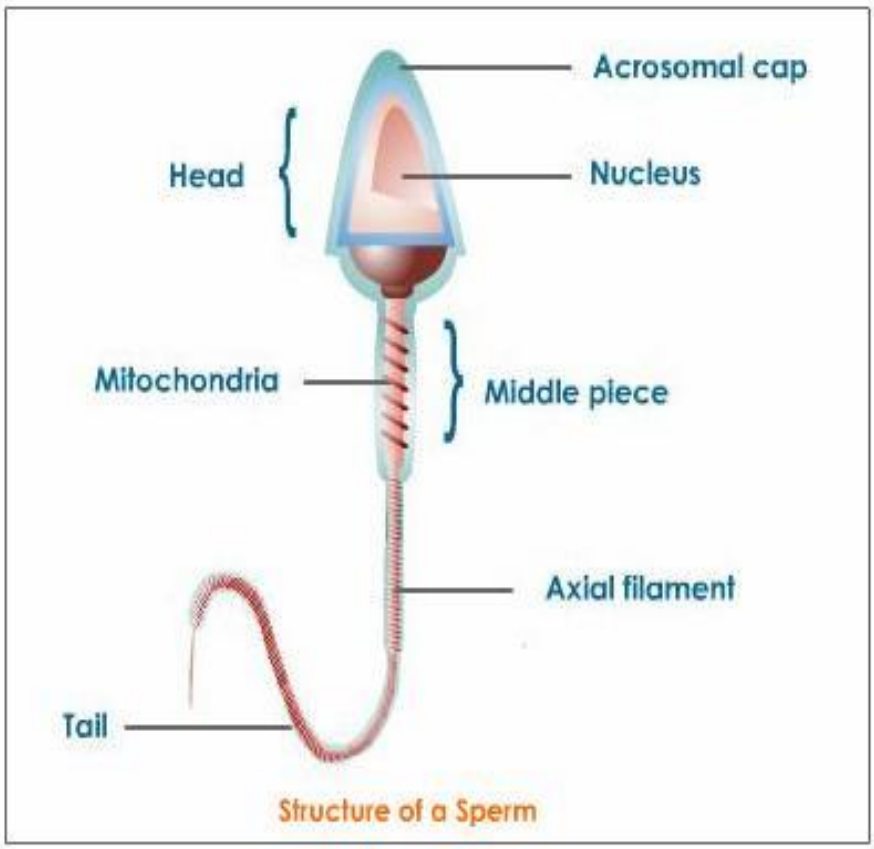


Figure 1.11 The mature spermatozoon consists of three main structures, the head, the midpiece and the tail. From www.uwstudentweb.uwyo.edu

1.10 Sperm Maturation

The spermatozoon then enters the epididymis and travels along in order for motility etc. to be acquired. Spermatozoa are first inserted into the caput epididymis where approximately 90% of the fluid secreted by the testis is absorbed (Aitken *et al.* 2007). Throughout the migration from caput epididymis to corpus and eventually to cauda epididymis where spermatozoa are stored for ejaculation, the sperm gains the ability of forward progressive movement in response to elevated cAMP, ATP Mg^{2+} and protein carnitine. In addition, sperm head acquires proteins from epididymis in order to be able to recognize and bind to the zona pellucida of the oocyte (Kirchhoff 1998; Dacheux *et al.* 2003). Through this migration, the sperm plasma membrane undergoes great changes as proteins are being modified by proteases and epididymal proteins bound to the surface stabilizing the membrane. Also some proteins have a decapacitation function to prevent premature sperm activation. Epididymal proteins such as fertilin proacrosin, membrane-bound hyaluronidase, 1,4-galactosyltransferase, sp56 and p59 represent some of the proteins which are bound to the sperm plasma membrane. In humans

this process takes approximately 2 days until the mature sperm is eventually stored at cauda epididymis (Lasserre *et al.* 2001; Dale *et al*. 2010).

Mature spermatozoa remain stabilized in the corpus epididymis at low pH in seminal fluid. To be capable of fertilization they need to become motile which is achieved only after ejaculation and the capacitation stage (Gadella *et al.* 2008). Once spermatozoa are inserted into the female genital tract after ejaculation, their seminal plasma is removed by cervical mucus which also prevents infection in the uterus (Florman *et al.* 2008). Sperm plasma proteins with decapacitating function are removed allowing initiation of capacitation, which also involves extensive rearrangement of e.g. lipid structure, this enables sperm binding to the zona pellucida and triggers the acrosome reaction as seen in (Figure 1.12) (Gadella et al. 2008).

1.11 Sperm Capacitation

Studies have revealed that capacitation is initiated by cholesterol efflux from sperm, resulting in plasma membrane destabilization and increased permeability. Plasma membrane permeability leads to influx of bicarbonate (HCO_3^-) and Ca^{2+} which increase intracellular pH and initiate intracellular Ca^{2+} signalling respectively (Florman *et al*. 2008; Gadella *et al*. 2008; Abou-haila and Tulsiani 2009). Initiation of the signalling results in adenylate cyclase activity, adenosine monophosphate (cAMP) production, in protein kinase activation, protein tyrosine phosphorylation and sperm hyperactivation required for reaching the ampulla and undergoing the acrosome reaction. Also, studies have shown that during capacitation, plasma membrane phospholipids undergo methylation as phospholipids were found to be in different distribution (Abou-haila and Tulsiani 2009). Capacitation which is considered vital for fertilization requires approximately 6 hours in humans under optimal conditions (Dale *et al*. 2010).

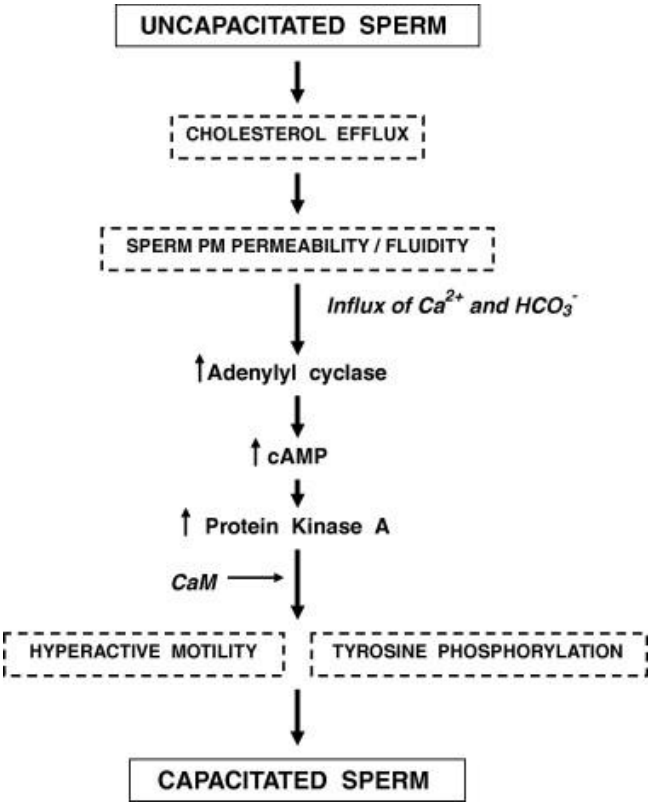


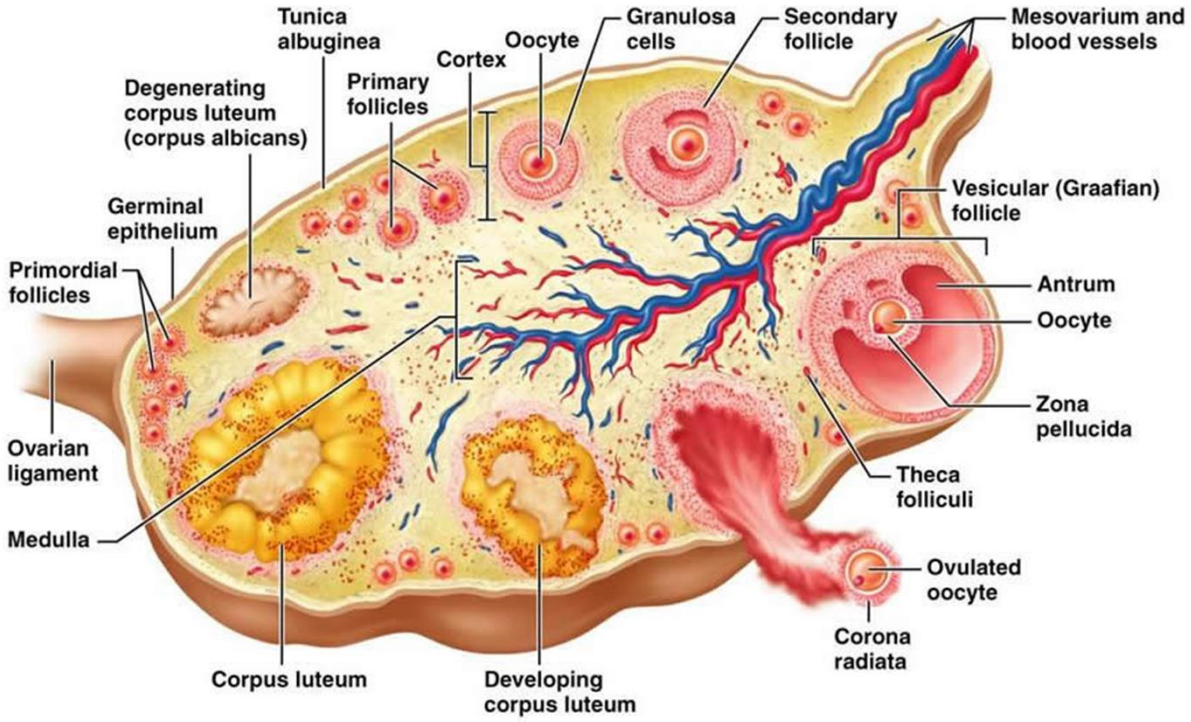
Figure 1.12 Sperm capacitation. Capacitation is required for the sperm to undergo the acrosome reaction and fertilization to occur. Capacitation involves cholesterol efflux and calcium bicarbonate influx such that sperm hyperactivation is achieved. From (Abou-haila and Tulsiani 2009). cAMP: Cyclic adenosine monophosphate, CaM: Calmodulin, HCO_3^-: Bicarbonate.

1.12 Acrosome reaction

The sperm then enters the oviduct and reaches the ampulla where the oocyte is located enveloped by cumulus cells. The cumulus contains hyaluronidase which is also present in the capacitated acrosome resulting in sperm penetration of the cumulus layer (Figure 1.12, 1.13). It has been suggested that the attachment of sperm to zona pellucida of the oocyte is the starting point of the acrosome reaction. The zona pellucida is a barrier which surrounds the oocyte and consists of three glycoproteins, (zona pellucida 1) ZP1, ZP2 and ZP3. The last glycoprotein has been found to be the first ligand between the zona and the sperm. The contact of sperm and egg through ZP3, results in opening of Ca^{2+} ions and increase of intracellular pH which, in the presence of progesterone, triggers the acrosome reaction. Zona pellucida also activates phospholipase C which regulates calcium influx and generates IP_3 and diacyglecerol (DAG) for acrosome reaction induction. Once induced, acrosomal granules

break down by acrosin enzyme resulting in fusion of zona pellucida and exocytosis of acrosomal compartments (Florman *et al*. 2008; Gadella *et al*. 2008; Abou-haila and Tulsiani 2009).

a



b

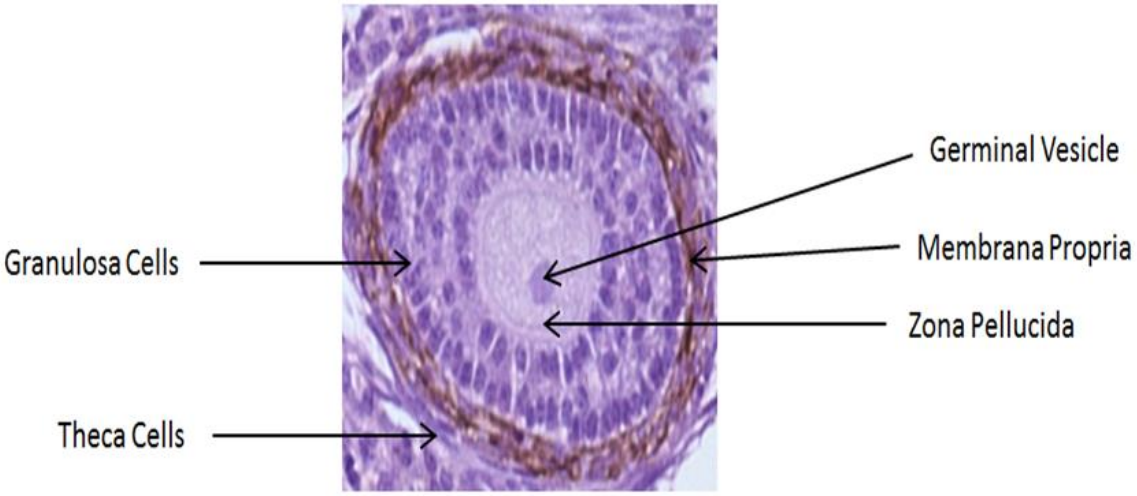


Figure 1.13 The maturation process of the human oocyte. a) The Maturation of the human oocyte Primordial is the first form of ovarian follicle which develops into primary and secondary form until it reaches either the preovulatory stage or undergoes atretic degeneration (corpus

luteum). Adapted from (http://classes.midlandstech.edu/). b) The structure of late primary follicle. In mammals have a defined structure, they consist of an oocyte which is surrounded by granulosa cells, membrane propria and a layer of theca cells which is located at the outer side of the follicle

1.13 Sperm-egg fusion

The penetration of the zona pellucida is followed by sperm oolemma binding and fusion. The sperm binds to the oolemma at the equatorial region where oolemma integrins interact with proteins of the ADAM family on the sperm. Also, it has been found that sperm fusion is regulated by CD9 protein which functions as a linker between sperm functional domain linker and glycosylphosphatidylinositol (GPI-) anchored proteins (Evans 2002; Kaji and Kudo 2004; Primakoff and Myles 2007).

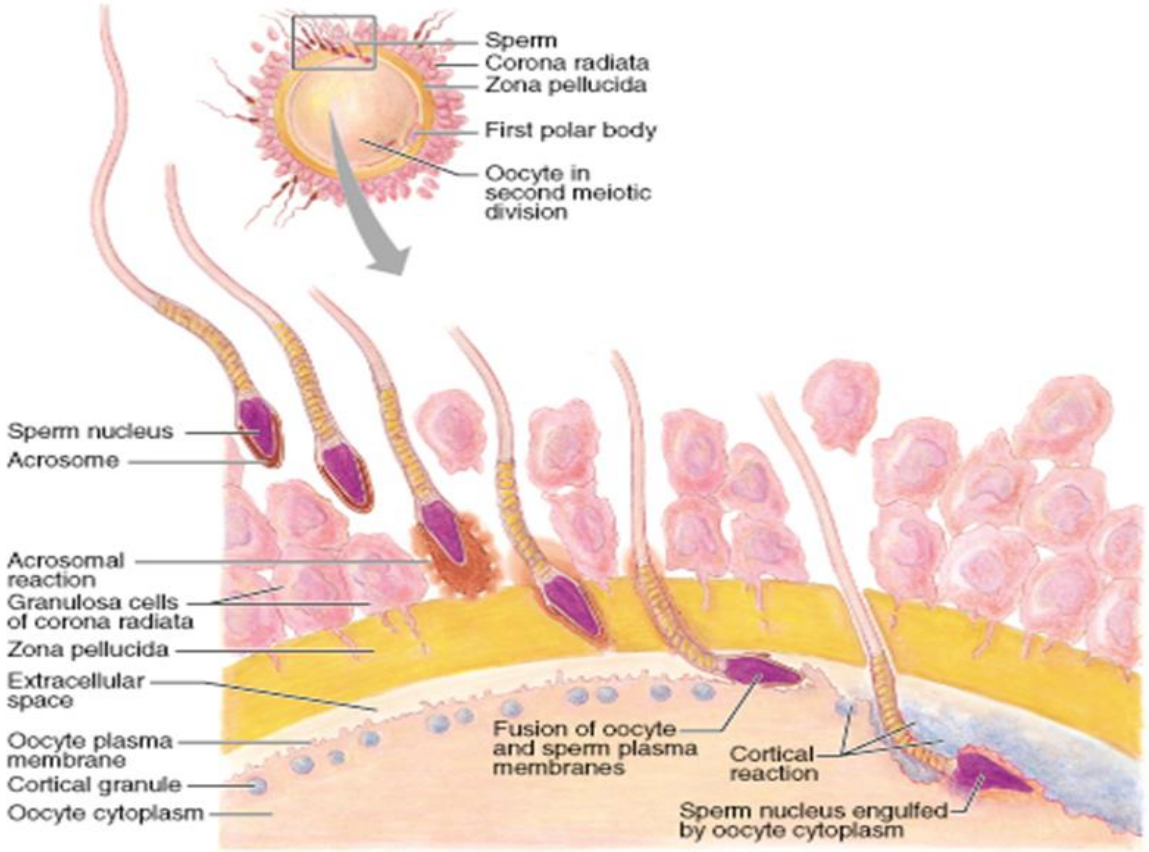


Figure 1.14 Acrosome reaction. Acrosome reaction allows the penetration of zona pellucida by the sperm and the equatorial binding and fusion of sperm to oolemma the completion of fertilization. From www.legacy.owensboro.kctcs.edu

1.14 Cortical reaction and embryogenesis activation

Sperm-oocyte fusion is followed by cortical reaction where enzymes released from the cortical granules of the oocyte cause rearrangement of the zona pellucida to avoid polyspermic fertilization. Once sperm fuses with the oocyte, there is a rapid increase in extracellular calcium which is followed by calcium oscillations. Recent studies have revealed that PLCζ, which is located in the post-acrosomal region, is responsible for the rapid calcium release. PLCζ regulates PIP_2 hydrolysis which results in calcium release regulated by IP_3 and initiation of embryonic development. Studies in mouse embryo have shown that PLCζ injection can initiate calcium oscillations and immunodepletion of it leads to inhibition of embryos development (Kashir *et al.* 2010). The naked sperm, having fused with the oocyte, rotates approximately 180 degrees and then forms the male pronucleus. The sperm midpiece and tail degrade and the embryo forms a zygote which is the beginning of embryonic development (Dale *et al*. 2010).

1.15 Therapeutic applications

Infertility affects approximately 15% of the couples and can be either due to male or female causes. *In vitro* fertilisation since the first successful birth in 1978 has become increasingly popular accounting for up to 7% of total births in some countries. Severe infertility affects approximately 19% of the infertile couples. In the case of severe male factor the standard approach in assisted reproduction is Intracytoplasmic sperm injection (ICSI) (Nomikos *et al*. 2013a). However, even the extremely successful ICSI does not solve the problem completely as about 5% of the cycles still fail (Ramadan *et al.* 2012). Lack of egg activation following ICSI treatment has been observed in a number of cases. Also in patients with mutated PLCζ and globozoospermic patients where PLCζ is absent, egg activation was abolished. Calcium ionophore has been used successfully to overcome this issue however, it causes a single Ca^{2+} transient rather than multiple oscillations as in fertilisation, failing to mimic physiological conditions. Also, such applications can be potentially toxic and teratogenic therefore PLCζ the physiological sperm factor has been suggested (Chapter 2 and 3) to be an effective potential treatment for such cases mimicking physiological conditions without affecting embryonic development (Nomikos *et al*. 2013a).

1.16 The role of PLCβ1 in health and disease

Phospholipases C have a key role in a range of biological processes, above it was described the role of PLCζ in infertility. In this project another PLC, the PLCβ1 was investigated for its role in thyroid cancer.

1.16.1 The structure of PLCβ1

PLCβ is an important phospholipase as there is a number of studies indicating that it has a role in a wide range of biological processes such as cell growth, differentiation and metabolism (Suh *et al*. 2008). PLCβ is a member of the PLC family consisting of 4 isozymes. It is activated by G - protein coupled receptor (GPCR) at the plasma membrane. It has a conserved structure comprised of a PH domain which is considered to be involved in the plasma membrane localization of the protein and its activation as it interacts in a cooperative manner with phosphatidylinositol 3 phosphate and the Gβγ subunit (Razzini *et al.* 2000; Wang *et al.* 2000). The series of EF hands, which function as a scaffold, regulate the initiation of GTP hydrolysis following the binding of Gαq (Dowal *et al.* 2006; Lyon *et al.* 2013). The X and Y domains form the (TIM) α/β-barrel and are considered to be the most conserved domains among PLCs. Studies have indicated that Gβγ interacts with both PH and X and Y catalytic domains (Rhee 2001) (Figure 1.14). A C2 domain which contributes to inter-intra molecular binding sites (Lyon *et al*. 2013) and a C terminal extension, where Gaq binds in order to activate PLCβ1 and is also important for the intracellular localization of the protein (Rhee 2001; Fukami 2002).

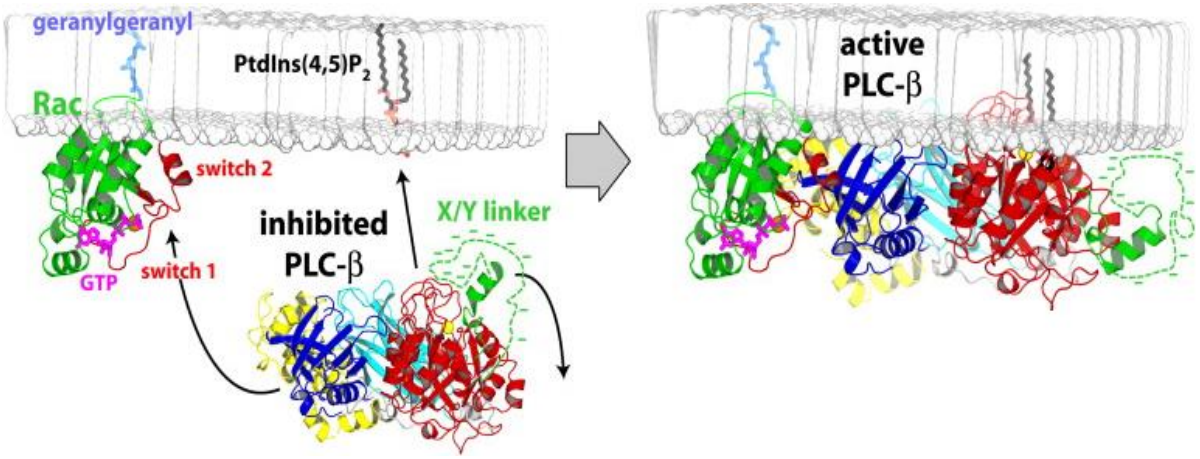


Figure 1.15 The binding of PLCβ1 to PIP2. In PLCβ the X-Y linker is negatively charged inhibiting the binding of the molecule to PIP_2, only a conformational change following its

activation by Rac allows the binding of PLC to the negatively charged membrane. Adapted from (Hicks *et al.* 2008).

1.16.2 The PLCβ1 isozymes

PLCβ isozymes differ from the other members of the PLC family as they contain an elongated C- terminus. This C terminal extension consists of about 400 amino acids and is highly conserved. The C terminal extension, is subdivided into three sections, the N-terminus, the 28-61 residue linker region and the elongated coiled coil domain of approximately 300 amino acids (Lyon *et al.* 2013). PLCβ1 is present in two forms as result of alternative splicing, PLCβ1a and PLCβ1b. The two splice variants PLCβ1a (150 kDa) and PLCβ1b (140 kDa), differ in a region of the C- terminus (Suh *et al.* 2008; O'Carroll *et al.* 2009). They share a 94% homology with PLCβ1a being 43 amino acids longer than the PLCβ1b. PLCβ1a has cytoplasmic localization containing a nuclear export sequence at its C terminal extension, in contrast with PLCβ1b which is mainly localized within the nucleus (Figure 1.15) (Piazzi *et al.* 2013). The last 3 amino acids of the C terminal region of PLCβ1a constitute a consensus post synaptic density protein/Drosophila disc large tumour suppressor/ zona occludens-1 protein motif. This motif is considered to have a key role in the localization of the enzyme. On the other hand the 10 amino acids at the C terminus of PLCβ1b contain a proline rich sequence (Grubb *et al.* 2008; Lyon *et al.* 2013).

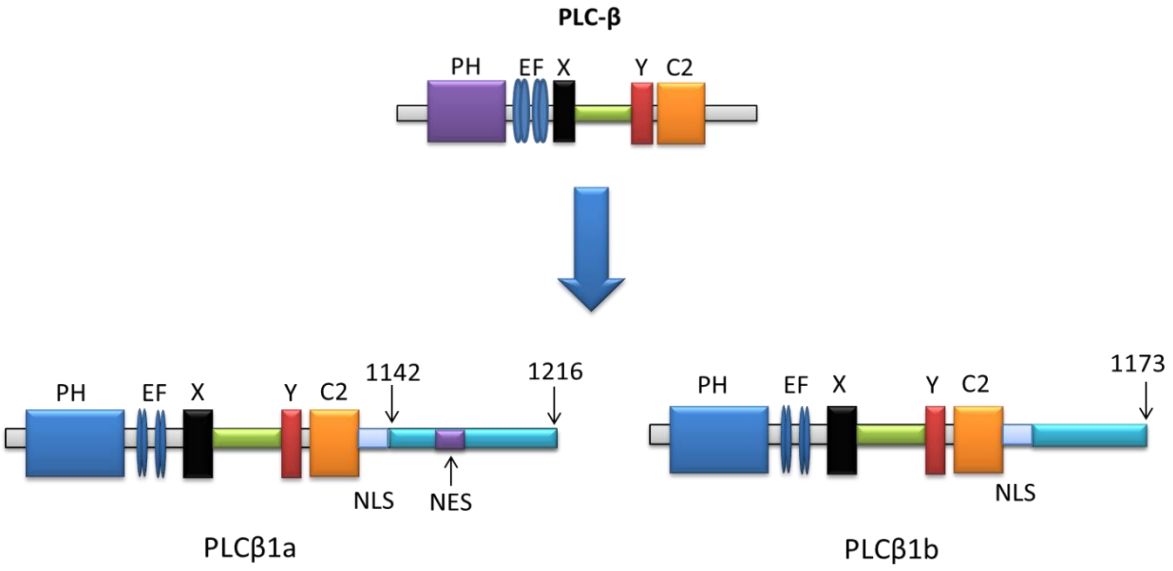


Figure 1.16 The PLCβ1 splice variants. PLCβ1 has two splice variants PLCβ1 α (150kDa) and β (140kDa). They differ at their C-terminal extension as PLCβ1α contains an NES.

1.16.3 PLCβ1 activation

PLCβ1 has multiple activation mechanisms. The cytoplasmic activation of PLCβ1 requires both Gaq and Gβγ subunits. In an inactive stage Gβγ is attached tightly to the GDP bound Gaq, keeping the molecule inactive (McCudden *et al.* 2005; Golebiewska and Scarlata 2010). Gβγ acts as guanine nucleotide dissociation inhibitor (GDI) and consequently it decreases the rate of GDP release significantly. Upon activation via an extracellular signal which binds to the G protein coupled receptor (GPCR), a conformational change occurs on the receptor which in turn initiates the exchange of GDP to GTP. The conformational change enables the interaction of Gaq with the receptor which dissociates from the Gβγ subunit. The dissociation of Gβγ exposes protein regions on Gaq which in turn binds to PLCβ1 and activates it. Gaq has an intrinsic GTPase activity, causing hydrolysis of GTP to GDP which in turn terminates the signal and re-association with Gβγ takes place returning the complex to its initial inactive state (Figure 1.16) (Dowal *et al.* 2006; Golebiewska and Scarlata 2010).

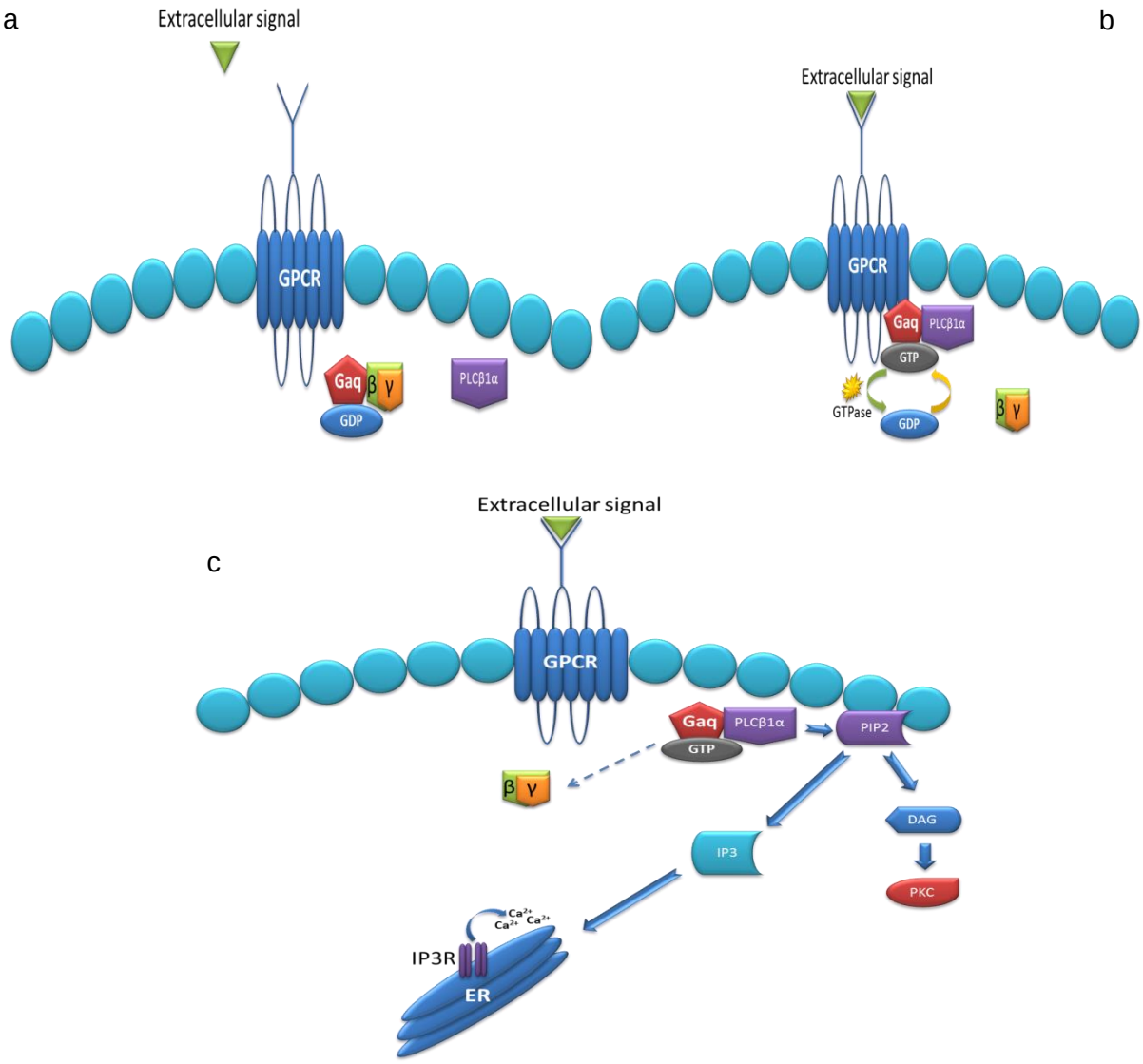


Figure 1.17 a) Activation of PLCβ1. At inactive state Gβγ is bound tightly to the GDP bound Gaq. b) The GPCR undergoes conformational change once an extracellular signal is attached and enables the binding of Gaq to the receptor and the exchange of GDP by GTP. The Gβγ dissociates from Gaq allowing the interaction of PLCβ1a leading to its activation. c) Activated PLCβ1a hydrolyses PIP_2 generating IP_3 and DAG.

1.16.4 Subcellular distribution mechanism

The mechanism of how the two splice variants are localized within the cell is still unclear. A recent study (Piazzi *et al*. 2013) reported that the variants interact with kpna and kpnb1 proteins, also known as importins. PLCβ1 splice variants as previously mentioned, contain a nuclear localisation sequence (NLS) at their C terminal extension. It has been suggested that

importin α (kpna) binds to this NLS sequence and promotes the binding of importin b (kpnb1) forming a complex which in turn is transferred into the nucleus. Nuclear importation occurs when importin b binds to the proteins of the pore complex leading to the dissociation of the complex and the insertion of the protein into the nucleus (Figure 1.17a). It is thought that both splice variants are imported however PLCβ1α, due to its nuclear export sequence (NES) which is present at its exclusive 75 amino acid region, is exported back to the cytoplasm. This also explains the fact that whilst both variants are found in the nucleus, the highest concentration of PLCβ1a and PLCβ1b is found in the cytoplasm and the nucleus respectively. The export of PLCβ1a is possibly operated by the binding of exportin 1 via its binding at the leucine rich region of the NES as observed with PLCδ1 (Figure 1.17b) (Okada *et al.* 2002).

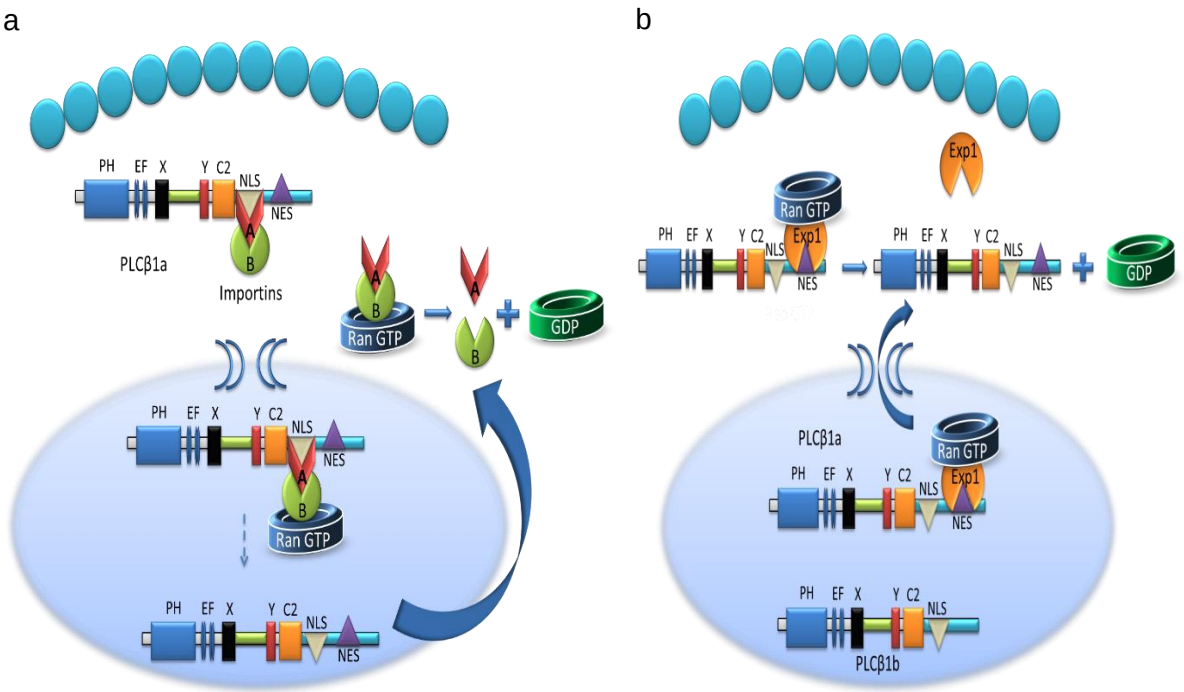


Figure 1.18 The nuclear import and export of PLCβ1. a) The import mechanism of PLCβ1, where importin a (A) binds to the nuclear localization sequence promoting the association of importin b (B) which in turn translocates the complex into the nucleus. The complex via a Ran dependent reaction, is dissociated with importins a and b being transferred back to the cytoplasm. b) The nuclear export of PLCβ1a is possibly operated by the exportin1 via its binding to the leucine rich region of the nuclear export sequence.

The two splice variants not only differ in their C terminal domain but also in their intracellular localization and mode of activation (Molinari *et al.* 2012). PLCβ1a has been found to be predominantly in the cytoplasm in contrast with PLCβ1b which is located in the nucleus and

more specifically at the nuclear speckles. This nuclear organelle is composed of clusters of interchromatin granules and is known to be rich with pre messenger RNA splicing factors. Little is known about the exact function of speckles, however a number of studies have indicated that transcriptionally active genes localise to the speckle area suggesting an active role in transcriptional pre mRNA regulation of the cell (Lamond and Spector 2003; Hall *et al.* 2006; Hu *et al.* 2009; Zhao *et al.* 2009; Tripathi *et al.* 2012). A cluster of lysines at their C-terminal have been shown to have a key role in the nuclear translocation of the protein as they provide interaction sites which retains PLCβ1 inside the nucleus (Piazzi *et al*. 2013). Also, another study indicated that the phosphorylation on S867 by PKC also regulates the localization of the protein as PLCβ1 remained within the nucleus due to the lack of phosphorylation. The subcellular localization mechanism of the PLCβ1 isoforms therefore includes translocation of both isoforms into the nucleus with the assistance of α/β importins with which they form complexes and are transported inside the nucleus. Then PLCβ1 interacts with negatively charged proteins and is retained within the nucleus, whereas PLCβ1a is exported back to the cytoplasm via its NES predicted by NetNES 1.1 (Piazzi *et al*. 2013).

1.16.5 The presence of PLCβ1 in human cancer

1.16.5.1 The role of PLCβ1 in Cancer

According to world health organization (WHO) and the International Agency for Research on cancer (IARC), cancer is the main cause of death globally with more than 7 million deaths for the year 2008 and more than 12 million new cases for the same year worldwide. In 2010, more than 1.5 million new cases were reported in the United States (Jemal *et al.* 2010). The molecular pathology of the disease is considerably complex and not fully yet understood. Carcinogenesis arises after the disruption or malfunction of the cell cycle control mechanisms, resulting in uncontrolled cell division, growth and death (Fresno Vara *et al.* 2004; Knowles *et al.* 2005; Gabriel 2007). The processes involved in carcinogenesis have long been extensively studied forming a complex network where signalling pathways and proteins control gene expression, cell growth and differentiation. A human phospholipase and more specifically PLCβ1 has been shown in many studies to be involved in such processes and to have a central role in various types of human cancer.

The gene for PLCβ1 is located at the short arm of chromosome 20 (Figure 1.18). Interestingly, chromosome 20 has been shown to be altered or to be involved in a variety of cancers, such as prostate, colon, breast, leukaemia and ovarian cancer (Friederichs *et al.* 2000; Kristiansen

et al. 2002; Follo *et al.* 2009; Molinari *et al*. 2012; Ribeiro *et al.* 2012; Jia *et al.* 2013). PLCβ1, is present in most somatic cells and being involved in a range of different cancers makes it a good candidate for the study of cell proliferation and differentiation, (Table 1.1).

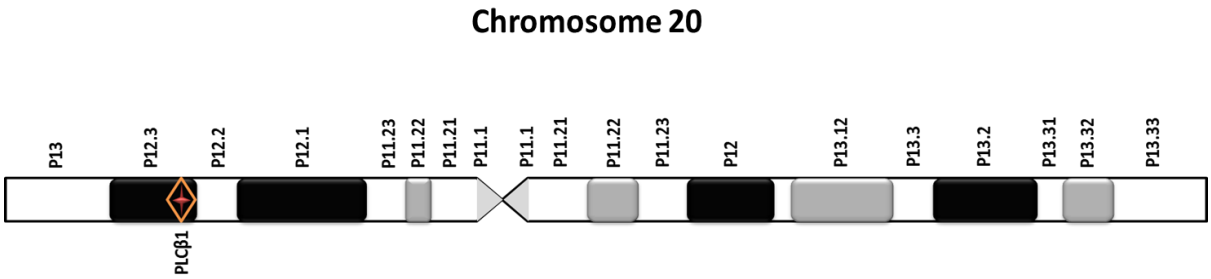


Figure 1.19 Chromosomal location of PLCβ1. PLCβ1 gene is located at the short arm of chromosome 20. Copy number variation at the long arm have been associated with the presence of different types of cancer such as breast, colorectal and prostate cancer (Friederichs *et al*. 2000; Kristiansen *et al*. 2002; Follo *et al*. 2009; Molinari *et al*. 2012; Ribeiro *et al*. 2012; Jia *et al*. 2013).

A recent study examining a number of patients with breast cancer identified that the majority had aneusomies and genetic amplifications in the PI-PLCβ1 locus (Molinari *et al*. 2012). In addition, PLCβ1 has been found to be overexpressed in colorectal cancer tissue as a specific SNP has been detected to be present in colorectal cancer patients (Gundem *et al.* 2010; Jia *et al*. 2013). It has also been found to be overexpressed in the periprostatic (PP) tissue of patients with prostate cancer (Ribeiro *et al*. 2012). Furthermore, PLCβ1 has been suggested to have a role in ovarian cancer (Kristiansen *et al*. 2002) and adenocarcinomas (Friederichs *et al*. 2000) due to its effect on CD24 which in turn promotes cell cycle progression (Fiume *et al.* 2005). Moreover, a mono-allelic cryptic deletion in PLCβ1 gene was identified to worsen the clinical profile of patients with myelodysplasmic (MDS) syndromes and increased the risk in developing acute myeloid leukaemia (Follo *et al*. 2009).

Table 1.1 List of cells expressing PLCβ1. PLCβ1 is expressed in most somatic cells of the human body including cells from the immune, nervous, muscle, internal, secretory and reproductive system.

	Microarray-BioGPS	RNAseq (Illumina Human Body Map)
Bone Marrow	+	
Whole Blood	+	
White Blood Cells		+
Lymph node	+	+
Thymus	+	
Brain	+	+
Cerebellum	+	
Retina	+	
Spinal Cord	+	
Heart	+	+
Smooth Muscle	+	
Skeletal Muscle	+	+
Small Intestine	+	
Colon	+	+
Adipocyte	+	+
Kidney	+	+
Liver	+	+
Lung	+	+
Pancreas	+	
Thyroid	+	+
Salivary Gland	+	
Adrenal Gland	+	
Breast		
Skin	+	
Ovary	+	+
Uterus	+	
Placenta	+	
Prostate	+	+
Testis	+	+

1.16.7 PLCβ1 in Thyroid Cancer

A family with high prevalence of multinodular goitre (MNG), progressing to papillary thyroid cancer (PTC) will be described in more detail in chapter 4. Previous studies conducted by Dr. Ameen Bakhsh using Genome Wide Linkage Analysis (GWLA) and non-parametric linkage analysis revealed a high LOD (3.01) score region on chromosome 20. Further investigation using techniques to assess copy number variation in the high-LOD score region indicated that affected family members are heterozygous for an insertion-deletion (InDel) in the PLCβ1 gene. In addition patients carrying the PLCβ1 InDel exhibited elevated PLCβ1 transcript levels in

their thyroids compared with subjects without the InDel. The high prevalence of patients progressing from MNG to PTC and the elevated transcript levels of thyroid PLCβ1 led to the hypothesis that the PLCβ1 inDel could be involved in benign thyroid proliferation.

1.16.8 Structure and Function of the Thyroid Gland

The thyroid gland is located in the neck forming two lobes connected by an isthmus. Its name is derived from the Greek ‘’θυρεός’’ which stands for shield and describes its shape. The thyroid gland is the largest among the endocrine organs responsible for the synthesis of three hormones, the 2 thyroid hormones and calcitonin. The synthesis of these hormones takes place in follicular cells and parafollicular C cells respectively. Follicular cells which is the dominant cell population constitute the epithelium and originate from the endoderm (Werner *et al.* 2005; Kondo *et al.* 2006).

1.16.8.1 Thyroid hormone synthesis

The thyroid gland is responsible for the synthesis of Thyroxine (T_4), Tri-iodothyronine (T_3) and Calcitonin. The synthesis of T4/T3 occurs in the lumen of follicular cells with Iodide being initially transported by the sodium iodide symporter (NIS) into follicular cells. Iodide which is a rate limiting substrate is inserted partially into the lumen via the activity of Pendrin (PDS) transporter in a process known as Iodide trapping. Iodide is then attaching to tyrosine molecules of the thyroglobulin. However, prior to this binding, Iodide undergoes oxidation in the apical lumen under the activity of thyroperoxidase (TPO). The oxidation reaction is operated in the presence of hydrogen peroxidase (H_2O_2) which in turn is generated by the enzymatic activity of dual oxidase (DUOX1/2). Thyroglobulin which is synthesized in the lumen and contains numerous tyrosine residues at its N and C termini, is iodinated following the attachment of oxidised iodide to the tyrosine molecules which contain iodide acceptors in excess to prevent free iodide accumulation. The binding is catalysed by TPO in a process known as organification. The binding of iodine molecule to tyrosine lead to the formation of Monoidotyrosine (MIT-T1) and Diiodotyrosine (DIT-T2). In physiological conditions more DIT is generated than MIT. The following step is known as the coupling, it is noteworthy that only 20% of the tyrosine molecules which have been iodinated go through the coupling process. The reaction is catalysed by TPO and results in in the formation of Tri-iodothyronine (T3) following the coupling of one MIT and one DIT and Tetra-iodothyronine or thyroxine (T4) with the coupling of two DIT (Goodman 2003; Werner *et al*. 2005) (Figure 1.19).

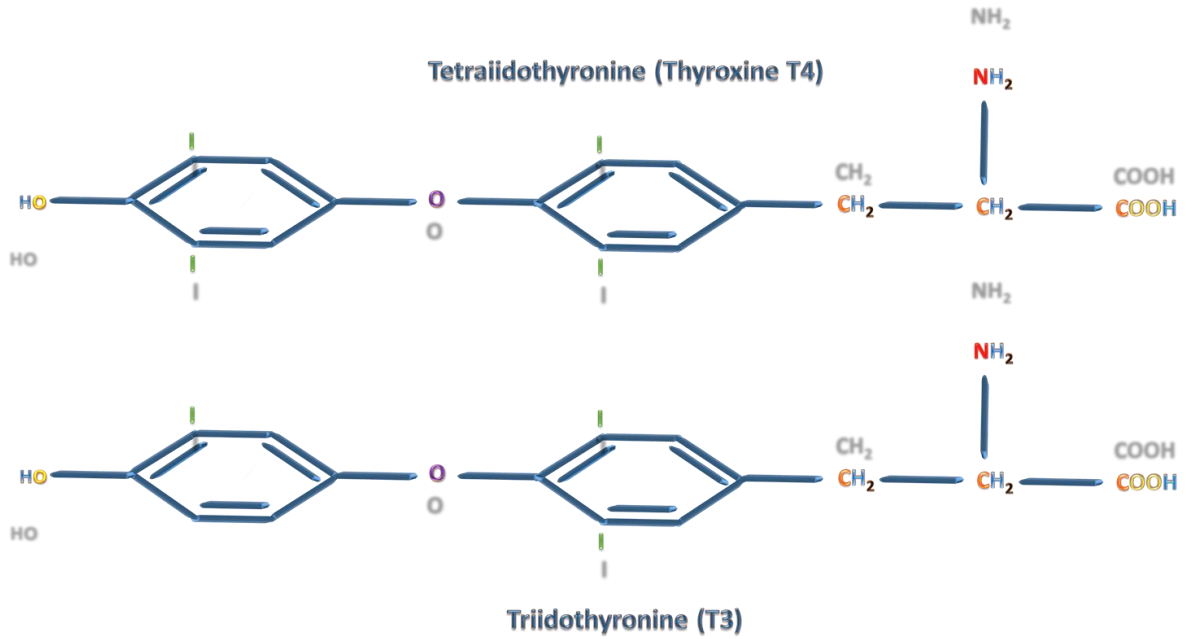


Figure 1.20 The chemical structure of the thyroid hormones. The thyroid hormones synthesized in the follicular cells of thyroid gland. T3 and T4 are the result of the coupling process where T1 and T2 are coupled T1+T2, T2+T2 to form T3 and T4 respectively.

Iodinated Tg is then fused with lysosome which results in the proteolytic degradation and release of T4, T3 and iodotyrosines (Wermer and Ingbars 2012). The degradation of iodinated Tg and subsequent release of T4 and T3 hormones lead to their release into the bloodstream via the transporter monocarboxylate 8 (MCT8). Thyroid hormones in the circulation are bound to thyroid hormone binding proteins (THBP) which maintain them in an inactive state and act as thyroid hormone storage for uninterrupted tissue supply (Bhatkar *et al.* 2004) (Figure 1.20). The small amount of iodotyrosines released MIT and DIT are deiodinated by dehalogenase enzyme (DEHAL1) which recycles iodide for thyroid hormone production (Werner and Ingbars 2012). The main thyroid hormone which is released is T4 with 80% of the resulting T3 to emerge from the deiodination of T4 by type I and type II deiodinase (D1), (D2) (Chi *et al.* 2013). D2 is considered to have greater impact than D1 and have an essential role in T3 homeostasis in conditions such as hypothyroidism. On the other hand deiodinase type III (D3) deactivates T3 and T4 generating T2 and inactive reverse triiodothyronine (rT3) respectively in conditions such as hyperthyroidism (Williams and Bassett 2011).

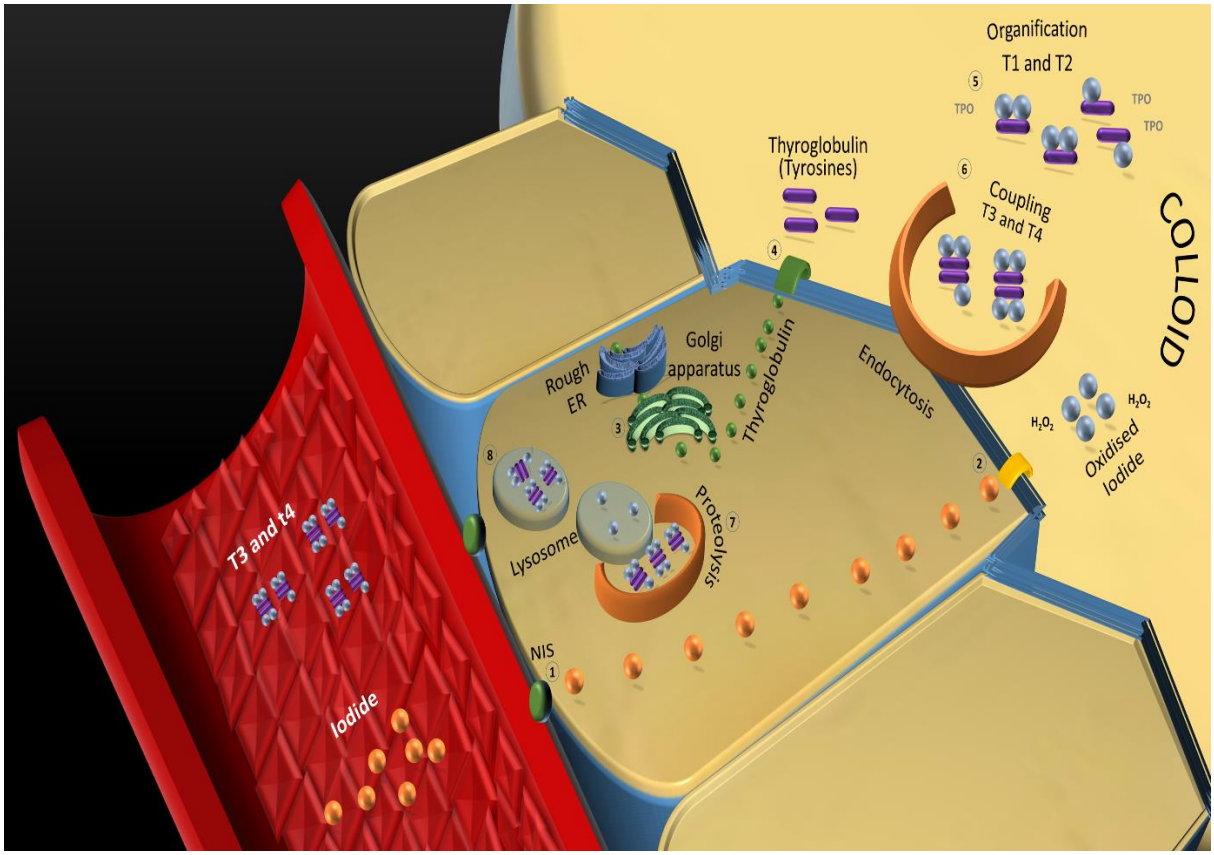


Figure 1.21 Thyroid hormone synthesis. 1) Iodide is taken up by follicular cells through NIS transporter and 2) is oxidised in the presence of H_2O_2.3) Thyroglobulin is synthesized by the Endoplasmic Reticulum, 4) contains tyrosine molecules with iodide acceptors and is 5) iodinated with oxidised iodide by TPO forming T1 and T2. 6) The coupling of T1 and T2 result in the formation of T3 and T4. The thyroid hormones are then inserted through pinocytosis into the cell. 7) The vesicle containing the Thyroid hormones is fused with lysosome where proteolysis takes place.8) Thyroid hormone is then released into the bloodstream.

1.16.9 Thyroid homeostasis

Thyroid stimulating hormone (TSH) is the essential regulatory protein of thyroid function. It is an approximately 30kDa glycoprotein secreted by the anterior pituitary gland. It is synthesized by the thyrotrophs and has a central role in maintaining the circulating levels of thyroid hormones, iodine uptake and thyroid growth (Szkudlinski *et al.* 2002). It is regulated by hypothalamic- thyrotropin- releasing hormone (TRH) and the level of circulating thyroid hormones by positive and negative feedback loops respectively (Ock *et al.* 2013). TSH acts via its binding to thyroid stimulating hormone receptor (TSHR), a 7 transmembrane G protein linked receptor leading to dissociation of G complex and the release of Gα and Gβγ subunits

(Szkudlinski *et al*. 2002; Davies *et al.* 2005; Werner *et al*. 2005). The dissociation of the G subunits activates cAMP and DAG pathways regulating thyroid gene expression, proliferation and thyroid hormone synthesis (Szkudlinski *et al*. 2002) (Figure 1.21).

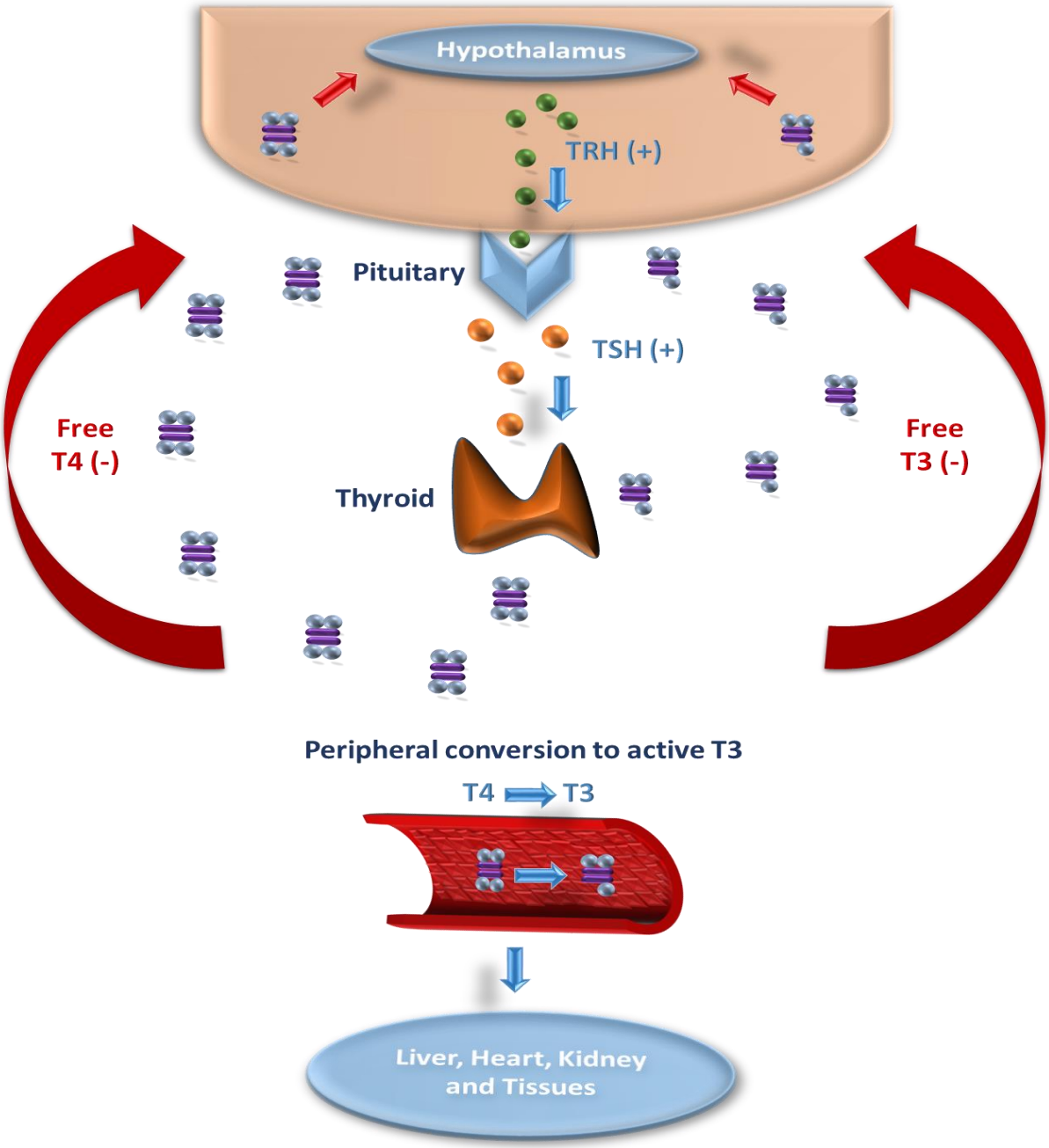


Figure 1.22 The euthyroid state. TRH is synthesized by the hypothalamus and triggers the secretion of TSH from the pituitary. TSH in turn following it's binding to TSHR in the thyroid follicular cells, initiates thyroid hormone synthesis. Thyroid hormone in the bloodstream is bound to thyroid hormone binding protein (THBP) remaining inactive. On the other hand the unbound –free thyroid hormone forms a negative feedback loop with TSH and TRH synthesis.

Circulating thyroid hormones (TH) bind with high affinity to thyroid receptors (TR) which have predominantly nuclear localisation. The thyroid receptors are in turn binding to thyroid response elements (TREs) in gene promoters, altering the transcriptional activity in a range of different genes involved in insulin signalling, cell proliferation, apoptosis and lipogenesis. In the absence of TH, basal transcription is repressed by TRs in a process known as transcriptional silencing (Sinha *et al.* 2012).

1.16.10 Thyroid Dysfunction

Dysfunctions of the thyroid gland are often caused by excess or limited levels of thyroid hormones known as hyper and hypo-thyroidism respectively. Thyroid dysfunction predominantly affects greatly women with a female: male ratio of about 4:1 (Truter 2011).

1.16.10.1 Hyperthyroidism

Hyperthyroidism also known as thyrotoxicosis, is excess circulating thyroid hormone levels due to hyperactivity of the thyroid gland. Subsequent symptoms include increased blood pressure, anxiety and weight loss. Hyperthyroidism is most commonly caused (up to 80%) by the autoimmune condition Graves' disease (GD). In (GD) thyroid stimulating antibodies (TsAb) are produced which bind the TSH receptor mimicking the activity of TSH and leading to an excess of thyroid hormone production (Amballi 2007) and goitre. Hyperthyroidism often occurs following an infection. Inflammation can occur after a viral infection resulting in excess thyroid hormone generation and temporary hyperthyroidism which can last even months. Also postpartum thyroiditis has been found to affect up to 10% of females where following pregnancy exhibit increased levels of thyroid antibodies (Truter 2011). Excess iodine intake from diet and some more rare tumours have also been associated with the cause of hyperthyroidism.

1.16.10. 2 Hypothyroidism

Hypothyroidism also known as myxoedema is caused by the down regulation of thyroid hormone production. This leads to an increase of body mass due to disruption of metabolism; hypothyroidism also results in cold intolerance and dull facial expressions. Hypothyroidism similarly to hyperthyroidism can be caused by thyroiditis and unbalanced iodine intake from

diet (Truter 2011). Hashimoto's syndrome is also a common cause of hypothyroidism, this autoimmune disease causes the apoptosis of thyroid cells due to the increased expression of cytokines which activate apoptotic and other cell death pathways (Hiromatsu *et al.* 2013). Congenital hypothyroidism which occurs in 1 in 2000 new-borns, is responsible for the poor development or function of the thyroid gland which consequently lead to hypothyroidism and can potentially impair the new-born's growth and promote mental retardation (Rastogi and LaFranchi 2010; Truter 2011).

1.16.11 Thyroid nodules and goitres

TSH has been also associated with thyroid growth and nodule formation. A prolonged increased TSH can lead to thyrocyte proliferation and the formation of diffuse goitre (Ock *et al*. 2013). Thyroid disease has a particularly increased prevalence in iodine deficient regions and can be classified as solitary nodular and multinodular thyroid disease. Nodules result from dysregulated growth of a single thyroid cell whereas diffuse goitre is often enlargement of all follicular cells. Nodules can be characterised histologically as true adenomas and adenomatous nodules which have encapsulated lesions and no lesions respectively. Morphologically they can be subdivided into cold nodules which illustrate decreased radioactive iodine uptake, normal and hot nodules which illustrate increased uptake. In the vast majority nodules are cold in a proportion of about 85%, 10% of nodules appear normal and only 5% hot. Hot nodules are usually caused by activating mutations in the TSHR or Gsq (Krohn *et al.* 2005).

Enlarged thyroid entities with normal thyroid function are known as euthyroid goitres and are classified into diffuse goitre, multinodular goitre (MNG) and solitary thyroid nodule (*Adriana* 2015). Multinodular goitre is a thyroid containing a large number of enlarged thyroid nodules. In iodine deficient areas multinodular goitre can be endemic and the enlarged thyroid gland is considered to be due to the lack of thyroid hormones which in turn cause the generation of excess TSH which stimulates thyroid growth. The thyroid gland in order to respond to such changes increases its volume. Multinodular goitres can often cause dyspnoea, dysphagia and orthopnoea and in about 10% can lead to malignancies (Nada *et al.* 2011). Initial hypotheses regarding the formation of multinodular goitres suggested that thyroid goes through 2 distinct phases, a phase of hyperplasia and a resting phase. Iodine deficiency with its consequent rise of TSH will result in a hyperplasia phase. The theory suggests that a potential decrease in requirement for thyroid hormone will eventually lead to a resting phase where storage of colloid takes place. A potential repetition of these two cycles has been suggested to cause the

formation of non- toxic multinodular goitre which could possibly explain the histological colloid goitre (Moulick 2008). Also, three main thyroid specific transcription factors are considered vital for thyroid development and function, NKX2-1, PAX-8 and FOXE-1. NKX2-1 has a key role in C-cells and thyroid follicular cells proliferation and survival, PAX-8 in proliferation of thyroid follicular cells and FOXE-1 regulates the differentiation of the thyroid. Studies in mice have revealed that these transcription factors are essential for thyroid development as NKX-2 and FOXE1 null mice have illustrated thyroid agenesis and PAX-8 null mice were born with severe hypothyroidism (Kimura 2011). Another important regulator of thyrocyte proliferation and nodule formation is insulin-like growth factor 1 (IGF-1). IGFs have been found to be involved in a range of different diseases such as breast cancer. A number of studies have shown that overexpression of IGF-1 is associated with increased thyroid volume and supressed TSH, indicating the interaction between IGF-1 and TSH (Ock *et al*. 2013). Also, investigation of IGF-1 levels on follicular adenoma, goitre and papillary thyroid cancer patients has revealed that both IGF-1 and IGF-1R had a significantly increased transcript levels compared to healthy patients (Liu *et al.* 2013).

1.16.11.1 Risk factors

Iodine deficiency is considered the major factor for nodular thyroid disease causing endemic goitre. Studies have shown that the prevalence of endemic goitre in areas with iodine deficiency was approximately 30% in males and 40% in females (Knudsen *et al.* 2002). In addition to this another important factor appears to be age as the prevalence increases up to 3 fold in older populations. Smoking is considered another risk factor which can potentially decrease iodine availability and organification due to the elevated levels of thiocyanate (Knudsen *et al*. 2002; Krohn *et al*. 2005). As with all thyroid conditions, there is a higher prevalence of diffuse goitre in women compared to men (ratio 4:1) and a role for oestrogens in goitre formation has been proposed (Manole *et al.* 2001; Vanderpump 2011).

The prevalence of thyroid malignancy has been estimated to vary according to different area from 0.9% to 13%. In the past MNG was considered as a relatively low risk factor for thyroid malignancy however, recent studies have estimated a prevalence ranging from 7 to 17%. The aetiology for MNG formation is not well established yet, with various factors such as iodine deficiency, decrease in thyroid hormone synthesis and reduction of circulating iodide due to increased iodine clearance from the kidneys are some potential reasons responsible for the disease (Hanumanthappa 2012). The number of thyroid cancer cases has been increased up to 5 fold in the last 6 decades constituting approximately 2% of all cancer. (Hanumanthappa 2012; Cossu *et al.* 2013; Xing 2013).

1.16.12 Thyroid cancer

Thyroid cancer has a varied cellular origin and histological classification. The majority of thyroid cancers derive from follicular cells such as papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), poorly differentiated thyroid cancer (PDTC) and anaplastic thyroid cancer (ATC). On the other hand, medullary thyroid cancer (MTC) originates from parafollicular C cells and has a relatively low prevalence (Xing 2013). PTC accounts for 80% of thyroid cancers with a survival rate of approximately 90% (Brown *et al.* 2011; Tufano *et al.* 2012). Papillary thyroid cancer along with follicular thyroid cancer (FTC) the second most frequent type of thyroid cancer with prevalence up to 11% are often referred together as differentiated thyroid cancer (DTC) (Schneider and Chen 2013). The least aggressive and lethal thyroid cancer is considered to be the DTC with a 90% 10 year survival rate and excellent prognosis (Table 1.2). On the other hand, anaplastic thyroid cancer (ATC) with prevalence up to 2% is one of the most dangerous lethal and rapidly developing cancers. ATC is thought to be derived either from de-differentiated DTC cells or to arise *de novo* (Brown *et al*. 2011; Schneider and Chen 2013).

Table 1.2 Different types of thyroid cancer: their prevalence, patient target group and lethality. Adapted from (Kondo *et al*. 2006).

Tumour type	Prevalence	Sex ratio (female: male)	Age (years)	Lymph-node metastasis	Distant metastasis	Survival rate (5 year)	References
Papillary thyroid carcinoma	85–90%	2:1–4:1	20–50	<50%	5–7%	>90%	1,4,7
Follicular thyroid carcinoma	<10%	2:1–3:1	40–60	<5%	20%	>90%	1,4,7
Poorly differentiated thyroid carcinoma	rare–7%	0.4:1–2.1:1	50–60	30–80%	30–80%	50%	1,4,9,10
Undifferentiated thyroid carcinoma	2%	1.5:1	60–80	40%	20–50%	1–17%	1,4,8
Medullary thyroid carcinoma	3%	1:1–1.2:1	30–60	50%	15%	80%	1,4
Mixed medullary and follicular-cell carcinoma	rare						4

1.16.13 Aetiology

1.16.13.1 Sporadic cancer

Aetiology of thyroid cancer can be either environmental and/or genetic causes. Exposure of the head and neck region of the body to radiation in the past for the treatment of tonsillitis and acne and nowadays for treatment of Hodgkin disease can lead to potential PTC occurrence. Also following the Chernobyl nuclear accident an increased prevalence of thyroid cancer had been recorded in children aged 12 or less at the time of the I^{131} release. Diet is also another important factor as papillary thyroid cancer and follicular thyroid cancer can be caused in areas with high iodine content and iodine deficiency respectively. Consequently, the level of TSH has a key role in thyroid malignancy as increased TSH has been linked with increased prevalence of DTC (Schneider and Chen 2013).

1.16.13.2 Familial cancer

Genetic causes are also considered to have a key role in DTC occurrence. Familial non-medullary thyroid cancer accounts for the approximately 9% of the DTC cases. Linkage analysis in different families has identified three different loci MNG1 (Neumann *et al.* 1999), MNG2 (Capon *et al.* 2000) and MNG3 (Takahashi *et al.* 2001) associated with recurrent euthyroid goitre, an X- linked loci with increased prevalence of MNG in females and euthyroid MNG with elevated serum TSH respectively. Also a recent study has revealed another form of familial MNG exhibiting high risk of PTC progression (Bakhsh *et al.* 2006). Thyroid cancer can be inherited in patients suffering from Cowden, Gardner and Werner syndromes. Also a number of molecular biomarkers have been identified to be directly related with DTC pathogenesis and development.

1.16.13.3 Proto-oncogenes and growth factors

Rearranged during transfection (Ret) proto-oncogene is expressed solely in parafollicular C cells. Although in some papillary carcinomas receptor tyroisine kinase (RTK) was found to be expressed which is encoded by ret. Chromosomal rearrangements lead to activation of ret gene in these carcinomas. The ret rearrangements produce a fusion of the tyrosine kinase domain of ret with gene regions expressed in thyroid follicular cells resulting in expression ret/PTCs proteins. The consequent ret/PTCs are in turn causing ligand independent

expression of ret tyrosine kinase activity in thyroid follicular cells leading to PTC. Also, it is noteworthy that 70% of the cancers caused by the Chernobyl nuclear accident found to carry the ret/PTC rearrangement (Moretti *et al.* 2009; Schneider and Chen 2013).

V-Raf murine sarcoma viral oncogene homolog B (BRAF) mutation is another molecular biomarker found in approximately 45% of the PTC patients. T1799A transversion mutation is the most common BRAF mutation accounting for 80% of the mutations in BRAF. The T1799A leads to an amino acid substitution V600E and in turn MAP kinase /ERK pathway activation. The amino acid substitution is believed to disrupt the ATP binding P loop responsible for the inactivity of BRAF, leading to a catalytically active form. BRAF is the most potent among the Raf kinases and overactivation of the MAP kinase/ERK pathway is known to be responsible for cell proliferation and tumour formation (Xing *et al.* 2005; Schneider and Chen 2013).

Chromosomal translocation between PAX8 and PPARγ have been found to affect 36% of FTC, 11% of FA and 13% of follicular variant of PTC. PAX8 is a thyroid specific transcription factor, vital for follicular cells genesis and has been reported to be present in benign thyroid neoplasias (Moretti *et al*. 2009). PPARγ it is a nuclear hormone receptor and transcription factor with tumour suppressor activity. The fusion of these two genes leads to the formation of PPFP protein. Studies have suggested that expression of PPFP interrupts PAX8 activity leading to the downregulation of thyroid genes such as Thyroglobulin (Tg), TPO and NIS. In addition, PPFP has been suggested to have also oncogenic activity as it was found to abolish PPARγ activity leading to cell proliferation and decrease of apoptotic cells (Placzkowski *et al.* 2008; Braunstein 2012).

FOXE-1 is forkhead transcription factor, with nuclear localisation and expression and as mentioned above it is an essential thyroid transcription factor for thyroid function. It regulates thyroid differentiation and gene expression of thyroglobulin (TG) and thyroperoxidase (TPO). Studies have illustrated that SNPs within and near FOXE-1 are associated with increased susceptibility to thyroid cancer. Studies have shown that SNPs near the FOXE-1 gene are linked with increased risk up to 7 fold of developing thyroid cancer (Gudmundsson *et al.* 2009). Also, SNPs within the 11-22 polyalanine tract in the DNA-binding domain of the gene have confirmed the association of that domain with PTC (Bullock *et al.* 2012).

1.17Thesis Aims

Phospholipases C enzymes have been investigated for their role in health and diseases. The first two chapters of this thesis were dedicated to PLCζ and its potential role in assisted reproduction. The aim of Chapter 2 was the design of 24 different plasmid constructs of recombinant human PLCζ for optimisation of protein expression in a range of different bacterial cells. In Chapter 3, recombinant human PLCζ was purified with the use of affinity chromatography and microinjected into mouse eggs for *in-vivo* activity assessment.

In Chapter 4, 5 and 6 the role of PLCβ1 InDel in patients suffering from MNG who are likely to develop PTC was investigated. In chapter 4, a qPCR based genotyping tool was designed for estimation of the prevalence of the InDel in different group of patients. Also the presence of an active ERα promoter within the InDel using a luciferase reporter assay was investigated. In Chapter 5, we investigated the effect of altered PLCβ1 transcript levels on thyrocyte proliferation. In Chapter 6, the region of Chr.20 with high LOD score was investigated by next generation sequencing for the potential presence of SNPs that may be responsible for the disease.

Chapter 2. Expression of Recombinant hPLCζ for Biomedical Application

2.1 Introduction

2.1.1 Biomedical applications

hPLCζ is a sperm specific protein with the unique function of activating the oocyte following sperm - egg fusion (Saunders *et al*. 2002). The egg activation occurs due to the PLCζ ability to trigger Ca^{2+} release from the intracellular calcium stores in the form of Ca^{2+} oscillations. hPLCζ is also considered to be linked with certain forms of male infertility (Nomikos *et al.* 2011a). Experiments using mutant hPLCζ (H398P) identified in a patient who failed intracytoplasmic sperm injection (ICSI), revealed that the mutant sperm was unable to hydrolyse PIP2 and activate the egg. This signifies the central role of PLCζ in fertility and its potential biomedical therapeutic application in similar clinical cases (Nomikos *et al*. 2011a). Another potential biomedical application of hPLCζ could be in globozoospermia which is estimated to affect approximately 0.1% of the population, and it is characterised by the rounded head of the spermatozoa (Dam *et al.* 2007). Studies have shown that round-headed spermatozoa that have low levels or absence of hPLCζ, resulted in failure of egg activation following ICSI. Current applications such as calcium ionophore A23187 are not causing physiological oscillations as in natural fertilisation as they trigger a single Ca^{2+} spike instead of multiple. The role of PLCζ in the fertilization process as sperm factor that mimics physiological fertilisation can have a great potential in therapeutic application in the future (Yoon *et al.* 2008).

2.1.2 Bacterial Expression system

The bacterial expression system is currently the most popular protein expression system, because it is inexpensive and can produce particularly high yields of recombinant protein. However, eukaryotic proteins expressed in bacterial cells are often insoluble or non- functional due to the lack of post translational modification machinery. It can often jeopardise protein solubility producing inclusion bodies, promoting disulphide bond formation, incorrect protein folding and mismatch of the use of codons between the bacterial cell and the eukaryotic protein (Costa *et al.* 2014). The most common problem encountered is the codon usage between the host cell and expressed eukaryotic protein. The lack of necessary number of tRNA pools can lead to serious problems such as premature termination of translation and amino acid frameshift (Terpe 2006; Makino *et al.* 2011). Therefore, genetically modified *E.Coli* strains are often used to increase protein expression and solubility allowing co expression of tRNAs for rare codons (Table 2.1).

Table 2.1 Different bacterial strains that are commonly used to enhance expression and solubility of eukaryotic proteins.

Bacterial Strains	**Key Features**
Rosetta De3	Designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli.*
BL21 DE3	This strain is deficient of lon and omp-t proteases and is therefore suitable for expression of non-toxic genes.
Rosetta pLysS	In Rosetta (DE3) pLysS the rare tRNA genes are present on the same plasmids that carry the T7 lysozyme.
BL21 pLysS	The pLysS plasmid produces T7 lysozyme to reduce basal level expression of the gene of interest. Thus it is suitable for expression of toxic genes.
Rosetta gami	Rosetta-gami host strains are Origami derivatives that combine the enhanced disulphide bond formation with enhanced expression of eukaryotic proteins that contain codons rarely used in *E. coli.*

2.1.3 Protein solubility tags

The use of recombinant proteins with the incorporation of peptide fusion partners for protein purification is becoming increasingly popular. Such partners known as affinity tags can be either small peptide tags or large peptides and proteins. The choice of the ideal protein solubility tag is dependent on the type of the protein of interest and the application that the purified protein is destined for (Terpe 2003).

The small peptides are a popular choice of affinity tags as they may not be immunogenic and in some biomedical applications, removal of the tag may not be necessary. On the other hand, large peptides or proteins offer the advantage of increasing the solubility of the recombinant protein. Although, removal of the solubility partner often is required for number of applications such as crystallography or biomedical applications (Terpe 2003).

The most commonly used solubility tags in *E.Coli* expression systems are the Maltose binding protein (MBP) and the Glutathione-S-Transferase (GST). They offer the unique advantage of

binding to amylose and glutathione resin respectively, providing an extra purification choice. MBP is considered as the most efficient solubility tag achieving high solubility in recombinant proteins when fused at the N terminal of the protein of interest. MBP promotes protein solubility as it exhibits intrinsic chaperone activity enhancing the correct protein folding of the recombinant protein (Costa *et al*. 2014). Thioredoxin (Trx) and NusA even though they don't offer a particular purification option, have been shown to enhance solubility (Esposito and Chatterjee 2006). NusA increases protein solubility as it delays translation at transcriptional pauses allowing the correct folding of the target protein. Similarly Trx prevents inclusion bodies formation due to its oxidoreductase activity which in turn reduces disulphide bond formation (Costa *et al*. 2014).

2.1.4 Human Thioredoxin as a solubility tag

Thioredoxin is classified as a redox active protein with a size of 12kDa. The protein was initially discovered (Laurent *et al.* 1964) in *E.Coli* and it was identified to be important cofactor of ribonucleotide reductase which in turn is known to form the deoxyribonucleotide precursors of DNA (Holmgren 1985). Although, further studies have revealed that Thioredoxin protein has an important role in a wide range of biological phenomena such as cell growth (Gasdaska *et al.* 1995), apoptosis (Matsui *et al.* 1996), immune response (Rosen *et al.* 1995), virology, cancer cellular responses to oxidative stress and many others (Holmgren 1985). Thioredoxin is also present naturally in the oocyte (Lopata *et al.* 2001). Recent studies have shown that overexpression of Thioredoxin can potentially promote foetal growth (Umekawa *et al.* 2008). Therefore , this protein which is naturally present in the oocyte environment and the same time can serve as a solubility tag, was chosen to be sub-cloned along with hPLCζ and retain its solubility when the other bacterial solubility partners have been removed. The use of human Thioredoxin as solubility tag has given a great potential for hPLCζ in future biomedical applications due to its human origin and its solubility retaining abilities (Figure 2.1).

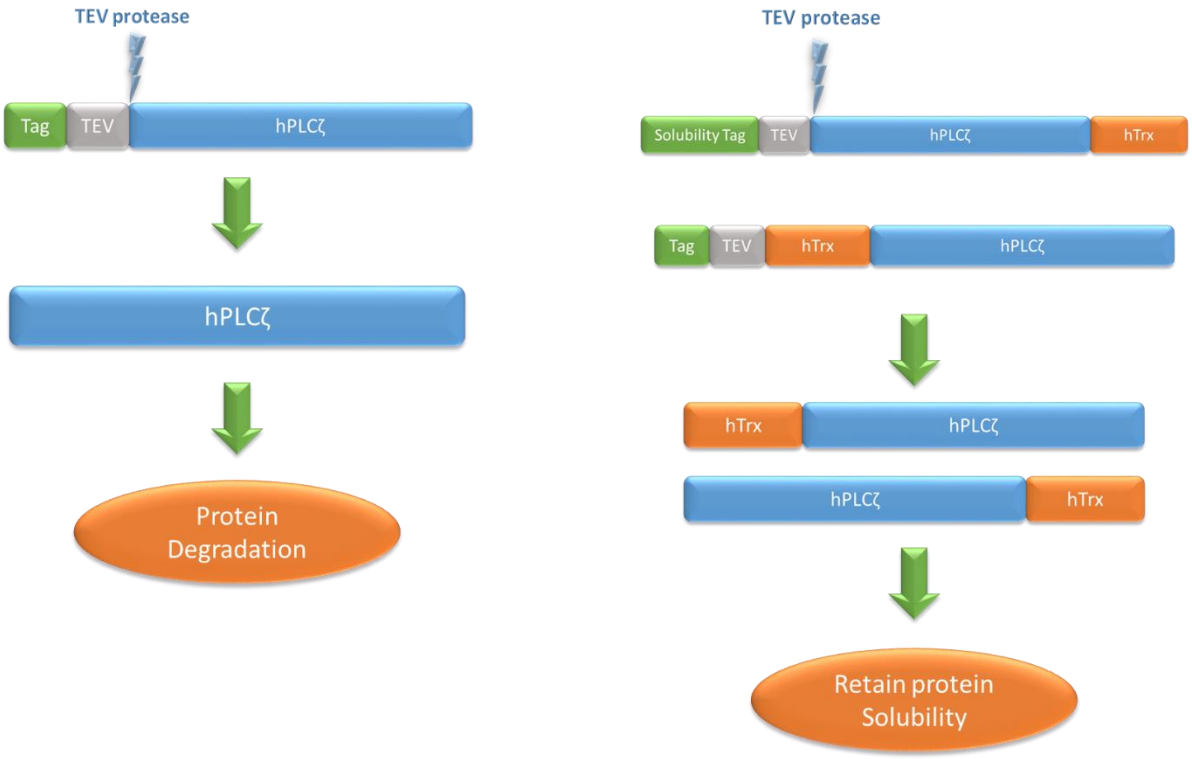


Figure 2.1 The principle of using hTrx retaining protein solubility following removal of bacterial solubility tags.

2.1.5 Bacterial lysis

Bacterial lysis is vital for the recovery of the protein of interest in soluble form. In the current study bacteria have been lysed by either Ultrasonication or High pressure homogenisation. In both cases lysis was enhanced by enzymatic lysis with the use of lysosome. Ultrasonication is the current gold standard method for bacterial lysis, it uses sonic waves of high frequency (>16 kHz) to disrupt the cell walls due to the thousands of atmosphere pressure caused by the shock waves of the sonication process. On the other hand, the high pressure homogenizer uses high pressure for passing the bacterial culture through a narrow orifice valve followed by an instant release into a nozzle outlet. The fluctuation of the high pressure in the orifice valve and the instant release into the nozzle outlet result in cell wall disruption (Ho 2008). The choice of cell lysis method was dependent on the volume of the bacterial culture as Ultrasonication could lyse cultures up to 100ml where the same time high pressure homogenization could lyse bacterial cultures up to 2L. In addition, hPLCζ is a very sensitive protein to temperature, sonication often generates heat thus it could potentially contribute to the protein degradation and low yield.

2.1.6 Aims of the project

The aim of this project was the identification of the optimum bacterial cell line and PLCζ protein construct for successful expression and solubility. To optimize protein expression and solubility hPLCζ was sub- cloned into the pETMM series of plasmids which contain various protein solubility tags (Table 2.2). Protein expression and solubility was assessed using a range of different *E. Coli* strains for the identification of the most competent protein construct. The identified optimal protein construct was used in Chapter 3 for purification and *in-vivo* enzymatic activity assessment (Figure 2.2).

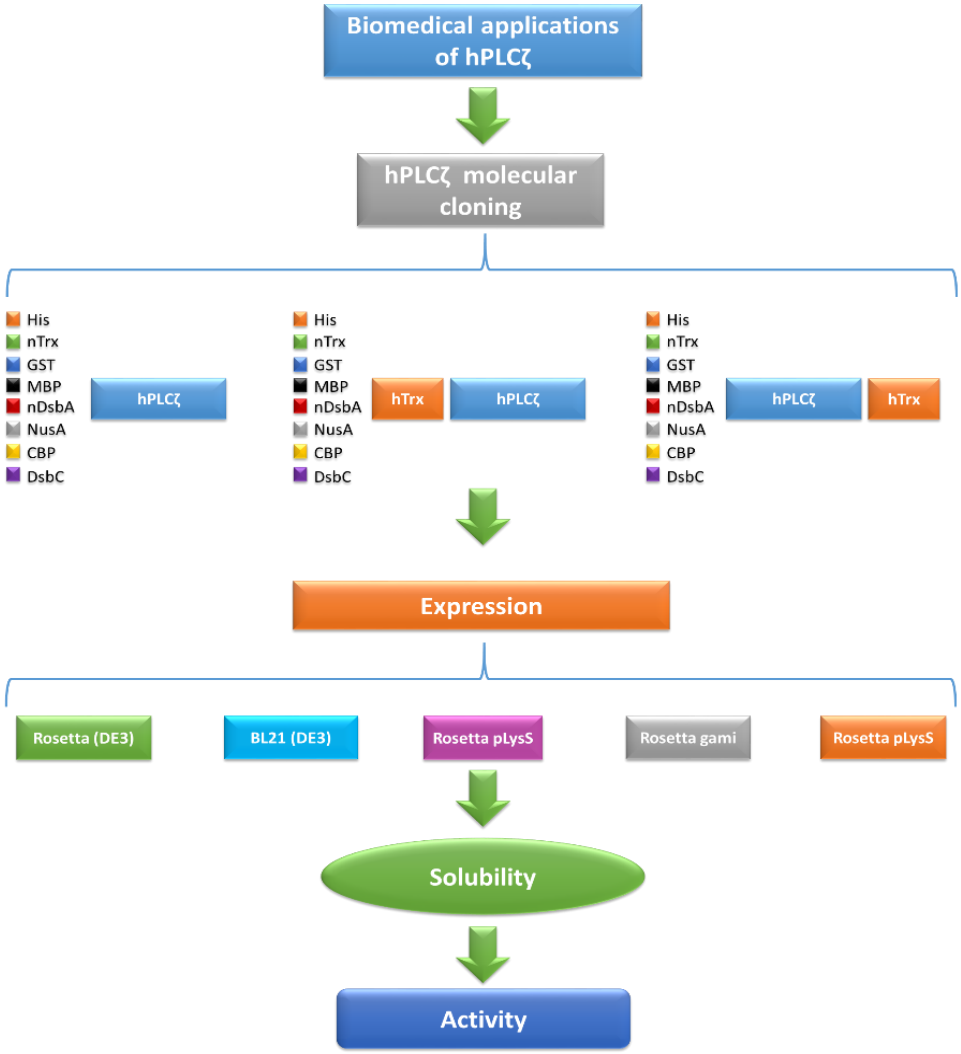


Figure 2.2 Aims of the chapter. The selection for the optimum condition and protein construct of hPLCζ. The first strategy included the sub cloning of hPLCζ into the pETMM series vectors which in turn included a range of different solubility tags to retain protein solubility and

providing different purification options. The second strategy included that subcloning of hTrx as a solubility tag of human origin at the N-terminal of hPLCζ and the cloning of the hTrx-hPLCζ into the pETMM series vectors as previously described. In the third strategy hTrx was cloned at the C terminus of hPLCζ following its cloning into the pETMM series vectors. The plasmid constructs were then assessed for protein expression and solubility with the use of 5 different bacterial strains. Activity was assessed following protein microinjection into mouse oocytes.

2.2 Materials and Methods

2.2.1 Cloning strategy

Three different cloning strategies were used to improve protein expression and solubility and then compared to determine the optimum conditions for in vivo assays and later biomedical applications of hPLCζ (Figure 2.3).

In the first strategy hPLCζ was cloned into the pETMM series of vectors which contained a range of different protein solubility tags that not only increase both protein expression and solubility but also provide additional options for protein purification.

The second strategy was to clone hTrx, a protein which is naturally present in the embryonic environment into the pETMM-hPLCζ vectors to improve and potentially retain protein solubility following cleavage of the bacterial solubility partners.

The third strategy included the cloning of the previously mentioned hTrx into the C-Terminus of the hPLCζ in pETMM-hPLCζ vectors.

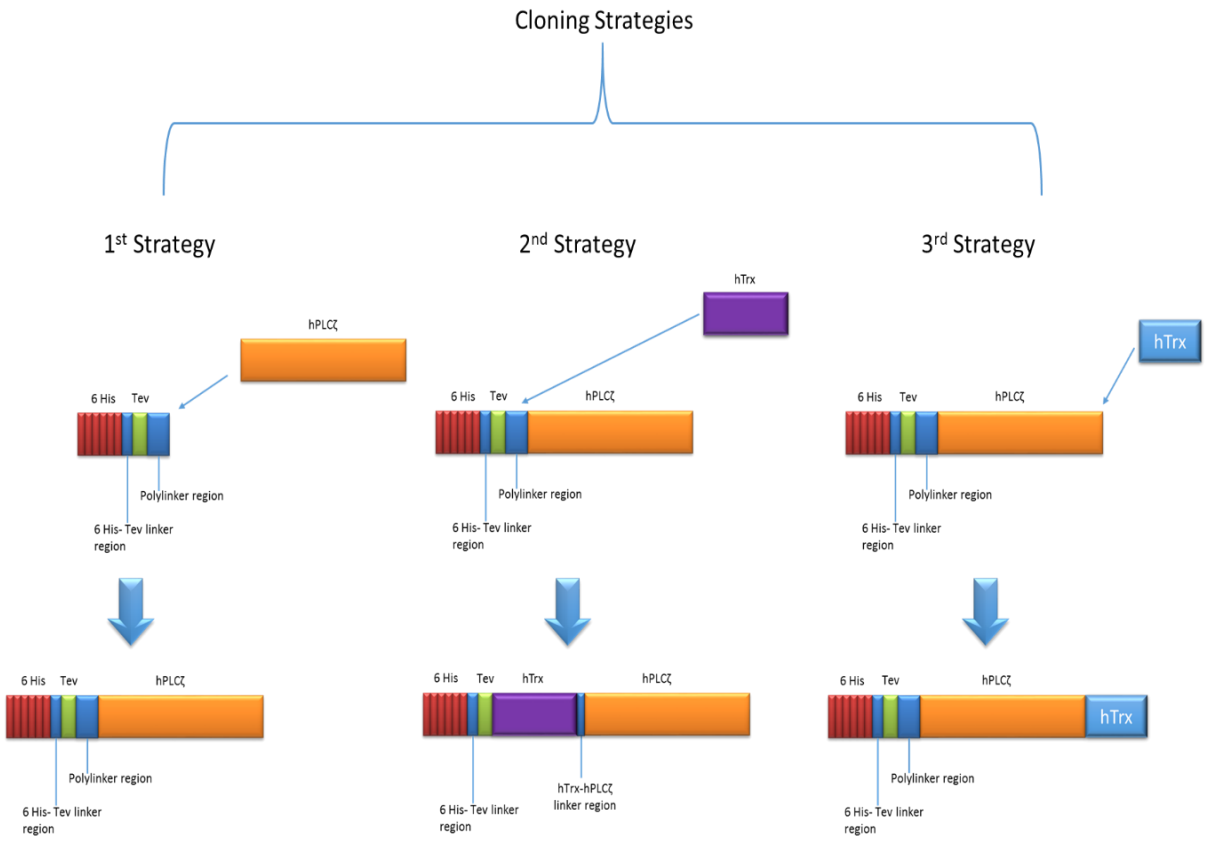


Figure 2.3 The cloning strategies included in this project. 1) Cloning of hPLCζ into the pETMM series vectors. 2) Cloning of hTrx at the hPLCζ pETMM series vectors. 3) Cloning of hTrx at the C terminus of hPLCζ in hPLCζ pETMM series vectors.

2.2.2 Molecular sub-cloning

2.2.2.1 PetMM series with hPLCζ

A pair of primers was designed using Primer 3 online analysis tool (http://bioinfo.ut.ee/primer3/), including a SalI (5' G/TCGAC 3') and NotI (5' GC/GGCCGC 3') restriction enzyme recognition site (Table 2.2). The restriction enzymes have been selected following screening for naturally occurring restriction sites within the gene with the use of NEB cutter software (http://nc2.neb.com/NEBcutter2/). The designed primers were also screened by BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure that there are no other unrelated significant homology sequences. Target DNA (coding sequence of hPLCζ) was amplified from the original cDNA clone (GenBank accession number AF435950) using phusion DNA proofreading polymerase (New England Biolabs). Amplification was performed with the addition of the following reagents into a 500 µl reaction tube: 2.5µl of 2x phusion master mix (NEB), 0.5µM forward primer, 0.5 µM reverse primer, 1µl of template DNA and nuclease-free water (Fisher Scientific) up to 50µl total reaction volume (Table 2.3). Amplification took place at PE Applied Biosystems 9700 thermocycler.

Table 2.2 Primers designed for hPLCζ amplification

Primer Name	**Tm**	**Sequence**	**Tm**
F(hPLCζ)/*R(hPLCζ)*	60	CCTAGTCGACATGGAAATGAGATGGTTTTTGTC / CTAAGCGGCCGCTCATCTGACGTACCAAACATAAAC	65

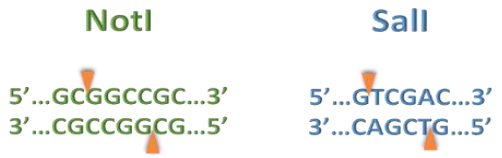
Table 2.3 PCR amplification reaction

	Temperature ºC	**Time**
	98	30 seconds
30 cycles	98	30 seconds
	55	1 minute
	72	2 minutes
	72	7 minutes
	4	∞

2.2.2.2 Purification of PCR products

The amplification product was separated with an 0.8% agarose gel (1.6g of Agarose (Sigma Aldrich UK) and 200ml 1xTAE (40 mM Tris-acetate, 1mM EDTA) and 1µg/ml ethidium bromide. 100 bp DNA Ladder (NEB) was prepared and run in parallel to the samples) via gel electrophoresis (90V for 40minutes) and then was gel extracted using the QIAquick Gel Purification Kit (QIAGEN. Crawley, UK). According to the manufacturer's protocol, the DNA visualised in 0.8% agarose gel was extracted with the use of a clean scalpel and placed into a 1.5ml Eppendorf tube. The gel slice was weighted and 100µl QC buffer per 100mg of gel were added. The mix was then heated at 50 ^{o}C for 10 minutes to dissolve the gel. Isopropanol was then added 100µl per 100mg of gel and the mixture was transferred into a QIAquick spin column. The column was centrifuged for 1 minute at 13,000rpm with flow through being discarded. Additional 500µl QC buffer was added to the QIAquick column and centrifuged for 1 minute at 13,000rpm. Washing steps included addition of 750µl of washing buffer PE followed by centrifugation for 1 minute at 13,000rpm. The column was then centrifuged for an additional minute followed by addition of 40µl of nuclease free water (Fisher Scientific) at the centre of the membrane and incubated for 5 minutes. The column was then centrifuged for 1 minute at 13,000rpm and eluate was collected. Quantification of the DNA was performed with the use of Nanodrop Lite spectrophotometer (Thermo Scientific).

Purification of the fragment was followed, by restriction digestion of both vectors and insert with NotI and SalI restriction enzymes at 37 ^{o}C water bath for 4 hours (Figure 2.4). The vector was then linearized at 80 ^{o}C for 20 minutes and then dephosphorylated with alkaline phosphatase calf- intestinal (CIP) 10U/µl for 1 hour (New England Biolabs, Herts, UK).



Reagents	Volume (µl)
BSA (NEB)	0.5
NotI 10U/µl (NEB)	1
SalI 20U/µl (NEB)	1
DNA 1µg	5-10
10x NEB buffer 3 (NEB)	5
H_2O (Fisher Scientific)	Reaction up to 50
Total	50

Figure 2.4 Restriction digestion reaction.

2.2.2.3 Purification of digested DNA

Insert and vector DNA were then purified using the PCR purification kit (QIAGEN, Crawley, UK). According to the manufacturer's protocol, the DNA was mixed with 5 volumes of PB buffer to 1 DNA sample volume and then placed into a QIAquick column followed by centrifugation for 1 minute at 13,000rpm. The flow through was discarded and 750µl of PE washing buffer were added followed by centrifugation for 1 minute at 13,000 rpm. The QIAquick column was transferred into a clean 1.5ml Eppendorf tube. Elution step included addition of nuclease water (Fisher Scientific) and incubation at room temperature for 5 minutes followed by centrifugation for 1 minute at 13,000rpm. Quantification of the DNA was performed with the use of Nanodrop Lite spectrophotometer (Thermo Scientific).

2.2.2.4 Ligation and transformation

The insert was then ligated into the vector using T4 DNA ligase (New England Biolabs, Herts, UK) (Table 2.4). The ligated DNA was transformed into competent *E.Coli* TOP10 One-shot cells (Invitrogen, Paisley, UK). A vial of competent cells from -80 ^{o}C was thawed on ice for 1 minute. Then 3µl of ligated vector were added to the competent cells followed by incubation on ice for 30 minutes. The cells were then heat shocked at 42 ^{o}C for 30 seconds and the immediately on ice for 5 minutes. Then 300µl of pre-warmed Luria Broth (LB) medium (Sigma-Aldrich Company Ltd, Dorset, UK) was added to the cells followed by incubation at 37 ^{o}C for 1 hour at 225 rpm in an orbital shaker incubator. LB/agar plates with 50µg/ml Kanamycin were prepared and 100µl of cells were spread onto the plate and incubated overnight at 37 ^{o}C. The following day plasmid DNA was isolated for successful cloning confirmation.

Table 2.4 Ligation reaction with T4 Ligase

Reagents	Volume (µl)
T4 Ligase 3U/µl (NEB)	2
10x T4 Ligase Reaction Buffer (NEB)	3
DNA Molar Ratio- (insert : vector 3:1)	5-10
H_2O (Fisher Scientific)	Reaction up to 30
Total	30

2.2.2.5 Miniprep DNA purification

LB medium (Sigma Aldrich) was prepared containing 25mg/ml Kanamycin and 5ml were added to falcon tubes (Corning). Then 20 colonies were selected from the agar plate and a single colony was added to each falcon tube. The culture was incubated at 225 rpm, 37 ^{o}C in an orbital shaker incubator for 14 hours. DNA was purified with the Wizard plus SV miniprep kit (Promega, Southampton, UK). According to the manufacturer's protocol, the bacterial cultures were centrifuged for 30 minutes at 3000 rpm and the pellet was resuspended in 250µl cell resuspension solution followed by the addition of 250µl of cell lysis solution to each sample. Then 10µl of alkaline protease solution was added and the mix was incubated for 5 minutes at room temperature. Incubation was followed by addition of 350µl of neutralisation solution and centrifugation at 13,000rpm for 10 minutes at room temperature. The cleared lysate was then transferred into a spin column and centrifuged for 1 minute at 13,000rpm at room temperature. Then 750µl of wash solution was added and centrifuged at 1 minute at 13,000rpm. The washing step was repeated with the addition of 250µl of wash solution and centrifugation at 13,000rpm for 2 minutes. The spin column was transferred into a clean 1.5ml Eppendorf tube and 100µl of nuclease free water (Fisher Scientific) was added. The DNA was eluted by centrifugation at 13,000rpm for 1 minute. Quantification of the DNA was performed with the use of Nanodrop Lite spectrophotometer (Thermo Scientific).

2.2.2.6 Confirmation of successful cloning

The correct insert was confirmed using PCR with the designed primers described above, restriction digestion with the restriction enzymes that have been used for the cloning of the different construct and DNA Sanger sequencing. The latter required initial PCR amplification with primers that flank the sequence of interest. The amplified DNA was then purified using the PCR purification kit (Qiagen) and quantified by nanodrop Lite spectrophotometer (Thermo Scientific). Then 50ng of the target DNA were used for 3.1 BigDye sequencing reaction performed using a PE Applied Biosystems 9700 thermocycler) (Table 2.5). The constructs with the correct insert were transformed into *E.coli* Rosetta (DE3) cells as described above to assess protein expression (Figure 2.5).

Table 2.5 Sanger sequencing. a) 3.1 Big Dye sequencing reaction mix, b) 3.1 Big Dye sequencing reaction

a

DNA	40ng
Primer	3.2pmol
5x buffer	2µl
Big Dye Reaction mix	2µl
Final Volume	10µl

b

Cycle	Steps	Temperature °C	Time
25 x	1	96	1 sec
	2	96	10 sec
	3	50	5 sec
	4	60	4 mins
	5	4	∞

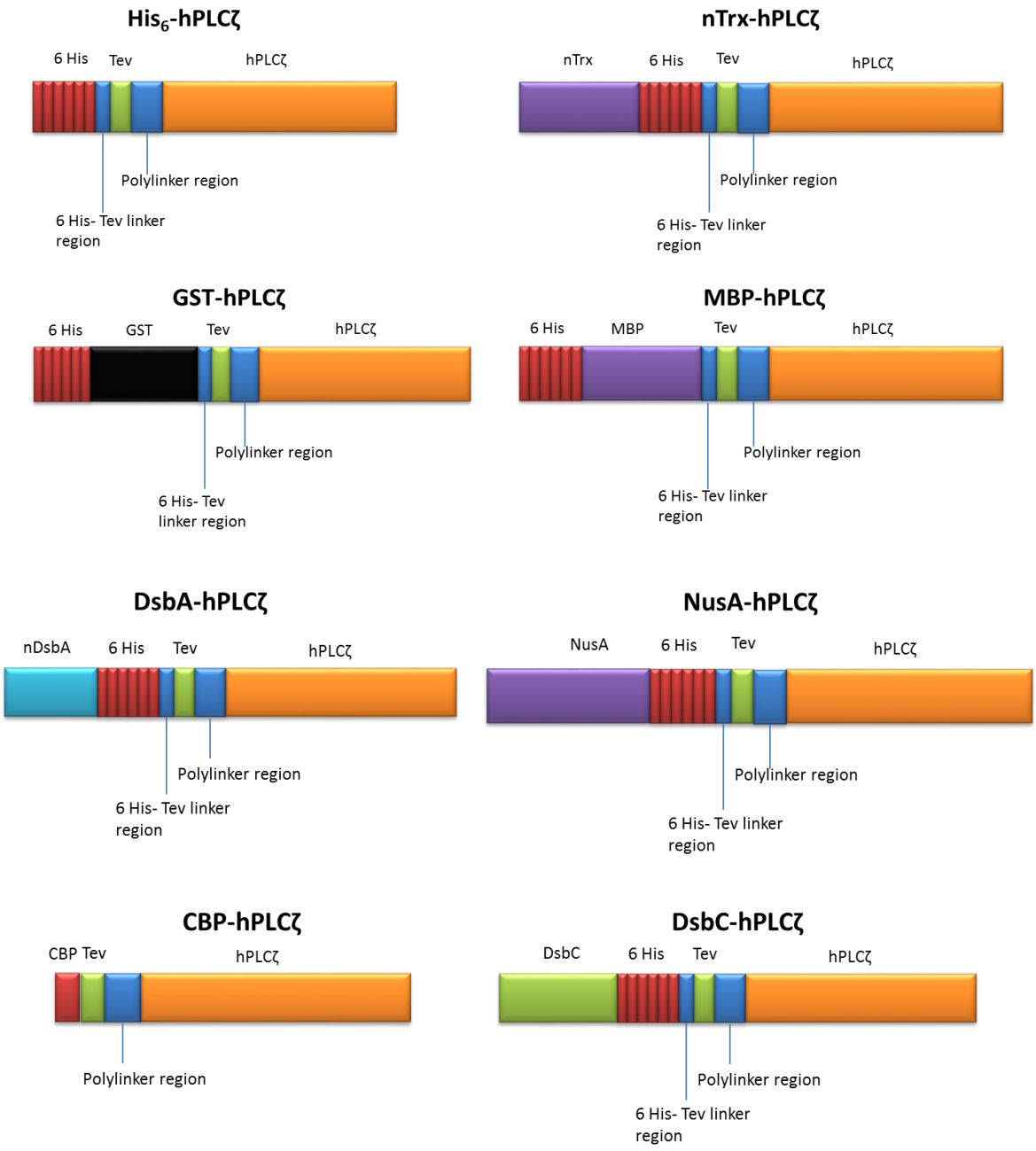


Figure 2.5 The cloning of hPLCζ into pETMM vectors. The cloning of hPLCζ into pETMM vectors resulted in 8 different vector constructs with different solubility tags. All vector constructs but pETMM70 included a 6xHistidine tag for protein purification with affinity chromatography.

Table 2.6 The size of the various bacterial tags included in the vectors of the pET-MM series and the size of the total protein after introduction of hPLCζ as illustrated in (Figure 2.5).

Bacterial Tag	Size-kDa	Bacterial tag + hPLCζ size-kDa
nTrx	12	83
GST	26	97
MBP	44	115
nDsbA	27	98
NusA	54	125
CBP	4	75
DsbC	26	97

2.2.3 Thioredoxin

Thioredoxin has been shown by Lopata *et al*. 2001 Umekawa *et al*. 2008 to occur naturally in the egg and that overexpression of hTrx can potentially promote foetal growth. Therefore, hTrx was subcloned into the following constructs as potential solubility partner which could stabilize hPLCζ and be used in biomedical applications.

2.2.3.1 Cloning of hTrx into hPLCζ-pETMM series vectors

A pair of primers was designed as described above, including with a BamH1 and EcoR1 restriction enzyme recognition site. The cloning process was performed as described above and expression procedure was applied as described below followed by SDS-PAGE gel and western blot confirmation.

2.2.3.2 Cloning of hTrx at the C terminus of hPLCζ-pETMM series vectors

Two pairs of primers were designed using Primer 3 online analysis tools, the first for hTrx and the second for hPLCζ. The forward hTrx primers included a SalI and NotI restriction enzyme recognition site and the hPLCζ primers a KpnI and a SalI restriction site (Table 2.7). Target DNA (hPLCζ) and (hTrx) were amplified using phusion DNA polymerase. The amplification product was separated using a 0.8% agarose gel and gel electrophoresis and gel extracted

using the QIAquick Gel Purification Kit (QIAGEN. Crawley, UK). The cloning process was completed as previously described.

Table 2.7 Primers designed for the amplification of hTrx and hPLCζ

Primer Name	Tm	Sequence	Tm	Gene
F(hPLCζ)/R(hPLCζ)	63	CAGCGGTACCGATGGAAATGAGATGGTTTTTGTC / TTGGGTCGACTCTGACGTACCAAACATAAAC	61	hPLCζ
F(hTrx-N)/R(hTrx-N)	65	ATCGGATCCGATGGTGAAGCAGATCGAGAGC/ TTGGAATTCGGGACTAATTCATTAATGGTGGCTTCAAG	62	hTrx
F(hTrx-C)/R(hTrx-C)	65	GGTTGTCGACATGGTGAAGCAGATCGAGAGC/ TTAAGCGGCCGCTCAGACTAATTCATTAATGGTGGC	65	hTrx

2.2.3.3 Bicinchoninic Acid (BCA) protein assay kit

The BCA protein assay kit was performed according to the manufacturer's protocol. Initially a set of protein standards and working reagent were prepared (Table 2.8). To prepare the working reagent 50 parts of BCA reagent A were mixed with 1 part of BCA reagent B. Then 260µl of working reagent was added to each well of a 96 well microplate (Thermo Scientific) and 9µl of the protein standards and target protein were added to the appropriate wells. The plate was incubated at 37 ^{o}C for 30 minutes followed by absorbance measurement at 595nm by EnsPire Perkin Elmer plate reader.

Table 2.8 The preparation of BCA protein standards for BCA protein Assay kit (Adapted from Thermo Scientific)

Vial	Diluent Volume (µl)	BSA Source and Volume (µl)	Concentration (µg/ml)
A	0	200 of stock	2,000
B	66	200 of stock	1,500
C	100	100 of vial A	1,000
D	100	100 of vial B	750
E	100	100 of vial C	500
F	100	100 of vial E	250
G	100	100 of vial F	125

2.2.4 Bacterial culture and protein expression

The plasmids of interest were transformed into the various bacterial cells used in this study as described above. Cells were cultured in LB medium (Sigma Aldrich) containing the appropriate antibiotic (Table 2.9). All flasks contained 25mg/ml Kanamycin as pETMM vectors contain a Kanamycin resistance gene the cells were grown at 37 ^{o}C at 180rpm in an orbital shaker until OD_{600} is approximately 0.6. Then 1ml culture was collected for uninduced protein expression assay and 100μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Life Technologies) was added to bacterial culture. The cells were grown at 16 ^{o}C in an orbital shaker at 180 rpm overnight and harvested by centrifugation at 3,400rpm for 15 minutes at 4 ^{o}C. The samples collected prior and post induction, were centrifuged at 10,000rpm for 1 minute and were lysed using 5x SDS loading buffer (0.3M Tris-HCl, pH 6.8, 10% w/v SDS, 50% glycerol, 0.25% w/v bromomethol, 5% β- mercaptoethanol) followed by heating for 5 minutes at 100 ^{o}C. The lysate was then analysed on by 7% SDS-PAGE polyacrylamide gel electrophoresis with protein stained with Coomassie brilliant blue (1g Coomassie, 480ml Methanol, 420ml H_2O and 100ml Acetic Acid) and destained with destain buffer (450ml Methanol, 450ml H_2O and 100 Acetic Acid).

Table 2.9 The various bacterial cell lines used in this study.

Bacterial cell line	Antibiotic resistance
Rosetta De3	Chloraphenicol 34mg/ml
BL21 De3	-
Rosetta gami	Chloraphenicol 34mg/ml
Rosetta pLysS	Chloraphenicol 34mg/ml
BL21 pLysS	Chloraphenicol

2.2.5 Western Blotting

Soluble fractions of the bacterial lysates were quantified with Pierce BCA Protein Assay Kit (Thermo Scientific) and data were acquired with the use of the EnsPire Perkin Elmer plate reader. Approximately 10μg of protein were loaded into a 7% SDS-PAGE gel. The protein prior to the loading was mixed with 5x sample buffer (0.3M Tris-HCl, pH 6.8, 10% w/v SDS, 50% glycerol, 0.25% w/v bromomethol, 5% β- mercaptoethanol). The protein was then denatured at 100 ^{o}C for 3 minutes and loaded onto the 7% SDS –PAGE gel (Separating gel (5ml): (1.5M Tris, pH 8.8, 0.4% sodium dodecyl sulphate (SDS), 40% polyacrylamide, 10% ammonium persulfate (APS), 2.55ml H_2O, 5μl TEMED. Stacking gel (4ml): 0.5M Tris, pH 6.8, 0.4% SDS, 40% polyacrylamide, 10% ammonium persulfate (APS), 5μl TEMED, 2.55ml H_20). The gel run in 1x running buffer (15g/L Tris, 72g/L glycine, 5g/L SDS) for 1 hour at 150 mV). It was then transferred onto a Hybond^{TM}-P polyvinylidene fluoride (PVDF) membrane (GE Healthcare) with the use of Trans-blot SD Semi-Dry Transfer Cell (Bio-Rad) soaked in transfer buffer (190mM of Glycine, 25mM of Tris, pH of 9.2, 20% Methanol). The protein gel transfer onto the membrane was performed at 12V for 1 hour and then the membrane was transferred into blocking buffer 1xTBST (0.2M Tris, 1.37M NaCl, pH 7.5, 0.1% Tween 20 and 10% milk powder (Marvel). The membrane incubated for 1 hour in blocking buffer and then incubated with primary antibody at 1x TBST with 10% milk powder (Marvel) for 1 hour. The membrane was then washed three times for 10 minutes with 1x TBST and then the membrane incubated with the secondary antibody for 1 hour (Table 2.10). The incubation was followed by 3 washes of 10 minutes in 10x TBST. The washes were followed by 1 minute incubation with Amersham ECL Plus western blotting detection system (GE Healthcare). The acquisition of the picture was performed with the use of GE Healthcare Amersham Hyperfilm.

Table 2.10 List of antibodies used in this project for protein confirmation by Western Blotting

Primary Antibody	Concentration	Secondary Antibody	Concentration
Anti-hPLCζ (Monoclonal)(National Centre For Scientific Research – Demokritos)	1:100	Anti-Rabbit	1:10,000
Anti-hTrx (Monoclonal)(Origene)	1:10,000	Anti-Rabbit	1:10,000

2.2.6 Protein solubility assay

The constructs with the best expression were further analysed for solubility. The frozen pellets of the bacterial cultures were thawed at 4◦C and the cells were lysed with the addition of lysis buffer (50Mm Na_2PO_4 and 500Mm NaCl, Ph:8, 100mg of lysozyme, 3 tablets of protease inhibitors (Roche Life Sciences) and 1mM Dithiothreitol (DTT)) and sonication. Sonication was performed on ice in three 15 second bursts following 2 minutes incubation on ice. Cell lysis was followed by centrifugation of the crude lysate at 18,000RPM for 30 minutes at 4°C. Supernatant was collected and soluble fractions were assayed using denaturing 7% polyacrylamide gel SDS-PAGE with proteins being stained with Coomassie brilliant blue staining. Protein solubility was also confirmed by western blot.

2.3 Results

2.3.1 hPLCζ cloned into the pETMM series vectors

2.3.2 Confirmation of successful cloning

The hPLCζ gene was amplified with hzF/hzR primers generating a 1.8 kbp product. The DNA fragment was digested with SalI and NotI restriction enzymes and cloned successfully into pETMM series vectors. Successful cloning of hPLCζ gene was confirmed by restriction digestion of the cloned plasmid DNA with SalI/NotI enzymes which successfully digested the 1.8kbp DNA insert for hPLCζ (Figure 2.6). Successful cloning was also confirmed by Sanger sequencing.

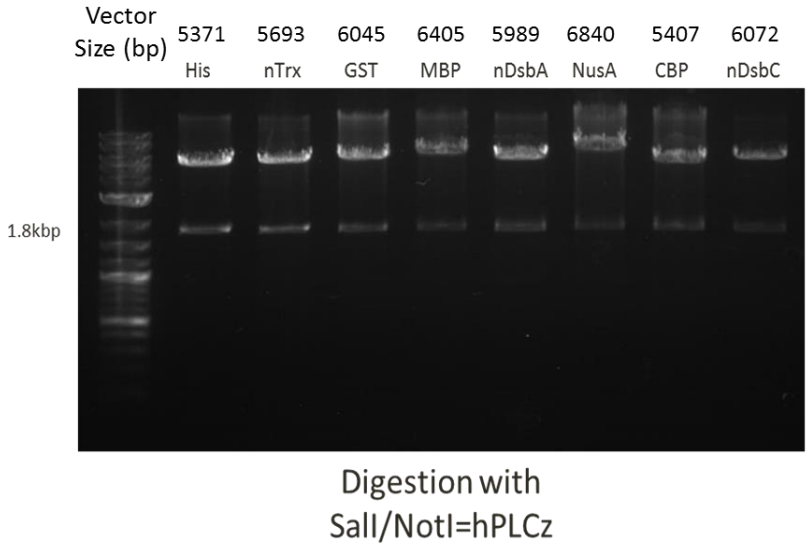


Figure 2.6 Confirmation of successful cloning with restriction digest with SalI and NotI enzymes for human PLCζ gene (1.8 kbp) into the pETMM series vectors.

2.3.3 Protein expression and solubility assessment of hPLCζ

The confirmed pETMM-hPLCζ constructs were transfected into 5 different bacterial strains (described in Table 2.1) and assessed for its protein expression and solubility (Figure 2.7). The protein expression was evaluated with the comparison of bacterial lysates prior and post protein expression induction with IPTG. The protein solubility was evaluated following bacterial lysis and centrifugation. Supernatant was then collected and quantified by BCA protein assay (Promega) and equal volumes were loaded into each well for direct protein solubility comparison (Figure 2.8).

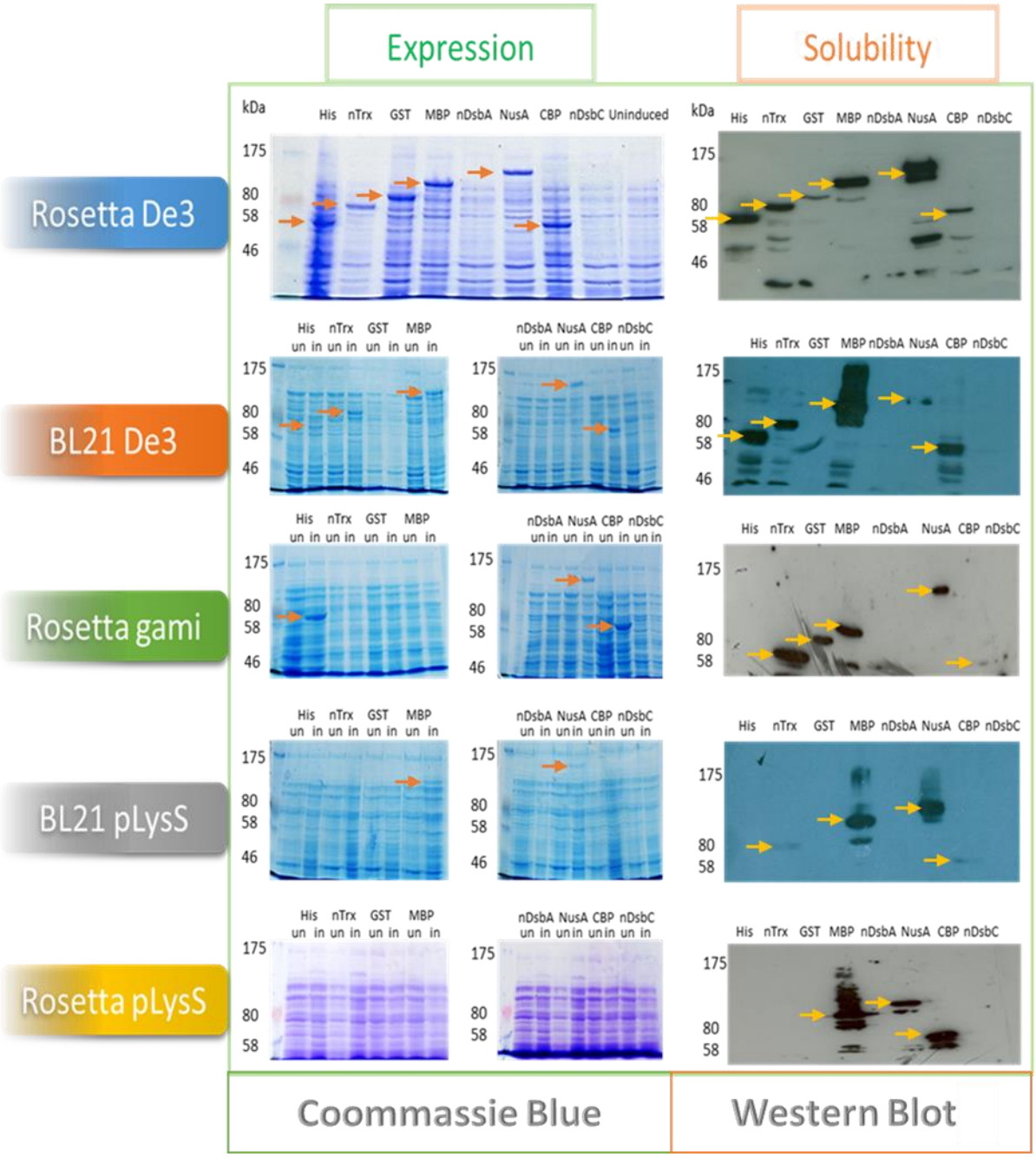


Figure 2.7 Assessment of protein expression and solubility of hPLCζ-pETMM protein constructs in a range of different bacterial cells. Expression was assessed by direct comparison between induced (In) and uninduced (Un) protein expression of bacterial culture lysates using Coommassie blue stained gels. Protein solubility of soluble fractions of bacterial culture lysates was evaluated by western blotting using anti-hPLCζ monoclonal antibody.

hPLCζ-pETMM										
	Expression					Solubility				
	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS
His	+	+	+	-	-	+	+	-	-	-
nTrx	+	+	-	-	-	+	+	+	+	-
GST	+	+	-	-	-	+	+	+	-	-
MBP	+	+	-	+	-	+	+	+	+	+
nDsbA	-	-	-	-	-	-	-	-	-	-
NusA	+	+	+	+	-	+	+	+	+	+
CBP	+	+	+	-	-	+	+	+	+	+
nDsbC	-	-	-	-	-	-	-	-	-	-

Figure 2.8 Evaluation of protein expression and solubility of hPLCζ-pETMM proteins. Expression and solubility as illustrated in (Figure 2.7).

2.3.4 hTrx is cloned into hPLCζ-pETMM series vectors

2.3.4.1 Confirmation of successful cloning

The successful cloning of hTrx into hPLCζ-pETMM vectors was confirmed by double restriction digestion. Restriction digestion with SalI/NotI and BamHI/ EcoRI enzymes confirmed cloning of hPLCζ (1.8kbp) and hTrx (300bp) respectively (Figure 2.9). The incorporation of hTrx aimed to serve as a solubility partner for hPLCζ as it is present in human oocyte eliminating potential ethical issues in biomedical applications.

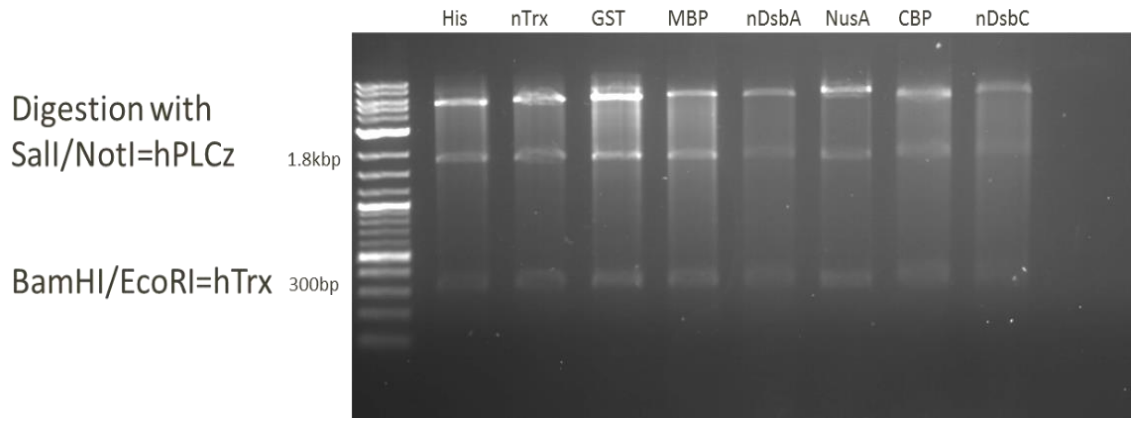


Figure 2.9 Confirmation of successful cloning with restriction digest with SalI/NotI and BamHI /EcoRI restriction enzymes for human PLCζ gene (1.8 kbp) and hTrx (300bp) into the pETMM series vectors respectively.

2.3.4.2 Protein expression and solubility assessment of hTrx-hPLCζ-pETMM constructs

Protein expression of the 8 different constructs was assessed by gel electrophoresis with SDS PAGE gel and Coommassie blue staining and protein solubility with western blot confirmation using anti-hPLCζ and anti-hTrx antibodies (Figure 2.10, 2.11).

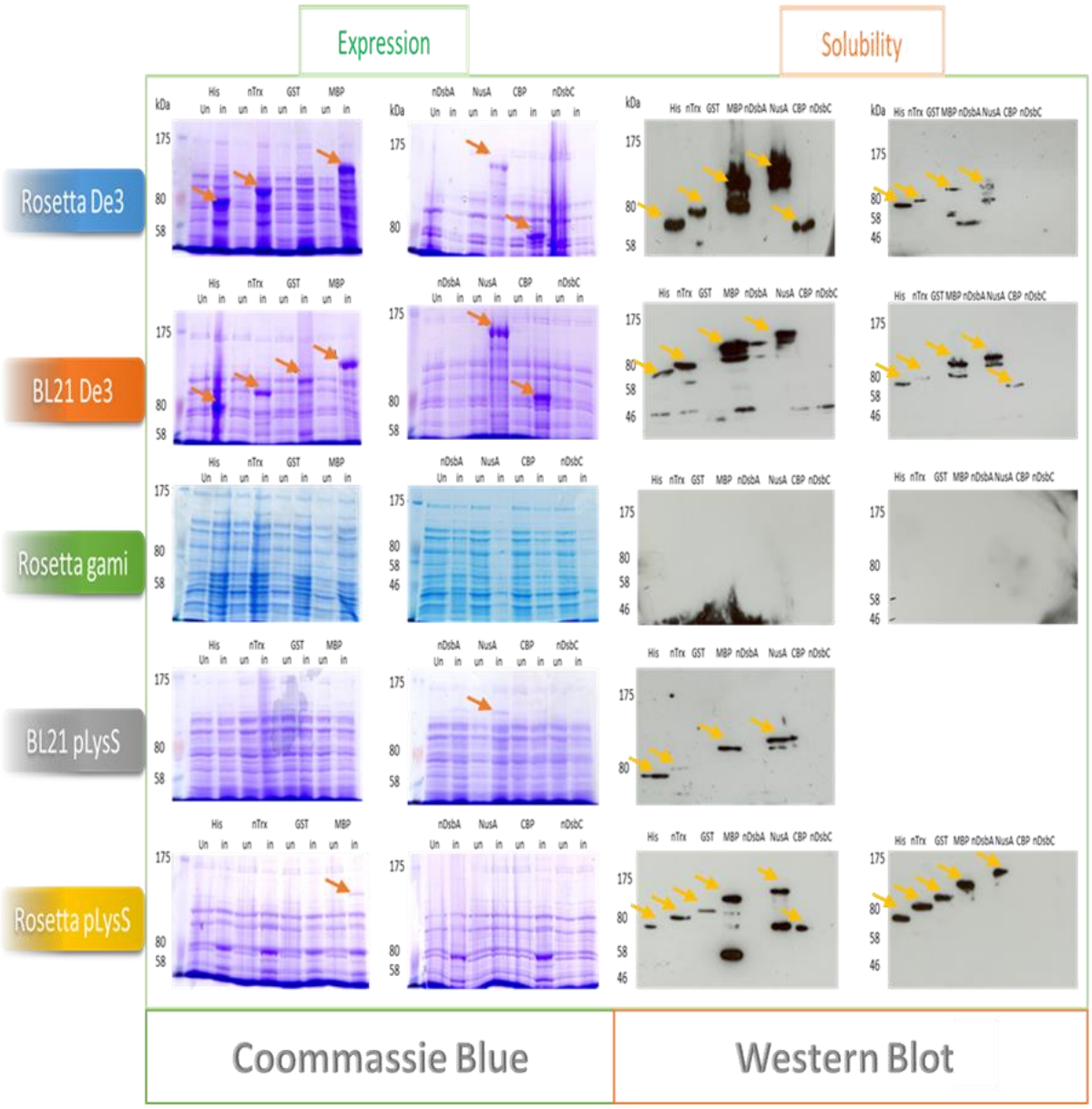


Figure 2.10 Assessment of protein expression and solubility of hTrx-hPLCζ-pETMM protein constructs in a range of different bacterial cells. Expression was assessed by direct comparison between induced (In) and uninduced (Un) protein expression of bacterial culture lysates using Coommassie blue stained gels. Protein solubility of soluble fractions of bacterial lysates was evaluated by western blotting using anti hPLCζ monoclonal antibody (left) and anti-hTrx antibody (right).

‘

hTrx-hPLCζ-pETMM										
	Expression					Solubility				
	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS
His	+	+	-	-	-	+	+	-	+	+
nTrx	+	+	-	-	-	+	+	-	-	+
GST	-	+	-	-	-	-	-	-	-	+
MBP	+	+	-	-	+	+	+	-	+	+
nDsbA	-	-	-	-	-	-	-	-	-	-
NusA	+	+	-	+	-	+	+	-	+	+
CBP	+	+	-	-	-	+	-	-	-	-
nDsbC	-	-	-	-	-	-	-	-	-	-

Figure 2.11 Evaluation of protein expression and solubility of hPLCζ-pETMM proteins Expression and solubility as illustrated in (Figure 2.10).

2.3.5 Cloning of hTrx at the C terminus of hPLCζ-pETMM series vectors

2.3.5.1Confirmation of successful cloning

The cloning of hTrx at the C terminus of hPLCζ-pETMM vectors aimed to alter the stability and folding of the protein and potentially increase its solubility. The successful cloning was confirmed by double restriction digestion (Figure 2.12) and protein expression and solubility was evaluated using 5 different bacterial cells (Figure 2.13, 2.14).

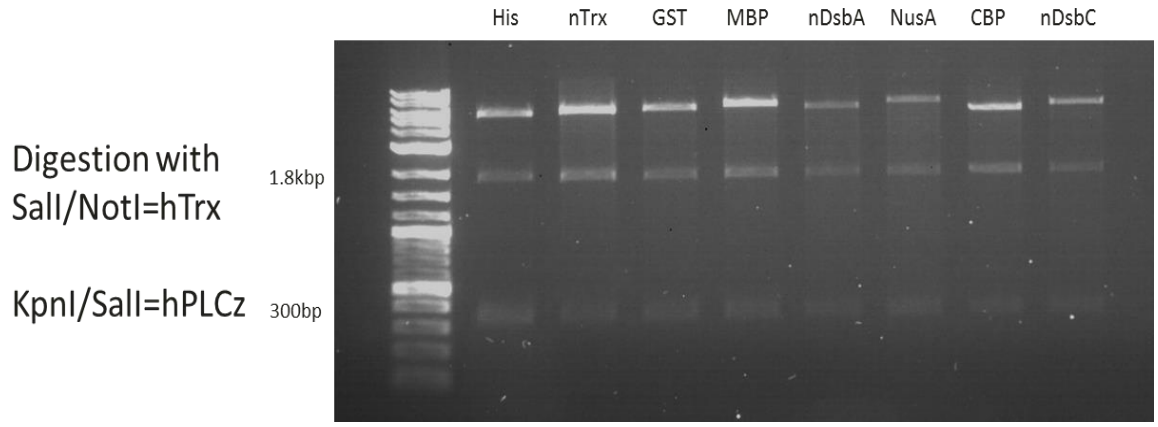


Figure 2.12 Cloning of hTrx at the C terminus of hPLCζ-pETMM series vectors. hTrx was cloned upstream the C- terminus of hPLCζ and sucessful cloning was confirmed by double restriction digest with KpnI/SalI for hPLCζ and SalI/NotI for hTrx at 1800 and 300bp respectively.

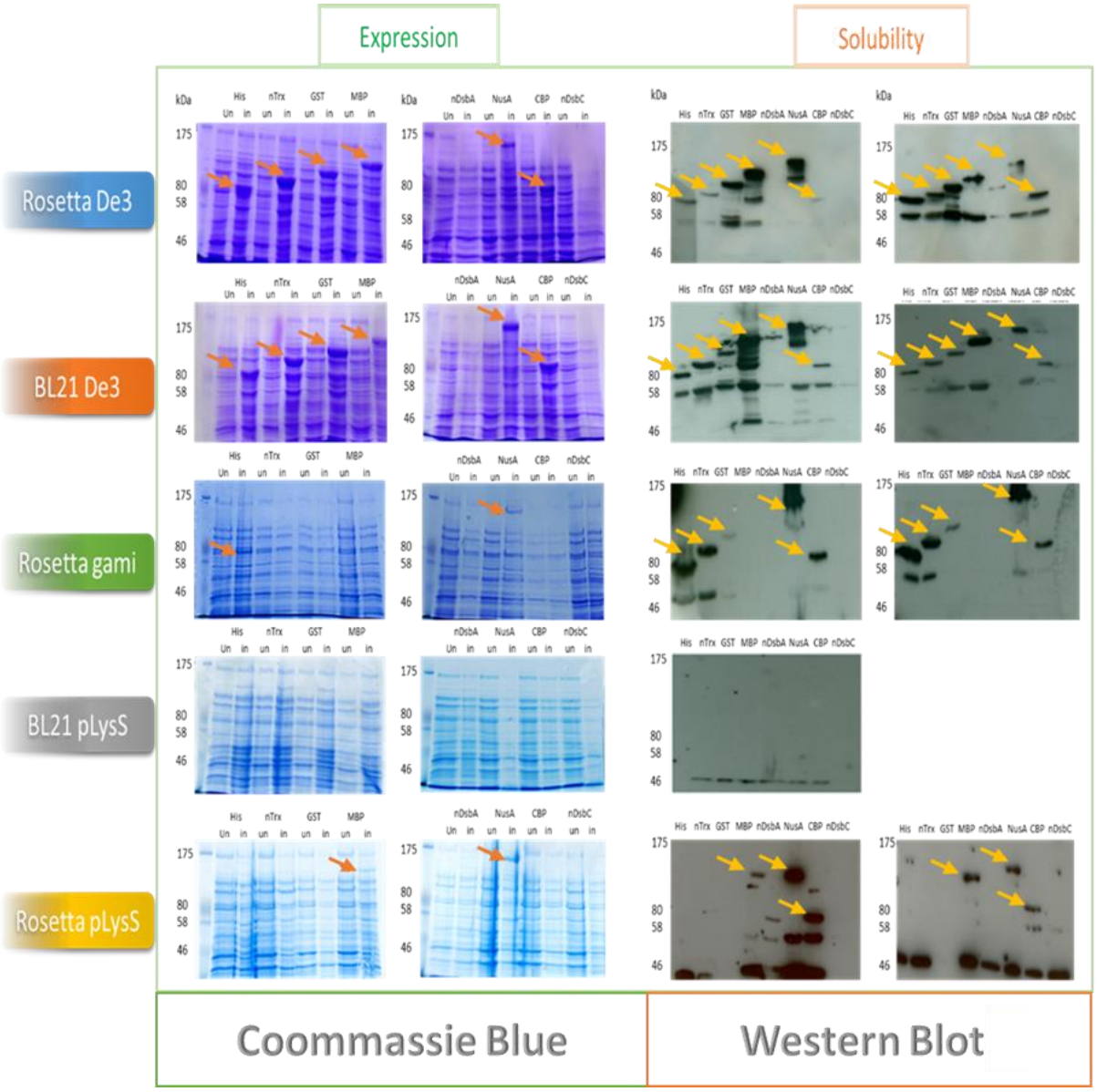


Figure 2.13 Assessment of protein expression and solubility of hPLCζ- hTrx pETMM protein constructs in a range of different bacterial cells. Expression was assessed by direct comparison between induced (In) and uninduced (Un) protein expression of bacterial culture lysates using Coommassie blue stained gels. Protein solubility of soluble fractions was evaluated by western blotting using anti hPLCζ monoclonal antibody (left) and anti-hTrx antibody (right).

hPLCζ-hTrx-pETMM										
	Expression					Solubility				
	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS	Rosetta De3	BL21 De3	Rosetta gami	BL21 pLysS	Rosetta pLysS
His	+	+	+	-	-	+	+	+	-	-
nTrx	+	+	-	-	-	+	+	+	-	-
GST	-	+	-	-	-	-	-	+	-	-
MBP	+	+	-	-	+	+	+	-	-	+
nDsbA	-	-	-	-	-	-	-	-	-	-
NusA	+	+	+	-	+	+	+	+	-	+
CBP	+	+	-	-	-	+	+	+	-	+
nDsbC	-	-	-	-	-	-	-	-	-	-

Figure 2.14 Evaluation of protein expression and solubility of hPLCζ-pETMM proteins. Expression and solubility as illustrated in (Figure 2.13).

2.3.6 Evaluation and selection of the bacterial cell line

Expression and solubility has been assessed by the direct comparison between protein constructs. The optimal bacterial cell line for each protein construct has been selected by the direct comparison of the solubility in different bacterial cells for each protein construct (Figure 2.15, 2.16). The cells were grown at equal volumes and lysed by sonication. The supernatant following centrifugation was loaded at equal volumes for protein solubility assessment. The protein constructs with the highest expression and solubility assessed in experiments described above (Figure 2.10, 2.13) were selected for the bacterial cell line optimisation assay.

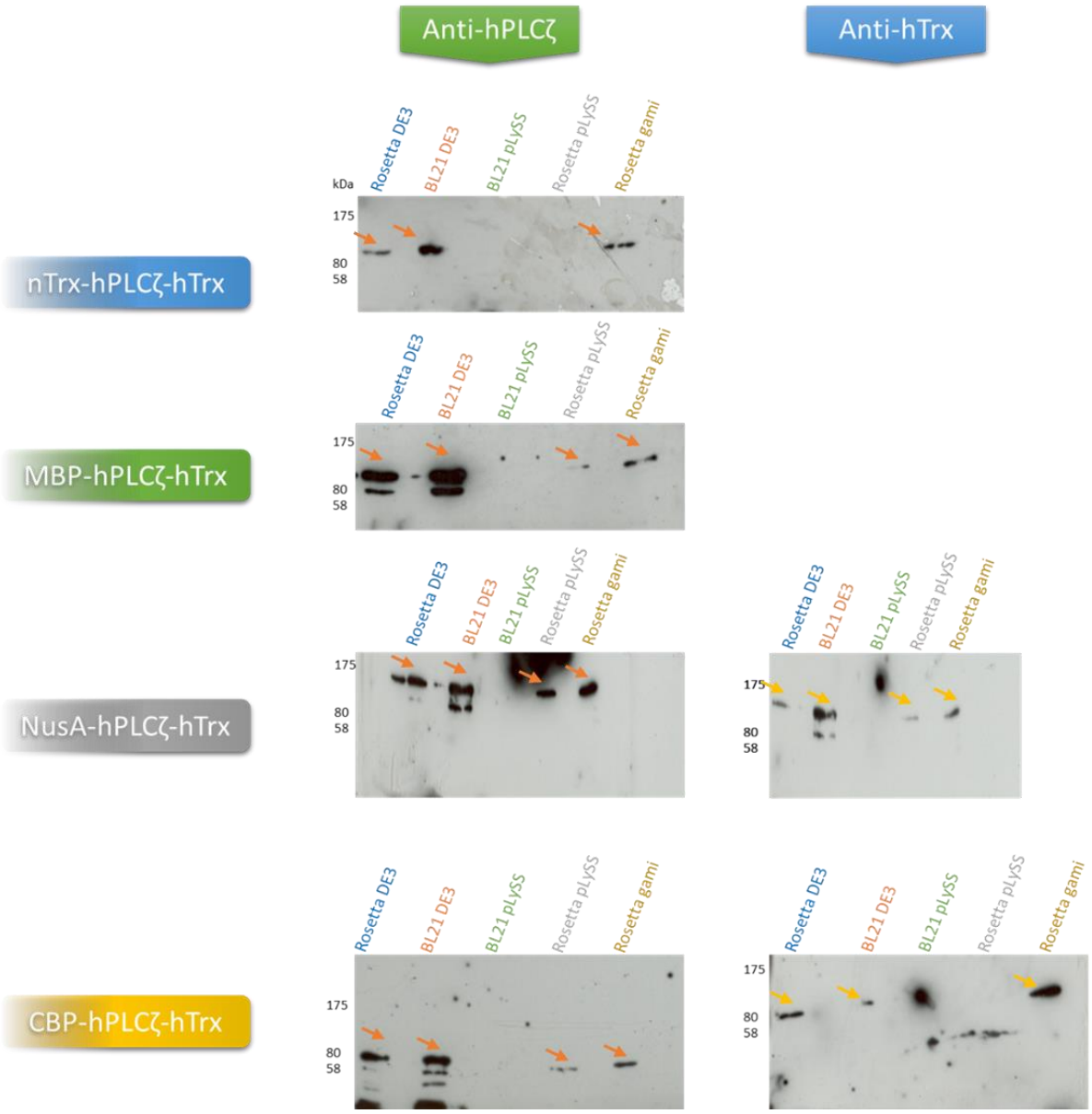


Figure 2.15 The optimal bacterial cell line for the solubility of each protein construct (hTrx-hPLCζ). The solubility was assessed by western blotting with anti-hPLCζ monoclonal antibody (left) and anti-hTrx (right).

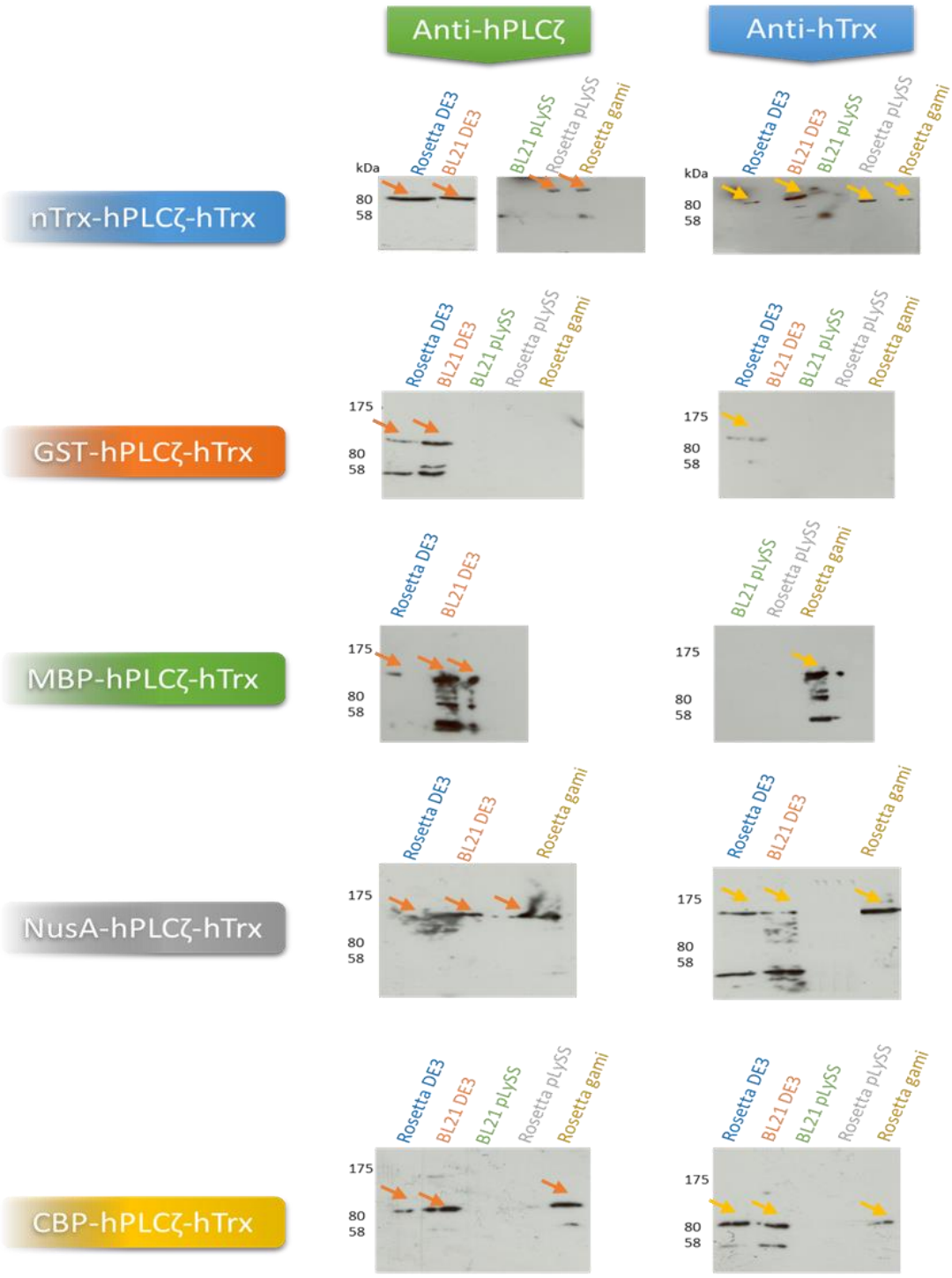


Figure 2.16 The optimal bacterial cell for the solubility of each protein construct (hPLCζ-hTrx C-terminus). The solubility was assessed by western blotting with anti-hPLCζ monoclonal antibody (left) and anti-hTrx (right).

2.4 Discussion

2.4.1 PLCζ constructs

The aim of the experiments performed was to express the protein in high yield retaining solubility and eventually activity and then find the constructs which will be subsequently used for cloning of hTrx-hPLCζ. hPLCζ it is a highly sensitive protein in both temperature and concentration which degrades almost immediately following protein purification. The cloning of hPLCζ in pETMM vectors which contain series of different solubility tags, was to first increase protein expression and second and most important to retain protein solubility as the majority of the pETMM tags included were chaperones and heat shock proteins. The results illustrated that soluble and active form of hPLCζ has been created using a range of different solubility tags. Also, a range of different bacterial cells were used to achieve optimum expression and solubility. The results indicate that Rosetta De3 and BL21 De3 cells express the highest amount of protein retaining solubility. The expression of hPLCζ with a range of protein solubility tags has shown that nTrx, MBP, NusA and CBP express hPLCζ with relatively high protein solubility (Figure 2.17). The assays have been performed with the use of soluble fractions of bacterial lysates. Investigation of the protein present in the insoluble fractions indicated that only approximately 20 – 30% of the total protein expressed was soluble therefore maximizing and retaining the remaining soluble protein was the key for future investigations. The method that was used in this project was not quantitative and therefore accurate conclusion regarding the proportion of protein solubility and expression cannot be drawn. However, it has provided information regarding the different protein expression and solubility potential of the various construct and could give clear indication which bacteria and protein constructs are more likely to retain protein solubility.

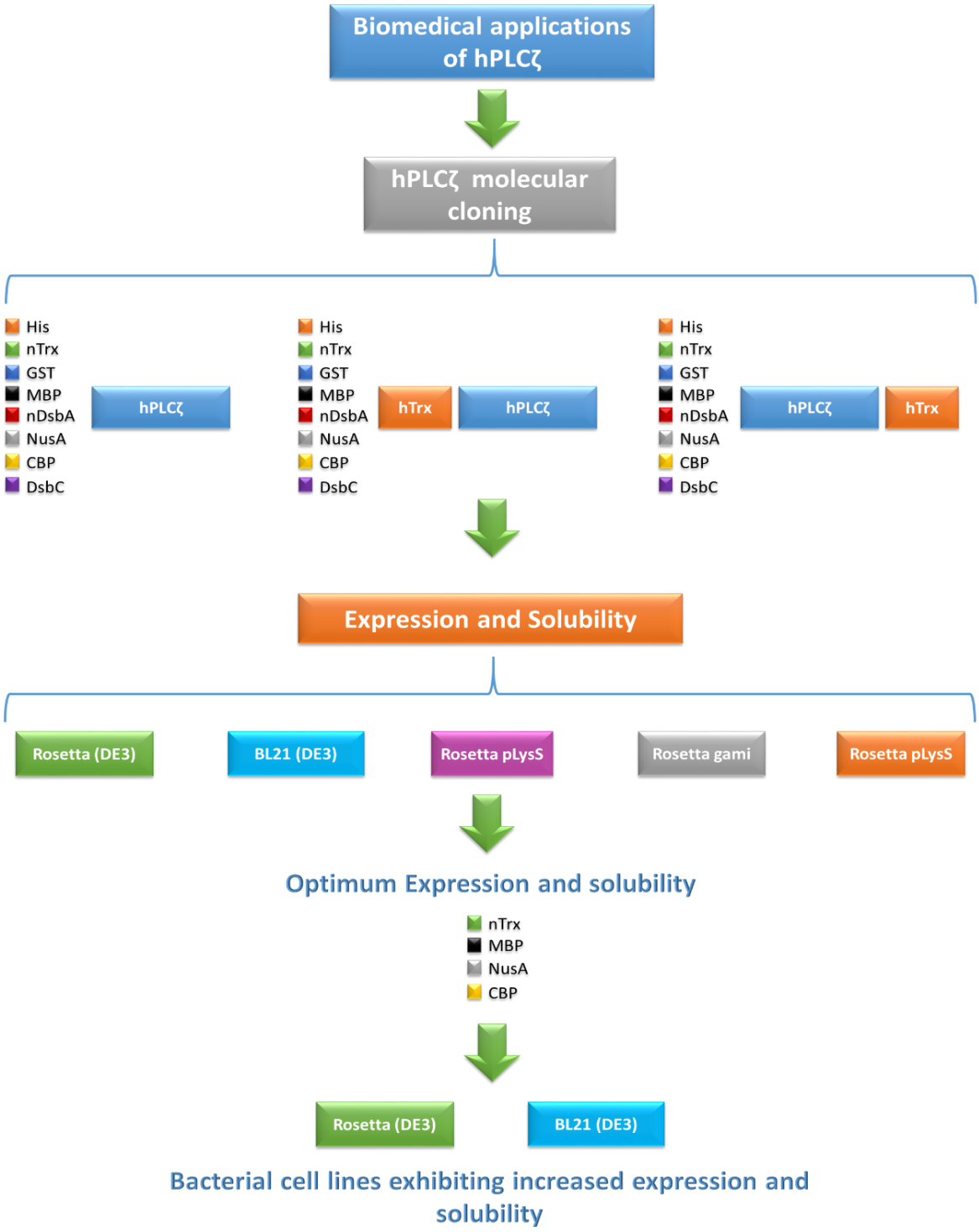


Figure 2.17 Summary of the results obtained by the assessment of various protein constructs and bacterial cell lines for protein expression and solubility optimization. Optimisation strategies as described at Figure 2.2, revealed that 4 of the plasmid constructs exhibited better protein expression and solubility when expressed in Rosetta (DE3) and BL21 (DE3) bacterial strains.

Chapter 3. Activity Assessment of Purified Recombinant hPLCζ

3.1 Introduction

3.1.1 *In-vivo* PLC activity assay

In vivo assessment of human PLCζ was performed in mouse eggs (Figure 3.1). Mouse is an ideal animal model for the study of mammalian egg activation and assessment of human PLCζ activity as multiple Ca^{2+} oscillation that trigger egg activation are similar to those occurring in the human oocyte. Ca^{2+} oscillations are required for egg activation in both mammalian and non-mammalian species however, non-mammalian eggs are generating only a single transient. In mammals, Ca^{2+}oscillations exhibit different multiple oscillation patterns with mouse eggs being similar to the human. Also, mouse as an animal model is generally very popular due to its low maintenance cost and the short generation time providing sufficient number of eggs.

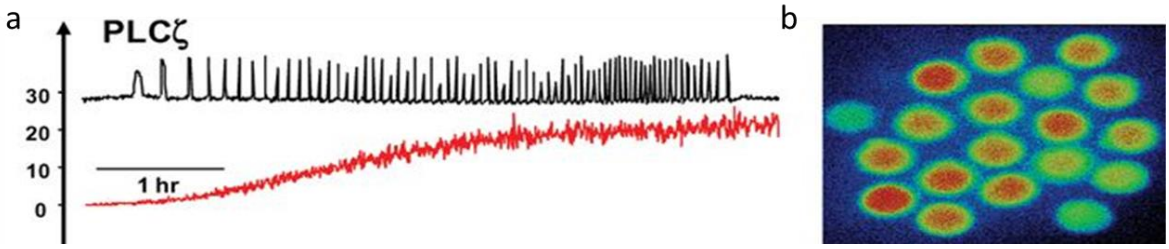


Figure 3.1 Calcium oscillations observed in mouse eggs. a) cRNA hPLCζ-Luc injection into mouse eggs leading to multiple Ca^{2+} oscillations and egg activation. Red line illustrates luciferase emission upon hPLCζ expression. b) Increase in Ca^{2+} concentration was monitored with the use of Ca^{2+} fluorophores. Adapted from (Theodoridou *et al.* 2013).

3.1.2 PLC *in-vitro* activity assay

The PLC activity assay in vitro utilises the ability of PLCs to hydrolyse PtdIns(4,5)P_2 generating IP_3. In PLC activity assays, radiolabelled PtdIns(4,5)P_2 is used for the measurement of the released radioactive phosphates following PtdIns(4,5)P_2 hydrolysis by PLCs (Huang *et al.* 2013). Although this assay can provide quantitative results of PLC activity it cannot illustrate hydrolysis in physiological cell conditions and often can be technically challenging (Rusten and Stenmark 2006; Huang *et al.* 2011).

3.1.3 Calcium Fluorescent dyes

The forefathers of modern calcium imaging were bioluminescent calcium binding photoproteins. The most well-known is aequorin. This calcium indicator has a relatively low calcium sensitivity and a rapid signal decay. The breakthrough in calcium imaging occurred in 1980 by (Tsien 1980). The first generation of fluorescent calcium dyes were more sensitive and versatile. The group of first generation fluorescent dyes included quin-2, fura-2, indo-1 and fluo-3 dyes. The first among them to be developed and applied in vivo experiments was quin-2 with ultraviolet light excitation (339nm). However its requirement for high intracellular dye concentrations due to low brightness led to the development of fura-2 which is excited at 350 or 380 nm and is more sensitive to calcium changes than quin-2 allowing its use in greater range of biological experiment (Grienberger and Konnerth 2012). The third generation of fluorescent dyes have a wider excitation spectra such as Oregon Green BAPTA and Fluo-4 fluorescent dyes (Table 3.1) (Grienberger and Konnerth 2012). Oregon Green BAPTA which was the fluorescent dye of choice in this project offers the advantage of application with lower dye concentration preventing phototoxic effects and relatively high sensitivity allowing detection of small changes in calcium concentration close to resting levels (Paredes *et al.* 2008).

Table 3.1. Characteristics of the various Ca2+ indicators. K_d for Ca^{2+} illustrate the binding affinity to Ca^{2+}. Adapted from (Paredes *et al*. 2008).

Indicators	**K_d for Ca^{2+}**	**Excitation nm, Emission nm**
Fura-2	145	363/335 ex, 512 em
Fluo-3	325	506 ex, 526 em
Fluo-4	345	494 ex, 516 em
Indo-1	230	488 ex, 405/486 em
Oregon Green 488 BAPTA-1	170	488 ex, 520 em

3.1.4 Intracytoplasmic sperm injection (ICSI)

Intracytoplasmic sperm injection is the current gold standard for treatment of male infertility. ICSI is performed in patients where spermatozoa are unable to fertilise the egg due to lack of mobility or increased percentage of abnormal spermatozoa present in the semen. Then a single spermatozoa is injected into the ooplasm. The process of ICSI is operated with the use of a micromanipulator equipped with a microscope with x20 and x40 objectives. Microinjection

of the spermatozoa into the oocyte it is an invasive method which has to be operated with caution. The polar body of the oocyte has to be either at 6' or 12 o'clock position to avoid disruption of the meiotic spindle which can lead to chromosomal aneuploidy and cell death (Figure 3.2). ICSI has approximately 70% success rates but as an invasive technique it can potentially damage the oocyte and a possible selection of an abnormal spermatozoa can lead to chromosomal abnormalities or fertilisation failure (Coward and Wells 2013).

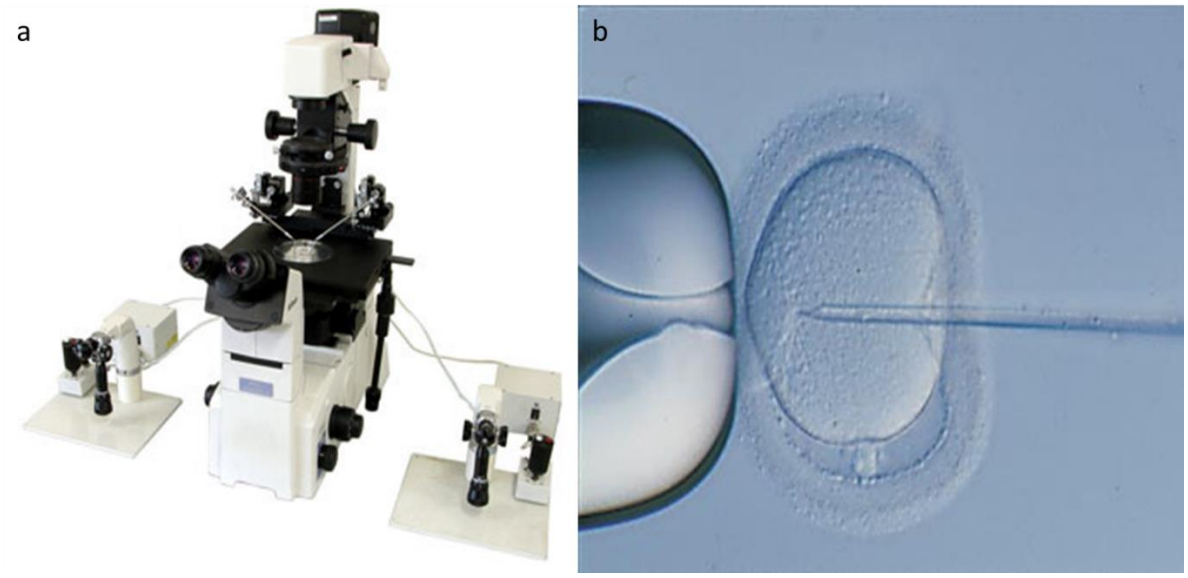


Figure 3.2 Microinjection of a single spermatozoa into the oocytes. a) The micromanipulator with which the microinjection takes place (Adapted from http://www.origio.com/). b) Single sperm injection (ICSI) of a spermatozoa into the oocyte, the polar body has to be at 6 or 12 o' clock to avoid damage of meiotic spindle (Adapted from http://www.nuffieldhealth.com/).

3.1.5 Cleavage of protein solubility tag

Solubility tags have the ability to overexpress soluble recombinant protein. However, often they have to be removed so the protein of interest can be used in various applications such as crystallography and *in vivo* assays. Thus, a cleavage site for binding of a protease is usually designed between the protein tag and the protein of interest so that the purified protein can be removed from its solubility tag (Esposito and Chatterjee 2006). The most popular proteases for solubility tag removal are the Thrombin and the Tobacco etch virus protease (TEV), with the latter being the selected protease in this project. TEV protease has a recognition site E-N-L-Y-F-Q-S, cleaving between glutamine and serine (Terpe 2003) (Figure 3.3). The major drawback of proteolytic cleavage of solubility tags is the rapid degradation of the purified protein following tag removal. There is yet no perfect solution to prevent protein degradation,

although addition of detergents or chaperones are considered to prevent protein degradation (Esposito and Chatterjee 2006).

Figure 3.3 The cleavage site of TEV protease removing the solubility tag from the target protein

3.1.6 Affinity chromatography

The principle of affinity chromatography is based on the ability of the protein of interest to bind to a ligand on the chromatography matrix. The binding is regulated either via antibody or an Immobilized metal-affinity chromatography (IMAC) (Costa 2014). The IMAC technique which has been used in this project can employ tags solely for affinity chromatography such as the polyhistidine tag or proteins which can act as solubility tags such as MBP and GST and the same time can bind to amylose and glutathione resin respectively. Elution of the target protein can be achieved by either alteration of pH or with addition of a competitive ligand such as free Imidazole (Bornhorst and Falke 2011).

The use of polyhistidine tag is the most popular method for protein purification utilizing the ability of histidine to bind strongly to immobilised metal ion matrices such as (Ni^{2+} Co^{2+}, Cu^{2+} and Zn^{2+}). The interaction takes place as electron donation from the imidazole rings on the histidine molecules interacting with the immobilized metal. The purification method using histidine tag is known as Immobilized metal-affinity chromatography (IMAC) (Terpe 2003). The current gold standard of Ni^{2+} purification is with nitrilotriacetic acid (NTA). Nickel ion (Ni^{2+}) has 6 ligand binding sites four of which are occupied by NTA and the remaining two by the His_6 sequence (Khan *et al.* 2006) (Figure 3.4). The drawback of the Ni-NTA purification is the low binding affinity thus often a double His_6 cloning strategy is preferred.

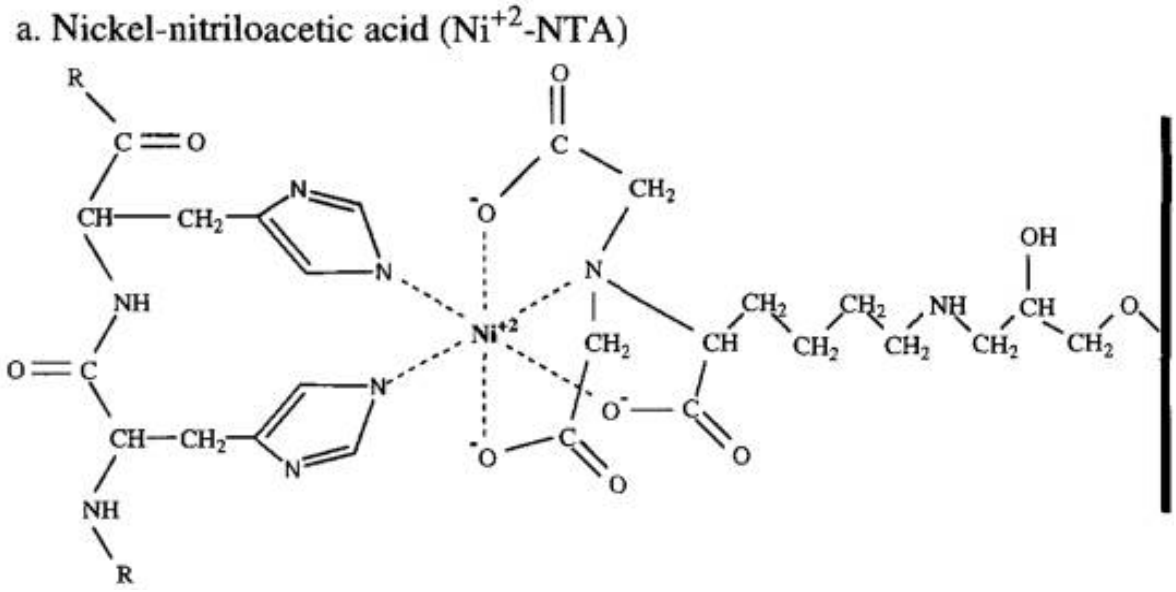


Figure 3.4 The binding of Ni-NTA resin to His6. The His_6 binds to Ni^{2+} -NTA with two binding sites. Adapted from (Bornhorst and Falke 2011).

3.1.7 AIMS

The aim of the chapter was the assessment of purification and PLCζ activity of the constructs found to exhibit the highest protein expression and solubility in Chapter 2. The recombinant proteins were expressed in Rosetta De3 cells in large culture volumes and were purified by affinity chromatography. The purified protein of the selected constructs was microinjected into mouse oocytes along with Ca^{2+} fluorophores and PLCζ activity was monitored. The second aim was the removal of the bacterial solubility tags from the recombinant protein maintaining protein solubility. Thus recombinant MBP-hTrx-hPLCζ was selected as hTrx can potentially retain hPLCζ solubility following the removal of MBP. The natural presence of hTrx in human oocytes could potentially allow biomedical applications of purified hPLCζ (Figure 3.5).

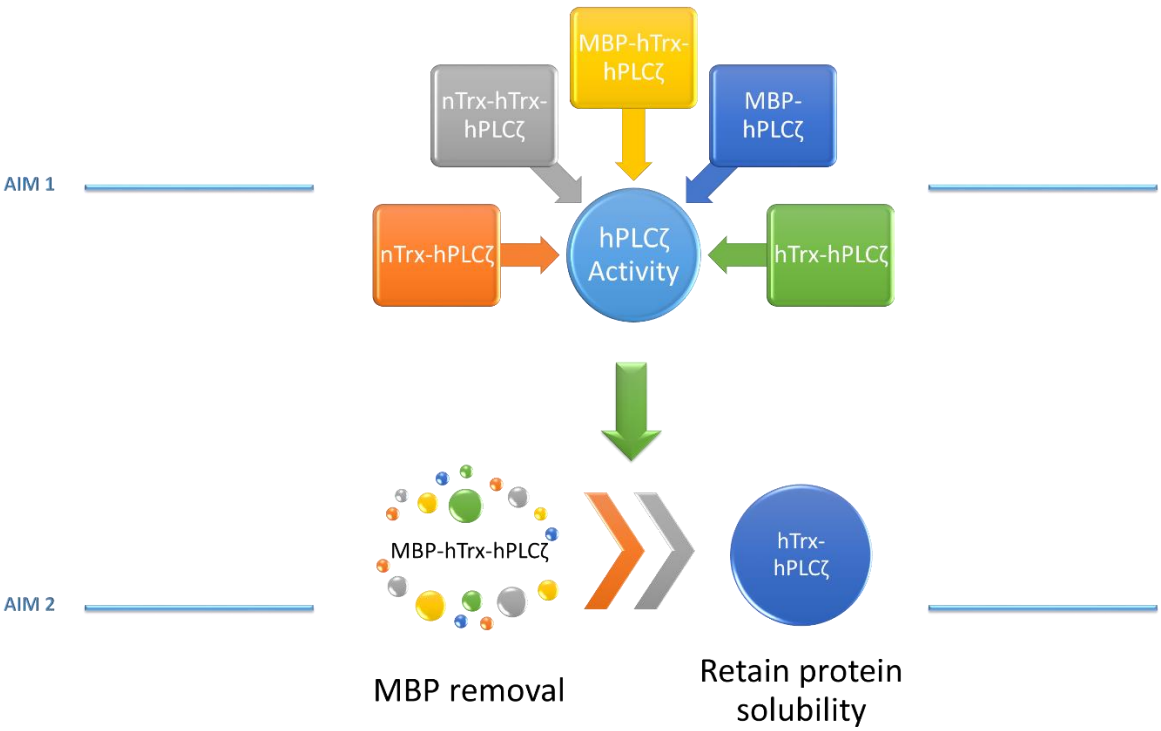


Figure 3.5 Aims of the chapter. The chapter assessed hPLCζ constructs *in-vivo* activity and protein solubility of hPLCζ following bacterial solubility tag removal, an essential requirement for biomedical applications.

3.2 Material and Methods

3.2.1 Purification of hPLCζ with His_6 tag

The cells (Rosetta De3) were cultured as described in chapter 2, using (4L) of culture for each construct. The cells were lysed with the addition of lysis buffer (50Mm Na_2PO_4 and 500Mm NaCl, pH:8, 100mg of lysozyme, 3 tablets of protease inhibitors-EDTA free (Roche Life Sciences) and 1mM Dithiothreitol (DTT)). Sonication was performed in three sets of 15 seconds with each set followed by 2 minutes incubation on ice. Crude lysate was centrifuged at 18,000RPM for 30 minutes at 4°C. The supernatant was collected and incubated with Ni-NTA Agarose beads (QIAGEN) for 1 hour. Prior to the purification step a Bradford assay (described below) was prepared for the quantification of the washed and eluted protein in each purification stage. The lysate was passed through a gravity column where a series of washings took place with phosphate buffer (50Mm Na_2PO_4 and 500Mm NaCl, pH:8,) and addition of 20mM and 35mM of Imidazole for the 2 washings respectively with 75mM and 250mM Imidazole for protein elution. Purified protein was assayed using denaturing SDS-PAGE gel as described in chapter 2. The protein was then concentrated with the use of 3K Amicon tubes (AMD Millipore) as described below and buffer exchanged with HEPES buffer following its microinjection into mouse eggs, where Ca^{2+} oscillations were monitored as described below.

3.2.2 Bradford Assay

A quick and straight forward method was used to measure protein levels during washing and elution steps. Bradford reagent (Bio-Rad) was diluted 4 fold, 2.5ml Bradford (Bio-Rad) and 7.5ml nuclease-free water (Fisher- Scientific) were mixed and 200µl were placed into each well of 96 well-plate. Protein presence was detected by visual detection of colour change without spectrophotometry assay.

3.2.3 Purification of hPLCζ with MBP solubility tag

Bacterial cells were lysed with lysis buffer (20Mm Tris/HCl, 200Mm NaCl, 1Mm EDTA, 1Mm DTT, pH: 7.5, 3 tablets of protease inhibitors-EDTA free (Roche Life Sciences) and 1mg/ml lysozyme. The cells were sonicated and the soluble fraction was collected as described above. The soluble fraction was incubated with amylose resin (NEB) 1.5ml per 30ml of supernatant and incubated for 1 hour at 4 ^oC on a mechanical shaker. The mix was then applied to a gravity column and washed with washing buffer (20mM Tris/HCl, 200mM NaCl, 1mM EDTA, 1mM DTT, pH: 7.5) until there was colour change in Bradford assay indicating absence of protein.

Flow through fraction of 25µl were tested every minute by Bradford assay. Once the washing flow through was almost protein-free the protein was eluted with the addition of elution buffer (20mM Tris/HCl, 200mM NaCl, 1mM EDTA, 1mM DTT, 10mM Maltose). Then 3 ml elution flow through was collected and assessed by SDS-Page gel electrophoresis.

3.2.4 Tev protease expression, purification and storage

The Tev protease construct kindly provided by Dr. Nomikos (Cardiff University) was expressed using 1L of bacterial culture (BL21 De3) which was grown at 37◦C and induced at O.D.=0.6 with 100µM IPTG overnight at 16 ◦C. The cells were lysed with 25ml of lysis buffer 20mM Tris pH-7.5, 200Mm NaCl, 1mM EDTA, 1mM DTT) and sonication as described above. Once lysed and centrifuged at 18,000RPM for 30 minutes, the supernatant was allowed to bind to 1.5ml of Ni-NTA beads for 1 hour. The protein was washed with 20 and 40mM of Imidazole and eluted with 70 and 250 mM of Imidazole. The protein was then quantified with bicinchoninic acid assay (BCA) assay 3µg/µl, aliquoted in 50% glycerol and stored in -80◦C.

3.2.5 TEV cleavage of MBP-hTrx-hPLCζ

1.5L of MBP-hTrx –PLCζ was expressed as described above and lysed with 20ml of lysis buffer (20mM Tris PH-7.5, 200Mm NaCl, 1mM EDTA, 1mM DTT) containing protease inhibitor mixture (EDTA-free, Roche) and 1mg/ml lysozyme). The cells were centrifuged at 18,000RPM for 30minutes. The protein was then allowed to bind to amylose beads for 1hour. It was then applied to a gravity column and washed with (20Mm Tris pH 7.5, 200mM NaCl, 1mM EDTA, 1mM DTT). The protein was then eluted with (20Mm Tris pH-7.5, 200mM NaCl, 1mM EDTA, 1mM DTT and 10Mm Maltose). The eluted protein was then dialyzed in 4L of dialysis buffer (5mM DTT, 1mM EDTA, and 50mM Tris pH-7.5) with Tev protease in ratio of 1:100 overnight at 4◦C. The following day the cleaved protein was buffer exchanged with (50mM Na_2PO_4, and 500mM NaCl, pH-8) and incubated with 200µl of Ni-NTA beads for 1 hour. The protein was then centrifuged at 600RPM for 5 minutes and the supernatant was collected. A second extraction was then applied by exchanging the buffer of the protein with (20Mm Tris pH 7.5, 200Mm NaCl, 1mM EDTA, 1mM DTT) and the protein was then incubated with 200µl of amylose beads for 1 hour. Following incubation the protein sample was centrifuged at 600rpm for 5 minutes and the supernatant was collected. The purity of the protein was assessed by western blot with 2 different antibodies (Anti-EF hand-hPLCζ and Anti- 6x Histidine) (Table 3.2).

Table 3.2. List of antibodies used in this project for protein confirmation by Western Blotting

Primary Antibody	Concentration	Secondary Antibody	Concentration
Anti-hPLCζ(Monoclonal)	1:100	Anti-Rabbit	1:10,000
Anti-hTrx (Monoclonal)	1:10,000	Anti-Rabbit	1:10,000
Anti-His_6	1:5000	Anti-Mouse	1:10,000
Anti-MBP	1:25,000	Anti-Mouse	1:10,000

3.2.6 Purified protein concentration and buffer exchange

Purified protein eluted in approximately 2ml of elution buffer (250mM Imidazole) was loaded into a 3K Amicon tube (AMD Millipore) to reduce the volume of the eluate to 50µl. The concentrated purified protein was then buffer exchanged in 3k Amicon tubes (AMD Millipore) with microinjection buffer (120mM KCL, 20mM Hepes, pH 7.4) to reduce Imidazole concentration from 250mM to 2mM and eliminate potential cytotoxic effects on oocytes. Equal volumes of microinjection buffer to the concentrated eluate were added to the 3K Amicon tube followed by centrifugation at 14,000g for 15minutes until final Imidazole concentration reached 2mM. To recover the protein from the filter the 3K Amicon tube was inverted and placed in a clean collection tube and centrifuged at 5000g for 5 minutes. Concentrated purified and buffer exchanged protein was quantified with Pierce BCA protein assay kit (Thermo Scientific).

3.2.7 Microinjection and Calcium imaging

Mouse eggs were initially washed in M2 media (Sigma) and were microinjected with the purified PLCζ protein mixed with an equal volume of 1mM Oregon Green BAPTA dextran (Molecular Probes) in the injection buffer. Calcium oscillations were recorded using a Zeiss Axiovert 100 microscope equipped with a cooled intensified CCD camera (Photek Ltd, UK). Recording of Ca^{2+} oscillations was performed for 1 hour post injection with measuring intervals of 20 seconds. Microinjection and Ca^{2+} imaging experiments were performed by Dr. Yu.

3.3 Results

3.3.1 *In-vivo* activity assessment of nTrx-hPLCζ

Expression levels of nTrx-hPLCζ as illustrated in Chapter 2, were encouraging, therefore, the expression, solubility, purity and activity were further evaluated. This protein has the advantage of containing bacterial thioredoxin (nTrx) which is a solubility tag with size approximately 12 kDa and a His_6 region which allows purification with affinity chromatography using Ni-NTA beads. The purified and concentrated protein was buffer exchanged following microinjection into mice oocytes along with calcium fluorescent dye for the monitoring of calcium oscillations. Each calcium oscillation line represents activity in different mouse oocytes. The Ca^{2+} oscillations varied among different egg as each of them has different development and survival potential and due to the potential variability of the amount of microinjected hPLCζ. The construct displayed good protein expression and solubility but following protein purification the eluate was found to contain multiple products, possibly due to protein degradation (Figure 3.6).

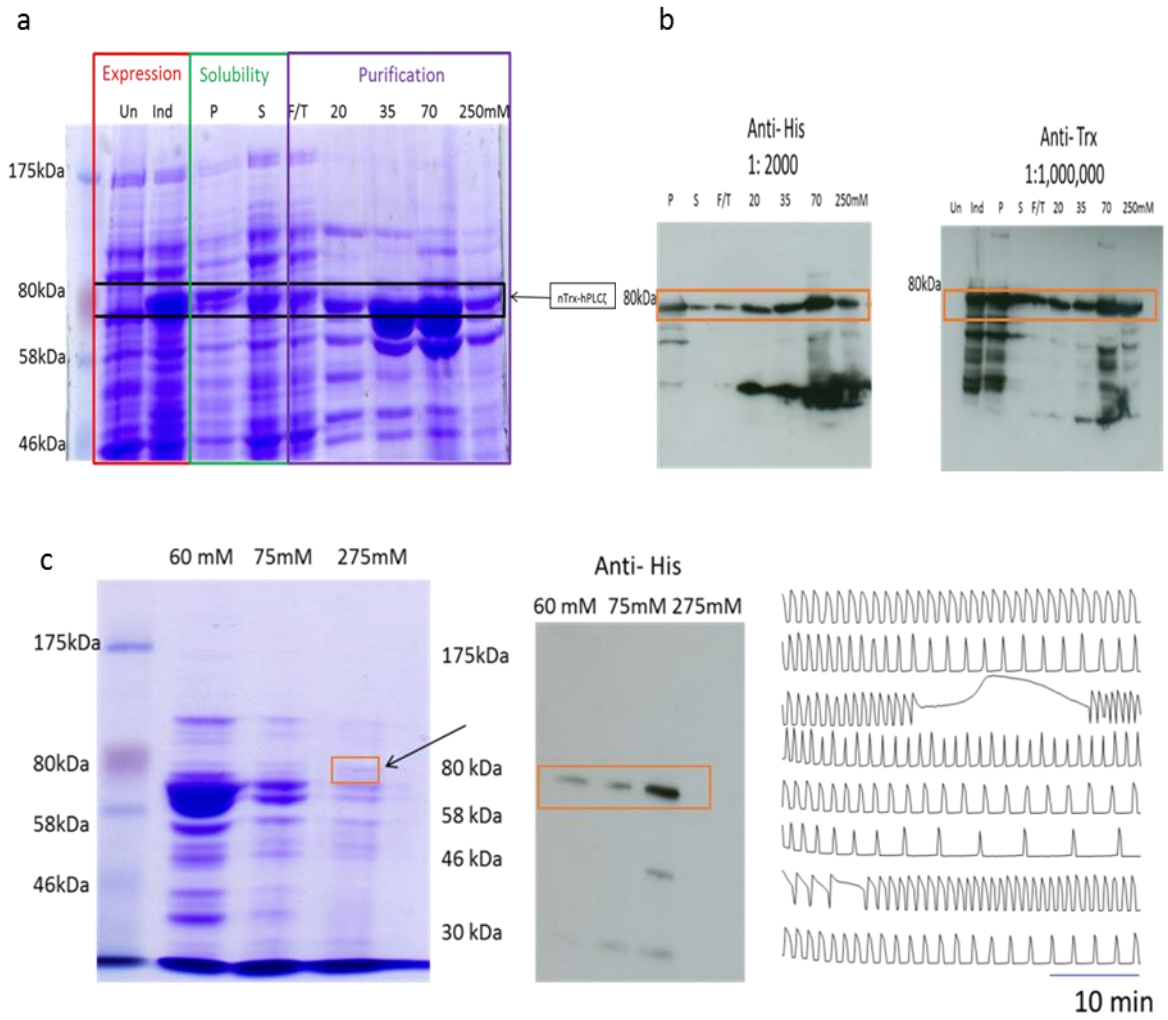


Figure 3.6 Evaluation of expression, solubility and activity of isolated recombinant nTRx-hPLCζ a) nTrx-hPLCζ has shown to have good expression Un:Uninduced, In:Induced and solubility p:pellet, s:soluble franction, F/T: flow through mM:Imidazole concentration. Although in the final 250mM elution multiple products found to be present. **b)** The presence of nTrx-hPLCζ was confirmed with Western blot using two different antibodies. **c)** nTrx-hPLCζ was concentrated using 3K Amicon centrifuge tubes. The coomassie blue stained gel illustrate the concentrated flow through at 60mM 75mM and 275mM Imidazole. The presence of the protein was confirmed with Western Blot and quantified with BCA protein assay illustrating the optimum concentration of Imidazole for protein elution. The elution was dialysed in HEPES buffer and injected to a number of mouse eggs resulting in calcium oscillations and therefore egg activation. Each Ca^{2+} oscillation trait represent oscillations in different mouse egg. Activity assay was performed injecting hPLCζ from the 275mM elution fraction. The microinjection experiments have been performed by Dr. Yuansong Yu.

3.3.2 Activity assessment in vivo of MBP-hPLCζ

MBP-hPLCζ has the advantage of containing MBP which is a protein solubility tag of size approximately 45kDa. In addition, this solubility partner allows the protein to be purified with amylose beads offering more purification options. In chapter 2 the expression and solubility have shown to be encouraging therefore it was further assessed for purity and activity. Activity assay following injection of purified MBP-hPLCζ illustrated multiple oscillations and active PLCζ protein (Figure 3.7).

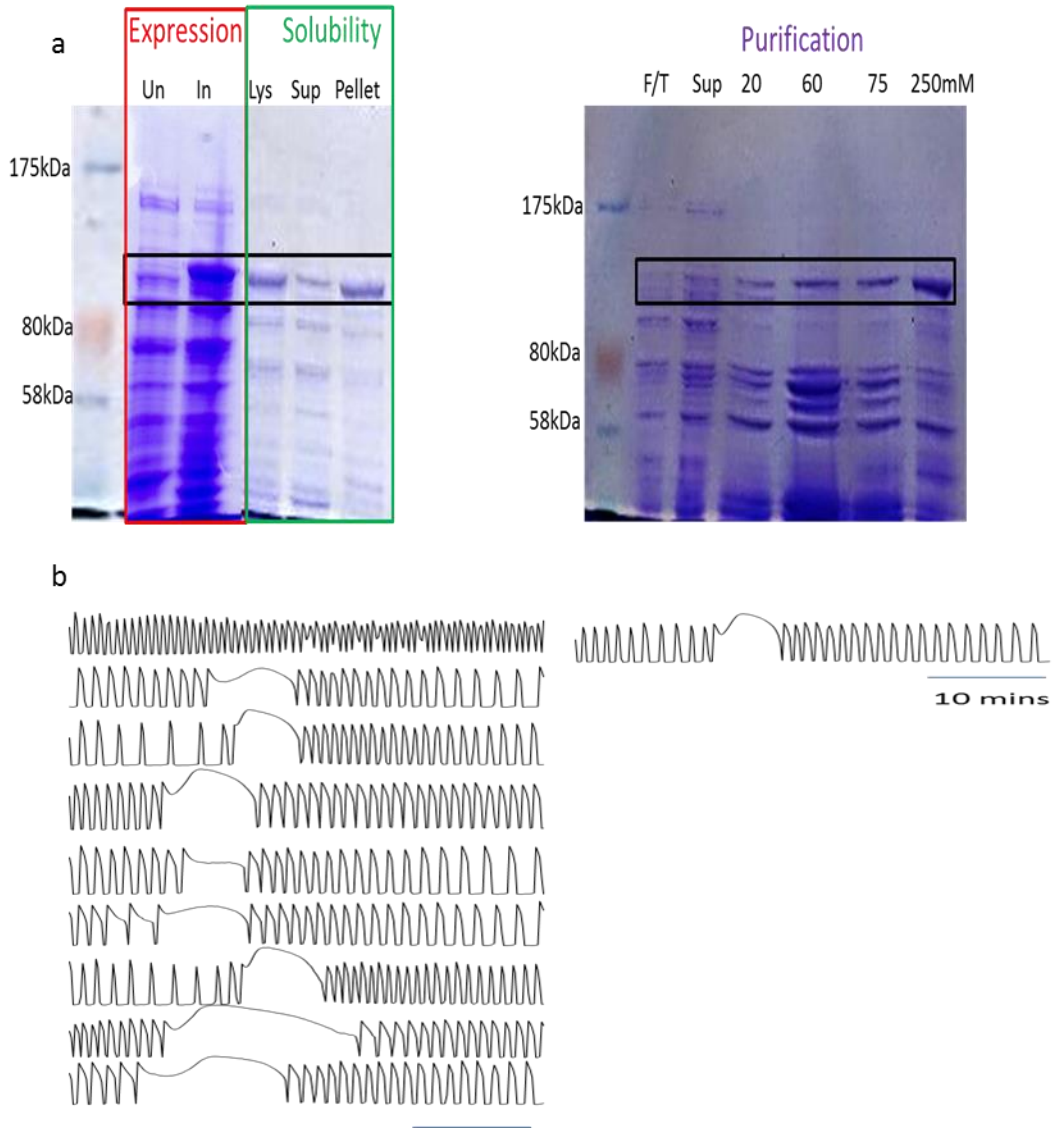


Figure 3.7 Evaluation of expression, solubility and activity of MBP-hPLCζ a) MBP-hPLCζ has been shown to be expressed well and with high solubility Un: Uninduced, In: Induced, Lys: bacterial lysate, Sup: Supernatant (soluble fraction). Although, multiple products were present in the 250mM elution indicating degradation during the purification and cell lysis, F/T: flow-through, Sup: supernatant-soluble fraction, mM: Concentration of Imidazole. **b)** MBP-hPLCζ was concentrated and dialysed with HEPES buffer and injected to a number of mouse eggs

resulting in calcium oscillations and therefore egg activation. Each Ca^{2+} oscillation trait represent oscillations in different mouse egg. Activity assay was performed injecting hPLCζ from the 275mM elution fraction. The microinjection experiments have been performed by Dr. Yuansong Yu.

3.3.3 Activity assessment in vivo of hTrx-hPLCζ

hTrx was sub-cloned into pET-MM11 which contains only the 6xhistidine tag for protein purification. hTrx is used to serve a bilateral purpose, initially a protein solubility tag but most importantly for its potential use to biomedical applications as hTrx is present naturally in the oocytes (Lopata *et al*. 2001). The protein was expressed using Rosetta De3 cells and purified with Ni-NTA affinity chromatography. The purified protein was concentrated and buffer exchanged following injection along with Fluo-3 calcium imaging dye into mouse oocytes. The protein was active exhibiting calcium oscillation for over an hour (Figure 3.8).

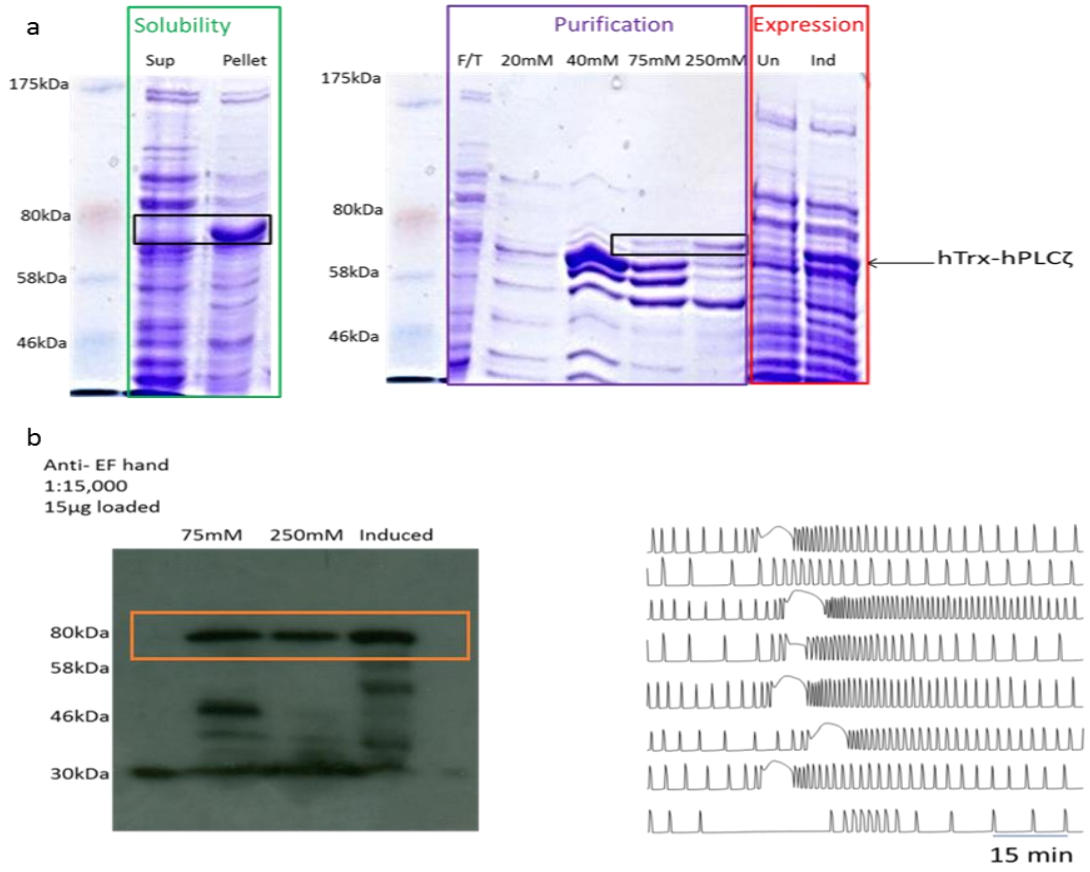


Figure 3.8 Evaluation of expression, solubility and activity of hTrx-hPLCζ. a) hTrx-hPLCζ was shown to have good expression, however the solubility was found to be poor Un: Uninduced, In: Induced, Sup: Supernatant (soluble fraction), F/T: flow- through, mM: Concentration of Imidazole. c) The presence of hTrx-hPLCζ was confirmed by Western blot

and microinjection of the protein in a number of mouse eggs, indicated its activity. Each Ca^{2+} oscillation trait represent oscillations in different mouse egg. Activity assay was performed injecting hPLCζ from the 250mM elution fraction. The microinjection experiments have been performed by Dr. Yuansong Yu.

3.3.4 Activity assessment in vivo of nTrx-hTrx-hPLCζ

The protein derived from the sub-cloning of hTrx into the His_6-hPLCζ vector was shown to have limited solubility. Therefore, hTrx was sub-cloned into nTrx-hPLCζ which contains both 6xHistidine tag and nTrx solubility tag, to combine the advantage of hTrx as a solubility tag that could potentially have a use in biomedical applications and the expression and solubility benefits of nTrx. The protein was purified, however a number of proteins appeared to be present possibly due to protein degradation as the bands appear to be recognised by both anti-Trx (N-Terminal of hPLCζ) and anti-hPLCζ (EF-hands) antibodies. Activity assay has illustrated multiple oscillation in a number of mouse oocytes injected with the protein (Figure 3.9).

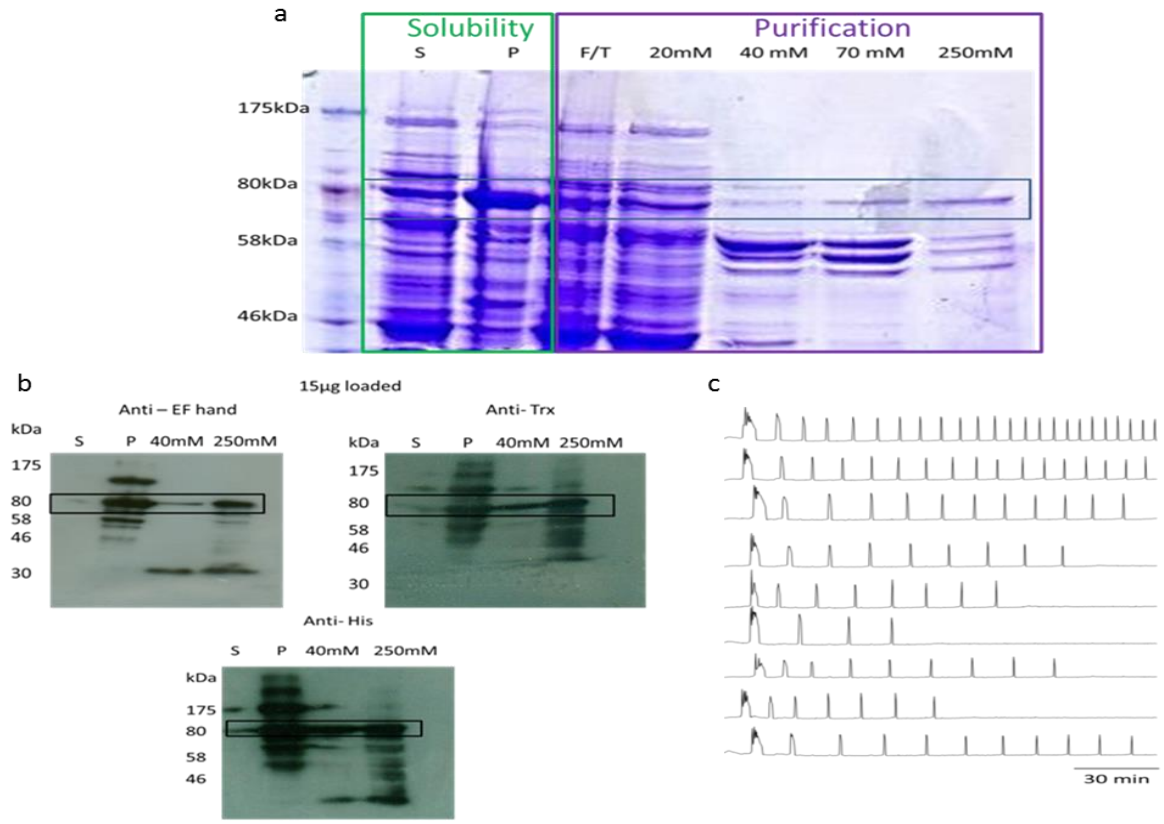


Figure 3.9 Evaluation of expression, solubility and activity of nTrx-hTrx-hPLCζ a) The protein is shown to be expressed in soluble form although at the 250mM Imidazole elution,

multiple products are present indicating degradation during purification p:pellet , s:soluble franction, F/T: flow through mM:Imidazole concentration. **b)** The presence of nTrx-hTrx-hPLCζ was confirmed with the use of two different antibodies the western blot was assessing for supernatant (soluble protein), pellet (insoluble protein) and 40mM and 250mM of Imidazole elution. **c)** The sample was concentrated, dialyzed in HEPES buffer and injected into mouse eggs resulting in calcium oscillations. Each Ca^{2+} oscillation trait represent oscillations in different mouse egg. Activity assay was performed injecting hPLCζ from the 275mM elution fraction. The microinjection experiments have been performed by Dr. Yuansong Yu.

3.3.5 Purification and TEV cleavage of MBP-hTrx-hPLCζ

The experiments illustrated above have shown that active soluble protein has been successfully expressed. The next step was to maximize the purity of the protein. Thus, hTrx was sub-cloned into MBP-hPLCζ vector which contains the MBP solubility tag. The presence of MBP allows the protein to be purified with more than one affinity chromatography strategy. The purified protein appears to have multiple products due to protein degradation (Figure 3.10). Although full length protein was still able to retain its solubility and structure.

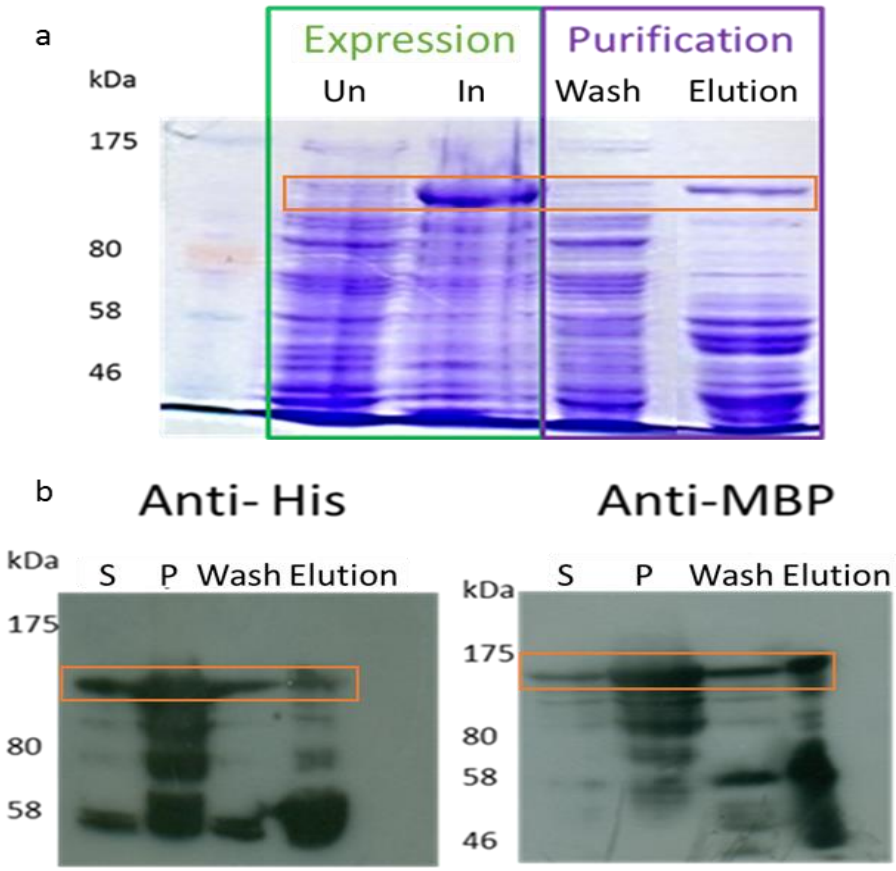


Figure 3.10 Purification and TEV cleavage of MBP-hTrx-hPLCζ a) The MBP solubility tag seemed to enhance expression and to offer a second purification option with amylose beads Un: Uninduced, In: Induced, S: Soluble fraction, P: Pellet. MBP-hTrx-hPLCζ was purified using amylose beads. **b)** The presence of MBP-hTrx-hPLCζ was confirmed with the use of 2 different antibodies S: Soluble fraction, P:Pellet.

3.3.6 TEV cleavage

TEV protease was expressed, purified and used for removal of the solubility tag that was attached to the protein. The protein after the cleavage seemed to remain soluble, with purity to be the last challenge before the protein could be potentially suitable for biomedical applications (Figure 3.11). The hTrx solubility partner has retained solubility following the removal of MBP and has been purified by Ni-NTA and Amylose beads which bind to His_6 and MBP respectively removing them from the elution as they have been cleaved from hPLCζ. The second 58kDa product that appears to be present is suggested to be bacterial protein possibly bound at the XY linker region.

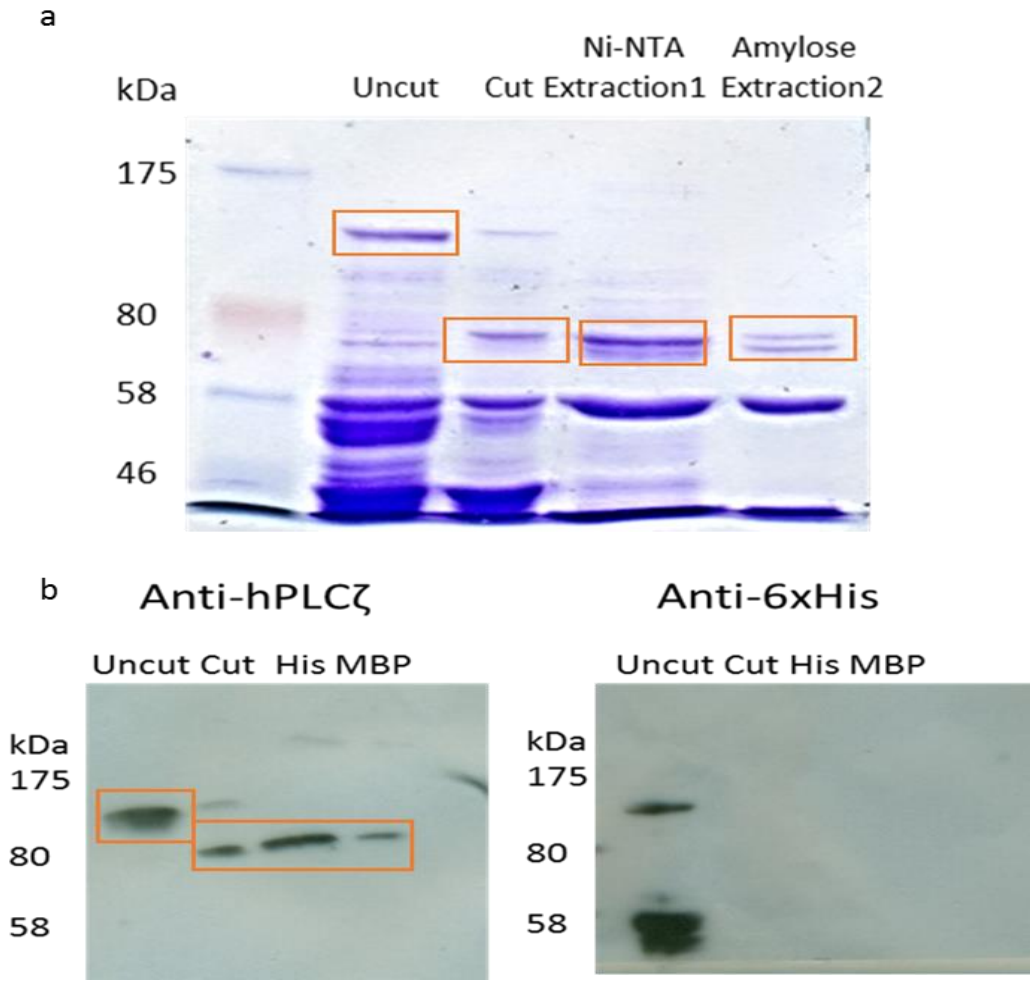


Figure 3.11 Removal of bacterial protein solubility tags a) TEV protease cleaves at the designed cleaving site of the protein removing the bacterial solubility partners from hPLCζ, resulting in a final construct of hTrx with hPLCζ. b) Protein was purified with amylose beads and then cleaved using TEV protease. The remaining hTrx-hPLCζ protein was then extracted using Ni-NTA beads and further extracted in order to maximize purity with amylose beads to remove any MBP related contaminants.

3.4 Discussion

In severe male factor causes of infertility intracytoplasmic sperm injection is performed with great results with over 70% success rates. However, about 30% of failing cases seem to be associated with absence of egg activation (Miao *et al.* 2012). hPLCζ is a protein that is solely present in spermatozoa and capable of triggering a series of calcium oscillations following its introduction into the egg (Saunders *et al*. 2002). Studies with calcium ionophores and IP3 revealed that they activate the egg by causing a single calcium peak and therefore fail to mimic physiological fertilisation (multiple oscillations), rising questions regarding the embryonic development and survival. Recombinant hPLCζ has been shown to trigger calcium oscillations similar to those of fertilisation and therefore mimic physiological conditions (Yoon *et al*. 2008). However, the drawback of recombinant hPLCζ was the presence of solubility tags of non-human protein origin. Eukaryotic protein expressed in bacterial cells often lose their functionality and solubility due to the lack of post translational modification machinery. Therefore, the presence of a solubility partner is inevitable. The presence of such bacterial solubility partners prohibit its use in fertility applications due to ethical issues.

3.4.1 Thioredoxin and TEV cleavage

The strategy of expressing hPLCζ with human Thioredoxin aimed to maximize protein expression and solubility without the necessity of removing protein solubility tags. Also, an expression of soluble hPLCζ with human Thioredoxin could potentially be ethically approved for biomedical applications, as Thioredoxin is naturally present in the human oocyte. The successful removal of the bacterial solubility partners and the increased solubility of the remaining hTrx-hPLCζ protein gives hPLCζ a great potential for future use in biomedical applications. Although, the construct was shown to express soluble protein, multiple products were found to be present following affinity chromatography protein purification. The causes could be potential degradation of the protein during expression and purification or the binding of a bacterial protein to the XY linker region. Possible solutions for achieving better purity could be further optimization of the purification process by using multiple purification steps such as Ni-NTA (6x His –tag) and Amylose (MBP-tag) or gel filtration in order to isolate the protein of interest.

3.4.2 Ca^{2+} oscillations

Physiological egg activation following sperm-egg fusion is characterised by multiple Ca^{2+} oscillations. The proteins successfully expressed and purified in this study were assessed for their activity *in-vivo* following microinjection of the purified protein into mouse oocytes. Activity assay has confirmed the active recombinant hPLCζ generating multiple oscillations similar to the physiological condition of egg activation. The concentration of recombinant purified protein microinjected was normalised however due to the purification impurities and contaminants, accurate target protein quantity measurement was impossible to be estimated. Thus Ca^{2+} oscillation patterns differ in each protein construct. The different pattern could be due to the different protein concentration injected or due to the different stability and folding of the protein with each solubility tag. In most of the constructs a characteristic loop has been observed approximately 15 minutes following the first spike which is hypothesised to be due to the sensitisation of $InsP_3R$. As the Ca^{2+} concentration it is increased is hypothesised that this may lead to an increase of $InsP_3R$ sensitivity to Ca^{2+} leading to a further increase of Ca^{2+} release *via* $InsP_3R$ (Hirose *et al.* 1999; Swann and Lai 2013) therefore the increased Ca^{2+} release may cause Ca^{2+} oscillations of increased frequency illustrated as a characteristic loop (Figure 3.12).

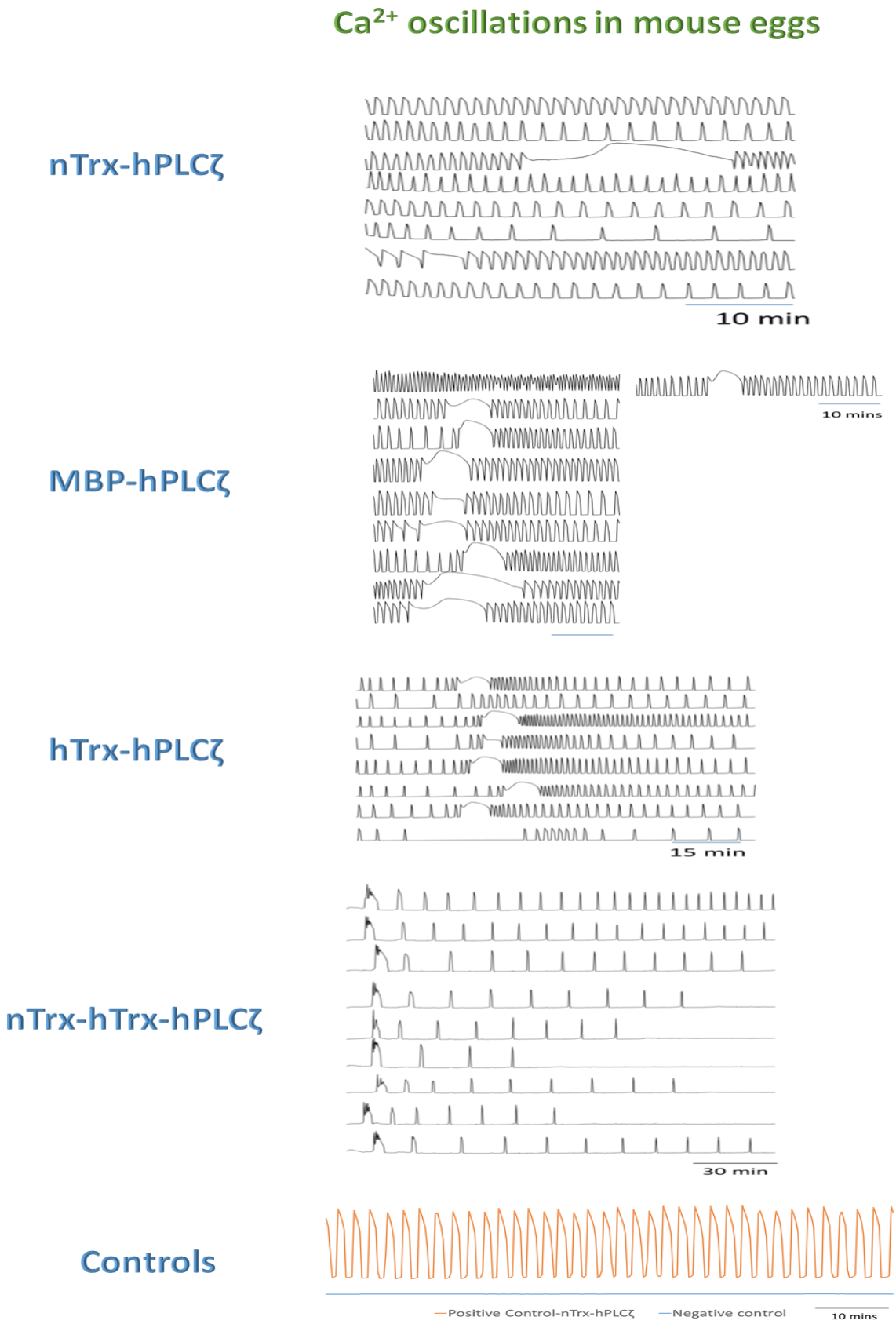


Figure 3.12 The different Ca^{2+} oscillation patterns observed in different protein construct of hPLCζ. Each oscillation line illustrates different oocyte. Microinjection experiments were performed by Dr. Yuansong Yu.

Chapter 4.The Design of a QPCR Based Genotyping Tool for a Novel InDel on Chr 20

4.1 Introduction

4.1.1 PLCβ1 deletion

A novel InDel in the 3rd large intron of the PLCβ1 gene on chromosome 20 (Chr20) was identified by Genome Wide Linkage Analysis (GWLA) followed by haplotyping and analysis of copy number variations plus DNA sequencing of a large kindred (described below) by Dr. Bakhsh. The InDel is present in a region on chr 20 which yielded a LOD score of 3.01 (described in chapter 6) in the GWLA. The deletion has a size of 1076bp and it is an InDel with an ATAA insertion at the junction (Figure 4.1). It is hypothesized that the insertion of the ATAA is a ''cut and paste'' event as an outcome of the transposable element activity. It has been found that there is an 11kb transposon cluster immediately upstream of the region of interest (Consortium 2012).

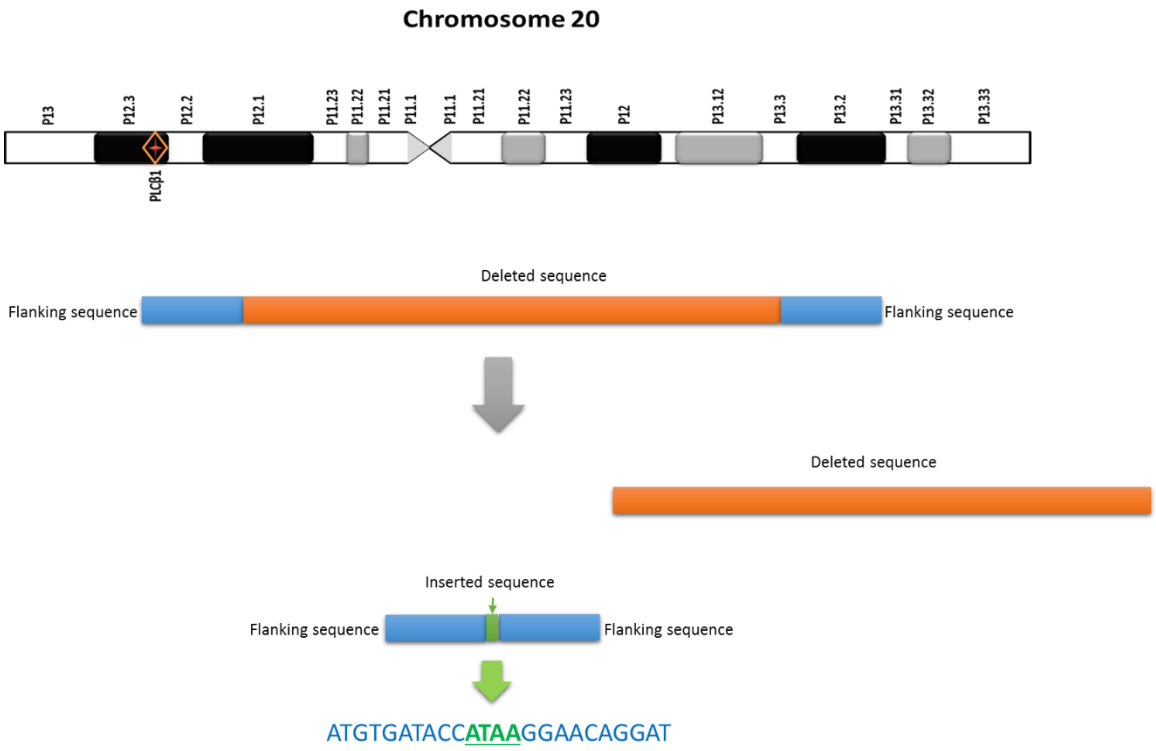


Figure 4.1 The InDel was identified at the 3rd large intron of PLCβ1 at Chr 20 with a 1076bp deleted region and an ATAA insertion at the junction.

4.1.2 Family History

The InDel was found to be present in a family with a strong history of Multinodular Goiters (MNG) (Figure 4.2) several members with MNG have subsequently developed papillary thyroid cancer (PTC). The mother of six children, III-2 suffered from MNG and has undergone partial thyroidectomy at the age of 17, followed by complete thyroidectomy for PTC at the age of 28. At the age of 41 she was diagnosed with breast carcinoma. Her son IV-1 was diagnosed with 9 hyperechoic cysts, fine needle aspiration (FNA) has suggested the development of papillary thyroid cancer thus a complete thyroidectomy was performed. He was also diagnosed with polycycstic kidney disease in infancy. His sister IV-2 was diagnosed with multiple hyperechoic thyroid nodules with FNA examination revealing the presence of multiple papilloid adenomata. Histopathological examination following total thyroidectomy has shown the presence of papillary hyperplasia. Her sister IV-5 was diagnosed with MNG but thyroid scan examination has shown small hypoechoic cysts and thus she was not referred for thyroidectomy. Her sister IV-6 was diagnosed with multiple hyper echoic nodules with FNA examination showing the presence of multiple papilloid adenomata. The presence of a large goitre, the multiple papilloid adenomata and the strong family history of thyroid cancer has led to a complete thyroidectomy at the age of 17. The sister IV-3 has undergone thyroidectomy for Graves' disease and her brother IV-4 had no thyroid malignancies or kidney lesions. The three relatives, I-1, II-2 and III-5 had undergone thyroidectomy for presumed benign thyroid disease with no further detail regarding their medical history or thyroid histology (Figure 4.2).

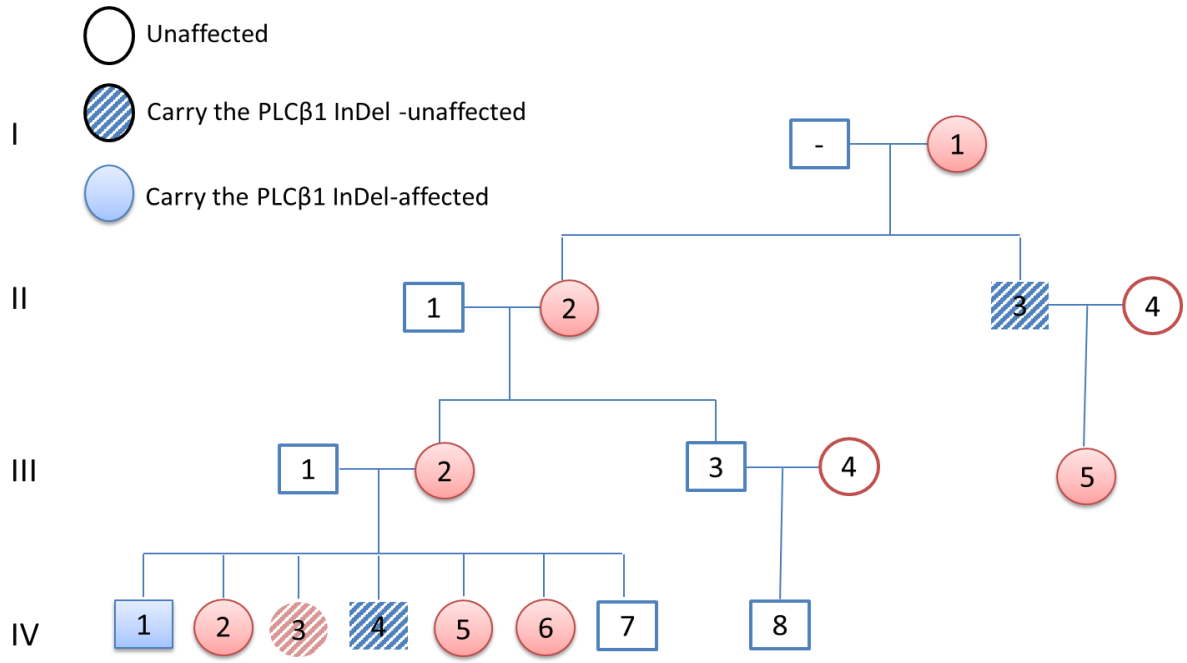


Figure 4.2 Family tree of the family identified with the PLCβ1 InDel. The patients with coloured background carried the InDel and were the same time affected by thyroid disease. Patients with a diagonal pattern carried the InDel but were not affected. The blue colour indicate males and the red females.

4.1.3 Prevalence studies

Previous study on the prevalence of the InDel by Dr. Bakhsh with standard PCR and *in silico* analysis has identified that the InDel was present in approximately 1% in the Caucasian population (3/285) (1/105 was identified following standard PCR genotyping and 2/180 by *in silico* analyses from databases for genomic variants) and in 5% of patients suffering with MNG (4/81) (Table 4.1). Of interest the InDel was not present in 70 patients having a solitary nodule that was subsequently diagnosed as papillary thyroid cancer.

Table 4.1. Prevalence studies performed on PLCβ1 InDel with PCR genotyping.

Group of Patients	Number of Patients	Number of patients found with the deletion	Prevalence (%)
Papillary thyroid cancer (PTC)	70	0	0
Caucasian – General Population	105	1	0.95
Multinodular Goiter (MNG)	81	4	4.94

4.1.4 Techniques to Measure Copy Number Variation (CNV)

There is a range of different techniques for detection and assessment of CNVs. The main methods include arrays, RFLP followed by Southern blot, qPCR, pyrosequencing, ligation detection reaction and the invader assay (Lee and Jeon 2008).

Microarray technology is widely used for CNV detection the use of which is either array-based comparative genomic hybridization or SNP-based microarrays (Lee and Jeon 2008). These two methods have been found to be ideal for detection of low level mosaicism being more accurate than cytogenetic tests. Although able to identify CNVs, microarrays seem to have a limited clinical application as they are less able to identify pathogenic duplications compared with their capacity to detect pathogenic deletions. In addition, microarrays cannot detect genomic alterations resulting in unchanged genomic material such as inversions and balanced translocations (Lee and Jeon 2008; Coughlin *et al.* 2012).

RFLP followed by Southern Blot analysis is considered the most traditional technique for screening CNVs ranging from 5 to 500kb. Even though it is a good method for detecting large CNVs within the duplication, it is an expensive, laborious and time consuming technique requiring more than a week for completion (Lee and Jeon 2008).

Pyrosequencing is another popular method for CNV screening. It has been developed by (Ronaghi *et al.* 1998) and its principle is based on the detection of the pyrophosphates (PPi) which are released following the binding of the dNTPs corresponding to the initial template located at the 3’ end of the newly synthesized strand (Lee and Jeon 2008). The limitations of this technique include its dependence on quality of the template DNA and the fact that it is an end point analysis and not a real time analysis with the consequent PCR requiring a thorough optimization (Lee and Jeon 2008).

Real time PCR on the other hand, is ideal for CNV detection as it is cost and time effective with high impact. The privilege of using an apparatus that combines a thermal cycle with the real time monitoring of spectral fluorescence, offers the unique advantages of assessing the PCR kinetics in real time without requiring further post – PCR analysis (Lee and Jeon 2008; D'haene *et al.* 2010) and so I opted to use this method.

4.1.5 Does the InDel have promoter activity? Reporter gene assay

In silico analysis of the PLCβ1 intronic InDel has found that there is an oestrogen receptor (ERα) α binding site within the deletion (Dunham *et al* 2012). The identification of the ERα binding site in combination with the increased prevalence of thyroid diseases, including MNG and PTC in the female population (Vanderpump *et al.* 1995) and the increased prevalence of the InDel in the female members of the family suggested a potential role for oestrogen in thyroid disease. These findings lead to a hypothesis that an ERα binding site could indicate the presence of silencer elements which can repress promoter activity and could either affect PLCβ1 transcript levels or regulate another transcript 'hidden' within the large intron. Also, an important finding that supports the hypothesis for the potential intrinsic role of ERα, was the discovery that 5% of the identified ERα binding sites (>1000) in the human genome are present within promoters in introns (Lin *et al.* 2007).

The reporter gene assay is the gold standard for this type of investigation. The principle of the technique is very simple and involves the cloning of the putative promoter sequence into an expression vector upstream of a reporter, for example the firefly luciferase gene. The vector is in turn transfected into cells so that on addition of the appropriate substrate, firefly luciferase can be directly detected indicating the presence of an active promoter. It is a very sensitive technique which can be used in a wide range of applications providing a quantitative luciferase signal (Allard 2008) (Figure 4.6).

4.1.6 MCF7

The cells selected for this study were the MCF7 breast cancer cell line. MCF7 were ideal for this study as they have the ability to respond to oestrogen in the form of estradiol, due to oestrogen receptors present in their nucleus. MCF7 breast cancer cell line originated from a 69 year old female from the Detroit area with metastatic breast cancer (Levenson and Jordan 1997). Characterisation studies with tamoxifen a drug with oestrogen competing activity has proved that MCF7 cells are hormone responsive and thus they provide a suitable cell line to study endocrine therapy in correlation with tumour response and the biological activity of oestrogen binding (Levenson and Jordan 1997).

4.1.7 AIMS

The aim of the project was first the generation of a qPCR based genotyping tool for the identification of the PLCβ1 InDel in large cohort of patients. Prevalence studies, along with the clinical information regarding the family affected by the deletion, suggested that it may cause benign proliferation and could potentially be used as a biomarker for MNG patients likely to develop PTC. However to confirm this it is necessary to analyse additional cohorts to provide a more accurate measurement of the prevalence of the InDel in healthy and diseased populations (Figure 4.3).

The second aim of this chapter was to investigate the putative promoter activity of the InDel and whether it can be modified by oestrogen. It has been identified that there is an oestrogen receptor α (ERα) binding site within the deletion, the increased transcript levels upon the deletion of that sequence suggested the potential presence of silencer elements of oestrogen receptor promoter, and thus promoter activity was investigated by luciferase assay (Figure 4.3).

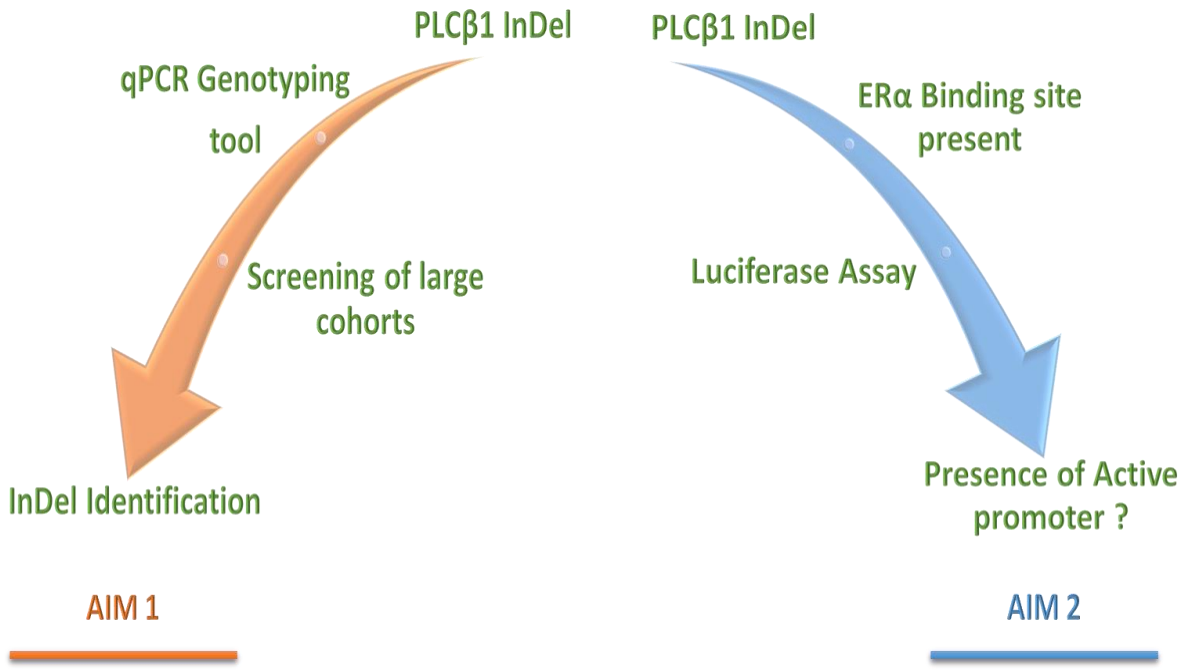


Figure 4.3 Aims of the chapter. The generation of a qPCR genotyping tool for the PLCβ1 InDel screening in patients and the identification of an active promoter within the InDel were the two aims of the chapter.

4.2 Materials and Methods

4.2.1 The design of a qPCR based genotyping tool

The necessity for a fast, cost effective and accurate tool for screening large cohorts of patients led to the optimisation and development of a qPCR based genotyping tool (Figure 4.4).

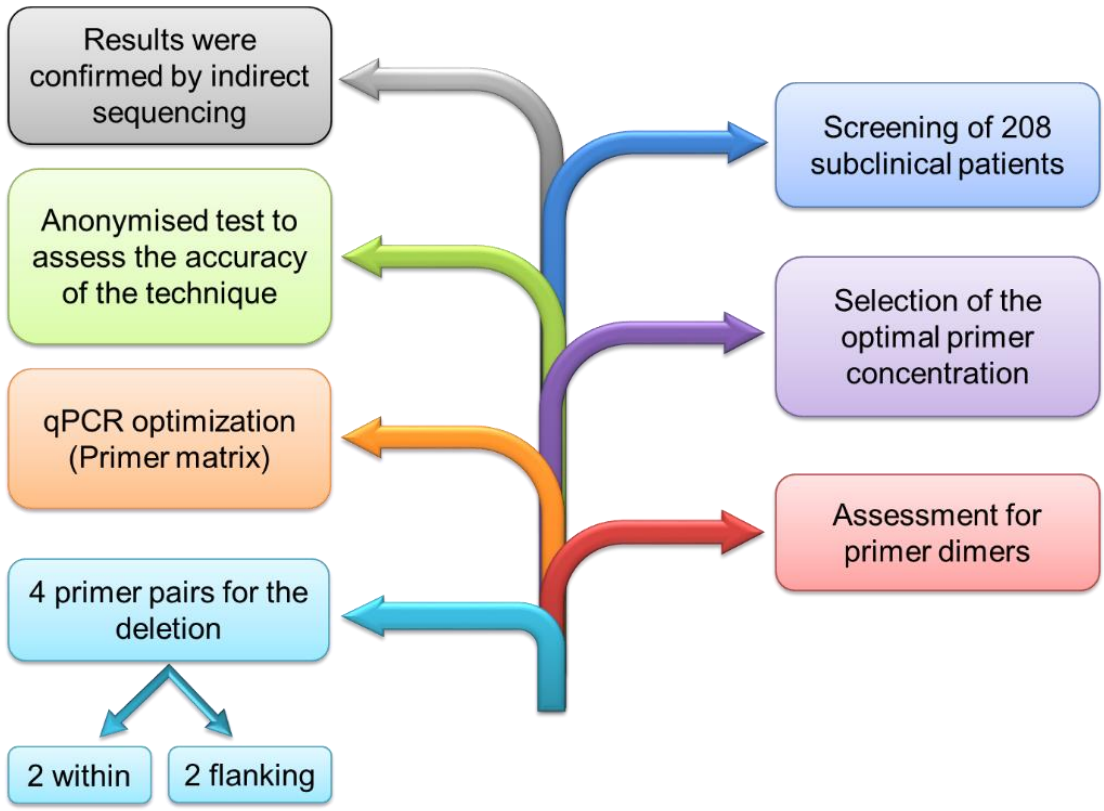


Figure 4.4 Flow chart of the genotyping of patients for the identification of the PLCβ1 InDel.

4.2.2 Extraction of Genomic DNA from blood

DNA was purified from whole blood with the use of QIAamp blood Midi Kit (Qiaqen) according to the manufacturer's protocol. Initially 100µl of Qiagen protease was placed in a 15ml centrifuge tube followed by 1ml of blood. Blood was mixed thoroughly for a minute and placed at 70 ^{o}C for 10 minutes. Then 1ml of 100% ethanol was added to the tube and mixed by inverting the tube 10 times. The mix was carefully transferred into the QIAamp Midi column and centrifuged at 3000rpm for 3 minutes. The QIAamp Midi column was removed and the

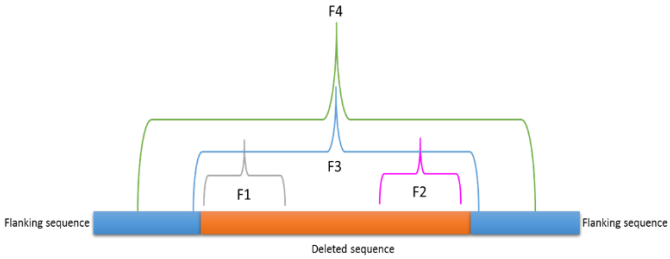
filtrate was discarded. The QIAamp Midi column was placed back in the 15ml centrifuge tube 2ml of Buffer AW1 was added and centrifuged at 5000rpm for 1 minute. Without moistening the rim, 2ml of AW2 buffer was added to the QIAamp Midi column and centrifuged at 5000 rpm for 15 minutes. The QIAamp Midi column was placed in a clean 15ml centrifuge tube and DNA was eluted by adding 200µl of distilled water and incubating for 5 minutes at room temperature. The filter was centrifuged at 5000 rpm for two minutes. The eluate was placed back in the column, incubated for another 5 minutes and centrifuged at 5000rm for two minutes to maximize DNA concentration.

4.2.3 Primer design and evaluation

Four pairs of primers were designed with the assistance of ‘’Primer 3’’ online primer design tool (http://bioinfo.ut.ee/primer3/), where two pairs of primers were designed within and two flanking the deletion (Table 4.3) and (Figure 4.5). The designed primers were also screened by BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure that there are no other unrelated significant homology sequences. Then a PCR amplification was performed (Table 4.2) on genomic DNA from family members with and without the deletion to evaluate the amplification efficiency. PCR amplification was performed according to (NEB) manufacturer’s protocol (2.5µl 10x ThermoPol reaction buffer, 0.5µl of 10mM dNTPs, 10mM forward primer, 10Mm reverse primer, 0.125µl of 5U/µg *Taq* DNA polymerase (NEB) and 0.1-1µg of genomic DNA template in a total reaction volume of 25µ) (Table 4.5). The amplified DNA products were evaluated following 1% agarose gel electrophoresis (Chapter 2). The DNA bands were extracted with the use of a gel extraction kit (Qiagen) and confirmed by Sanger sequencing as described in (Chapter 2).

Table 4.2 PCR amplification reaction

	Temperature °C	Time
	95	30 seconds
30 cycles	95	30 seconds
	60	45 seconds
	72	1 minute
	72	5 minutes
	4	∞



Primer set	Expected size (bp)
F1/R1	211
F2/R2	184
F3/R3	1159/79
F4/R4	1551/471

Figure 4.5 The design of the qPCR based genotyping tool. The four sets of primers were initially designed 2 within and 2 flanking the deletion, due to the heterozygosity of the deletion the primers flanking the deletion amplified two products as only one allele carries the deletion.

Table 4.3 List of primers designed for development of the qPCR based genotyping tool.

Primer Name	**Tm**	**Sequence**	**Tm**
F1/*R1*	60.5	GTGGATTGGGACCTGCAATT/TTGTTCTGATCATGCTCCTCA	57.8
F2/*R2*	55.7	ACACTGTGACTGGGTGTGAA/CTGTCCCCTAGGTCTTCGTG	58.18
F3/*R3*	59.5	TTTCAGTGCCATTGGTAGCC/AGACTTGCATTCTCAGCCAT	55.9
F4/*R4*	55.2	ACTACACCTTCCTGCCAGTC/CAGGCAAGAACCATGGTACC	58.9

4.2.4 qPCR optimization

The concentration for each primer pair, was optimized with the use of a primer matrix where 9 combinations of 3 different forward and reverse primer concentrations were assessed (Table 4.4). Each of the cDNA samples was quantified with Nanodrop (Nanodrop-Lite Thermo-Scientific) and 100ng were loaded in each well of a 96 well PCR plate (Agilent Technologies) for the qPCR reaction. The qPCR reaction mix was added to appropriate well of the 96 well plate with the genomic DNA and then, placed into the MX3000P® thermocycler for qPCR amplification (Table 4.5).

Table 4.4 Primer matrix with 9 different primer concentration combinations.

Forward (nM)	100	100	100	300	300	300	500	500	500
Reverse (nM)	100	300	500	100	300	500	100	300	500

Table 4.5 qPCR amplification conditions a) qPCR reaction mix b) Amplification programme

a

Reagent	Volume
Genomic DNA (100ng)	-
2xSyBR green master qPCR mix(Invitrogen)	12.5µl
Primer stock mix (10µM-Forward and Reverse)	0.5µl
Distilled Water	Up to 25µl

b

Cycles	Steps	Temperature (°C)	Time
1	1	50	2 min
	2	95	2 min
40 cycles	3	95	15 sec
	4	60	30 sec
1	5	95	1 min
	6	55	30 sec
	7	95	30 min

4.2.5 Anonymised test

The ability of the tool to identify the deletion in random patients was evaluated by performing an anonymised test. The test included six DNA samples from the affected family, three of which were known to harbour the deletion. The test was performed after anonymisation of each sample and performing qPCR amplification see (Table 4.5). The reaction contained 100ng of gDNA and both primers within and flanking the deletion were assessed for their ability to identify the InDel.

4.2.6 Additional cohorts

The genotyping tool was used for the screening of 208 subclinical hypothyroid patients the genomic DNA of which was provided by Dr. Anna de Lloyd and were screened for the presence of the 1076bp intronic deletion. The primer pair that was selected for its accuracy from previous optimisation experiments was the pair within the deletion. The cohort was screened twice with these primers, however due to the high number of positives, another two screenings were performed with primers flanking the deletion for the identification of potential false negative samples. Samples found to harbour the InDel by qPCR were confirmed by Sanger sequencing as described in (Chapter 2).

The previously optimized qPCR tool with primers flanking the deletion was used for the screening of additional cohorts comprised of a total of 36 patients. Genomic DNA from this group of patients was provided by Dr. Dagmar Fuhrer (Table 4.9).

4.2.7 Reporter Gene Assay

In Silico analysis of the deletion has identified an Oestrogen receptor α (ERα) binding site within the deletion. Therefore, a luciferase assay was designed to assess whether an active promoter is present within the InDel region (Figure 4.6).

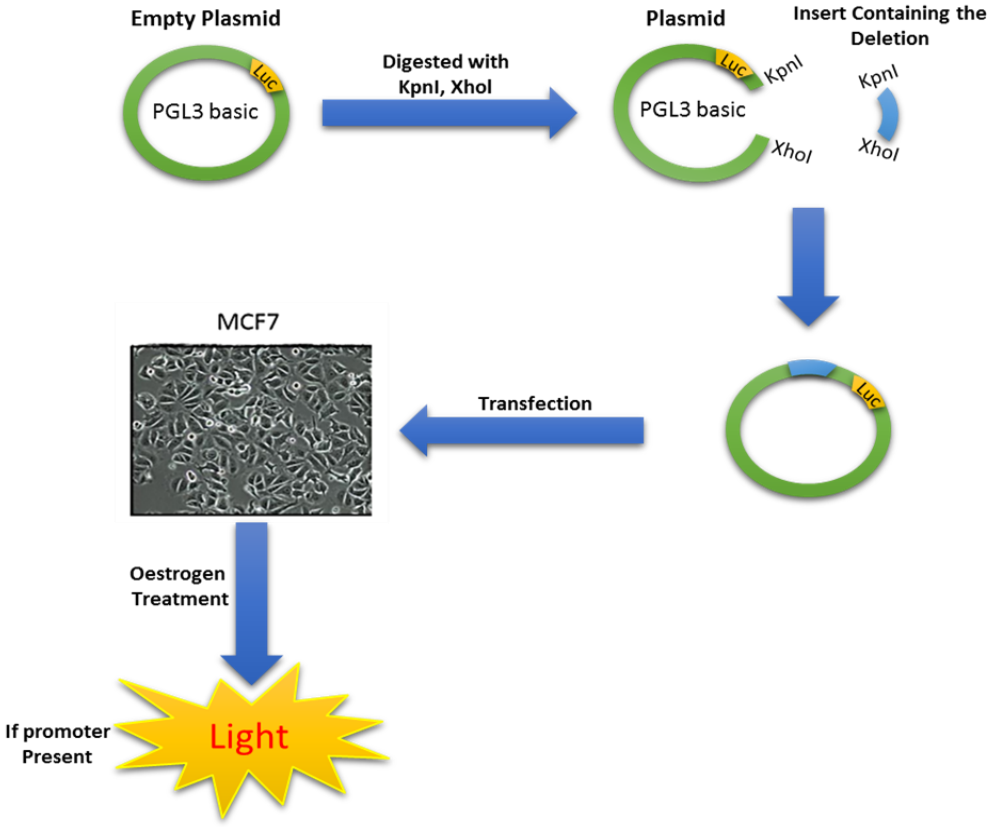
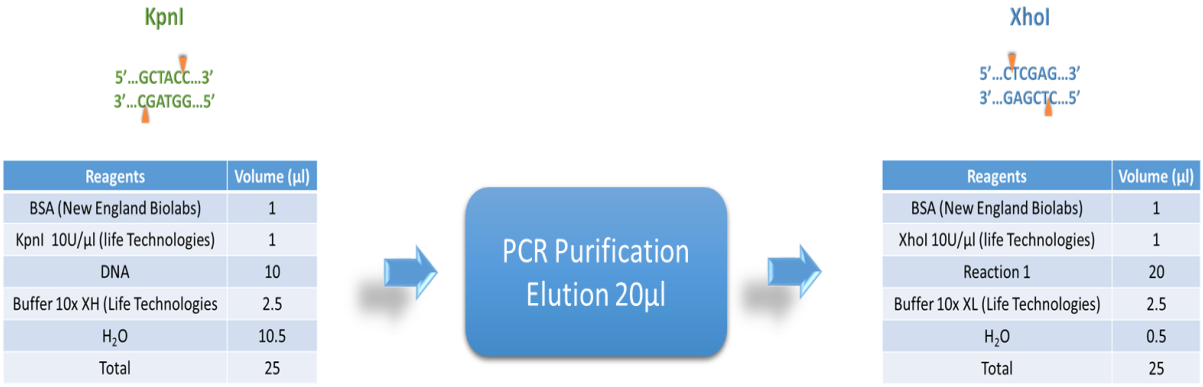


Figure 4.6 Flow chart of the reporter gene assay in MCF7 cells. Molecular cloning of the deletion region into the PGL3 basic vector and transfection of the plasmid into MCF7 cells was followed by Oestrogen treatment and Luciferase detection.

4.2.8 Molecular Cloning

The primers flanking the deletion F4/R4 (Table 4.2), amplifying a 1551bp product were used for the cloning of the PLCb1 deletion region into the PGL3 basic vector (Promega). Investigation with NEBcutter software (http://nc2.neb.com/NEBcutter2/) of the region of interest has found naturally occurring restriction sites for KpnI and XhoI, therefore both PGL3 basic vector and amplified DNA were digested with KpnI and XhoI enzymes (Figure 4.7). Initially, the region of interest was amplified with the addition of the following reagents into a

500 µl reaction tube: 2.5µl of 2x phusion master mix (NEB), 0.5µM forward primer, 0.5 µM forward primer, 1µl of template DNA and nuclease-free water (Fisher Scientific) up to 50µl total reaction volume (Table 4.2). Amplified products were then purified with the PCR purification kit (Qiagen). Molecular cloning was performed as described in (Chapter 2). Conditions for PCR amplification and Ligation are illustrated in (Table 4.6 and 4.7). Successful cloning was confirmed by Sanger sequencing. Successful incorporation of target sequence and the presence of Firefly Luciferase gene were confirmed by PCR amplification (similar conditions as described above) (Table 4.2). Firefly Luciferase primers (Table 4.8) were kindly provided by Dr. Martyn Bullock.



Reagents	Volume (µl)
BSA (New England Biolabs)	1
KpnI 10U/µl (life Technologies)	1
DNA	10
Buffer 10x XH (Life Technologies	2.5
H_2O	10.5
Total	25

Reagents	Volume (µl)
BSA (New England Biolabs)	1
XhoI 10U/µl (life Technologies)	1
Reaction 1	20
Buffer 10x XL (Life Technologies)	2.5
H_2O	0.5
Total	25

Figure 4.7 Sequential restriction digestion of the PLCβ1with KpnI and XhoI enzymes.

Table 4.6 PCR amplification programme

	Temperature °C	Time
	98	30 seconds
30 cycles	98	30 seconds
	55	1 minute
	72	2 minutes
	72	7 minutes
	4	∞

Table 4.7 Ligation conditions for cloning of the region of interest into PGL3 basic vector.

Reagents	Volume (µl)	Incubation
Ligation Buffer 10x (NEB)	3	16°C Overnight
T4 Ligase (NEB)	2	
Insert	4	
Vector	3	
Distilled H_2O	18	

Table 4.8. Firefly Luciferase primers

Primer Name	Tm	Sequence	Tm
LucF/*LucR*	62	GTGTTGGGCGCGTTATTTAT/CATCGACTGAAATCCCTGGT	63

4.2.9 Cell Culture

4.2.9.1 MCF7 cells

The cells were kindly provided by Dr. Illaria Muller (Cardiff University, UK). The cells were grown in 75 cm^2 Nuclon™ delta surface flasks (Nunc, UK) in 12 ml culture medium (DMEM - 4.5g/L glucose, L- glutamine without sodium pyruvate, 100U/ml potassium penicillin, 100µg/ml streptomycin sulphate, 1mM sodium pyruvate, 10µg/ml bovine insulin and 10% heat-inactivated foetal calf serum (FCS) (EU approved). The reagents were all acquired from (Lonza UK). The cells were incubated at 37 ^oC and 5% CO_2 incubator.

4.2.9.2 Trypsinisation

The cells were passaged at 80% confluency. Culture medium was removed and cells were washed with 10ml sterile PBS. Then 3ml of trypsin (Lonza UK) was added and cells were incubated at 37 ^oC for 10 minutes and cells were released following tapping of the flask. Then 5ml of culture medium (described above) was added and cells were placed into a 50ml falcon tube and centrifuged at 1000rpm for 5 minutes at room temperature. Supernatant was discarded and pellet was resuspended with 1ml of culture medium. In a new flask 20% of the cells were reseeded in 12ml of culture medium (described above) and incubated at 37 ^oC and 5% CO_2 incubator.

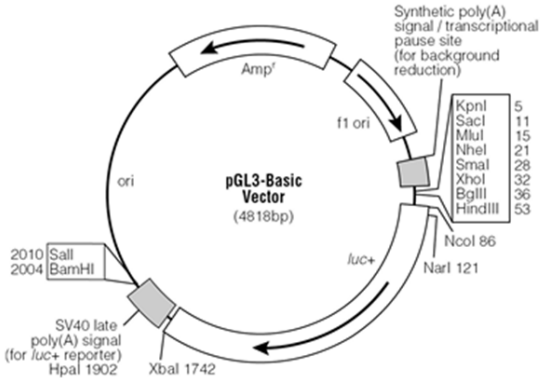
4.2.9.3 Cryopreservation of cells

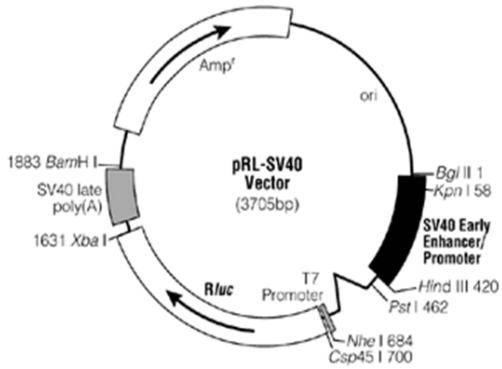
The cells were trypsinised as described above and cell pellet was resuspended with 1ml of freezing media (60% basal media, 20% FCS, 20% DMSO). The cells were aliquoted in 2ml cryotube vials (Thermo Scientific) which in turn were placed in insulating polystyrene freezing container with isopropanol (NALGENE). The freezing container was placed at -80 ^oC where

cells were frozen at a cooling rate of 1 ^{o}C/min. The cells following overnight freezing were transferred to liquid nitrogen in -190 ^{o}C.

4.2.9.4 Transfection

48 hours prior to transfection, MCF7 cells were transferred and maintained in phenol red free Dulbecco's modified Eagle's medium F12 containing 5% charcoal stripped foetal bovine serum (Lonza UK) and 2% (100U/ml potassium penicillin, 100µg/ml streptomycin sulphate). The cells 24h prior to transfection were trypsinised, counted, about 20,000 cells which is about 70% confluency were plated in the 96 well plate (Thermo Scientific) and incubated at 37 ^{o}C and 5% CO_2 incubator. For each transfection 150ng of PGL3-basic vector, 3ng of pRL vector (Promega) (Firefly Luciferase: Renilla Luciferase – 50:1) (Figure 4.8) and Transfast^{TM} reagent (Promega) with ratio (Transfast reagent: DNA – 3:1) were mixed with 50µl phenol red-free antibiotic-free and serum free QIBCO® Dulbecco's Modified Eagle Medium: Nutrient Mixture:DMEM/F-12 (Life Technologies). The mix was incubated at 15 minutes at room temperature. Culture medium was removed from cells following by washing with PBS. Then transfection mix was added on top of the cells drop-wise followed by incubation at 37 ^{o}C and 5% CO_2 incubator for 1 hour. Each well was then topped up with 150µl of phenol red-free QIBCO® Dulbecco's Modified Eagle Medium: Nutrient Mixture:DMEM/F-12 (Life Technologies) with 10% charcoal stripped fetal bovine serum (Life Technologies) (EU approved) and incubated at 37 ^{o}C and 5% CO_2 incubator overnight.





Strain/Plasmid	Description	Source/ Reference
pGL3-Basic Vector	Vector without promoter for assessment of promoters and enhancers activity with luciferase assay. Encodes Firefly luciferase	Promega
pRL-SV40 Vector	Encodes Renilla luciferase and is an Internal control for transfection efficiency of Firefly luciferase	Promega

Figure 4.8 List of plasmids that have been used in reporter gene assay

4.2.9.5 Luciferase Assay

The cells were incubated at 37^{o}C for 48h following medium change and treatment with β-Estradiol (Invitrogen) at 15, 30 and 60 pM. Estradiol was initially dissolved in absolute ethanol 1.83μM stock and sequentially diluted in Dulbecco's modified Eagle's F12 phenol red free medium. Treated cells were incubated at 37^{o}C for 24h, then the medium was replaced by 1x Passive Lysis Buffer (Promega) and stored at -80^{o}C overnight. The lysed cells were defrosted at room temperature with gently shaking with 50μl of cell lysate to be used for Dual Luciferase Reporter Assay System with the use of Promega Glomax multi detection system according to the manufacturer's protocol.

4.3 Results

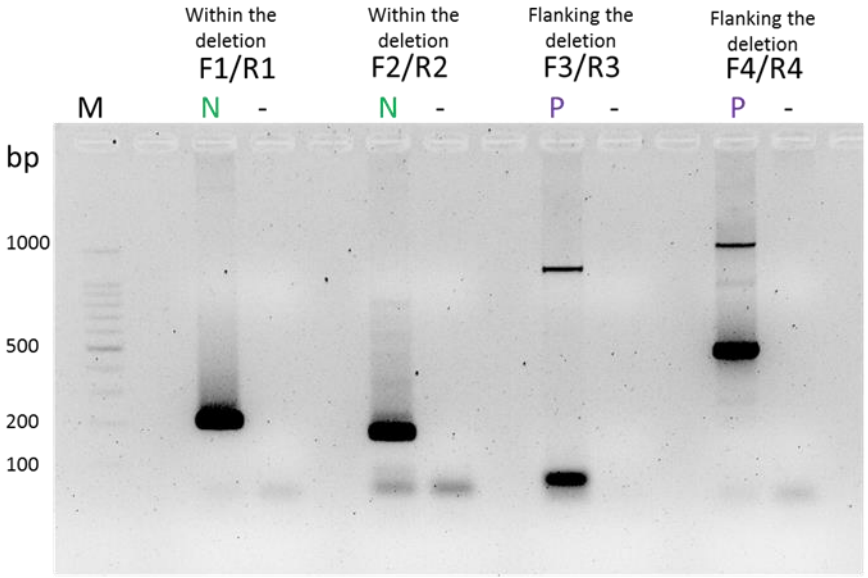
4.3.1 The design of a qPCR based genotyping tool

A novel InDel (1076 bp deletion with ATAA insertion at junction) has been identified in the large 3rd intron of the PLCβ1 gene on chromosome 20 using GWLA in a large kindred. The deletion has been shown to have a prevalence of about 1% in Caucasian population and approximately 5% in patients with euthroid multi-nodular goiter (MNG). Nodular thyroid disease is common and a risk factor for thyroid cancer. We hypothesize that this deletion could potentially be a marker for patients with MNG most likely to develop papillary thyroid cancer (PTC).

4.3.2 Primer pair Evaluation

A genotyping technique using qPCR was designed and optimized in order to screen a large number of patients and identify the prevalence among MNG & PTC patients and the general population. Four sets of different primers were designed; 2 within and 2 flanking the deletion. The qPCR was optimized using a primer matrix with 9 different primer concentration combinations. The expected amplification product size for the primers within the deletion was 211 (F1/R1) and 184 (F2/R2) base pairs and for the primers flanking the deletion was 1159 (F3/R3) and 1551 (F4/R4) (Figure 4.9).

a



b

N =Negative
P =Positive
- =Negative control
M=Marker

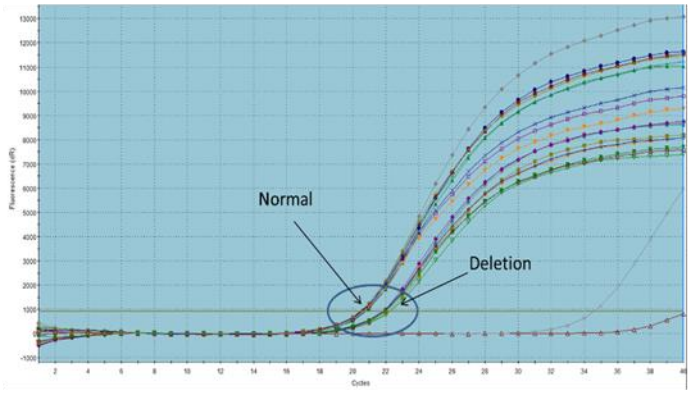
Primer set	Expected size (bp) Normal	Expected size (bp) InDel
F1/R1	211	-
F2/R2	184	-
F3/R3	1159	79
F4/R4	1551	471

Figure 4.9 Assessment of designed primers. a) DNA from patients carrying the deletion (P), patients without the deletion (N) and (-) negative PCR control without DNA template. The primer pairs flanking the deletion amplified two products due to the heterozygosity of the deletion. b) The expected sizes of the PCR amplified products. F: Forward primer and R: Reverse primer.

4.3.3 qPCR optimisation

The technique was then optimised using a primer matrix as described above using patients known to carry the deletion and patients without the deletion. The result was positive with all the true positives identified by the technique. It was found that the primers within the deletion produced a Ct value difference of 1 between subjects with and without the deletion. The optimal combination selected for F1/R1 primers were F1:500nM/R1:100nM. On the other hand, the primers flanking the deletion had a difference in Ct value of ~10 (Figure 4.10).

a



b

	1	2	3	4	5	6	7	8	9
Ct Value Difference F1/R1	2.01	2.03	1.48	1.92	1.9	1.275	1.945	1.57	1.815
Dissociation Curve	+++	+++	+--	++++	+++	+--	+++	+--	---
Ct Value Difference F2/R2	1.46	1.465	1.325	1.49	1.375	1.33	1.16	1.255	1.345
Dissociation Curve	+++	+++	+++	+++	---	---	+++	---	---

c

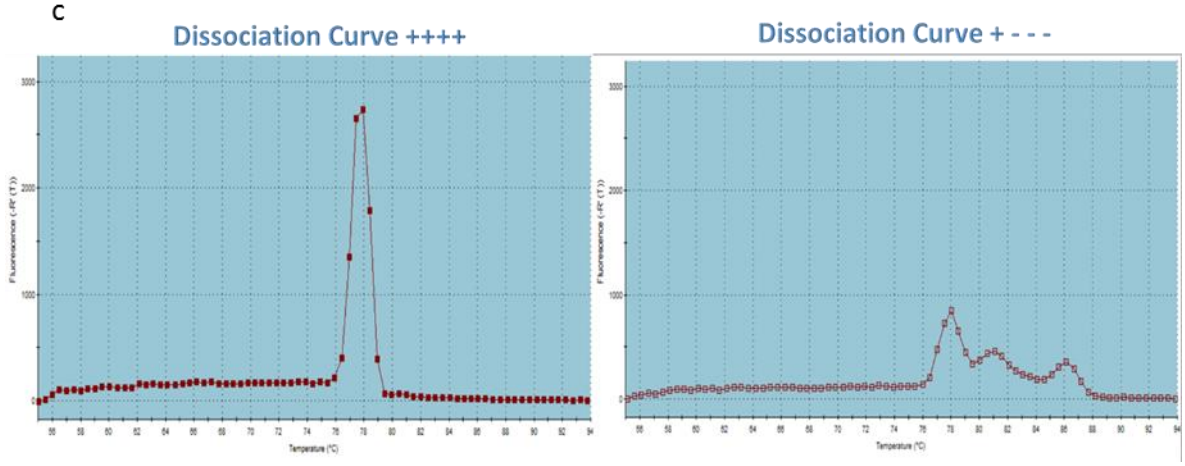
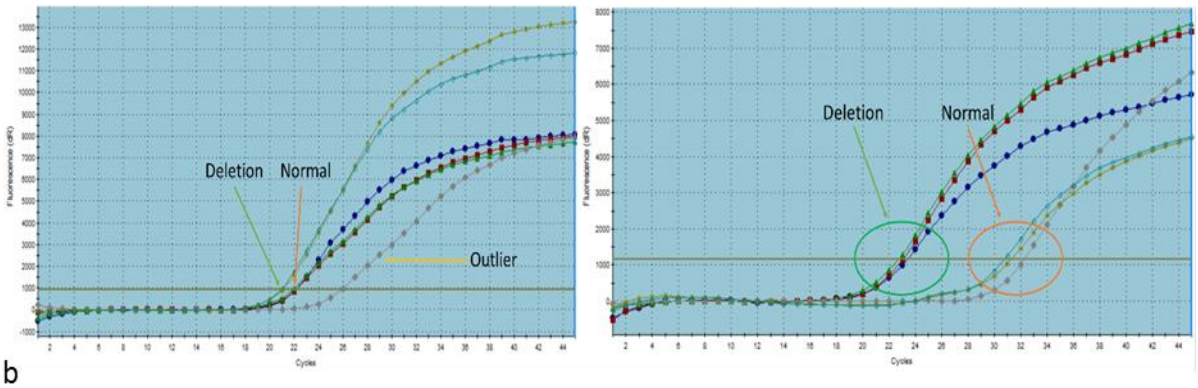


Figure 4.10 qPCR optimisation. A) Primer matrix with F2/R2 primers within the deletion. B) Evaluation of the primer concentration due to Ct value difference and quality of the dissociation curve. C) Example of the ideal dissociation curve for a primer pair.

4.3.4 Anonymised Test

The efficiency of the technique was then assessed using 6 anonymised samples, 3 with and 3 without the deletion. The samples carrying the deletion were successfully identified with only one outlier due to the old age of the blood samples (Figure 4.11). The primer set (F1/R1) within the deletion were then chosen to screen a large cohort of 208 subclinical patients.

a



b

Samples	A	B	C	D	E	F	Samples	A	B	C	D	E	F
F1/R1	22.26	22.23	22.02	26.02	21.18	21.02	F3/R3	23.35	22.96	22.66	32.26	31.25	30.70
F2/R2	22.19	22.13	22.06	26.37	21.26	21.05	F4/R4	21.78	21.55	21.50	35.11	23.79	25

Figure 4.11 Evaluation of the anonymised test. a) Anonymised test with primers within the deletion F1/R1 and flanking the deletion F3/R3 in 6 patients 3 of whom were positive for the deletion. b) The Ct value illustration for primers within (F1/R1 and F2/R2) and flanking (F3/R3, F4/R4) the deletion.

4.3.5 Screening of 208 subclinical hypothyroid patients

Two hundred and eight patients with subclinical hypothyroidism (SH) were genotyped using primers F1/R1 within the deletion. This identified 15 subjects having a difference of 1 Ct and who were presumed to be positive for the InDel. The 15 positives were screened with the primers (F4/R4) which flank the InDel, in order to confirm the true positives; only 2 out of 15 were confirmed. Consequently, the entire SH cohort was screened again using (F4/R4) primers which flank the InDel, for the identification of any potential false negative patient (Table 4.9). The screening identified one false positive, making the flanking the deletion primers ideal for screening large cohort of patients. The patients found carrying the deletion were 2 males and one female with one having a history of goitre and thyroid cancer in the family. The presence of the deletion was confirmed by Sanger sequencing as described in Chapter 2.

4.3.6 Additional Cohorts

Additional cohorts comprised 36 patients but subdivided into 4 groups, 17 with colloid nodules, 5 with follicular adenomas, 8 with papillary thyroid carcinoma, 6 with hot nodules were screened for the presence of the PLCβ1 InDel. The genotyping identified one patient with follicular adenoma carrying the deletion (Table 4.9). Previous studies by Dr. Grennan Jones, revealed that the deletion increased PLCβ1 transcript levels therefore, is hypothesized that the deletion may have an impact on the cell proliferation (benign) and this will be further investigated in chapter 5.

Table 4.9. The different cohort of patients genotyped by the qPCR based tool.

Group of Patients	**Total number of patients**	**Number of patients found with the deletion**
Number of Subclinical hypothyroid patients	208	3
Number of patients with colloid nodules	17	0
Number of patients with Follicular adenoma	5	1
Number of patients with papillary thyroid carcinoma	8	0
Number of patients with hot nodules	6	0

4.3.7 Reporter gene assay

4.3.7.1 Cloning of PLCβ1 deletion into PGL3 basic vector

The PLCβ1 deletion region was amplified with F4/R4 flanking the deletion primers generating a 1551bp product. The DNA fragment was digested with KpnI and XhoI restriction enzymes and cloned successfully into PGL3 basic vector (Figure 4.7). The region of interest was also confirmed by PCR using flanking the deletion primers F3/R3 generating an 1159bp product. Quality control of the plasmids used for luciferase assay included assessment of the presence of Luciferase gene by primers generating a 200bp product (Figure 4.12).

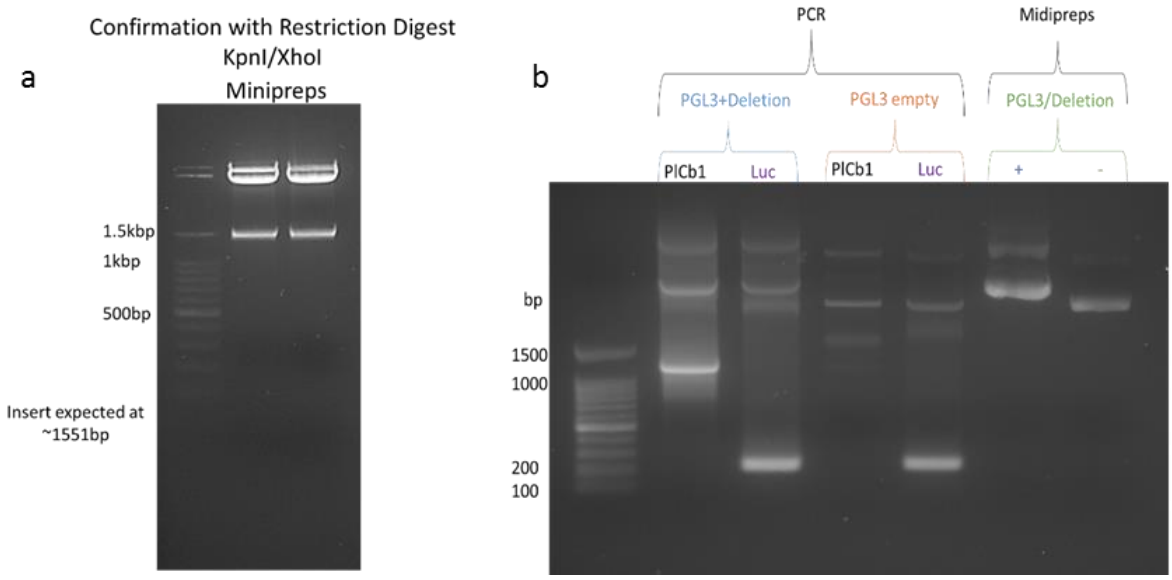


Figure 4.12 Cloning of PLCβ1 deletion into PGL3 basic vector. a) Confirmation of successful cloning of the deletion region into the PGL3 basic vector by restriction digestion with KpnI and XhoI. b) Confirmation of successful cloning and intact Luciferase gene by PCR amplification using F3/R3 PLCβ1 flanking the deletion and Luciferase primers respectively.

4.3.7.2 Luciferase assay

MCF7 cells which are known to be Oestrogen receptor alpha positive breast cancer cells, were co-transfected with PGL3 basic vector containing the deletion region and pRenilla for transfection efficiency positive control. Renilla Luciferase, αGCU-Luciferase which contains a glycoprotein hormone alpha subunit proximal promoter (Chatterjee *et al.* 1991) and PGL3 basic vector were used as positive controls for transfection efficiency. The significant luciferase emission from these positive control vectors confirmed the successful transfection (Figure 4.12). However, the results from the various treatments indicated either the absence

of an active promoter or the presence of silencer elements within the deletion region. Oestrogen was used in a dose dependent manner however the results confirmed that there is no significant light emission compared to PGL3 basic empty vector or the untreated cells suggesting the absence of an active promoter or the presence of silencer elements within the region (Figure 4.13).

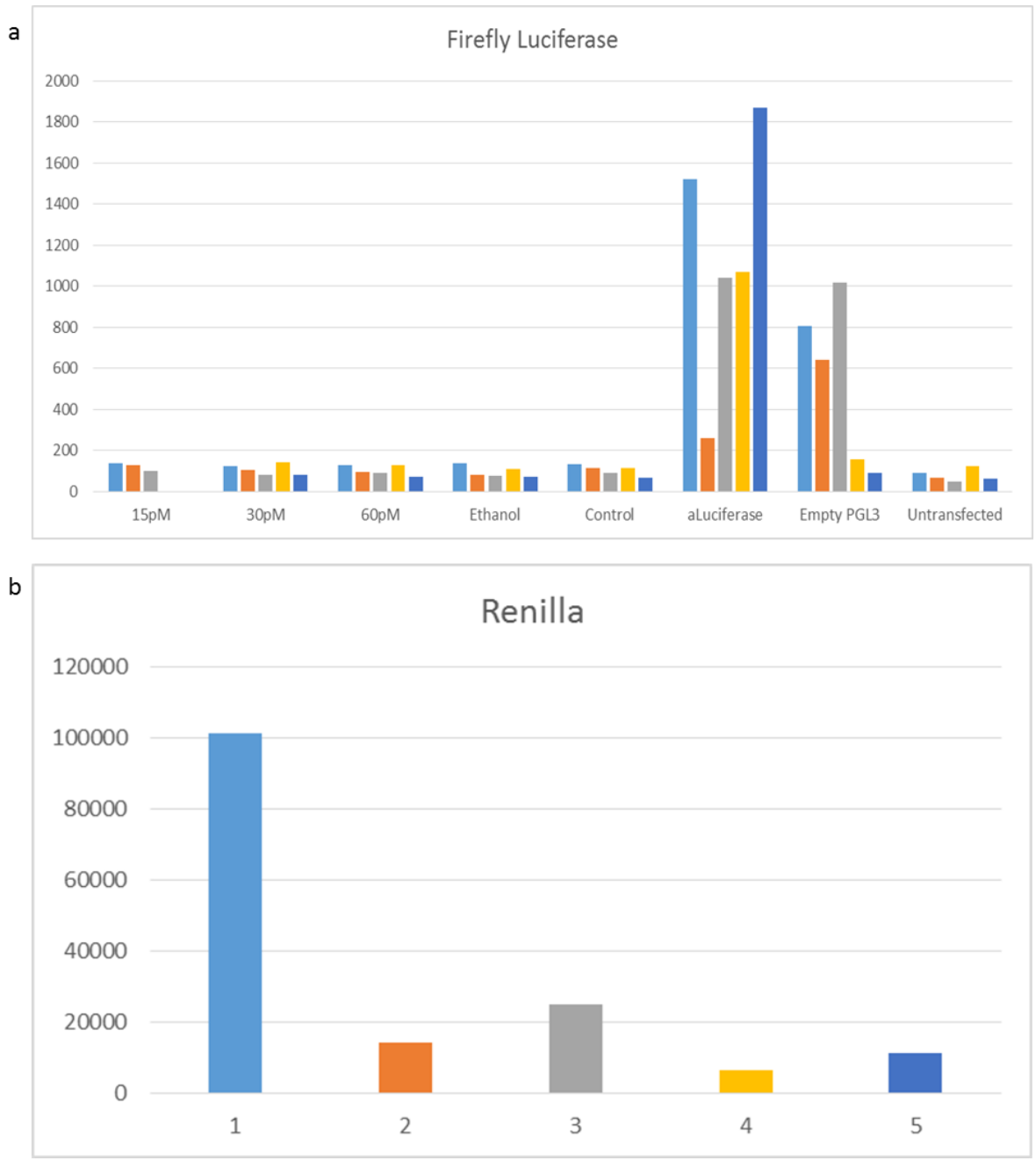


Figure 4.13 Luciferase assay for identification a potential promoter. a) Luciferase assay in 5 different single experiments under the same conditions on MCF7 cells indicating the absence of active promoter or the presence of silencer elements within the deletion. b) Renilla luciferase emission for the 5 same experiments illustrated at a) confirming the successful plasmid transfection into MCF7 cells.

4.4 Discussion

4.4.1 The design of a qPCR based genotyping tool

The development and optimisation of a qPCR based genotyping tool for the successful screening of large cohorts of patients has given the opportunity for the investigation of the deletion in a fast, cost-effective and efficient manner. It was hypothesized that the primers within the deletion, which during the anonymised steps had successfully identified patients carrying the deletion with a Ct value difference of 1, would be ideal for the screening of large cohorts of patients. Initially, the 208 subclinical hypothyroid patients were screened with the primers within the deletion resulting in 15 false positives. The screening of the 15 positive for the deletion patients with flanking the deletion primers confirmed that 2 were true positives. Therefore, screening of the total 208 patients was repeated with the use of flanking the deletion primers for the identification of potential false negatives. The investigation identified a false negative confirming that flanking the deletion primers were ideal for the screening of large cohorts of patients.

Screening of additional cohorts with flanking the deletion primers of a total number of 36 patients from 4 different groups identified and confirmed one more patient heterozygote for the deletion. In total, four patients carried the InDel 3 men aged 37 to 39 years old and 1 woman aged 47 years old. One of the male had follicular adenoma, the second male was noted with goitre and the woman had a family history of cancer. The 3 SH patients were contacted for a thyroid scan with the female patients arranging an appointment however she did not attend the appointed meeting.

Previous investigation by Dr. Bakhsh (unpublished data) about the prevalence of the InDel has revealed that none of the 70 papillary thyroid cancer patients harboured the deletion, it was present in approximately 1% of the Caucasian population 1 out of 105 patients and about 5% in patients with Multinodular Goiters (MNG) 4 out of 81. In the current study the prevalence of the InDel in the subclinical hypothyroid patient cohort which was similar to 1% previously found in the Caucasian population. On the other hand, 1 out of 5 patients with follicular adenoma found to harbour the deletion which is supporting our hypothesis that the InDel is promoting benign proliferation. Previous studies by Dr. Grennan Jones and Dr. Bakhsh confirmed that the patients carrying the PLCβ1 InDel exhibited significantly increased PLCβ1 transcript levels. Taking into account the role of PLCβ1 in cell cycle signalling and its effect on the BRAF and MEK/ERK pathway and the fact that it is present in patients with follicular

adenoma and MNG plus the original family it suggested that it could potentially have a role in benign cell proliferation.

4.4.2 Reporter gene assay

In silico analysis of the deletion region revealed that there is an Oestrogen receptor α (ERα) binding site within the deletion. Therefore, the deletion region was cloned into PGL3 basic vector, ideal for luciferase assays as it contains a luciferase gene. The cells used for the luciferase assay were MCF7 breast cancer cell line known to be Oestrogen receptor positive.

The results even though they were quite variable have confirmed that there is either no active promoter or the presence of silencer elements within the deletion region. Luciferase excitation was not induced following Oestrogen treatment with the signal being equivalent to the background luciferase signal. All positive controls for Firefly Luciferase emitted light and PCR amplification of the plasmids transfected into MCF7 cells with luciferase primers confirmed the presence of functional Firefly luciferase. In addition, Renilla Luciferase also emitted light confirming the transfection efficiency of the assay, as it is was co-transfected along with the luciferase assay plasmids. The deletion of the sequence has been reported to be correlated with elevated thyroid PLCβ1 transcript levels. These results in combination with the lack of oestrogen response in the reporter assays has suggested the potential presence of silencer elements within the deletion which may act as repressors of PLCβ1 expression.

Chapter 5.The Effect of PLCβ1 On Cell Proliferation

5.1 Introduction

5.1.2 The presence of PLCβ1 in human cancer

A number of studies have shown that PLCβ1 has a major role in a range of different cancers such as prostate, colon, breast, leukaemia and ovarian cancer. Increased PLC activity was also found in neoplastic thyroid membrane (Kobayashi *et al.* 1993). Also, it has both nuclear and cytoplasmic locations with its signalling cascade being involved in cell cycle progression. In this study, the impact of potential alterations of PLCβ1 transcript levels on cell proliferation will be assessed characterising its potential contribution in goitre and thyroid cancer (Friederichs *et al*. 2000; Kristiansen *et al*. 2002; Follo *et al*. 2009; Molinari *et al*. 2012; Ribeiro *et al*. 2012; Jia *et al*. 2013).

5.1.3 The effect of PLCβ1 InDel on the PLCβ1 gene, mRNA expression.

Unpublished work by Drs. Bakhsh, Zhang and Grennan Jones in our group identified that patients who carried the PLCβ1 InDel exhibited increased thyroid PLCβ1 transcript levels. More specifically, tissue extracted from thyroid tissue following thyroidectomy has shown that PLCβ1 transcript levels were significantly elevated in those affected with the PLCβ1 InDel. The increased PLCβ1 transcripts in patients carrying the InDel was up to 3 fold compared to unaffected individuals. In this study we have used the Fisher rat thyroid cell line (FRTL-5) to assess the impact of alteration in the PLCβ1 transcript levels on cell cycle progression and cell doubling time. As will be described below I have used techniques to reduce (siRNA) and over-express (lentiviral vectors) PLCβ1 transcripts.

5.1.4 FRTL-5 cell line

The target cell line was the FRTL-5 cell line which is thyroid stimulating hormone (TSH) dependent for cell growth and shares characteristics with thyroid epithelial cells, therefore considered an ideal cell model for thyroid cell studies (Iitaka *et al.* 2001). The FRTL-5 cell line was derived from epithelial follicular cells from the thyroid of Fisher rat. The cells are diploid and grow as a monolayer, able to express *in-vitro* a range of thyroid differentiation markers. It is a TSH growth dependent cell line with a morphology similar to primary epithelial cells. They demonstrate the ability to accumulate iodine *in-vitro* and generate thyroglobulin, although are unable to synthesize thyroid hormones (Stiblar-Martincic *et al.* 2002).

5.1.5 RNA interference (RNAi)

RNAi technology was discovered by (Fire *et al.* 1998) and since then it has become a popular tool for the evaluation of gene function by regulation of gene expression. The technology is based on the mechanism of naturally occurring endogenous siRNAs which derive from dsRNAs and have a regulatory role in gene expression (Watanabe *et al.* 2008). RNAi technology uses small interfering RNA sequences (SiRNAs), approximately 21bp double stranded RNA (dsRNAs), to interfere with gene expression. RNA interference operates at two distinct phases, the initiation phase and the effector phase. During the initiation phase long double stranded RNA is fragmented by RNAseIII endonuclease Dicer enzyme into RNA duplexes of approximately 21 nucleotides. The fragmented double stranded RNA has a dinucleotide overhang at the 3' end and a monophosphate group at the 5' end both having essential roles in recognition during integration of the RNA induced silencing complex (RISC). Dicer along with the human immunodeficiency trans-activating response RNA-binding protein (TRBP) load the RNA onto Argonaute (AGO2) to form RISC. Following the formation of RISC, double stranded RNA is separated as AGO2 selects the ''guide'' strand and cleaves the ''passenger strand''. The ''guide'' strand whilst bound with AGO2 is paired with mRNA with the rule of complementation. AGO2 then cleaves the bound mRNA resulting in depletion of a specific gene (Figure 5.1). (Carmell *et al.* 2002; Aagaard and Rossi 2007; Shreve and Carter 2009)

The main advantage of RNAi technology is the simplicity with which loss of function of a particular target protein can be induced. Also, even in comparison with other anti-sense techniques, RNAi has the ability to bind to mRNA and lead to its fragmentation with higher efficiency even at lower concentrations (Aagaard and Rossi 2007).

RNAi despite its wide application has a number of limitations such as the off target effect which may lead to depletion of other genes. Also, it is cleared very quickly from cells which makes it ideal only for short-lasting experiments. In addition, siRNAs are susceptible to enzymatic degradation having a short half-life and therefore can be very unstable in physiological conditions *in vivo*. (Sledz and Williams 2005; Aagaard and Rossi 2007).

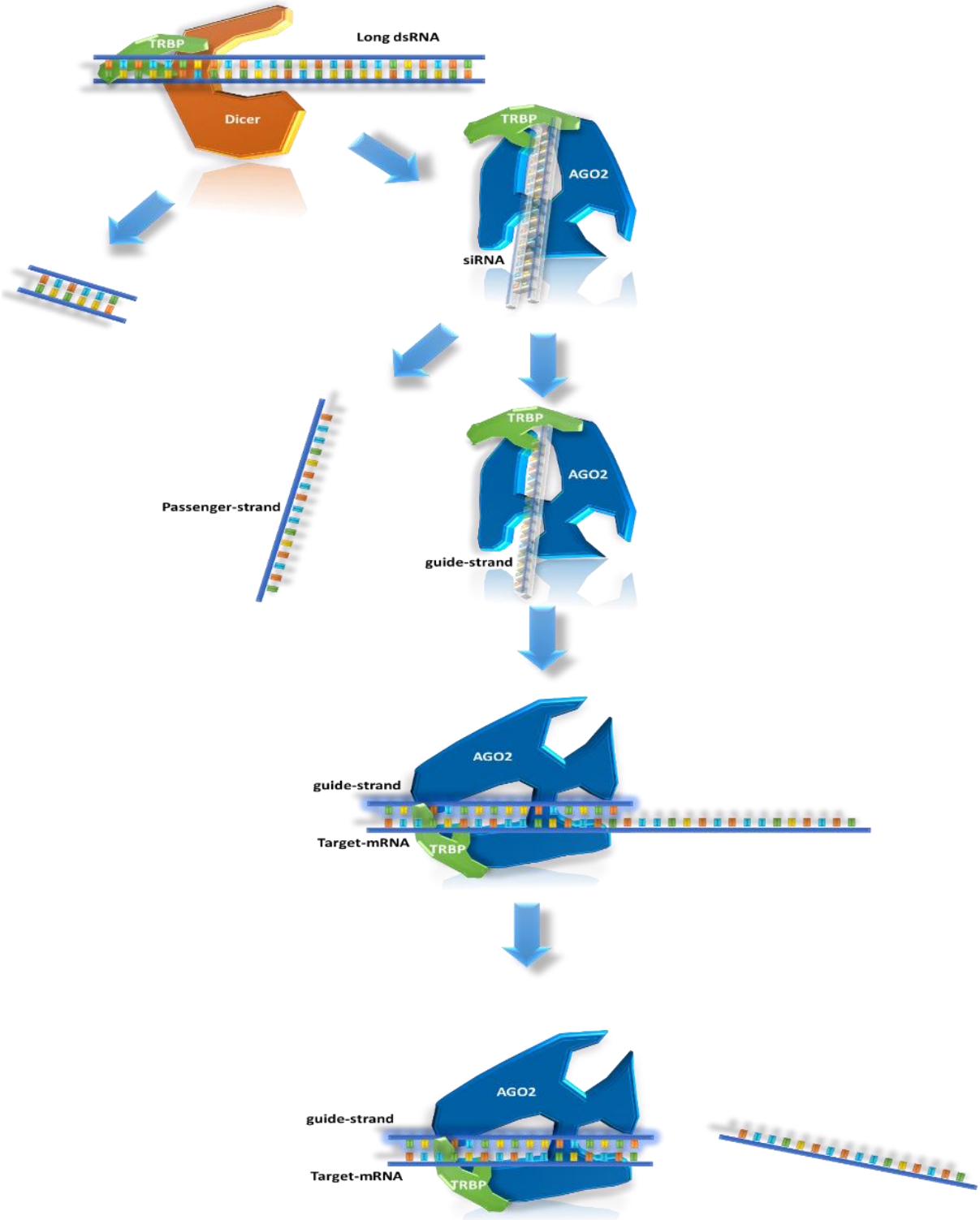


Figure 5.1 The principles of siRNA knockdown. Dicer protein binds to long double stranded RNA and cleaves RNA into 21bp fragments which along with trans activating response RNA-binding protein (TRBP) are loaded onto Argonaute 2 (AGO2) forming the RNA- binding protein complex (RISC). AGO2 cleaves the siRNA and selects the guide strand which eventually will bind to the target mRNA and AGO2 then cleaves bound mRNA leading to depletion of the specific gene.

5.1.6 Lentiviral vectors

The generation of viral vectors and their use for alteration of gene transcript levels offers unique advantages in research and gene therapy. The use of lentivruses allows the transduction of both dividing and non-dividing cells, without expression of viral proteins eliminating immunogenic responses. Also, a great advantage of using viral vectors compared to siRNAs is the long term transgene expression allowing long term experiments without clearance of the effect. On the other hand, their disadvantages include the possibility of low titres and the possible random insertion in the host genome which can potentially trigger tumorigenic or tumour suppressor activity.

5.1.4.7 Lentivirus generation

The lentivirus that is widely used is the human immunodeficiency virus type1 (HIV-1). The generation of a lentiviral vector requires three plasmids, the first contains the viral genes in *trans* such as Gag-Pol and rev which encode regulatory proteins required for packaging, replication and transduction. The second plasmid is called the envelop plasmid and contains pseudotyped viral particles, the Vesicular Stomatitis Virus (VSV-G) envelope gene, encoding for the glycoprotein VSV-G an envelope protein which is incorporated into the viral membrane and allows transduction in a plethora of different cells. The third plasmid is known as the transfer plasmid and contains the gene of interest aimed for packaging, reverse transcription and integration. Transfer vector also contains a *rev* responsive element (RRE) which increases the viral titre as it enhances the nuclear export of unspliced viral genomic RNA. The plasmids should be co-transfected in a pseudoviral particle producer cell line for increased titre of pseudoviral particles such as HEK293T cells which express SV40 large antigen and increases the protein expression from plasmids containing SV40 as replication origin (Figure 5.2) (Zufferey *et al.* 1999; Stone *et al.* 2000; Zennou *et al.* 2000; Kay *et al.* 2001).

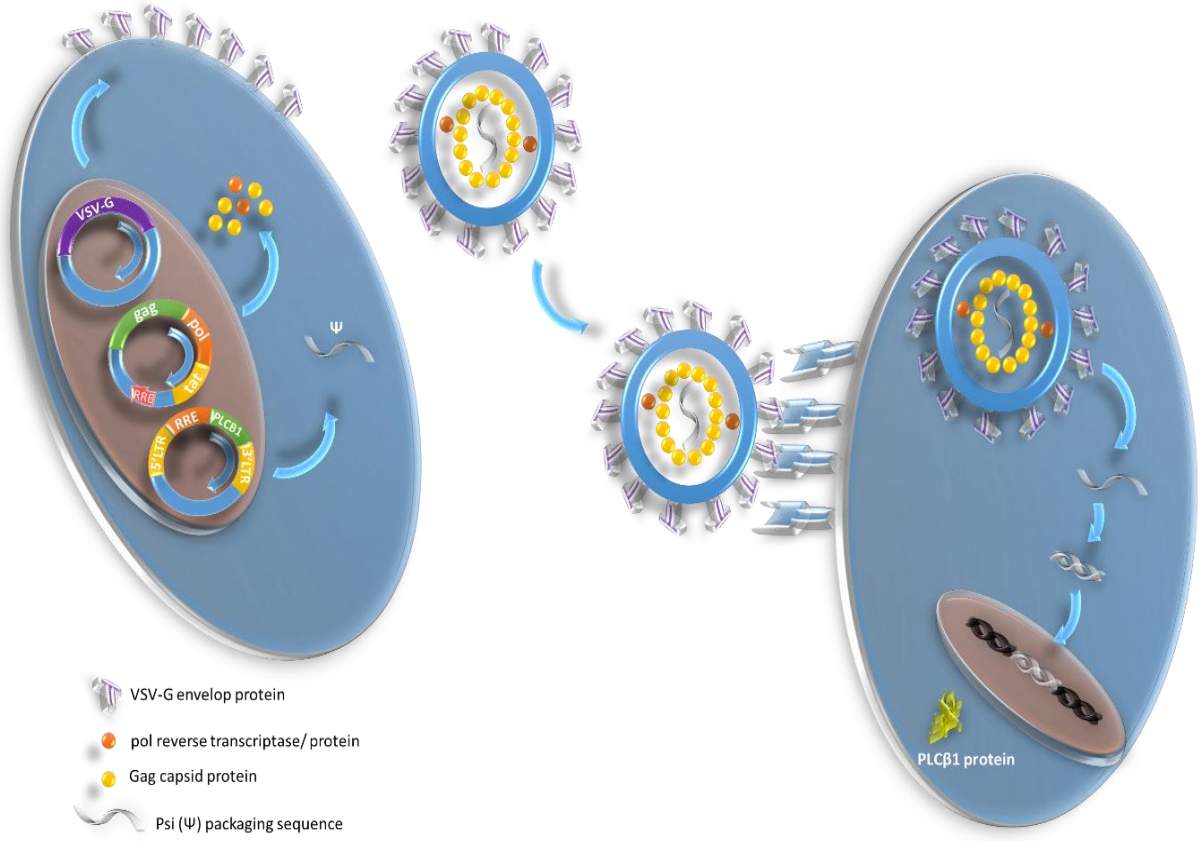


Figure 5.2 The principles of lentiviral generation and infection. Packaging, envelope and transfer plasmids are required for the generation of an active lentivirus on cell lines, such as HEK293T, capable of producing high titre pseudoviral particles.

5.1.8 AIM

PLCβ1 protein has been found to be involved in numerous cancers and is capable of altering the cell cycle progression. In this study, patients carrying the PLCβ1 InDel had increased PLCβ1 transcript levels with a high prevalence of transition from Multinodular Goiters to papillary thyroid cancer. Therefore the impact of alterations of PLCβ1 transcript levels on cell proliferation were investigated with the use of siRNA and lentiviral vectors for knockdown and overexpression respectively (Figure 5.3).

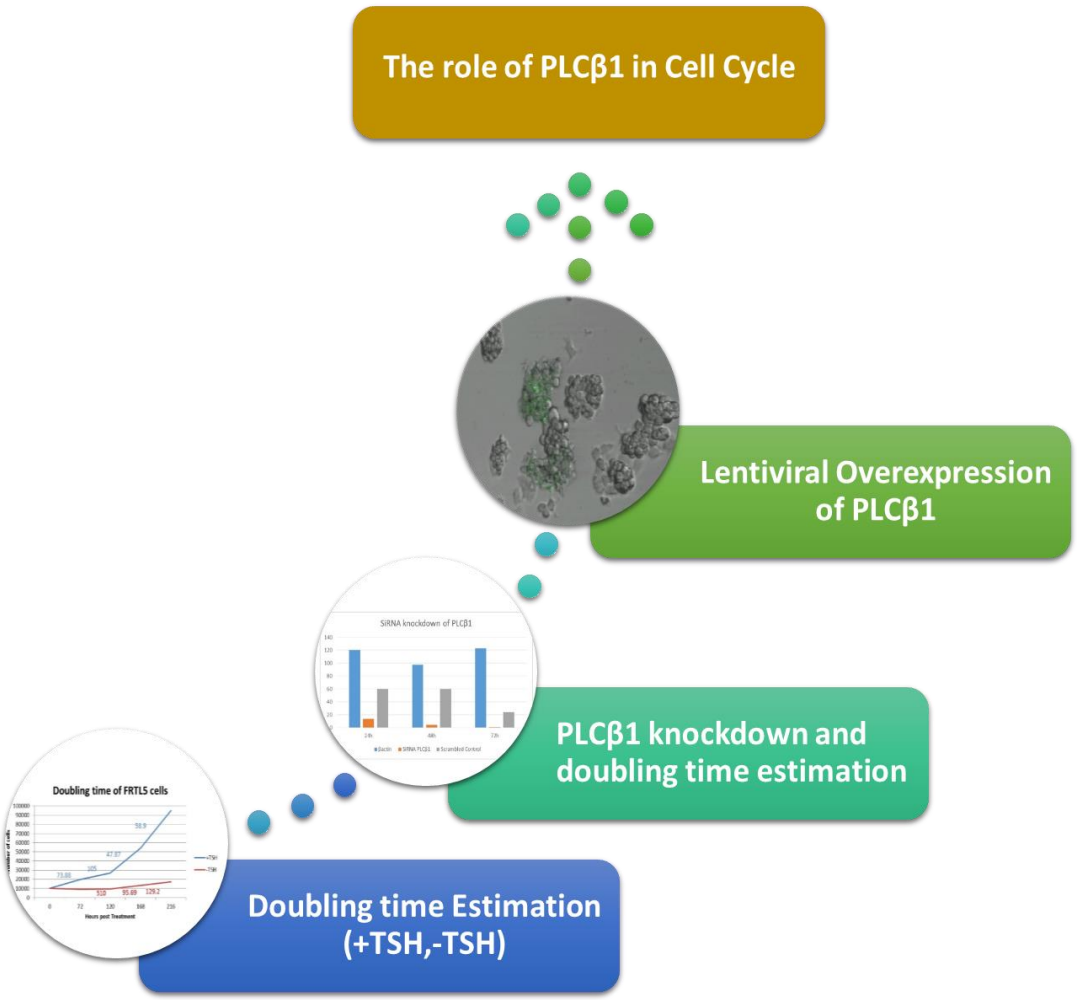


Figure 5.3 Aims of the chapter. The aim of this project was to characterise the role of PLCβ1 in cell proliferation. Initially, doubling time in TSH growth dependent FRTL-5 cells was estimated and then again following PLCβ1 knockdown with siRNA. Overexpression of PLCβ1 was attempted with the generation of active lentiviral vectors.

5.2 Materials and Methods

5.2.1 Cell culture of FRTL5

FRTL5 rat thyroid cells were kindly provided by Prof. Wynford- Thomas. The glycerol stocks were defrosted following 2 minutes incubation at 37 ^{o}C and mixed with 7ml of preheated HAMF'S F12 (Lonza Uk). Cells were then centrifuged at 1000rpm for 5 minutes and supernatant was discarded. Pelleted cells were resuspended with 7ml of 5 hormone (5H) medium (HAM'S F12(Lonza UK) with 5% bovine calf serum (BCS)(EU approved), 100U/ml potassium penicillin, 100µg/ml streptomycin sulphate (Lonza UK), 10mu/ml bovine TSH, 10nM Hydrocortisone, 10µg/ml Insulin, 5g/ml Transferrin and 45µg/ml Ascorbic acid)(Sigma–Aldrich UK). Cells grown at 37 ^{o}C and 5% CO_2 incubator.

5.2.2 Trypsinisation and Cryopreservation.

The passaging and cryopreservation of FRTL-5 cells was performed as described in (Chapter 4). The medium requirements for cell growth for FRTL-5 is the 5H culture medium which was described above.

5.2.3 Doubling Time estimation

Assessment of the doubling time was performed using FRTL5 cells grown in 5H and 4H (no bovine TSH). More specifically 10,000 cells were plated in duplicates in a 24 well plate (Thermo Scientific UK) and cultured in 5H and 4H medium for 9 days. Cell counting was performed on days 3, 5, 7, and 9 using an automated cell counter (Nexcelom Bioscience Cellometer Auto T4) (described in 5.2.5).

5.2.4 SiRNA knockdown

24h prior to transfection, FRTL-5 cells were trypsinised as described in (Chapter 4). FRTL5 cells were counted with the use of an automated cell counter (Nexcelom Bioscience Cellometer Auto T4) and 60,000 cells were plated in each well of a 12 well plate (Thermo Scientific UK). The cells were grown at 37 ^{o}C and 5%CO_2 incubator overnight. The following day the cells were transfected with 40nM of SiRNA or scrambled SiRNA (Santa Cruz) (Table 5.1). The SiRNAs were mixed with 50 µl of serum and antibiotic free medium HAM'S F12

(Lonza UK) (Figure 5.4). The Viafect (Promega UK) transfection reagent was mixed with 100µl of serum and antibiotic free medium for each of the SiRNA samples with ratio (Viafect reagent: DNA – 3:1). The samples were incubated at room temperature for 5 minutes and then the 50 µl of siRNA and scrambled siRNA were mixed with a 100µl of transfection reagent mix and incubated for 15 minutes at room temperature. The siRNA- transfection reagent complex was then added to the cells drop-wise with serum antibiotic free medium up to 250µl for 3 hours at 37 ^{o}C, 5% CO_2. 5H medium was added up to 500µl to each well and incubated for 48h at 37^{o}C, 5% CO_2 then it was replaced by 500ml of fresh 5H medium. Cell counting and cell lysis were performed at 24h, 48h, and 72h post transfection (Figure 5.4). The PLCβ1 siRNA cleaves PLCβ1 in three different positions (Table 5.1).

Table 5.1 PLCβ1 siRNA is cleaving PLCβ1 mRNA in three different positions.

	Sense	**Antisense**
sc-270424A:	GAAGAGCCAAUUGUAUUCAtt	UGAAUACAAUUGGCUCUUCtt
sc-270424B:	CAGAGAUGAUCCGAUCAUAtt	UAUGAUCGGAUCAUCUCUGtt
sc-270424C:	GAACUCCAUCGGACUCUAAtt	UUAGAGUCCGAUGGAGUUCtt

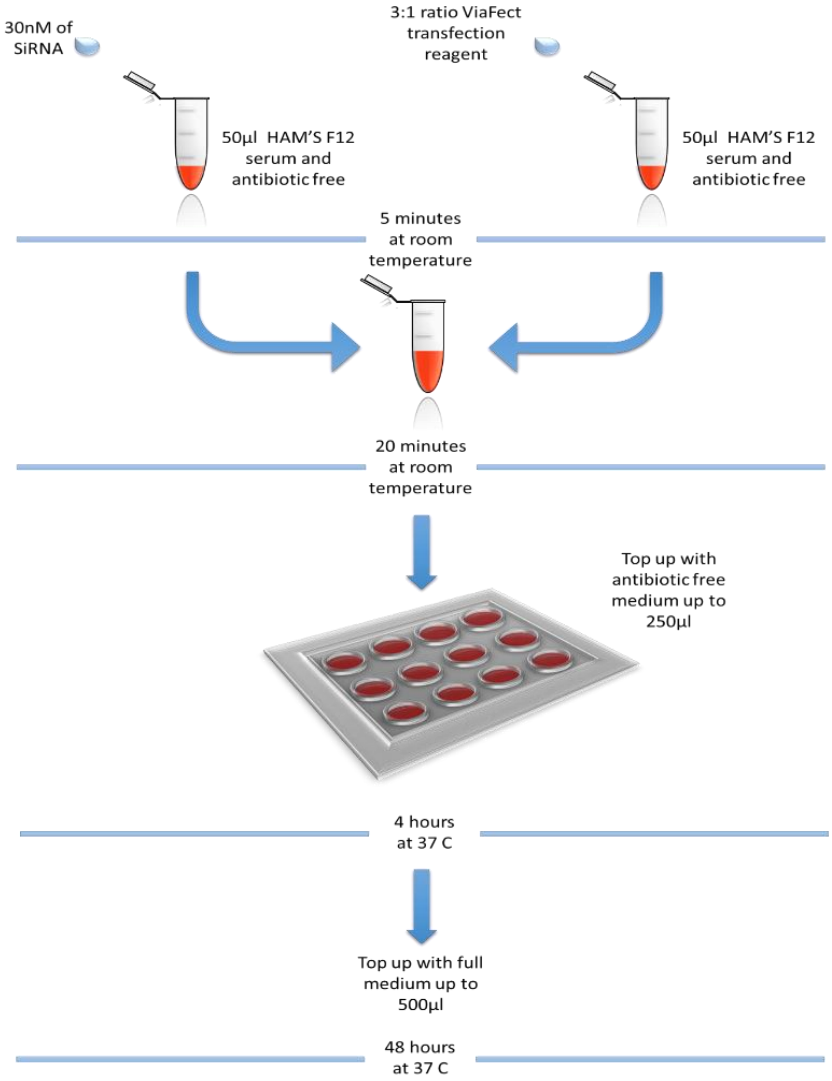


Figure 5.4 SiRNA treatment for PLCβ1 knockdown. FRTL-5 cells were treated with PLCβ1 siRNA and cell proliferation was assessed at 24, 48 and 72 hours post transfection.

5.2.5 Cell counting and Cell Lysis

Following trypsinization cells were centrifuged at 3000 rpm for 1 minute and supernatant was discarded. Pellet was resuspended in 100µl of HAMF'S F12 (Lonza UK) and 20µl were loaded into the cell counting chamber and cell number was estimated with the use of an automated cell counter. The remaining cells from the same well were lysed following the addition of 500µl of TRI reagent (Sigma- Aldrich UK) and stored at -80^{o}C.

5.2.6 mRNA extraction

Isolation of the total mRNA was performed according to the manufacturer's protocol (Sigma – Aldrich UK). The lysed cells stored at -80 ^{o}C were defrosted at room temperature and then, 0.2ml chloroform per ml of TRI reagent were mixed and incubated at room temperature for 15 minutes. The samples were then centrifuged at 13,000rpm at 4^{o}C. The top aqueous solution was transferred to a fresh 1.5ml Eppendorf tube and mixed with 0.5ml isopropanol (Fisher Scientific) per ml of TRI reagent were mixed and incubated at room temperature for 10 minutes. The mix was centrifuged at 13,000rpm for 10 minutes at 4^{o}C. Supernatant was removed and the pellet was washed with 1ml of 75% ethanol per ml of TRI (Sigma) reagent. The mix was vortexed and centrifuged at 8,000 rpm for 5 minutes at 4^{o}C. Supernatant was discarded and the pellet was briefly air dried and resuspended in 10µl of deionised water. RNA was quantified with Nanodrop-Lite (Thermo Scientific).

5.2.7 cDNA synthesis

A total volume of 6µl was prepared consisting of 1µg of RNA and distilled water up to 6µl. The mix was incubated at 60 ^{o}C for 10 minutes following incubation on ice for 5 minutes. RNA was then mixed with 4µl of M-MLV RT 5x buffer, 0.5µg/µl oligo(dT)-15 primer, 1mM of dNTPs (10mM each). The mix was vortexed and then mixed with 200units of M-MLV reverse transcriptase and 20units of recombinant RNasin®-Ribonuclease inhibitor. The total of 20µl reaction was centrifuged and placed into a thermocycler (Techne TC-512) at 37^{o}C for 1 hour, at 95 ^{o}C for 5 minutes and then hold at 4 ^{o}C. All reagents were acquired from (Promega UK).

5.2.8 PLCβ1 primer selection

A number of primers were designed but failed to amplify the rat PLCβ1 product. Therefore, primers cited in peer reviewed journals for either rat or mouse PLCβ1 were used for amplification optimisation (Table 5.2); primers for a rat β-actin housekeeping gene (kindly provided by Dr. Zhang) were used as a positive control.

Table 5.2 List of primers. A number of primers from studies where successful amplification of PLCβ1 was achieved in peer reviewed journals were selected and altered accordingly for rat PLCβ1 gene.

Primer Reference	**Forward Primer**	**Reverse Primer**	**Product Size (bp)**	**Exons**
Dwivedi et al. 2005	TTTTCGGCAGACCGGAAGGCA	TGCTGTTGGGCTCGTACTTCT	315	6-8
Ballester et al. 2004	CCACGTGAAACCAGGACAAC	ACGCTCTGAATCAGATCCTCTGT	149	22-23
Tvent et al. 2014	CAGAAGTAGAGGCGCAGACC	CTCGGTGGTTTTCTTGTGGT	127	23-24
Watkins and Sinclair 2014	TGGATGAGAAGCCCAAGC	GGCAGTCTTTTGAACTTGTC	67	29-30
β-actin	TGTCACCAACTGGGACGATA	GGGGTGTTGAAGGTCTCAAA	165	

5.2.9 qPCR and Densitometry analysis

PLCβ1 transcript levels were quantified by qPCR and also semi-quantitatively following gel electrophoresis and densitometry analysis. QPCR reaction included 1µl of cDNA, 12.5µl of 2x SyBR Green master qPCR mix (Invitrogen), 10µM primer stock and distilled water up to 25µl. Reaction cDNA included SiRNA (Santa Cruz), Scrambled-SiRNA (Santa Cruz), Control samples and a standard curve of PLCβ1($1.1x10^9$, $1.1x10^8$, $1.1x10^7$, $1.1x10^6$, $1.1x10^5$, $1.1x10^4$, $1.1x10^3$ copies) and β-actin ($1.4x10^9$, $1.4x10^8$, $1.4x10^7$, $1.4x10^7$, $1.4x10^6$, $1.4x10^5$, $1.4x10^4$, $1.4x10^3$ copies). The β-actin housekeeping gene was included in this study as internal control for normalisation of the qPCR quantitation as housekeeping genes have the same mRNA expression levels in all cells. The PLCβ1 silencing and ablation of PLCβ1 transcript levels resulted in generation of primer dimers during qPCR amplification due to the lack of specific DNA template, thus qPCR quantification could not be performed. Amplified samples were then loaded into a 2% agarose gel followed by gel electrophoresis (90V for 40minutes) as described in Chapter 2 and knockdown of PLCβ1 was confirmed with densitometry analysis with AlphaImager HP software.

5.2.10 Preparation of Lentivirus

5.2.10.1 Lentivirus preparation plasmids

Lentivirus preparation plasmids which included packaging, envelope plasmids and plasmids for GFP, PLCβ1a and PLCβ1b expression were kindly provided by Dr. Roberta Fiume from the University of Bologna. Also, active GFP lentivirus supernatant and Lentivirus packaging system with plasmid expressing GFP has been provided by Dr. Richard Darley from the University of Cardiff as a positive control (Table 5.3).

Table 5.3 List of vectors acquired from the collaborators from Cardiff and Bologna University

Dr. Fiume (Bologna University)	Dr. Darley (Cardiff University)
CMV gag/pol (packaging plasmid)	psPAX2 (packaging plasmid)
VSV-G (envelope plasmid)	pMD.2G (envelope plasmid)
GFP-SH (transfer vector)	250 pHIV GFP (transfer vector)
eEF-1a empty vector (transfer vector)	Control GFP lentivirus supernatant
eEF-1a +PLCβ1a (transfer vector)	
eEF-1a +PLCβ1b (transfer vector)	

5.2.10.2 Generation of Lentivirus

The cells required for the generation of lentivirus were the HEK293T cells which were grown in (DMEM -4.5g/L glucose, L- glutamine without sodium pyruvate, 100U/ml potassium penicillin, 100µg/ml streptomycin sulphate, 1mM sodium pyruvate, 10µg/ml bovine insulin and 10% heat-inactivated foetal calf serum (FCS) (EU approved) (Lonza UK) at 37 ^{o}C in 5% CO_2 incubator. The cells were passaged as described in Chapter 4.

72 hours prior to the start of transfection, HEK293T cells were seeded at 4x 10^6 cells in 75cm^2 flask (Nunc). Prior to the initiation of the experiment, flasks were coated with 2m of Poly-L-

lysine and allowed to stand for 30 minutes at temperature following rinsing with PBS. Then $4x10^6$ HEK cells were added to the $75cm^2$ flask and incubated overnight at 37°C. The following day the medium was replaced with fresh Dulbecco's Modified Eagle's Medium (DMEM) with 4.5g/L Glucose, L-Glutamine without Na Pyruvate with 10% FCS (EU approved) and 100U/ml potassium penicillin, 100µg/ml streptomycin sulphate (Lonza UK). The various components required to generate a lentivirus (8µg of envelope plasmid, 15µg of packaging plasmid), 20µg of transfer vector, 2.5M of $CaCl_2$, and 350µl of sterile H_2O) were mixed in a falcon tube. Then 450µl of 2x HEPES-buffered saline (HeBs) (12Mm Dextrose, 50Mm HEPES, 10Mm KCl, 280Mm NaCl, 1.5Mm Na_2PO_4, pH 7) solution is added to the mix dropwise while bubbling the solution with a stripette following brief vortex (supernatant of active GFP lentivirus and plasmids and reagents for the generation of active GFP lentivirus were kindly provided from Dr. Darley). The precipitate was incubated for 20 minutes at room temperature. Approximately 5 minutes prior to the transfection 25mM of stock chloroquine was added to the cells. The resulting precipitate was then added to cells and incubated at 37°C overnight. The following day the medium was changed with fresh DMEM with 10% FCS and 100U/ml potassium penicillin, 100µg/ml streptomycin sulphate (Lonza UK) and incubated at 37°C overnight. The following day the virus particles were harvested by removing supernatant which was frozen in liquid nitrogen and stored at -190 °C. The medium was replaced in cells and incubated at 37°C overnight. The following day the viruses were harvested for a final time.

5.2.10.3 Spinoculation

FRTL-5 cells were trypsinised as described in (Chapter 4) with $1x10^6$ cells being concentred in 100µl of medium. Then 900µl of 1:5 diluted virus supernatant was mixed with 100µl of cells and 10µl of 0.5mg/ml polybrene (Sigma-Aldrich UK). The mix was loaded into the designated well of a 6 well plate (Thermo Scientific UK) and centrifuged at 2200 rpm for 30minutes at room temperature. Following the centrifugation each well was filled with 1ml of complete medium. The cells were incubated at 37°C and 5% of CO_2 overnight. Initial experiments used control GFP lentivirus supernatant provided by Dr. Darley.

Plasmids provided by our collaborators from Bologna along with the plasmids from Cardiff were used to generate active GFP lentivirus in HEK293T cells. Generated viruses were compared for their efficiency following transduction of FRLT-5 cells and assessment for light emission following excitation at 488nm with an Argon laser of Leica SP5 microsystems confocal microscope. (The microscope imaging was performed by Prof. Hallett).

5.3 Results

5.3.1 FRTL5 Doubling time estimation

FRTL-5 cells are TSH growth dependent, therefore the doubling time of this cell line was estimated following growth in 4H and 5H culture medium for 9 days. The lack of TSH in the culture had a detrimental effect on the cell proliferation as cell number remained almost stable (Figure 5.5). On the other hand, TSH seemed to have a dose dependent effect on FRTL5 as seen a low dose of 1mu/ml and a high-recommended dose of 10mu/ml were used for cell treatment (Figure 5.5). An initial drop in cell number was observed at the first cell counting in day 3 post cell plating. Cell death was caused potentially during the plating process and due to stressing condition in culture without or low TSH which is essential for FRTL5 cell growth. Estimation of the doubling time illustrated that cells in optimal condition require approximately 72hours to divide. On the other hand, lack or low concentration of TSH resulted in ablation of cell growth with cells requiring about 21 days and 12.6 days for division respectively (Figure 5.5 and 5.6). The cell proliferation and doubling time was estimated as the average of two experiments. Also, lack of TSH affected the morphology of the cells as they are unable to form monolayers (Figure 5.7).

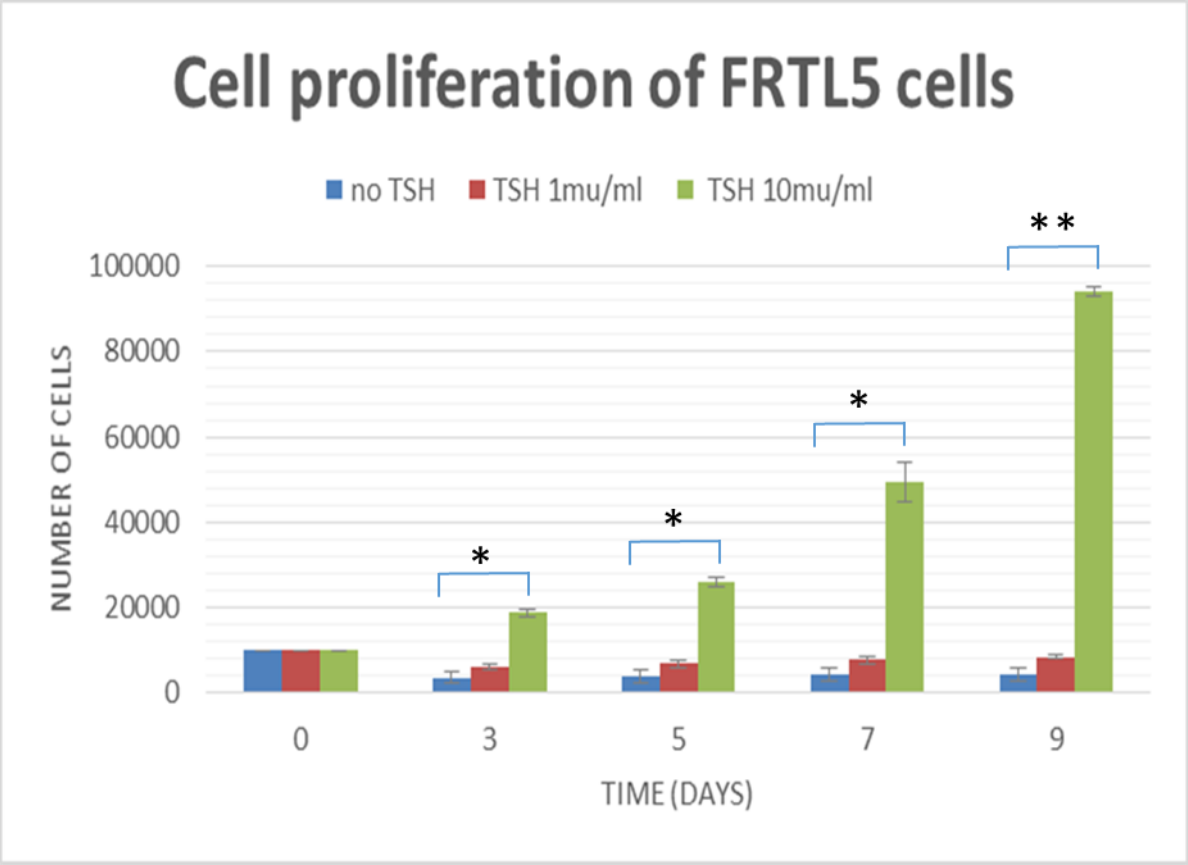


Figure 5.5 Cell proliferation of TSH growth dependent FRTL-5 cells. The cells were grown in 4H medium and 5H medium with 2 different TSH concentrations for 9 days. The data illustrated are the average of the measurements of two experiments both performed in duplicate. Statistical analysis used one way ANOVA comparison. The difference between the 10mu/ml of TSH and no TSH was significant $p<0.05$ (* represents $p<0.05$ and **$p<0.001$).

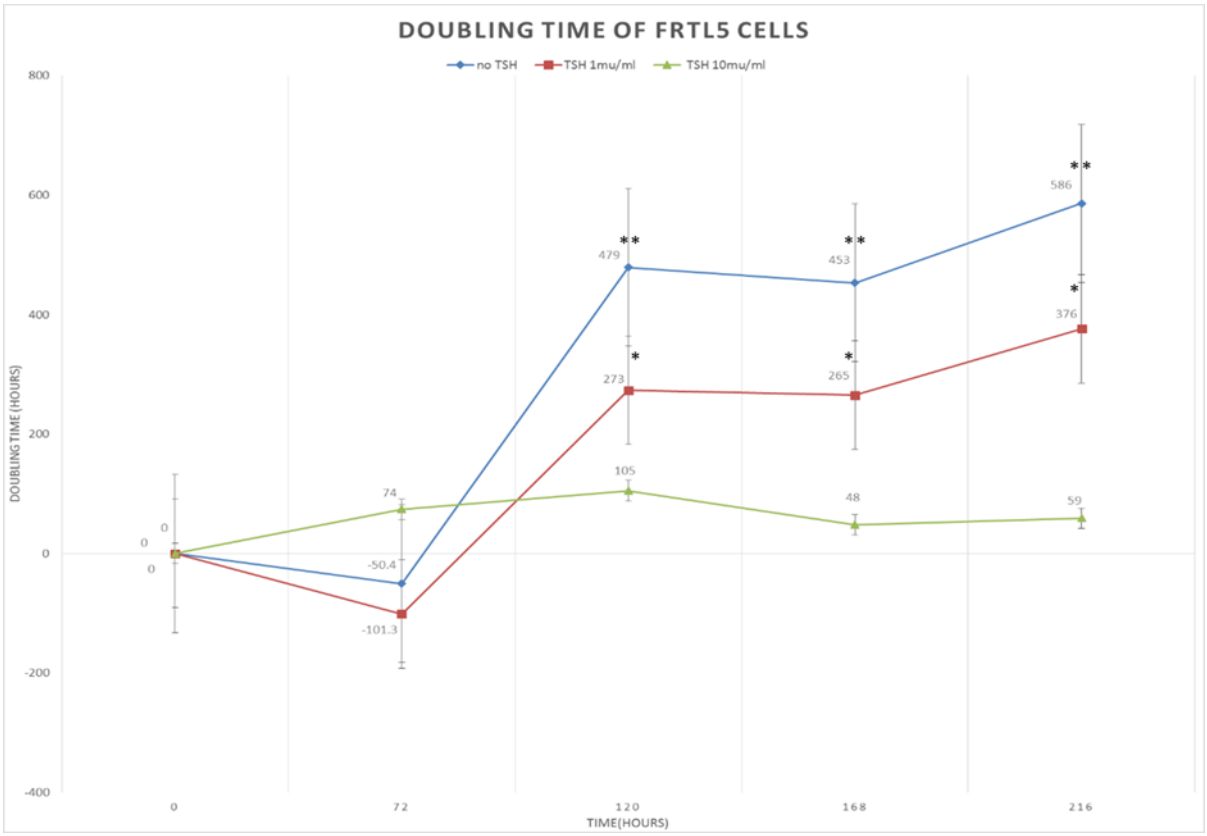


Figure 5.6 Doubling time estimation of TSH growth dependent FRTL-5 cells. Two different bovine TSH concentration were used for FRTL-5 treatment, which are TSH growth dependent. Measurements were taken every 3 days during a 9 day assay where cells have shown a dose dependent response to TSH. Representative of two experiments. Statistical analysis used one way ANOVA comparison. The difference of 10mu/ml compared to no TSH was significant P<001 (* represents p<0.05 and **p<0.001).

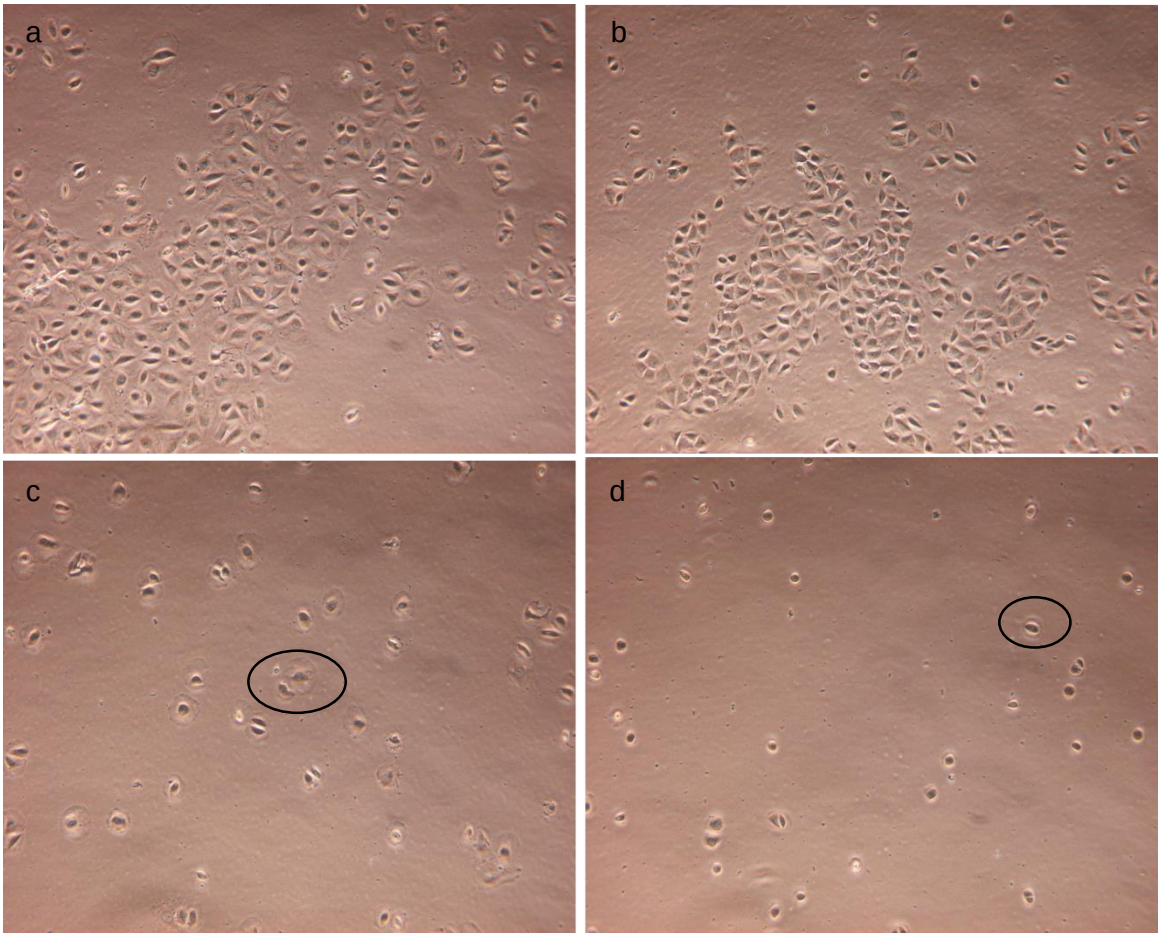


Figure 5.7 Lack of TSH abolishes cell proliferation and formation of monolayer. (a c) Cells treated with TSH seemed to be evenly distributed within the plate and display characteristic 'cobblestone' morphology**. (b, d)** in contrast with the cells without TSH which were found to form clusters and be tightly packed with a characteristic rounding shape.

5.3.2 SiRNA knockdown of PLCβ1 in FRTL-5 cells

5.3.2.1 Quantitative analysis

Equal numbers of FRTL-5 cells were plated and treated with PLCβ1, SiRNA, a scrambled SiRNA and a control untreated sample. SiRNA knocks down PLCβ1 by cleavage of the mRNA in three different positions. The level of PLCβ1 SiRNA silencing was estimated semi-quantitatively as the complete silencing of the gene caused primer dimer formation due to the lack of presence of a primer specific DNA template. Therefore, it was not possible to perform an accurate measurement with qPCR-Ct (Figure 5.8). Thus, the qPCR products were loaded into an agarose gel and densitometry analysis of amplified products was performed.

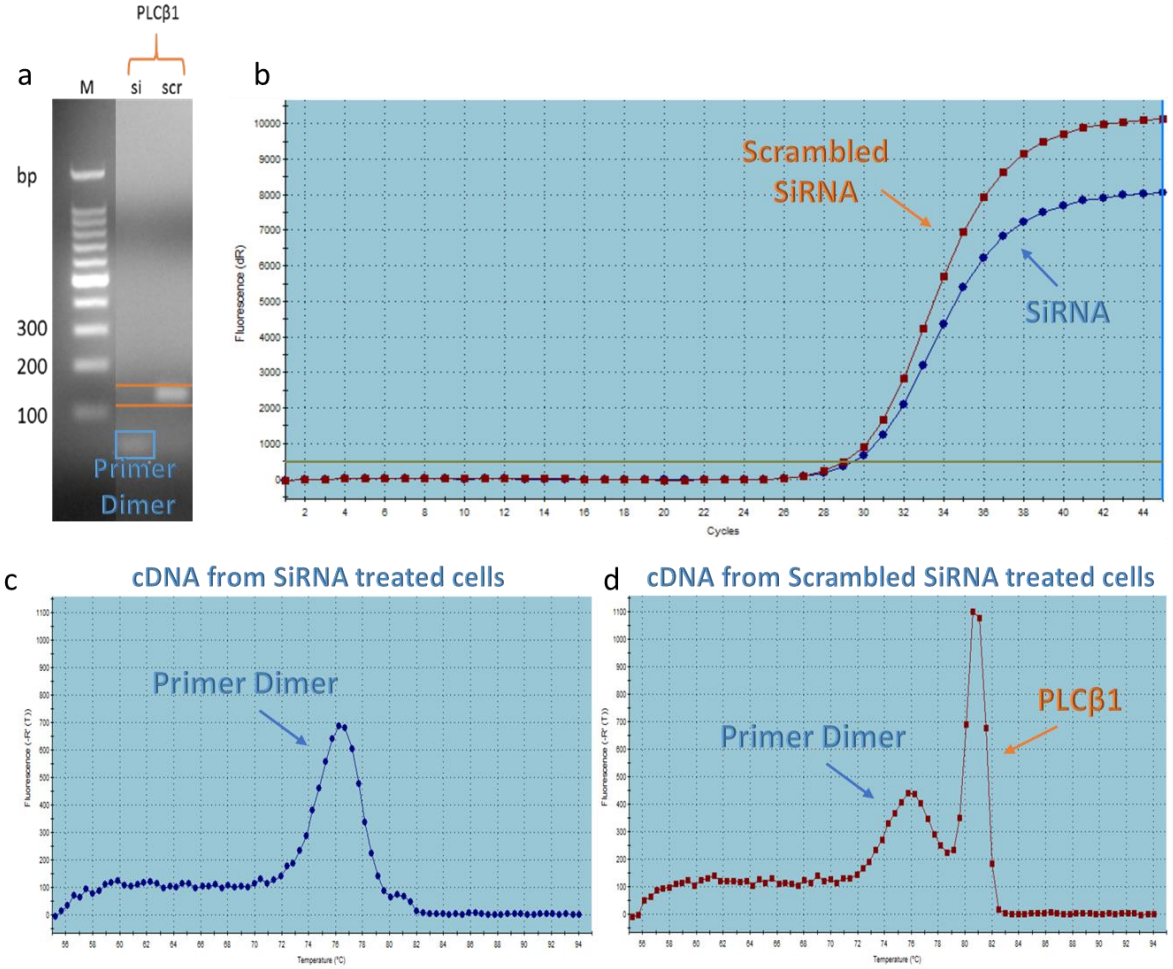


Figure 5.8 Quantitative analysis with qPCR Ct values could not be performed due to the presence of primer dimers. A) qPCR amplified products of SiRNA and Scrambled RNA treated cells. b) The qPCR amplification curve illustrating both SiRNA and Scrambled SiRNA treated cells with an almost identical Ct value of 29 cycles. c) The dissociation curve from cells treated with SIRNA, described in a) and b) illustrate the presence of a primer dimer interfering with the Ct value in b. d). The dissociation curve from cells treated with Scrambled SiRNA illustrating amplification of PLCβ1 gene.

5.3.2.2 Densitometry analysis

The successful gene silencing was unable to be quantified with qPCR as explained above therefore, qPCR products were evaluated by densitometry analysis. The amplified cDNA products confirmed the silencing of PLCβ1 gene as shown in (Figure 5.9). The estimation of PLCβ1 knockdown on cell proliferation revealed that ablation of PLCβ1 gene delayed cell proliferation (Figure 5.10). The cells treated with Scrambled SiRNA illustrated some delay in cell proliferation compared to untreated cells as transfection reagents are often cytotoxic

affecting cell proliferation rate. The highest delay in cell proliferation was observed at 72 hours post transfection as FRTL-5 cells in this project have a doubling time of approximately 71 hours thus, the effect of PLCβ1 knockdown at that time point was more prominent. The reduction of cell number following PLCβ1 SiRNA treatment illustrated 11%, 6% and 20% reduction at 24 hours, 48hours and 72hours post transfection time points respectively in comparison with the scrambled SiRNA control cells (Figure 5.10). The successful knockdown of PLCβ1 was confirmed by densitometry assay with levels of PLCβ1 found to be completely depleted (Figure 5.9b).

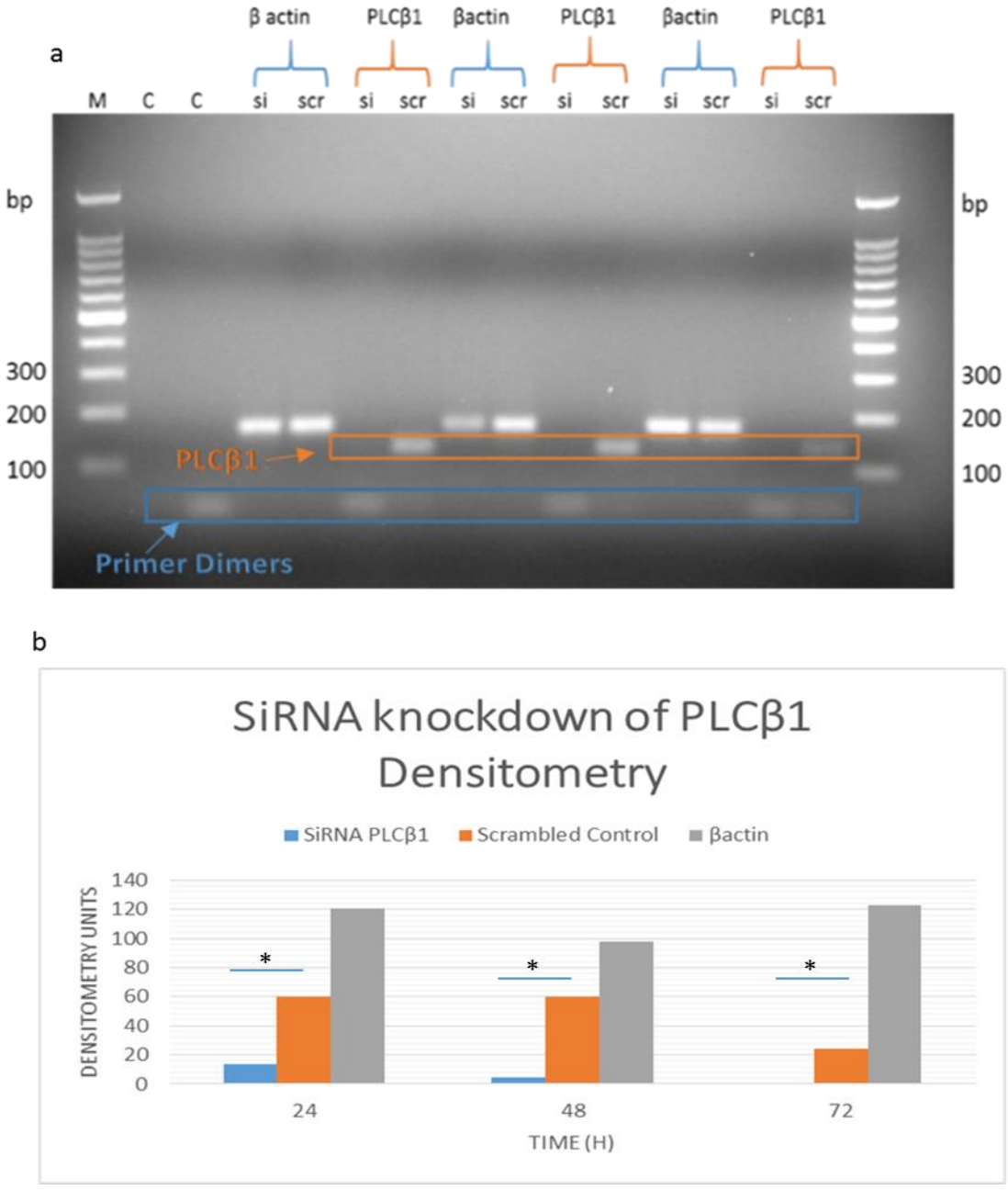


Figure 5.9 Confirmation of PLCβ1 knockdown with densitometry analysis. a) Amplified cDNA from cells treated with SiRNA (Si) and Scrambled SiRNA (Scr) using primers for β actin

(165bp) and PLCβ1 (127bp) confirmed the knockdown of PLCβ1 transcripts. b) The densitometry analysis of the agarose gel a) has provided a semi-quantitative measurement of PLCβ1 knockdown. Statistical analysis used one way ANOVA comparison. The difference between siRNA PLCβ1 and Scrambled Control was significant P<0.05 (* represents p<0.05).

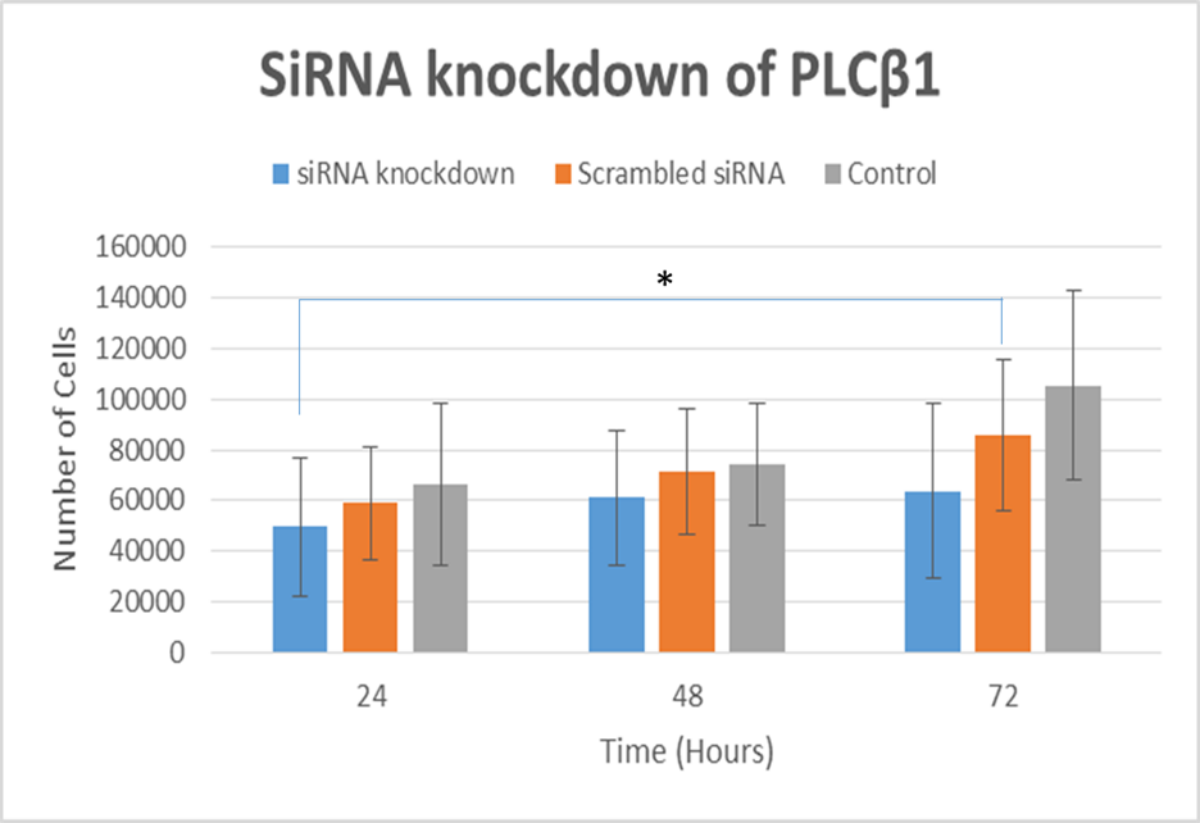


Figure 5.10 SiRNA knockdown of PLCβ1 and assessment of cell proliferation and doubling time. FRTL-5 cells following PLCβ1 knockdown were counted at 24, 48 and 72 hours post transfection. Cell number of FRTL-5 cells was estimated between 48 and 72 hours post transfection. Cell proliferation was decreased from 6% to 20% between 48h to 72h post transfection. The data illustrate the average measurement of three experiments performed in duplicate. Statistical analysis used one way ANOVA and the difference between 24h siRNA knockdown and the 72h scrambled siRNA treated cells was significant P<0.05 (* represents p<0.05).

5.3.2.3 Generation of active GFP lentivirus

Initially, generation of active GFP lentivirus was performed with the lentiviral set from Cardiff (Positive GFP virus) to successfully generate the virus in HEK293T cells (Figure 5.11) and transduce FRTL-5 cells with characteristic GFP light under the confocal microscope (Figure

5.12). On the other hand, the other set of plasmids from Dr. Roberta Fiume (Test GFP Lentivirus) (Table 5.3) was unable to generate the virus (Figure 5.11, 5.12). A direct transfection of the PLCβ1 expression plasmids into the FRTL-5 cells was attempted without success. The plasmids acquired from Dr. Roberta Fiume contained the PLCβ1 expression plasmids therefore, due to the inability to generate active lentivirus from the plasmids provided, overexpression of PLCβ1 could not be performed.

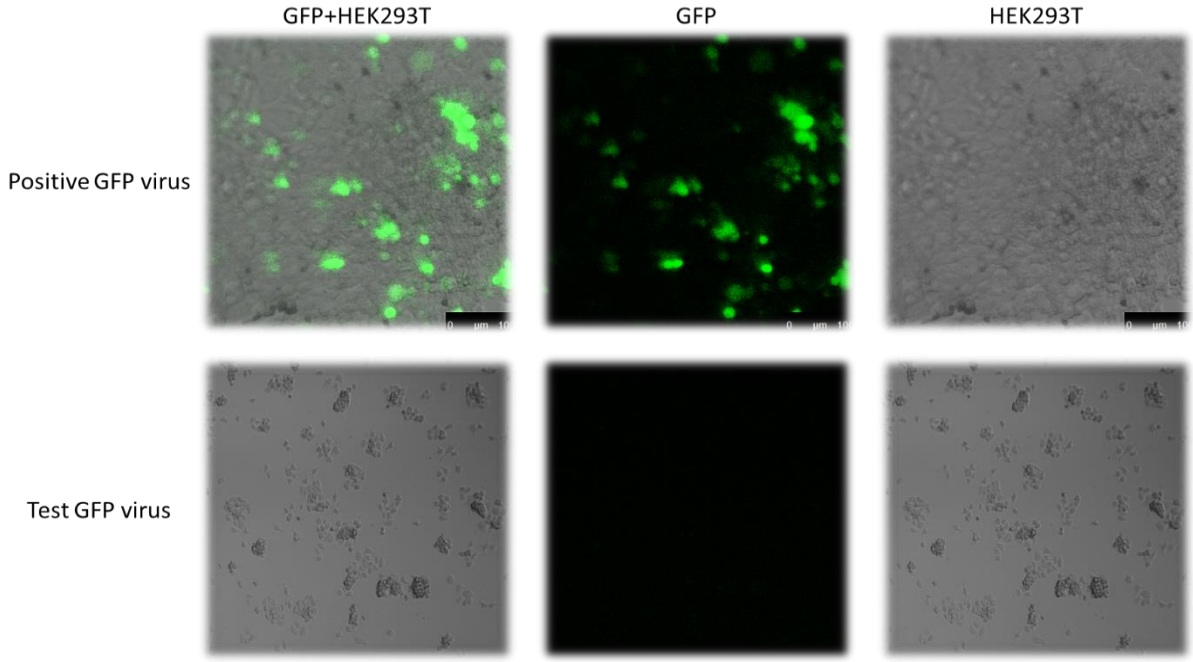


Figure 5.11 Active lentivirus generation. The lentiviral generation kit provided from (Cardiff collaborators) (Positive GFP virus) had successfully generated active GFP lentivirus using HEK293T cells. However, the second lentiviral generation kit (Dr. Fiume) (Test GFP virus) was unable to generate active lentivirus.

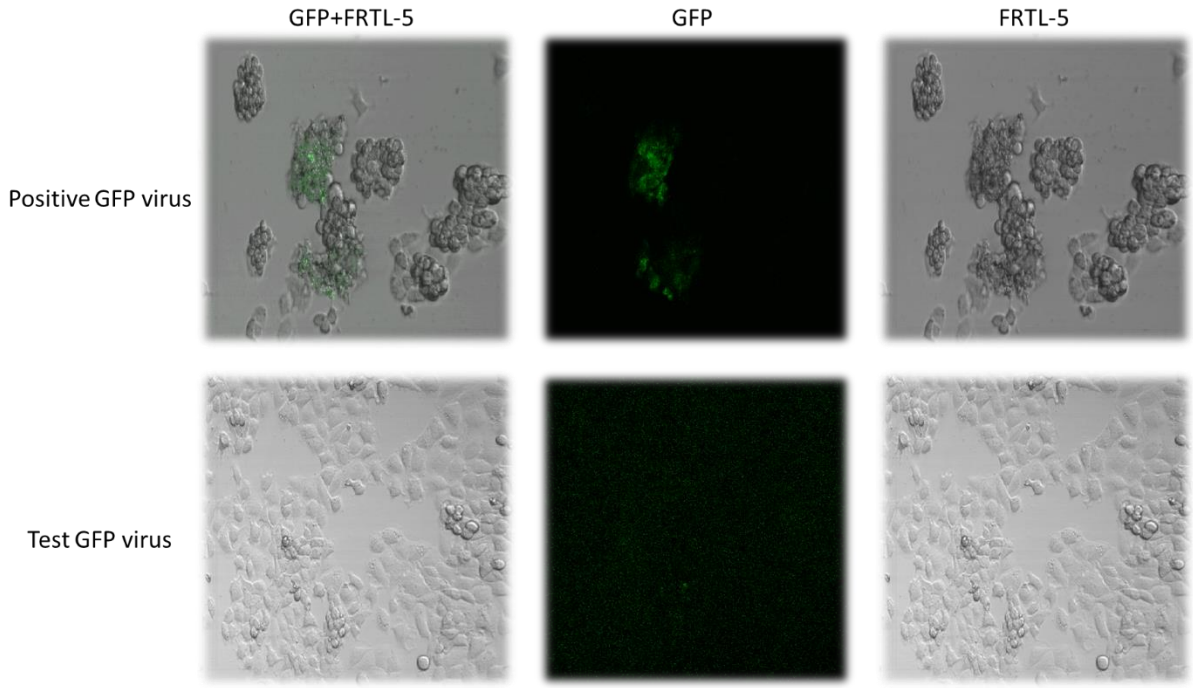


Figure 5.12 Infection of active lentivirus expressing GFP on FRTL-5 cells. Active lentivirus (Cardiff) (Positive GFP virus) generated from HEK293T cells expressing GFP was used for transduction of FRTL-5 cells. On the other hand, the second lentivirus (Dr. Fiume) (Test GFP virus) generation kit with was unable to generate active lentivirus.

5.4 Discussion

The ATAA InDel at the 3rd large intron of PLCβ1 gene has been found to affect a family with individuals suffering from MNG and with high prevalence of progression to papillary thyroid cancer. Following thyroidectomy of the papillary thyroid patients and mRNA extraction from their thyroid tissue, it was revealed that PLCβ1 transcript levels were significantly higher (up to 3 fold) compared with healthy patients who do not harbour the PLCβ1 InDel.

FRTL-5 cells are one of the most popular models for thyroid cell study as they share common characteristics with thyroid follicular cells. They are growing in monolayers and are relatively easy to be transfected. However they illustrate very high doubling time as they divide approximately every 3 days which may be a drawback for SiRNA experiments as their effect weakens with time.

In this chapter, in vivo studies have been conducted to investigate the impact of a potential alteration of PLCβ1 on cell proliferation. Depletion of PLCβ1 transcript levels led to a significant reduction in cell proliferation which agrees with the hypothesis that elevated PLCβ1 transcript levels due to the PLCβ1 InDel could trigger benign proliferation in affected patients. SiRNA technology was used for PLCβ1 knockdown on FRTL-5 cells and cell proliferation was assessed every 24h in a 72h treatment study. Complete depletion of PLCβ1 levels were confirmed with a considerable reduction on cell proliferation of approximately 11% at 24 hours, 6% at 48 hours and a 20% reduction at 72hours. The quantitation of PLCβ1 knockdown was performed semi-quantitatively using agarose gel electrophoresis and densitometry analysis, as qPCR –Ct value quantitation could not be performed due to the presence of primer dimers as PLCβ1 transcript levels were depleted. The absence of DNA template allowed the generation of primer dimers during the qPCR amplification which in turn interfere with the resulting Ct value. Thus a gel electrophoresis of the amplified products has clearly demonstrated the knockdown of PLCβ1 transcript levels.

Previous studies by Dr. Lei Zhang and Dr. Fiona Grennan Jones (unpublished data) have shown that overexpression of PLCβ1 causes an overexpression of D3 and cdk4 complex which in turn phosphorylates the tumour suppressor, retinoblastoma protein (pRb). The phosphorylation of pRb then causes the release of the E2F family proteins which are essential for the progression from G1 to S phase. Also, additional evidence indicated that overexpression of PLCβ1 prevents cell death due to oxidative stress as it suppresses apoptotic genes such as c-fos. That domino effect caused by the alteration of PLCβ1 transcript levels indicate its key role in cell proliferation and the possibility that depleted PLCβ1 transcript levels may result in cell arrest at G1 phase unable to progress to S phase and eventually leading to apoptosis thus explaining the significantly high reduction at the 72h interval.

Overexpression of PLCβ1 with the generation of lentiviral expression vector was attempted with plasmids for the generation of active lentivirus expressing GFP and PLCβ1 being provided from two different collaborators. The GFP lentiviral packaging system was found to be functional with active GFP lentivirus being generated. On the other hand, the plasmid system provided from the second collaborator for the generation of GFP and PLCβ1 lentivirus was unable to generate active lentivirus (Figure 5.11, 5.12). Direct transfection of FRTL-5 cells with the PLCβ1 expression plasmids was also unsuccessfully attempted. Therefore, overexpression of PLCβ1 on FRTL-5 cells could not be assessed and is considered as a potential future experiment for complete investigation of the role of PLCβ1 on cell proliferation.

Chapter 6. In Depth Investigation Of The High-LOD Score Chr 20 Region Using Next Generation Sequencing.

6.1 Introduction

6.1.1 Next Generation Sequencing

The first forefathers of next generation sequencing were developed during the 1970s involving chain termination and fragmentation protocols. The chain termination which was developed by Sanger and colleagues proved to be less laborious in comparison with the fragmentation protocol where a number of toxic chemicals and radioisotopes were used, establishing Sanger sequencing as the foundation protocol for DNA sequencing until the beginning of 21st century (van Dijk *et al.* 2014).

6.1.1.1 Sanger sequencing

Sanger sequencing was the first generation sequencing technology with which the sequencing of the entire human genome was achieved at the end of 2004 (van Dijk *et al*. 2014). The principle of Sanger sequencing is based on the activity of DNA polymerase to synthesize complementary DNA. The reaction requires dNTPS and ddNTPs. Each time a ddNTP has been incorporated the synthesis reaction is terminated generating fragments of different sizes. The ddNTPs are attached to different fluorescent dyes (originally P32 labelling was used) allowing the identification of the individual incorporated base (Morozova and Marra 2008).The use of ddNTPs was vital as they contain a 3'H instead of 3' OH, the lack of a hydroxyl group at the 3' end is responsible for the termination of the further elongation of the chain following the incorporation of a ddNTP. The different size fragments generated are identified with the use of a high-resolution polyacrylamide gel electrophoresis and laser excitation (Russell 2002; Primrose *et al.* 2003; Shendure and Ji 2008) (Figure 6.1).

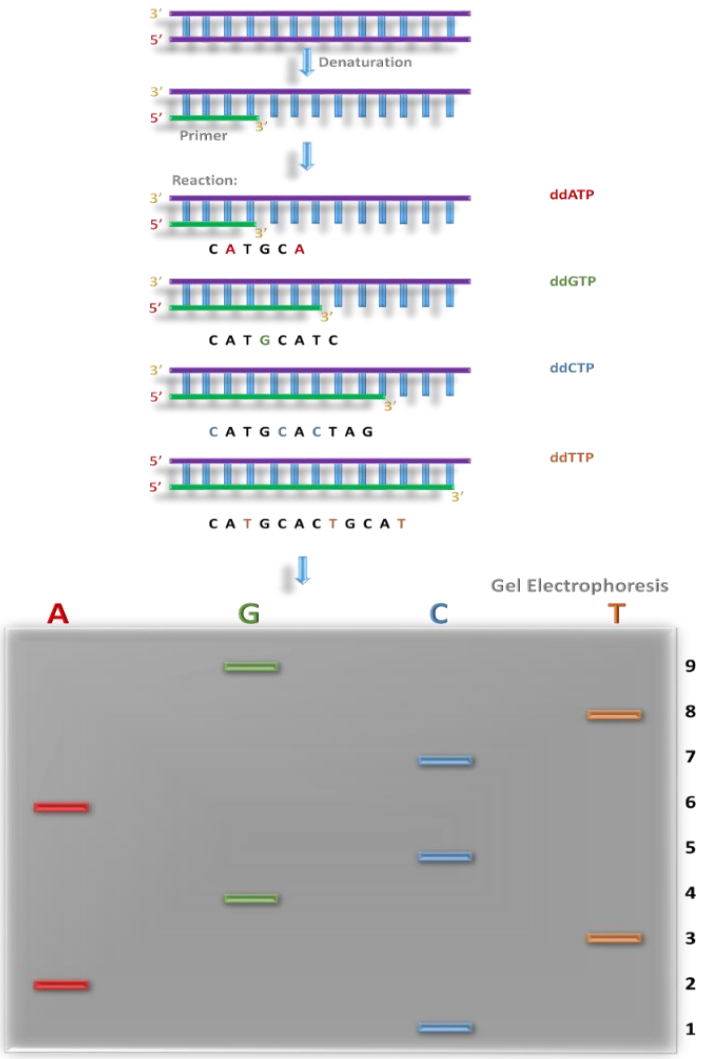


Figure 6.1 The principles of Sanger sequencing. The Sanger sequencing technology was the first sequencing technique achieving read lengths up to 1000bp. Each time a ddNTP is incorporated the chain is terminated and the nucleotide is identified by high resolution gel electrophoresis and laser excitation.

6.1.1.2 454 Pyrosequencing

The first among the next generation sequencing techniques was the 454 pyrosequencing, released in 2005 by Life Sciences. The technology utilises streptavidin beads bound to primers at one end and DNA fragments bound at the other end of the primers. Each streptavidin contains a single DNA fragment and thus following emulsion PCR each bead is bound to approximately 10^7 copies of the initial DNA fragment (Morozova and Marra 2008; Zhou *et al.* 2010). The enriched beads are then inserted into a picotitre plate where each well contains a single bead. The use of the picotitre plate offers the advantage of conducting simultaneous

pyrosequencing reactions generating high throughput results relatively fast (Mardis 2008; Morozova and Marra 2008). In the pyrosequencing reaction individual dNTPs are added sequentially and upon every nucleotide incorporation there is a release of pyrophosphate generating detectable light. The release of pyrophosphate is proportional to the amount of dNTPs incorporated each time. Following the pyrophosphate release the latter is converted into ATP by ATP sulfurylase and in the presence of adenosine 5' phosphorylase. The release of ATP is in turn responsible for the conversion of luciferin to oxyluciferin leading to generation of visible light (Shendure and Ji 2008; Ansorge 2009) (Figure 6.2). The detected light forms a pyrogram where the DNA sequence can be determined (Morozova and Marra 2008). The technology generates long reads approximately 1kb in less than 24hours, however it is low throughput with high error rate in homopolymer-rich regions as excitation of repetition of the same base is in general difficult to discriminate. The technology will no longer be supported by Life Sciences by the middle of 2016 (van Dijk *et al*. 2014).

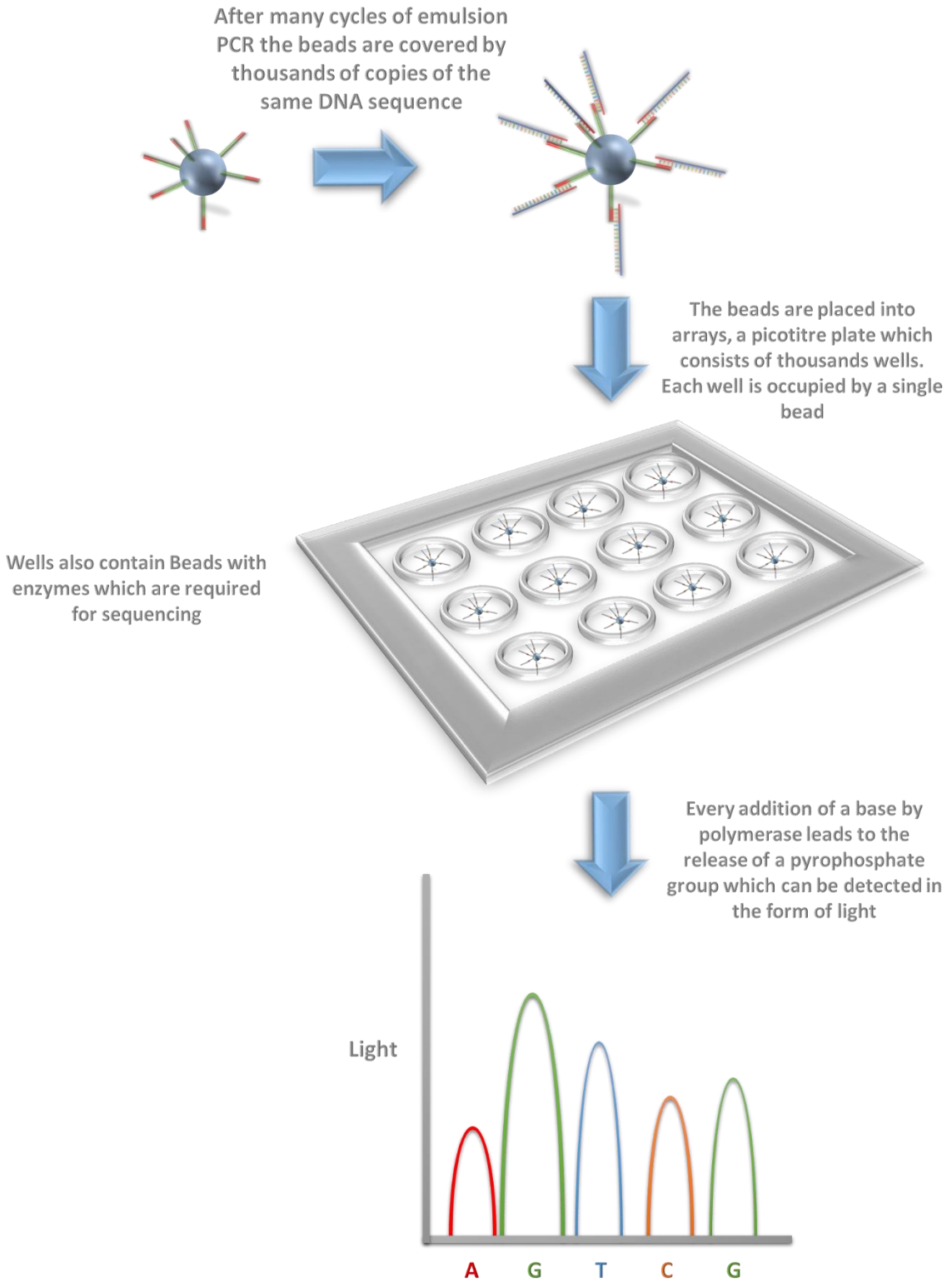


Figure 6.2 The principles of 454 pyrosequencing. 454 pyrosequencing is the first next generation sequencing technology based on a 'sequencing by synthesis' principle utilising light excitation detection for identification of incorporated nucleotides.

6.1.1.3 Illumina Solexa sequencing

Illumina Solexa sequencing technology was introduced in 2007 and is also a sequencing by synthesis based technology. The DNA fragments are bound at one end onto a solid surface which is coated with adapters with which the free end of the DNA fragment binds, due to complementarity forming a bridge (van Dijk *et al*. 2014). Solid phase amplification of these DNA fragments and denaturation generates a flow cell where each lane contains primers complementary to those on the DNA library (Figure 6.3). The sequencing occurs with the use of reversible terminators (Mardis 2008). The reversible terminators are equipped with four ddNTPs with cleavable fluorescent dyes. The synthesis is performed a base at a time with the detectable light being captured by a charge coupled device. The synthesis cycle is then repeated as both the terminator group and fluorescent label are removed (Shendure and Ji 2008; Ansorge 2009; Zhou *et al*. 2010; Liu *et al.* 2012). The technology is the most popular current approach for next generation sequencing offering high compatibility, it has the highest throughput with read length reaching up to 300bp. On the other hand it is more laborious than the other technologies with high error accumulation rate in elongated DNA strands (Zhou *et al*. 2010; van Dijk *et al*. 2014).

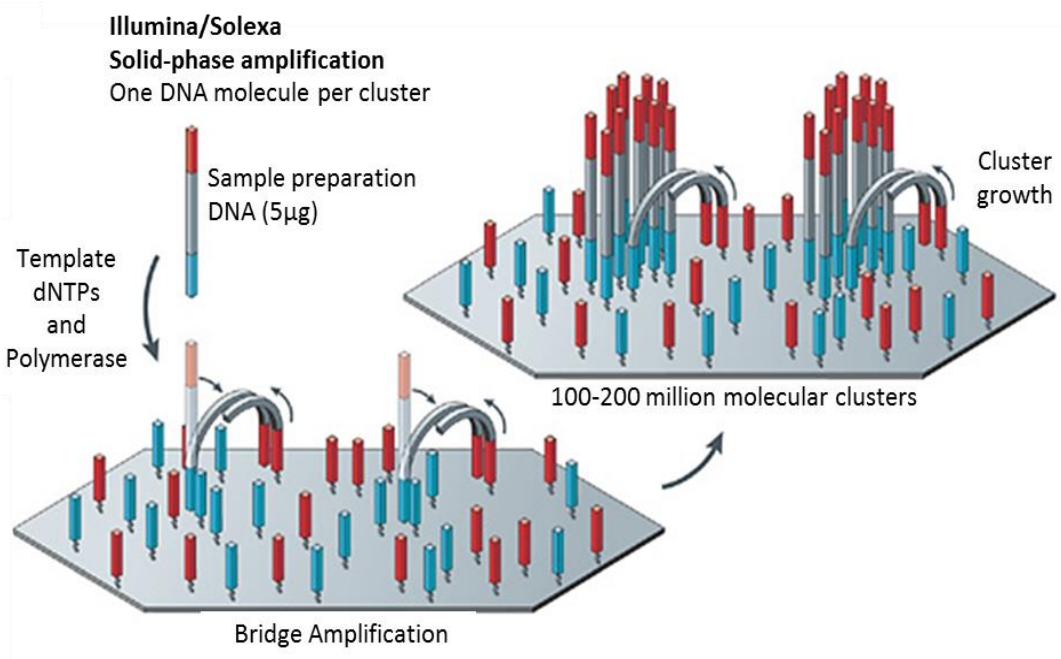


Figure 6.3 The principles of Illumina Solexa sequencing. Illumina Solexa technology is a 'sequencing by synthesis' technology where DNA is bound at one end to a solid surface while the other is binding to a complementary adapter forming a bridge. The bridge formation initiates the chain synthesis. Figure adapted from (Metzker 2010).

6.1.1.4 Supported oligonucleotide ligation and detection (SOLiD)

Supported oligonucleotide ligation and detection (SOLiD) was developed by Applied Biosciences in 2006 and is based on a two base ligation sequencing system. The DNA fragments is bound at its 5' end with an adapter where in turn a universal sequencing primer is hybridized. Octamers compete for their ligation to primers (Shendure and Ji 2008; Metzker 2010). The position one and two are designed to be interrogated and in position five there is a cleavage site as nucleotides five and six are bound with phosphorothiolate linkage which is cleaved by silver ions. The dinucleotides for the first and second position are labelled with four different fluorescent dyes enabling interrogation of 16 dinucleotides with the use of four different dyes. The process is repeated 10 times using a primer one base shorter (n-1) each time (Figure 6.4). SOLiD technology has the lowest error rate of all current technologies exhibiting approximately 99.95% accuracy, but the run is long lasting up to 14 days and the read length is relatively short at up to 75 base pairs (Liu *et al*. 2012; van Dijk *et al*. 2014).

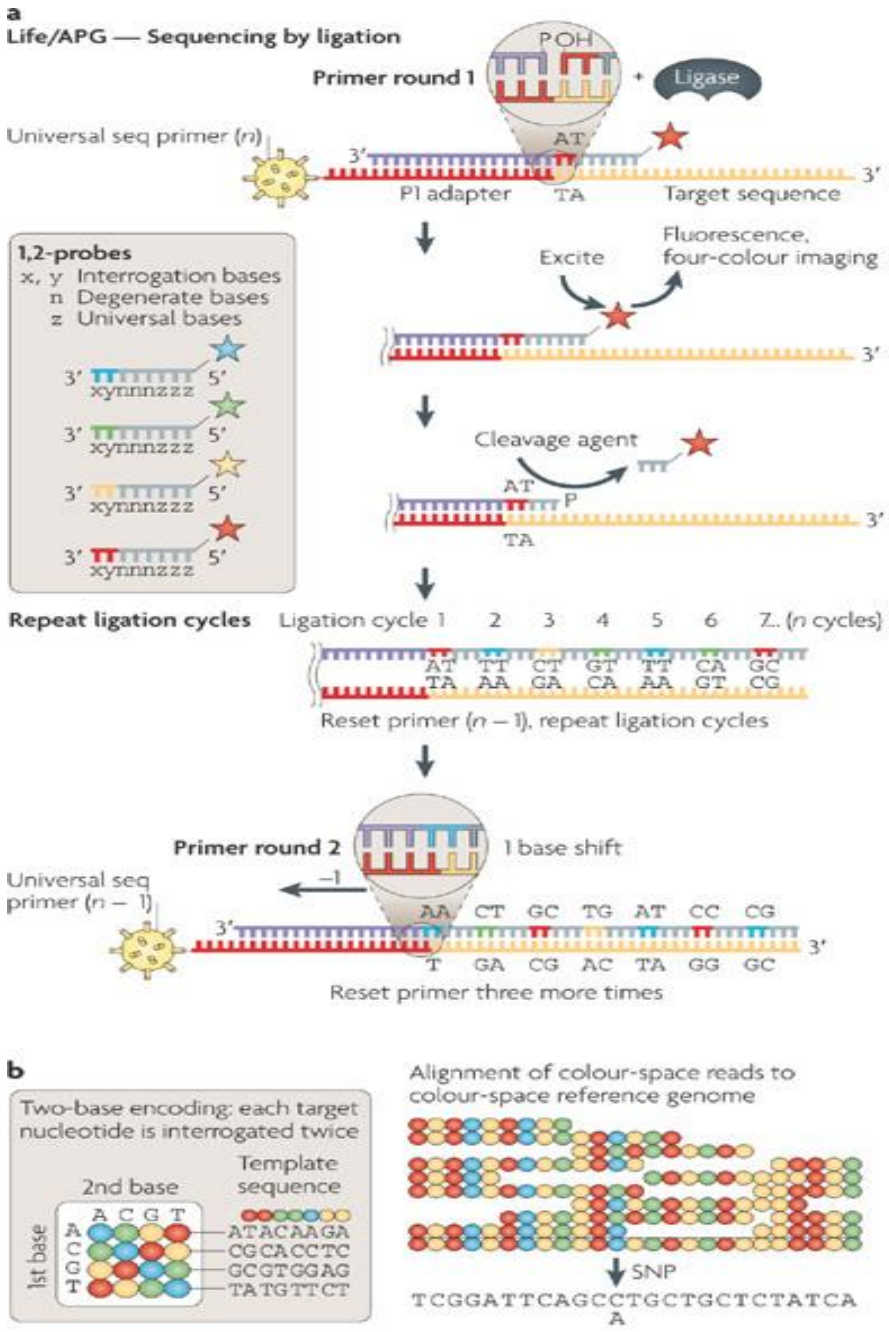


Figure 6.4 The principles of SOLiD sequencing. SOLiD sequencing is based on a two base ligation sequencing system. a) Octamers compete for binding next to the primer with the first two nucleotides to be designed for investigation. Following incorporation there is a cleavage site between position 5 and 6. b)The process is repeated 10 times each time with a primer 1 base shorter. Figure adapted from (Metzker 2010).

6.1.1.5 Semiconductor sequencing-Ion Torrent

Ion Torrent technology which is the sequencing method that was used in this chapter, is a relatively new advancement in next generation sequencing with the Personal genome analyser recently released in 2010. The technology is based on the 454 pyrosequencing principle with the innovation of being independent of light detection for the identification of incorporated nucleotides (van Dijk *et al*. 2014). In Ion Torrent sequencing DNA libraries are mixed in equal molarity with small beads following emulsion PCR. The beads contain specific sequences on their surface which are complementary to specific adapters. The DNA libraries are ligated to specific adapters and following the mix, each bead is bound to a different DNA library. Enriched beads are primed and placed into an Ion Chip which contains pH detection wells (Mardis 2008; Liu *et al*. 2012). Each bead is deposited into each well and following the incorporation of any nucleotide by the polymerase a proton is released causing a change in pH which is in turn detected by a sensor in the well (Figure 5). The technology does not require light detection and read length can reach up to 400bp in only a few hours. On the other hand, it shares the major drawback of 454 pyrosequencing as it is based on the same principle, therefore it has a high error rate in detection of homopolymer repeats (Mardis 2008; van Dijk *et al*. 2014).

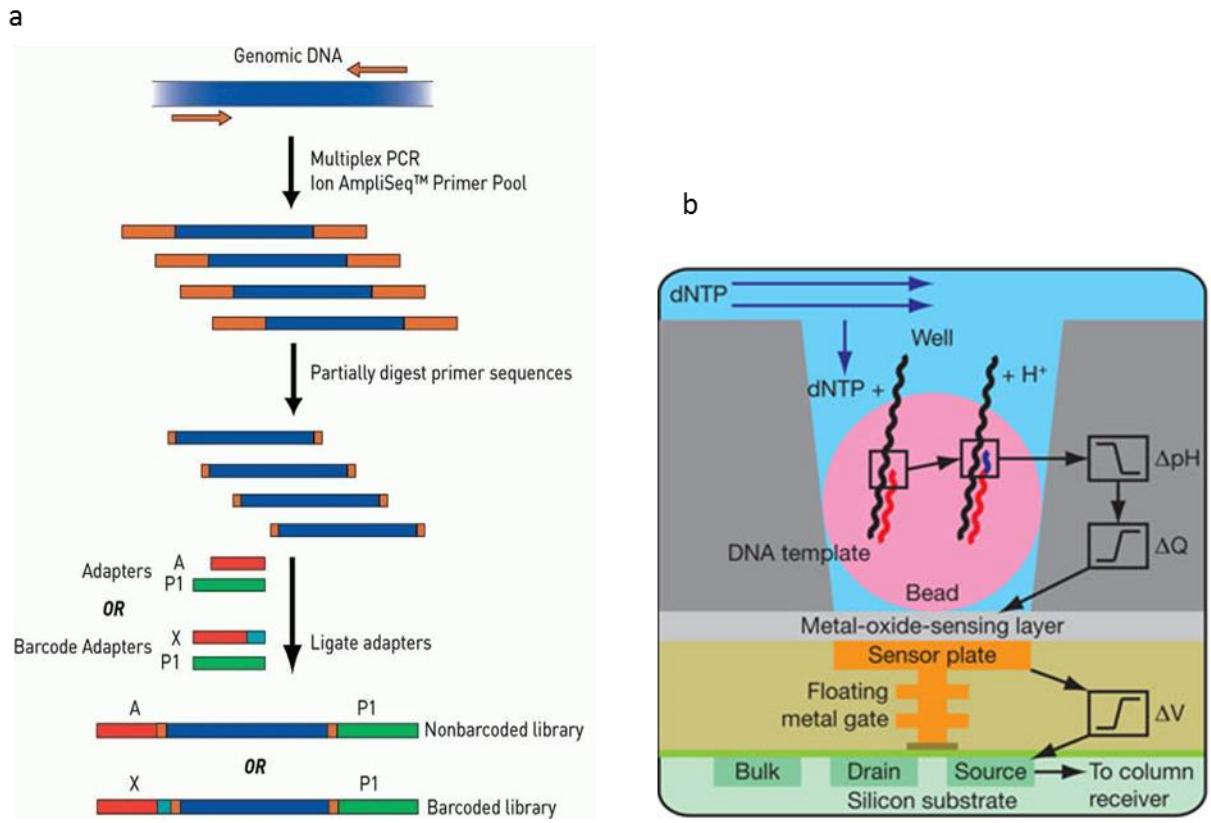


Figure 6.5 Semiconductor sequencing-Ion Torrent a) DNA library generation. Figure adapted from (Ion Torrent-Life Technologies), b) Incorporation of a nucleotide causes the release of a proton altering the pH. Small changes in pH are detected by sensors based in each well Figure adapted from (Rothberg *et al.* 2011).

6.1.1.6.1Ion Torrent platforms

Personal Genome Machine (PGM) was the first sequencing platform released by Ion Torrent. It is a powerful platform generating up to 2Gb output data. It is the cheapest sequencing platform available in the market. The PGM chip contains more than 1 million micro-machined wells generating more than 5 million reads per run (Figure 6.6). In 2012, Ion Torrent has recently released the Ion Proton Sequencer which can generate up to 10Gb of output data. The Ion Proton chip is larger than the PGM chip and contains more than 165 million micro-machined wells generating more than 80 million reads per run (Figure 6.6). Platform information was acquired from Ion Torrent, Life Technologies.

	PGM	Proton
Chip Types	**318**	**IP1**
Wells per Chip	>1 million	> 165 million
Mean Read bp	~200bp	~300bp
Number of Reads	~4-5.5 million	60-80 million
Yield/Q20, bases	2Gb per run	10Gb per run

Figure 6.6 Ion Torrent Chips. Information and Chip images were acquired from Ion Torrent Life Technologies.

6.1.1.7 Heliscope

Heliscope is a third generation sequencing technology which uses the same principle as illumina sequencing without requirement of DNA amplification. Sheared DNA hybridizes to a flow cell surface coated with primers (Thompson and Steinmann 2010). The primer extension and incorporation of labelled nucleotides generates visible light which is detected by a charge coupled device. Heliscope is a single molecule sequencing reaction which requires less material and doesn't suffer from GC content and size biases due to amplification but loss of synchronicity can lead to error accumulation during DNA synthesis (Figure 6.7) (Shendure and Ji 2008; Ansorge 2009; Thompson and Steinmann 2010).

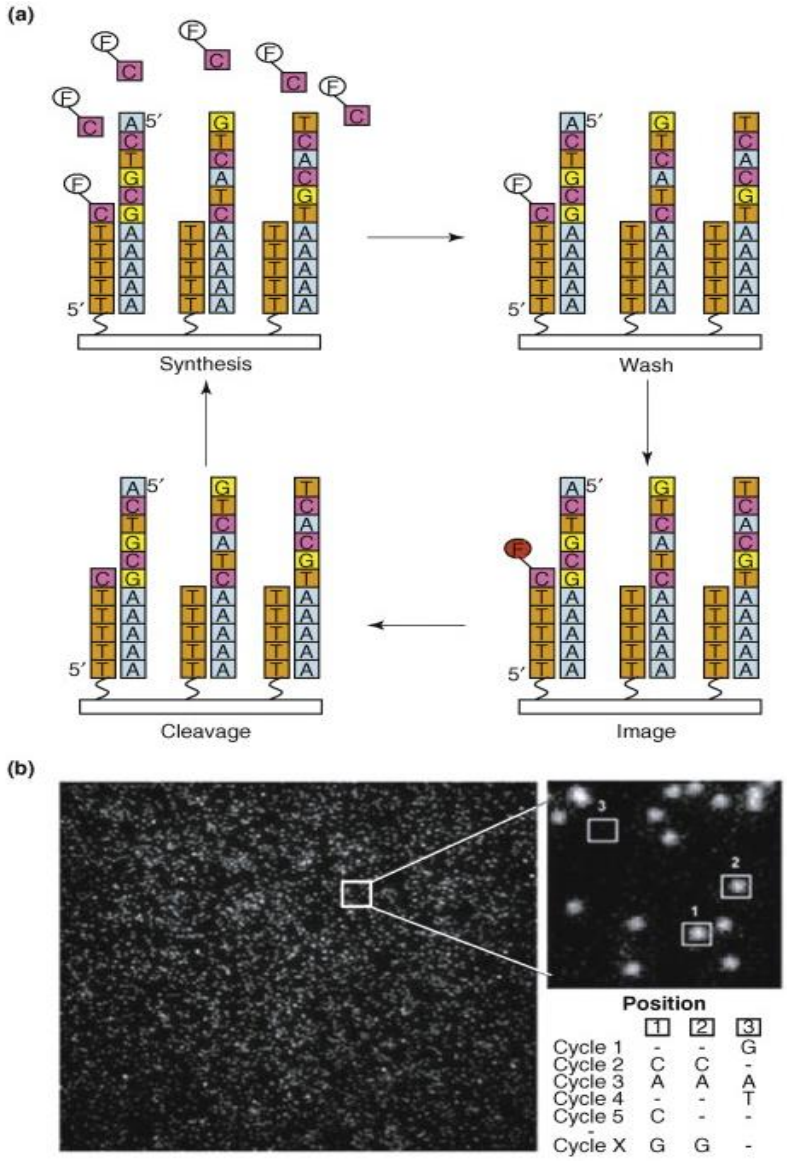


Figure 6.7 The principles of Heliscope sequencing. Heliscope is a third generation sequencing technology known as a single molecule sequencing technique. a) Its principle is based on Illumina solexa sequencing as a DNA fragment binds to flow surface coated with

primers. b) Incorporated nucleotides are detected by a charge coupled device. Adapted from (Gupta 2008).

6.1.2 Linkage analysis

Linkage analysis is a routine approach for the identification of a chromosomal location of specific genes associated with certain disease. The analysis is based on the principle that genes of the same chromosome which are located close to each other, following meiosis they remain linked to each other. The generation of genetic maps based on linkage analysis has been the foundation for the mapping of disease related genes (Pulst 1999).

6.1.3 Logarithm of the odds (LOD) score

The mathematical estimation of the likelihood of linkage is measured by the logarithm of the odds (LOD) score. The LOD score is calculated as the probability of two linked loci to the probability that the loci are unlinked.

$$\text{LOD score=Z= } \log_{10} \frac{Likelihood\ if\ the\ loci\ are\ linked}{Likelihood\ if\ the\ loci\ are\ unlinked}$$

Linkage is generally accepted to be significant with LOD score of 3 when the study include 100 to 300 markers. LOD score of 3 is translated as 1000:1 ratio in favour of genetic linkage a with a p value of 0.05. On the other hand, LOD score less than -2 is be significant against genetic linkage (Pulst 1999).

6.1.4 Parametric and Non-Parametric linkage analysis

Parametric and non-parametric linkage analysis are the two main computational solutions providing two different models of analysis. The decision for the selection of the correct tool is important. That depend on the genetic model that is decided for the analysis. Parametric analysis is more powerful than non-parametric when the genetic model is close to the real mode of inheritance of the disease (Kruglyak *et al.* 1996; Strauch *et al.* 2000). Parametric linkage analysis is a method based on a genetic model. It requires information such as gene frequency and penetrance parameters. However it can be misleading in cases of linkage model misspecification.

Non-parametric linkage analysis does not require a genetic model to operate and is considered essential for analysis of complex traits. It operates by evaluation of shared alleles between affected individuals independent from disease models (Liò and Morton 1997; Strauch *et al*. 2000; Dawn Teare and Barrett 2005).

6.1.5 Multipoint and single point linkage analysis

The most common linkage analysis method for linkage investigation in a small region with multiple markers is the multipoint linkage analysis. It is more powerful than the single point analysis when the identity by descent (IBD) is ambiguous as it utilises information from haplotypes from multiple markers in contrast to the single point analysis (Sullivan *et al.* 2003). However, multipoint analysis is sensitive to misspecification of distances between the markers in such small regions (Halpern and Whittemore 1999).

Single point analysis is uses the locus of a known genetic marker and the locus of an unknown disease marker thus is also known as a two point linkage analysis. Every variant is evaluated individually reducing the sensitivity due to misspecifications. Therefore, it is often considered important that both analyses be performed (Halpern and Whittemore 1999; Strachan and Read 2011).

6.1.6 MERLIN genetic linkage analysis software

The software selected in this chapter for the non-parametric linkage analysis was MERLIN, which was devised by the University of Michigan. It is a software that doesn't require knowledge of scripting, performing non parametric linkage analysis faster than any other software available on the market. It is considered an ideal solution for analysis of samples with large pedigrees due to its unique linear model for linkage analysis (Kong and Cox 1997; Dudbridge 2003; Neale *et al.* 2012).

6.1.7 Aim

Genome wide linkage analysis performed by Dr. Bakhsh (unpublished data) has revealed that there is a high LOD score region on chromosome 20 which includes PLCβ1 InDel. The analysis in this chapter has focused on the exome of the chromosome 20 high LOD-score region using next generation sequencing (NGS). Exome sequencing analysis aimed either to identify other potential genetic variants which may be responsible for the disease or to confirm that the PLCβ1 InDel is solely responsible for the pathogenesis of the patients (Figure 6.8). The initial focus was on the family (18 patients) in which the InDel and its linkage with MNG progressing to PTC was first identified. I also investigated other European patients who carry the InDel (3 patients identified in chapter 4) and subjects from families having similar pathogenesis to the original family. Finally I applied NGS to the cohort of 80 patients with Multinodular Goitres, 4 of whom are known to be InDel heterozygotes. PLCβ1 InDel carriers, were screened by NGS.

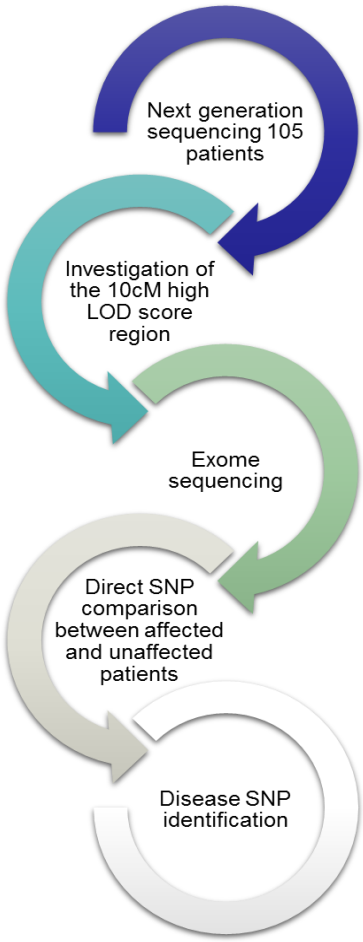


Figure 6.8 Aims of the chapter. Summary of the investigations for the identification of potential disease related variants across the high LOD score 10cM region.

6.2 Material and Methods

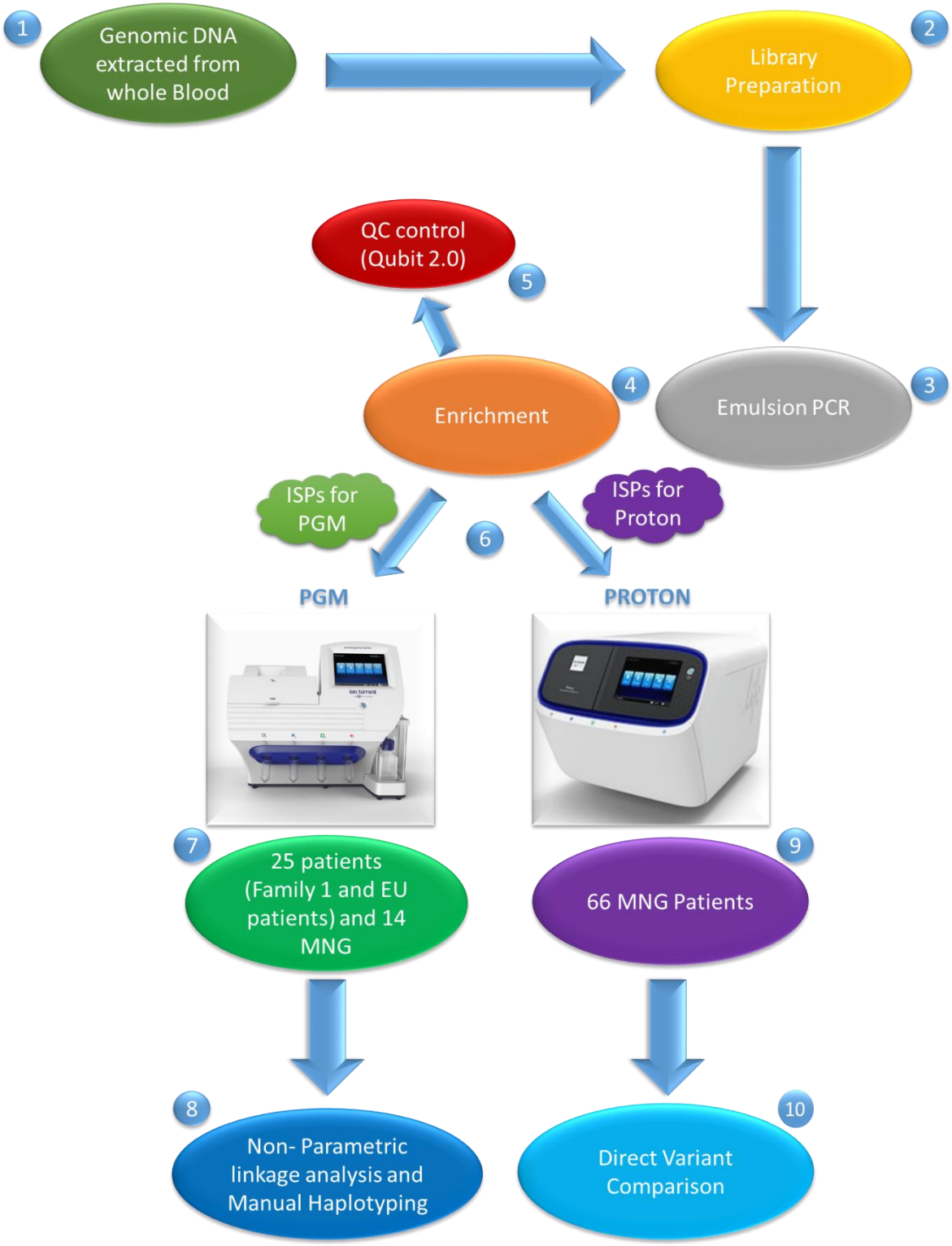


Figure 6.9 Summary of the work flow in this chapter. Quality Control (QC control), ISPs (Ion Sphere Particles), PGM (Personal Genome Machine), MNG (Multinodular Goitre). Ion Torrent platform pictures were acquired from Life Technologies.

6.2.1 Primer Design

The primer pools for the preparation of the DNA libraries were designed with the use of Ampliseq 3.0.1 software (https://ampliseq.com/) according to the manufacturer's protocol. A total number of 429 primers were designed generating 100-300bp amplification products. The designed primer pools covered the exome sequences (all coding regions, intron/exon boundaries, proximal promoters and 3' untranslated regions) of a region spanning from chr20: 8113337 to 11907302. A complete list of the primers used for the DNA library generation can be found in Appendix. Primer design was performed by Dr. Mantripragada.

6.2.2 Picogreen assay DNA quantitation

Genomic DNA was extracted from whole blood (as described in Chapter 4) and quantified with Quant-iT™ PicoGreen® dsDNA Assay kit (Life Technologies). Genomic DNA was initially diluted 1:4 in 1x TE buffer and 2µl were transferred to a 96 well plate and made up to 100ul using Picogreen (Invitrogen) diluted (1:200) in 1x TE buffer. In addition to the genomic DNA samples, a series of DNA standards (Lambda DNA standard) were prepared for the standard curve (Table 6.1). The plate was then loaded into the Fluoroskan Ascent fluorometer which measured absorbance between 485-538nm. Genomic DNA concentration was estimated with the use of the standard curve and the following formula (Fluorescence reading-c)/gradient)*Z, where c is the y-intercept and *Z the dilution factor.

Table 6.1 Picogreen DNA standard concentrations

Standard	**ng/µl**
1	75
2	50
3	25
4	12.5
5	6.25
6	3.125
7	1.5625
8	0

6.2.3 Ion AmliSeq Library preparation

6.2.3.1 PCR amplification of genomic DNA targets

Approximately 10ng of the genomic DNA of interest and an AmpliSeq primer pool of 429 primers were amplified in a 20µl reaction (Table 6.2).

Table 6.2 PCR amplification of genomic DNA targets. a) AmpliSeq (Life Teechnologies) reaction mix b) Amplification programme

a

Component	Volume (µl)
5X Ion AmpliSeq HIFI Mix	4µl
2X AmpliSeq Primer Pool	10µl
Genomic DNA 10ng	X
Nuclease- free Water	5-X
Total	20 µl

b

	Temperature (°C)	Time
Hold	99	2 min
Cycle (22cycles)	99	15 sec
	60	10 min
Hold	10	∞

6.2.3.2 Partially digestion of primer sequences and adapter ligation

The primers of the amplified DNA samples were partially digested by the addition of 2µl of FuPa reagent (Life Technologies) and placed in the thermal cycler (Table 6.3a). In the meantime barcode/adapter mix was prepared and the partially digested primer DNA samples were then ligated with adapters and placed on the thermal cycler for the ligation reaction (Table 6.3b, c, and d).

Table 6.3. Primer digestion and Adapter ligation. a) Partial digestion of primer sequences, b) Adapter ligation mix, c) Barcode adapter mix, d) Adapter ligation reaction

a

Temperature (°C)	Time (min)
50	10
55	10
60	20
10	Hold for up 1 hour

b

Component	Volume (µl)
Switch Solution	4
Barcode Adapter mix	2
Total Volume	28

c

Barcode Adapter mix for 4 reactions	
Component	**Volume (µl)**
Ion P1 Adapter	2
Ion Xpress Barcode X[1]	2
Nuclease-free Water	4
Total	8

d

Temperature (°C)	Time (min)
22	30
72	10
10	Hold for up 1 hour

6.2.3.3 Purification of unamplified library

DNA libraries were then purified using Agencourt AMPure XP beads (Beckman Coulter) at 1.5X sample volume. DNA libraries were initially treated with 45µl of 1.5x Agencourt AMPure XP Reagent, they were mixed and incubated for 5 minutes at room temperature. A magnetic rack was then placed underneath the DNA library plate and incubated for 2 minutes until solution appeared clear as the beads were magnetised by the magnetic rack to the side of the tube. Washing steps were then followed by the addition of 150µl of 70% ethanol while the plate was on top of the magnetic rack. The beads were allowed to mix by moving the magnet side to side and then the supernatant was discarded and washing step was repeated for a second time and the samples were then allowed to air-dry for 5 minutes at room temperature. The plate was then removed from the magnet and DNA was eluted from the beads with the addition of 50µl of Low TE (Life technologies) per samples which were in turn prepared for qPCR quantification. A 100-fold dilution was prepared for each of the samples with nuclease free water (Life Technologies). Series of dilutions were also prepared from *E.coli* DH10B control library for the generation of a DNA standard curve in qPCR amplification (Table 6.4). Reaction mix in a total volume of 10µl was then prepared and loaded into a 96 well plate. All samples including DNA standards were in triplicates and loaded on the qPCR thermocycler (StepOne Applied Biosystems) for qPCR amplification (Table 6.4c).

Table 6.4. Quantification of DNA libraries. a) *E.Coli* DH10B DNA standard concentrations, b) Ion Library TaqMan qPCR mix, c) qPCR amplification programme

a

Standard	Fold Dilution	Concentration (pM)
1	0.1	6.8
2	0.01	0.68
3	0.001	0.068
4	0.0001	0.0068
5	0.00001	0.00068

b

Component	Volume per 20µl reaction
Ion Library TaqMan qPCR Mix 2X	5
Ion Library TaqMan Quantitation Assay 20X	0.5
Nuclease free Water	2
Sample Library	2.5

c

	Temperature (°C)	Time
Hold	50	2 min
	95	25 sec
Cycle (45cycles)	95	1 sec
	60	20 sec

6.2.4 Preparation of template-positive Ion PGM Template Ion Sphere particles

Undiluted Ion AmpliSeq libraries were quantified with qPCR and based on the library concentration they were subsequently diluted with Low TE (Life technologies) to a final concentration of 100pM. Diluted libraries were then combined together. The library mix was then prepared for Emulsion PCR with Ion OneTouch 2 (Life Technologies). The preparation of the reaction mix (Table 6.5) was followed by the addition of 100µl of Ion PGM Template OT2 200 Ion Sphere particles. The mix was transferred to the Ion OneTouch Reaction tube which was filled with 1.5ml of Ion OneTouch Reaction Oil. The tubes were then placed on the Ion OneTouch 2 (Life Technologies) and emulsion PCR was performed.

Table 6.5 Emulsion PCR. a) Combining Ion AmpliSeq libraries, b) Emulsion PCR reaction mix

a

	Ion AmpliSeq DNA Library
Library Concentration	100pM
Volume of library	2µl
Volume of Nuclease-free Water	23µl
Total volume of diluted library to add to the amplification solution	25µl

b

Order	Reagent	Volume (µl)
1	Nuclease-free Water	25
2	Ion PGM Template OT2 200 Reagent Mix	500
3	Ion PGM Template OT2 200 PCR Reagent B	300
4	Ion PGM Template OT2 200 Enzyme Mix	50
5	Diluted library	25
-	Total	900

6.2.5 Ion Sphere particles recovery

The amplified product was collected from the Ion OneTouch and Ion OneTouch Recovery solution was removed leaving 50µl to avoid disturbance of the Ion Spheres. The ISPs were then resuspended and following the addition of 1ml of Ion OneTouch wash solution, were centrifuged for 2.5 minutes at 15,500xg and all but 100ul of supernatant was discarded. All reagents were acquired from Life Technologies.

6.2.6 Enrichment

The entire template-positive ISP sample (100µl) was transferred to the first well of an 8 well strip, 130µl of Dynabeads MyOne Streptavidin C1 beads washed and resuspended according to the Ion PGM OT2 200 protocol (Appendix) were added to the second well (Figure 6.11). Ion one touch wash solution (300ul) was added to wells 3, 4 and 5. Melt Off solution (300µl) (125mM NaOH and 0.1% Tween 20) was added on well 7and wells 6 and 8 remained empty (Figure 6.10). The 8 well strip was placed at Ion OneTouch ES which was initialised and run according to the Ion PGM OT2 200 protocol. Once enrichment was completed the sample was centrifuged at 15,500 x g for 1.5 minutes and then supernatant was removed leaving 10µl of solution followed by the addition of 200µl of Ion OneTouch Wash Solution. A total of 10µl were transferred in a separate PCR tube for polyclonality assessment by Qubit 2.0 Fluorometry. All reagents were acquired from (Life Technologies).

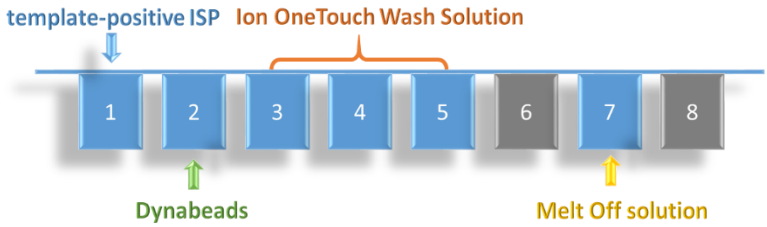


Figure 6.10 Loading order of 8 well strip for ISP enrichment

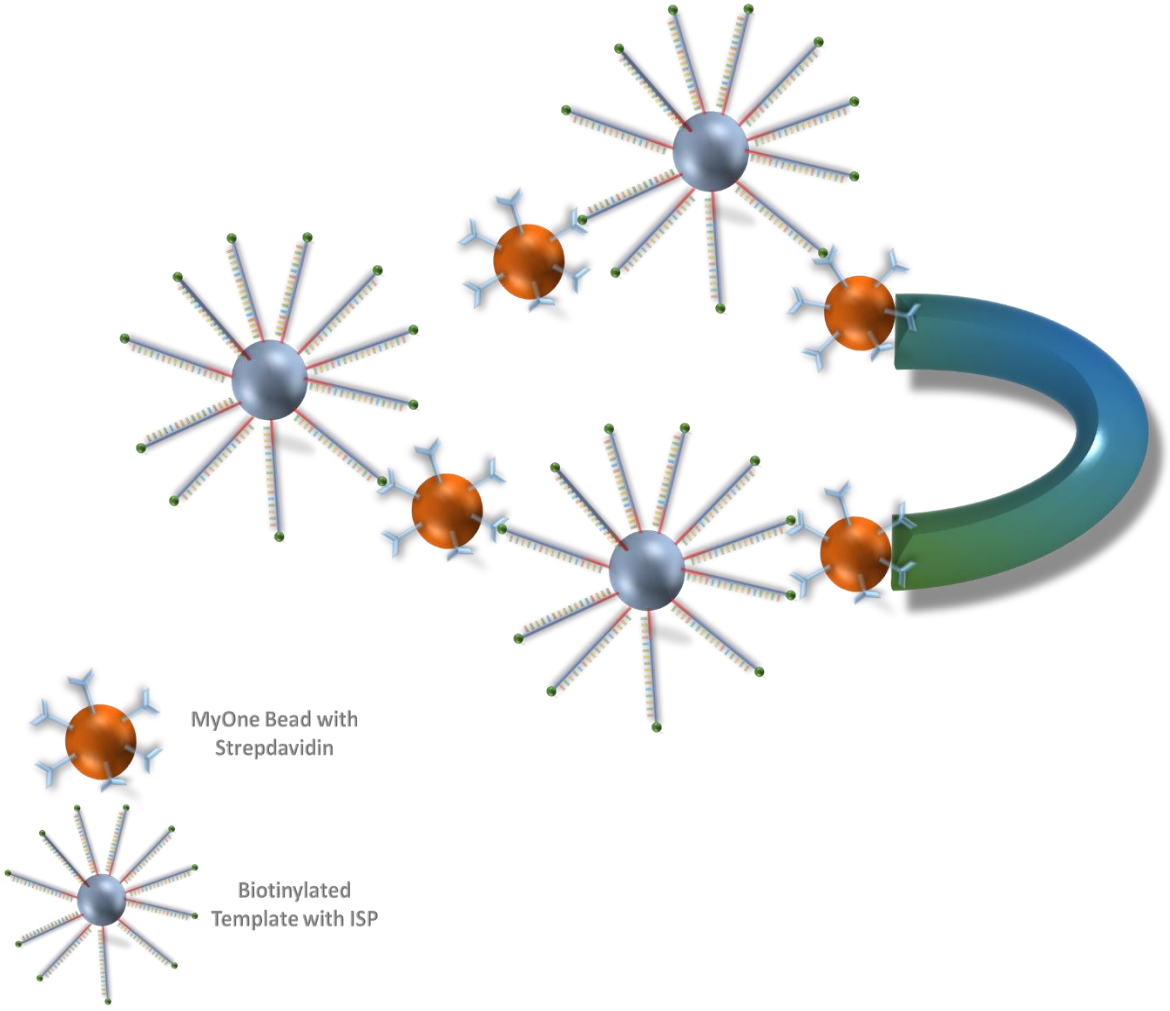


Figure 6.11 Enrichment of template-positive ISPs. The biotin attached at the end of DNA sequences binds to streptavidin on MyOne Beads which in turn are magnetised by a magnet.

6.2.7 Qubit 2.0 Fluorimetry

ISPs were mixed with 19µl of Annealing Buffer, 1µl of Ion Probes and 20µl of Ion Probe Master Mix and placed in a thermocycler (Table 6.6). Quality Control Wash Buffer (200µl) was added and following a vortex the mix was centrifuged at 15,500 x g for 1.5 minutes. Supernatant was then removed leaving 10µl to avoid disturbing the pellet. The washing steps were repeated two times and then the pellet was resuspended in a total volume of 200µl of Quality Control Wash Buffer. The sample was then inserted into the Qubit fluorimeter for quality assessment. Reading were recorded for both Alexa Fluor 488 and 647. All reagents were acquired from (Life Technologies)

Table 6.6. Annealing Ion probes programme

	Temperature (°C)	Time
Hold	50	2 min
Hold	95	20 sec
Cycle (40 cycles)	95	1 sec
	60	20 sec

6.2.8 Enriched template-positive ISPs preparation for PGM

Control Ion sphere particles were briefly vortexed and centrifuged for 2 seconds then 5µl were added to the, enriched-template positive ISPs. The mix was centrifuged for 2 minutes at 15,500xg then supernatant was removed leaving 15µl at the bottom of the tube. The pellet was resuspended by adding 12µl of sequencing primer. The total 27µl reaction mix was then placed in a thermo-cycler for 2 minutes at 95 °C and then 2minutes at 37 °C.

The ISP reaction mix (27µl) was then transferred at room temperature where a 3µl of Ion PGM Sequencing 200 v2 Polymerase to the ISPs. The mix incubated at room temperature for 5 minutes and then was ready for loading into the PGM Chip. The Ion 318™ Chip kit v2 was initialised and loaded as it was described at Ion 318™ Chip kit v2 User Guide, Appendix.

6.2.9 Template-positive Ion PI Ion Sphere Particles for Proton

Enriched template-positive ISPs were centrifuged at 15,500xg for 5 minutes. The supernatant was removed leaving 10µl at the bottom of the tube. The pellet was resuspended by adding 15µl of Ion PI Annealing Buffer. Also 20µl of Ion PI Sequencing primer was added to mix in a

total volume of 45μl. The mix was vortexed and placed in a thermos- cycler at 95°C for 2 minutes and then at 37°C for 2 minutes. The ISPs were then placed at room temperature following the addition of 10μl of Ion PI loading buffer. The Ion PI Chip V2 was initialised and loaded as it was described at Ion PI Sequencing 200 kit v3 User Guide, Appendix.

6.2.10 Calibration, Ion PI Chip loading and Sequencing run

The sequencer was calibrated prior to the chip loading according to the manufacturer's protocol (Appendix). The chip was then loaded, planned run was selected where each barcode was annotated for each sample. Once the predefined plan was selected, the sequencing run was performed. The sequencing data analysed by Ion Torrent Suite software (4.4.2), using the plug-in variant caller (v 4.2.10) and configuration with generic Personal Genome Machine (PGM) germ line settings and high stringency analysis mode. The output of the variant caller was analysed with Microsoft excel where a summary of each variant can be viewed and Integrative Genomics Viewer (IGV) where comprehensive details about the individual variant call can be assessed.

6.2.11 Linkage analysis

Merlin in order to operate requires three types of files a Pedigree, a Map and a Data file. The three files were created as it is described below and were converted into text files (tab delimited) so it can be uploaded in Merlin software and non-parametric linkage analysis to be performed.

The results from the next generation sequencing of the original family were processed in an excel file. A table was prepared with a list of the total number of different SNPs raised from the sequencing and the calls for each allele for the different SNPs for each patient. The SNPs which were not present for some patients and therefore had no calls were considered to be homozygous for the wild type. The PedFile which was required for Merlin software and non-parametric linkage analysis was prepared based on the excel file with the total list of different SNPs comprising the genotype for each patient. The new file included all the relevant information for each member of the family such as relation to each other, sex, disease status and genotype (Figure 6.12). Also for some patients only maternal SNPs were known as their father was not sequenced and therefore a dummy father was created where his genotype was illustrated as 0 0 for each SNPs. The genotype was then converted into number such as A, C, G, T to be 1, 2, 3, and 4 respectively (Figure 6.12).

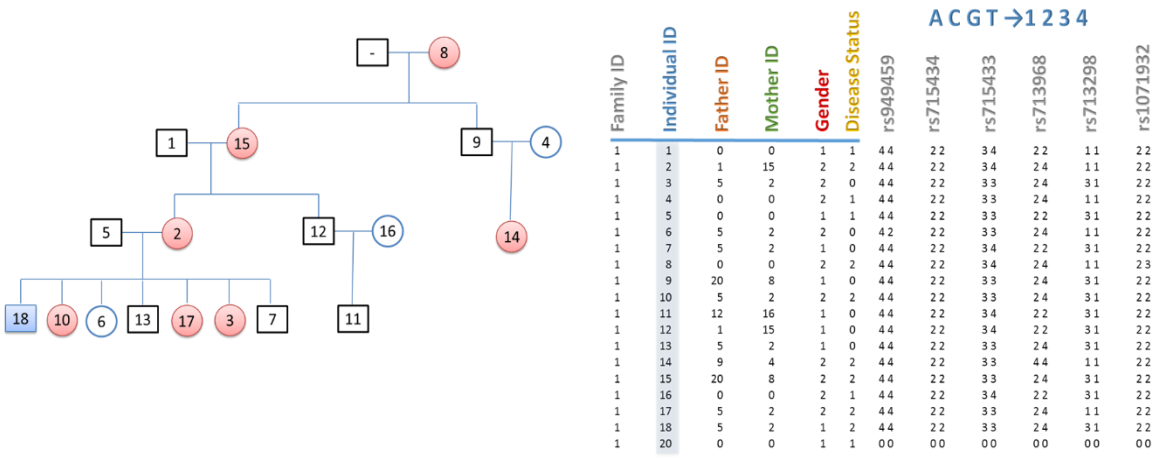


Figure 6.12 The pedigree file. An example of a Pedigree file where the first 6 columns illustrate family ID, individual ID, father ID, mother ID, gender and disease status followed by with genotype information for each SNP. The highlighted column illustrates the individual ID of the members of the family as seen in the family tree.

The second file that was required for Merlin software was the Map File. This file contained three columns including information regarding chromosome number, SNP name, and chromosomal position in centimorgan (cM) respectively, (Figure 6.13b).

The third file is called Data File it contains a list with the total number of SNPs found in sequencing. The first row starts with the disease title and the last with Age (Figure 6.13a).

a

	Thyroid
A	Cancer
M	r1
M	r2
M	r3
M	r4
M	r5
M	r6
M	r177
M	r178
M	r179
M	r180
M	r181
C	AGE

b

Chrom.	SNP	Position in cM
chr20	r1	8.113405
chr20	r2	8.626859
chr20	r3	8.626903
chr20	r4	8.637727
chr20	r5	8.665751
chr20	r6	8.665793
chr20	r7	8.678446
chr20	r8	8.696907
chr20	r9	8.698565
chr20	r10	8.703145

Figure 6.13 The data file. a) Data File with first row being the disease title (A) including a list of markers (M) and ending with Age (C) title. b) Map File consisted of three columns the first is the chromosome ID (chr20), the second SNP ID and the third cM position.

6.2.11. 1 PedCheck anaylsis for Mendelian error identification

PedCheck software has been used for the identification of potential Mendelian errors within the family. The analysis included 4 levels of assessment and revealed 4 inconsistencies across 181 SNPs. The inconsistencies were removed and the new file was used for Merlin analysis.

6.2.11.2 Graphical Representation of Relationships (GRR) assessment.

Potential labelling errors were investigated with the use of GRR software. The software performed an allele comparison between the individuals to show that relatives share also a pattern of alleles. The pedigree file prepared for Merlin was also used for GRR analysis.

6.2.11.3 Non parametric linkage analysis

The analysis was performed with the use of Merlin software. Merlin was downloaded from the University of Michigan centre of statistics. Analysis required three files, a pedigree file, a map file and a data file. A comprehensive list of the commands used for merlin can be found in Appendix. Merlin performed multipoint and singlepoint non parametric linkage analysis providing LOD score for the SNPs across the 10 cM region.

6.2.14 Manual Haplotyping

A comprehensive overview was performed by manual haplotyping. The process involved individual SNPs inheritance assessment from parents to children and identification of the disease-risk allele. Haplotyping can reveal chromosomal recombinations and in combination with Merlin indicates which part of the sequenced region is a potential disease region.

6.2.15 Identification of potential disease SNPs

6.2.15.1 Filtering Step

The filtering steps of the potential disease variants included allele frequency identification with the use of University of California Santa Cruz human genome browser, as only rare variants less than 2% in the general population could be considered as disease related (Figure 6.14).

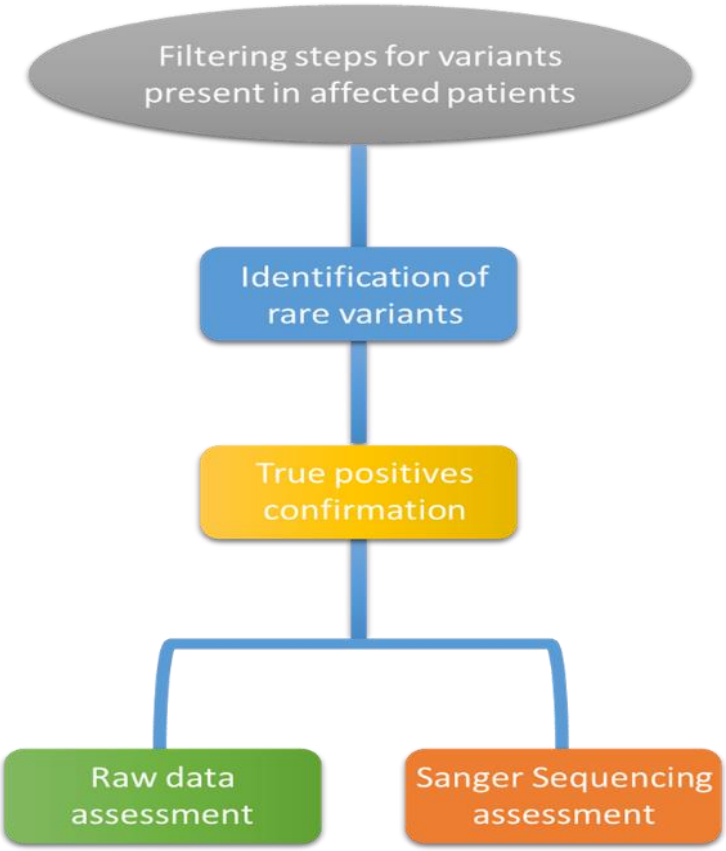


Figure 6.14 Summary of the variant identification filtering steps

6.2.11.4 Raw data assessment

The raw data were also assessed for potential presence of false negative variants. The raw data files were downloaded from the Ion Torrent server in IGV format and were viewed with use of IGV software a visualisation tool which allowed the comprehensive investigation of individual variant read depth in different calls.

6.2.11.5 Sanger sequencing confirmation

The identification of a single rare variant led to further investigation with Sanger sequencing for confirmation of true positives and true negatives from NGS results (as described in chapter 2). Primers were designed that flank the region of the variant with the use of primer 3 software (http://primer3.ut.ee/) (Table 6.7).

Table 6.7 Primers designed for the identification of true and false positives.

Primer Name	Tm	Sequence	Tm	Gene
F1/*R1*	64	CACACTTAGCTCGCCATCAA/TTTTAAGCCGCCAAGTTTTG	61	Intron SNAP-25

6.11.6 Study of Health in Pomerania (SHIP)

The Study of Health in Pomerania (SHIP) includes patients from the Pomerania region in East Germany which has shown particularly elevated mortality rate compared to West Germany. The lack of scientific studies has led to the necessity of the SHIP study. The aim of the study was the investigation of the prevalence of common diseases and disorder within this population. The first cohort recruited comprised from 4,308 patients the second from 3,300 and a third cohort is currently being recruited (Völzke *et al.* 2011).

The SHIP cohort of 986 patients was genotyped for the presence of ANKRD5 SNP. The cohort consisted of patients suffering from nodules and/or goitres and the analysis was kindly performed by Prof. Henry Völzke and Dr. Alexander Teumer.

6.3 Results

6.3.1 Quality control

6.3.1.1 TaqMan assay quantitation of Ion Torrent libraries

Genomic DNA was initially quantified using of PicoGreen which stains only double stranded DNA as it is exhibits emission at 530nM excluding readings from single stranded DNA. The samples were quantified twice and the average of the two measurements were used for the library preparation (Figure 6.15).

a

	A	B	C	D	E	F	G	H	I	J	K	L
1	21	45	49	72	16	250	14	21	40	254	258	29
2	90	32	-1	40	26	38	22	32	43	95	37	17
3	49	37	3	35	131	25	94	24	7	55	231	19
4	35	149	10	7	9	16	42	2	28	46	230	31
5	192	41	2	44	27	4	57	24	15	26	30	54
6	106	1	11	11	151	6	1	36	13	26	89	57
7	20	18	25	306	25	-4	8	86	76	7	51	59
8	29	36	18	-1	24	58	158	40	47	10	97	-4

b

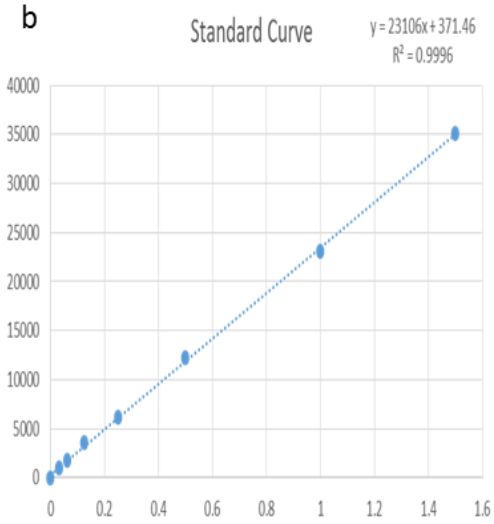


Figure 6.15 TaqMan quantitation. a) The concentration of genomic double stranded DNA was estimated (ng/μl) with the use of picogreen stain, b) DNA standards were used for standard curve and accurate calculation of genomic DNA concentration.

6.3.1.2 Amplification confirmation

Selected DNA samples were then amplified using of Ampliseq amplification kit. Confirmation of amplification efficiency was performed with the Bioanalyser. The Bioanalyzer chip has the ability to provide accurate measurement of the size and number of amplification products for 11 samples at the same time (Figure 6.16).

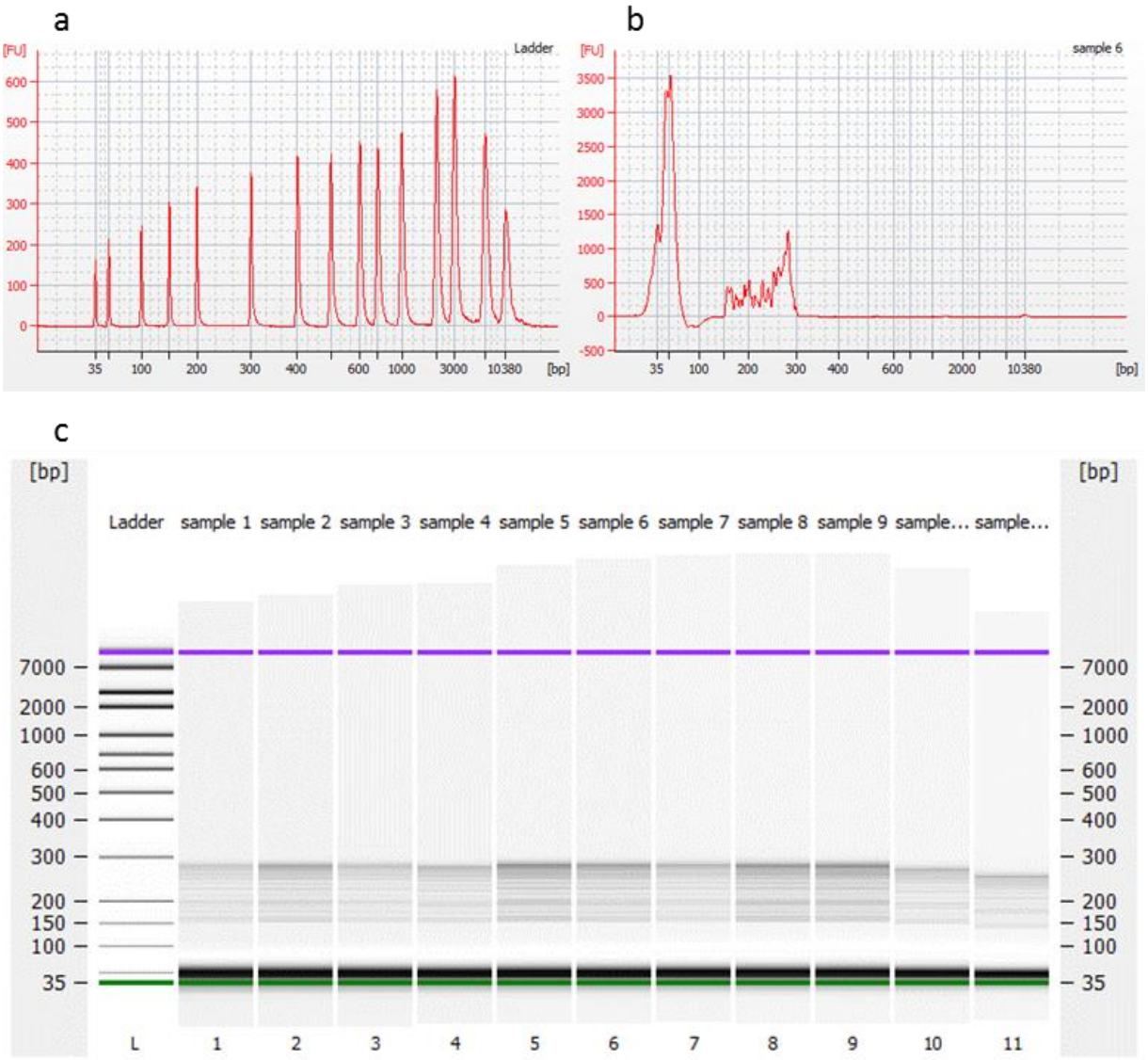


Figure 6.16 Assessment of Amplified genomic DNA. a) Electropherogram of a DNA ladder, b) and of a DNA library. c) Electrophoresis run of DNA libraries.

6.3.1.3 QPCR quantification of DNA libraries

Amplified libraries were partially digested and ligated with barcodes following purification. The purified libraries were then quantified and the concentration was estimated with the use of a standard curve.

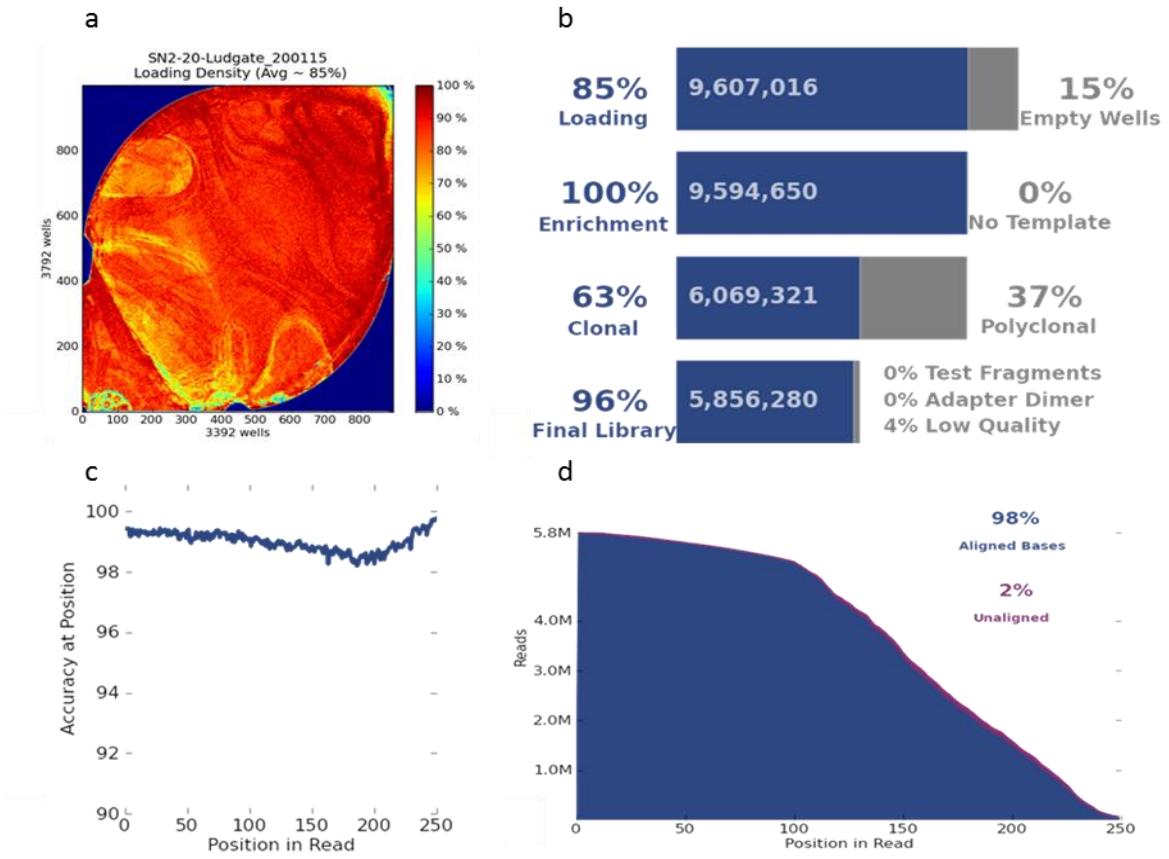
6.3.1.4 Polyclonality Assessment

The pools were pulled together amplified and enriched with ISPs. The level of polyclonality which indicated more than one template per ISP was estimated prior to the Chip loading with the use of the Qubit 2.0 fluorimeter. Measurements were taken both for Alexa Fluor 488 and 647 nm with levels of polyclonality between 20 to 30 % which is considered acceptable levels for sequencing.

6.3.2 Next Generation sequencing data quality assessment

6.3.2.1 Sequencing with Personal Genome Machine

The Personal Genome Machine has generated a total of 939Mbp of data, achieving a 99% accurately mapped sequences, over 91% of the percentage of target bases covered by at least 0.2 times the average base read depth (Figure 6.17).



e

	AQ17	AQ20	Perfect
Total Number of Bases [Mbp]	858 M	773 M	595 M
Mean Length [bp]	156	145	116
Longest Alignment [bp]	325	320	302
Mean Coverage Depth	0.3	0.2	0.2

Figure 6.17 Quality assessment of the generated data. a) Chip loading with approximately 85% occupied wells , b) breakdown of the final number of libraries screened, c) Average raw (1x) accuracy across each individual base position in a read, d) Total number of bases aligned to the reference sequence, e) Alignment quality (AQ17, mega bases of DNA with one mismatch in the first 50 bases relative to the reference strain, AQ20, mega bases of DNA with one mismatch in the first 100 bases relative to the reference strain)

6.3.3 Next generation sequencing of 25 patients

The first analysis included 25 patients (Figure 6.19); 18 were the members of the family carrying the PLCβ1 InDel, 4 European patients found to carry the InDel (The 22 patients mentioned are described in Chapter 4) and 3 patients from a family with similar phenotype to Family 1 with the II-2 female found to suffer from MNG, developing PTC. These 3 patients were genotyped for the PLCβ1 InDel and found not to be carriers (Figure 6.18).

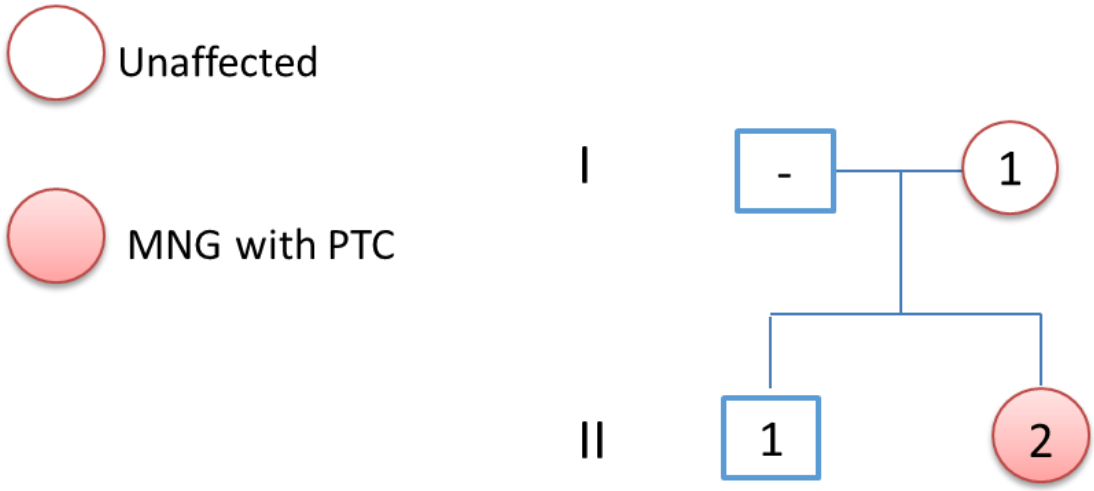


Figure 6.18 Family II. The family which illustrate similar phenotype with Family1 (described in Chapter 4). The II-2 female found to suffer with MNG developing PTC.

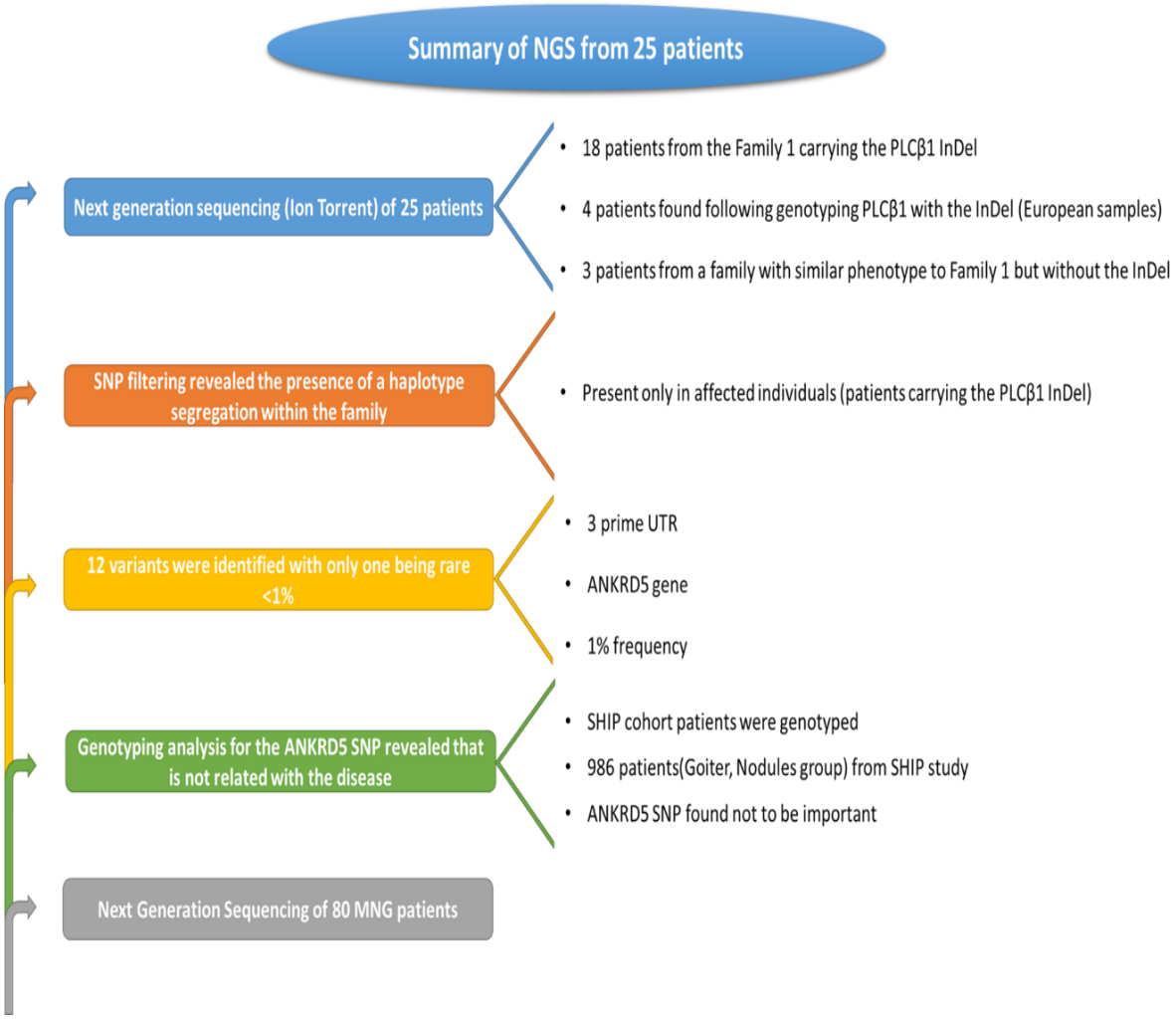


Figure 6.19 Summary of the data assessment from next generation sequencing with Personal Genome Machine. The screening included 25 patients 18 of whom were patients from the family carrying the PLCβ1 InDel.

6.3.4 Quality Control of files prepared for Merlin

The list of total different variants from all 25 patients was assessed for labelling and Mendelian errors by GRR and Pedcheck software respectively. The GRR software compared the alleles of the individuals for assessment of their relationships. The results indicated three different clusters in the first cluster, closer relatives such as brothers and sisters seemed to share most alleles. The second included parent- child and the third were unrelated individuals (Figure 6.20). Careful assessment of the relationship between the sequenced patients did not identify any labelling error. Also Mendelian error assessment with the use of PedCheck was performed at 4 levels and identified 4 inconsistencies across the pedigree file which were excluded from the study. The clusters appeared to be in close proximity to each other as the assessed

chromosomal region was relatively small with small number of SNPs therefore genetic similarities had a profound effect in allele comparison in that small region.

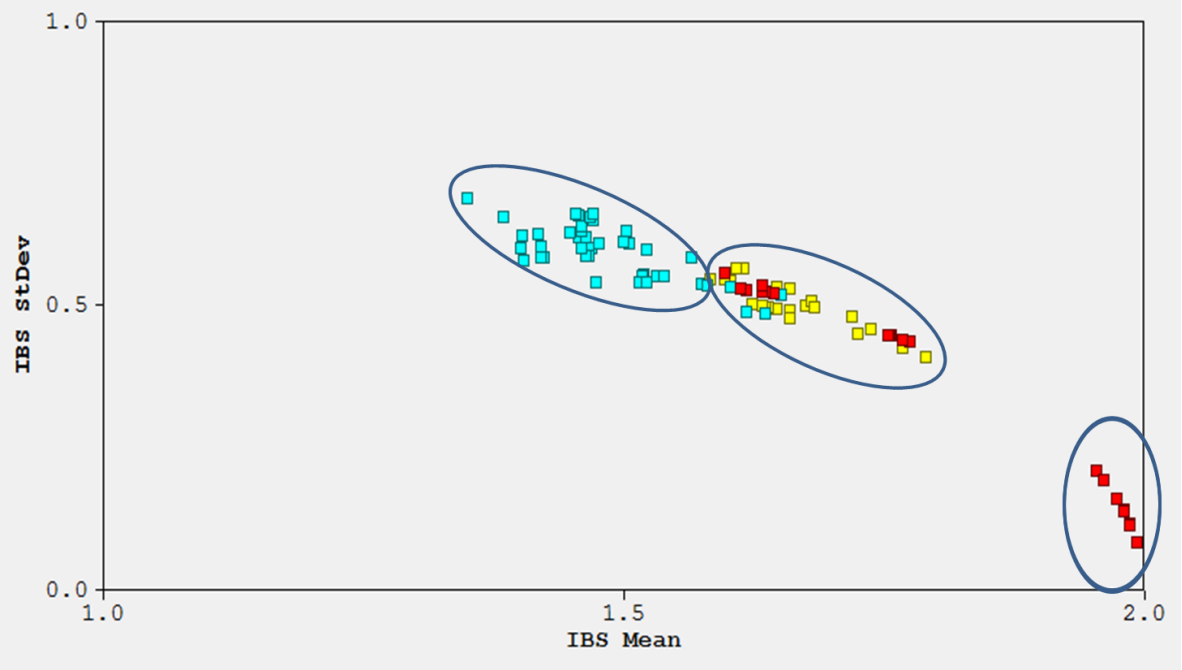


Figure 6.20 Summary of allele comparison with graphical representation of relationships between the 18 members of the family. Squares illustrate all the possible combinations of allele comparison between the chosen kindred. The squares annotated with red colour represent sibling-sibling relationship, with yellow parent-child relationship and with blue genetically unrelated relationship. Higher IBS mean indicate higher allele similarity with value of 2.0 to account for identical twins.

6.3.5 Non parametric linkage analysis

The variants from next generation sequencing were analysed and prepared for non-parametric linkage analysis with Merlin software. The linkage association was illustrated as a LOD score across the sequenced region measured in cM. Multipoint analysis has shown that linkage was not significant (LOD score >3.0) as LOD score was 2.39 but indicated that there is a major recombination right in the middle of the sequenced region which had to be thoroughly investigated (Figure 6.21a). In Single point analysis similarly to multipoint illustrated no variant with significant LOD score but there was a major cluster of variants present in the first half of the examined region that had to be investigated further (Figure 6.21b).

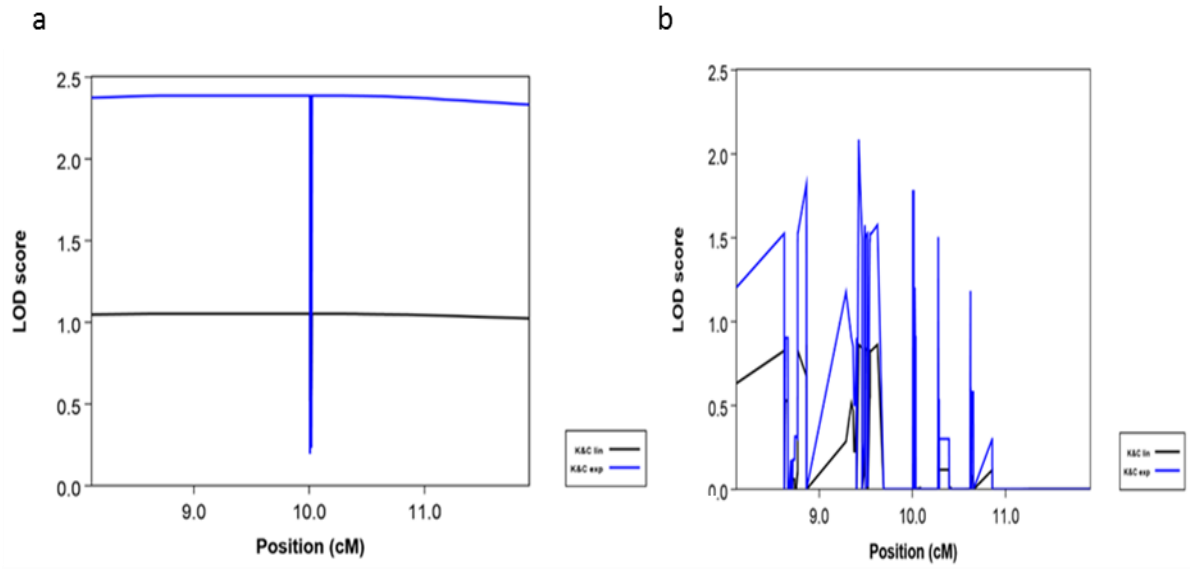
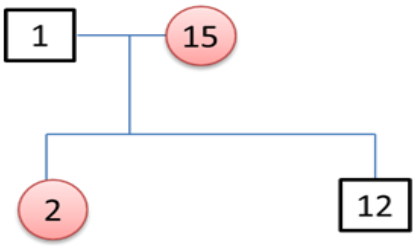


Figure 6.21 Non parametric linkage analysis summary. a) Multipoint analysis illustrating a linkage with a LOD score of 2.39 b) Singlepoint analysis has shown that none of the individual variants illustrated significantly LOD score > 3.0.

6.3.6 SNP Filtering

6.3.6.1 Manual Haplotyping

The results from Merlin have suggested a potential major recombination in the middle of the sequenced region suggesting that one of the two regions is associated with the disease. Therefore, the variants were manually haplotyped for the identification of minor allele or risk haplotype. The haplotype was processed by identifying the mode of inheritance from parent to the child and isolating the disease allele that has been inherited from one generation to the other (Figure 6.22). Manual haplotyping and direct comparison of SNPs between affected and unaffected-control patients has revealed the presence of SNPs which were unique for affected patients clustering in the first region prior to the recombination point as it was indicated by linkage analysis. The identification of these variants suggested that there is a haplotype segregating within the family (Figure 6.23) (Figure 6.24).



	Father			Mother			Son			Daughter	
	sample1	sample1		sample15	sample15		sample12	sample12		sample2	sample2
Position	a1	a2		a1	a2		a1	a2		a1	a2
8113405	T	T		T	T		T	T		T	T
8626859	G	T		G	G		G	T		G	T
8626903	C	C		C	T		C	C		C	T
8637727	A	A		G	A		G	A		A	A
8665751	C	C		C	C		C	C		C	C
8665793	T	T		T	T		T	T		T	T
8678446	G	A		G	G		G	A		G	A
8696907	T	T		T	T		T	T		T	T
8698565	T	C		T	T		T	C		T	C
8703145	G	G		G	G		G	G		G	G
8707900	A	A		A	A		A	A		A	A
8707927	C	C		C	C		C	C		C	C
8713842	A	A		A	A		A	A		A	A
8717713	G	G		G	G		G	G		G	G
8737734	A	A		G	A		G	A		A	A
8741027	C	C		C	T		C	T		C	C
8755243	C	C		C	C		C	C		C	C
8755243	C	C		C	C		C	C		C	C
8769180	T	T		T	T		T	T		T	T
8769180	T	T		T	T		T	T		T	T
8769423	C	C		T	C		T	C		C	C
8770108	C	C		C	C		C	C		C	C
8770130	G	G		G	G		G	G		G	G
8770282	G	G		G	A		G	A		G	G
8770822	T	T		C	T		C	T		T	T
8770932	C	C		A	C		A	C		C	C
8863438	G	G		G	G		G	G		G	G
8864177	C	C		T	C		T	C		C	C
8864203	G	G		A	G		A	G		G	G
8864449	ACTT	ACTT		ACTT	ACTT		ACTT	ACTT		ACTT	ACTT
8864592	GAA	GAA		GAA	GAA		GAA	GAA		GAA	GAA
8865006	T	T		C	T		C	T		T	T
8865567	T	T		G	T		G	T		T	T
9288522	G	A		A	A		A	A		A	A

Figure 6.22 Manual haplotyping identifying the disease allele. Manual haplotyping of the alleles from parents to children identifying the disease allele, annotated with **green**.

6.3.6.2 Identification of a haplotype segregating within the family

The Initial step of the identification of the risk haplotype was followed by the alignment of the disease related alleles from all the member of the family and with direct comparison between the affected and unaffected individuals. The comparison revealed chromosomal crossovers shared only among the affected members of the family (Figure 6.23).

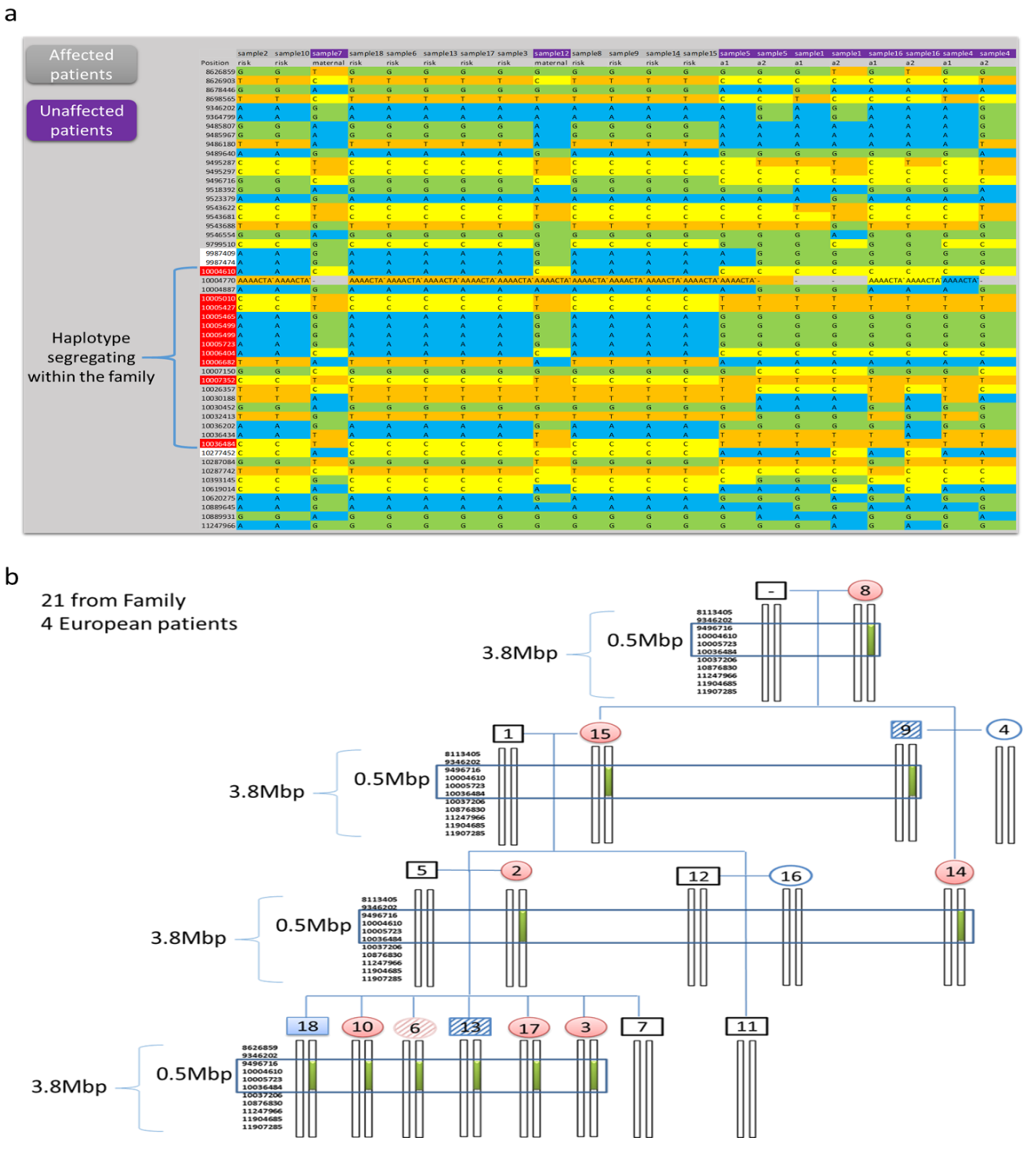


Figure 6.23 Manual Haplotyping. a) Alignment of the risk alleles from all members of the family has shown that there is a cluster of variants present only in affected individuals. Each base is highlighted with different colour and each column represents the risk allele of each patient. The cluster of variants identified and highlighted with red colour illustrated in affected

patients different risk alleles compared to healthy individuals. b) A cluster of variants illustrated in Figure 6.23a, was found to segregate within the family.

6.3.6.3 Filtering of Identified SNPs

The variants were then evaluated regarding their allele frequency in the general population. A variant is considered as a candidate when it is rare in the general population (<2%). The variants were assessed with the use of the University of California Santa Cruz (UCSC) Genome Browser. This filtering step has revealed that only 1 variant was rare with approximately 1.2% allele frequency (Table 6.8). The presence of the SNP within the family was also confirmed by Sanger sequencing as described in (Chapter 2).

Table 6.8. ANRD5 SNP was the only rare SNP present in the segregating variant cluster.

Chromosomal position	**Chr20:10036484**
Gene	ANKRD5
Gene position	3’ prime UTR
SNP change	T>C
Allele Frequency	T=98.8%, C=1.2%

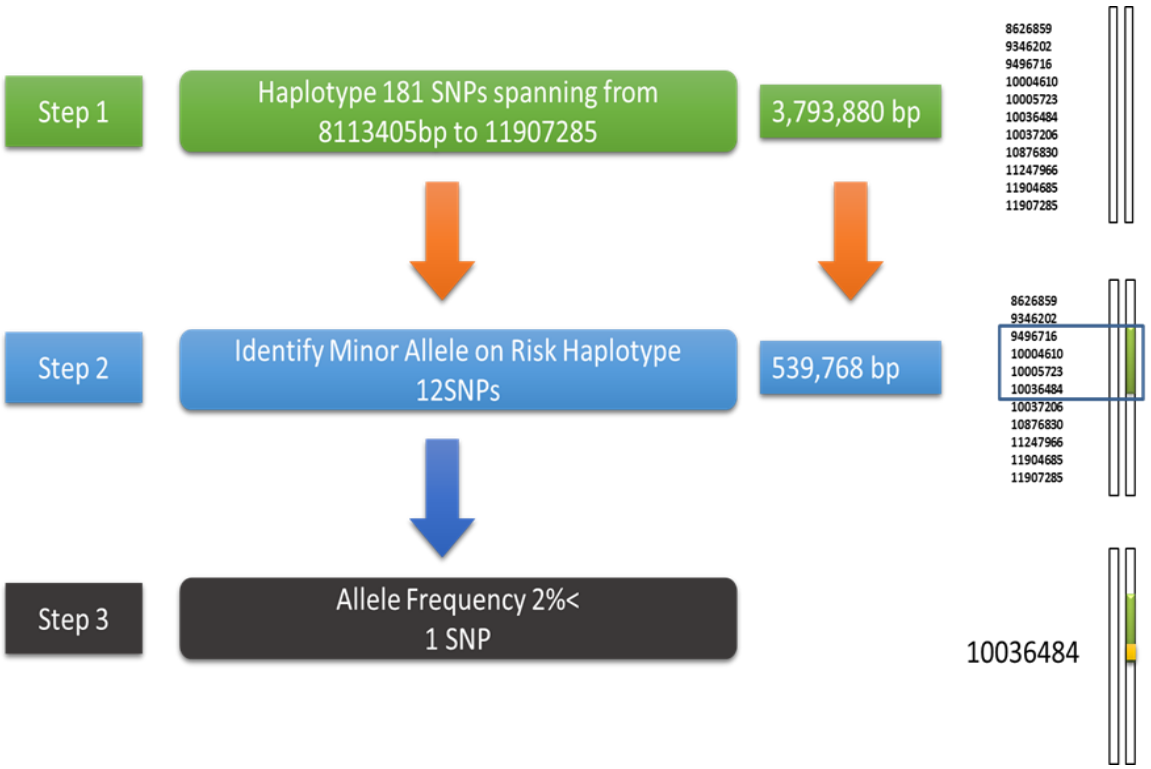


Figure 6.24 Summary of the filtering steps undertaken to identify the candidate ANKRD5 SNP.

6.3.6.4 Genotyping analysis of SHIP cohort for ANKRD5 SNP

The SHIP cohort of 986 patients was genotyped for the presence of ANKRD5 SNP. The cohort consisted of patients suffering from nodules and/or goitres and the analysis was kindly performed by Prof. Henry Völzke and Dr. Alexander Teumer. The genotyping screening has revealed that the SNP was not important for the disease as the prevalence in affected population (Goitres and Nodules) was lower than in the unaffected population (Table 6.9).

Table 6.9 Genotyping analysis of SHIP cohort for ANKRD5 SNP. a) Percentage of Nodules and Goitre patients with (affected) and without (unaffected the AKRD5 SNP. b) The total number of patients included in the genotyping screening from SHIP cohort

a

	Nodules	Goiters
Affected	2.54 %	1.91 %
Unaffected	2.85 %	2.79 %

b

	Missing Nodules	Nodules -	Nodules +	Sum
Missing Goiter	2			2
Goiter -	133	410	121	664
Goiter +	67	132	121	320
Sum	200	542	242	986

6.3.7 Next generation sequencing of 80 MNG patients

The screening of 25 patients including the 18 family members carrying the PLCβ1 InDel was followed by the screening of 80 MNG patients assessing the same region (Figure 6.25). The 80 patients were all suffering from MNG and previous genotyping studies by Dr. Bullock and Prof. Clifton-Bligh from the University of Sydney revealed that 4 out 80 carried the PLCβ1 InDel (also described in Chapter 4). The patients were unrelated to each other therefore, non-parametric linkage analysis could not be performed. Enriched libraries prior to the sequencing were assessed for polyclonality with Qubit fluorometer and measurements were taken at both Alexa Fluor 488 and 647nm. Levels of polyclonality were between 20 to 30 % which are acceptable levels for sequencing.

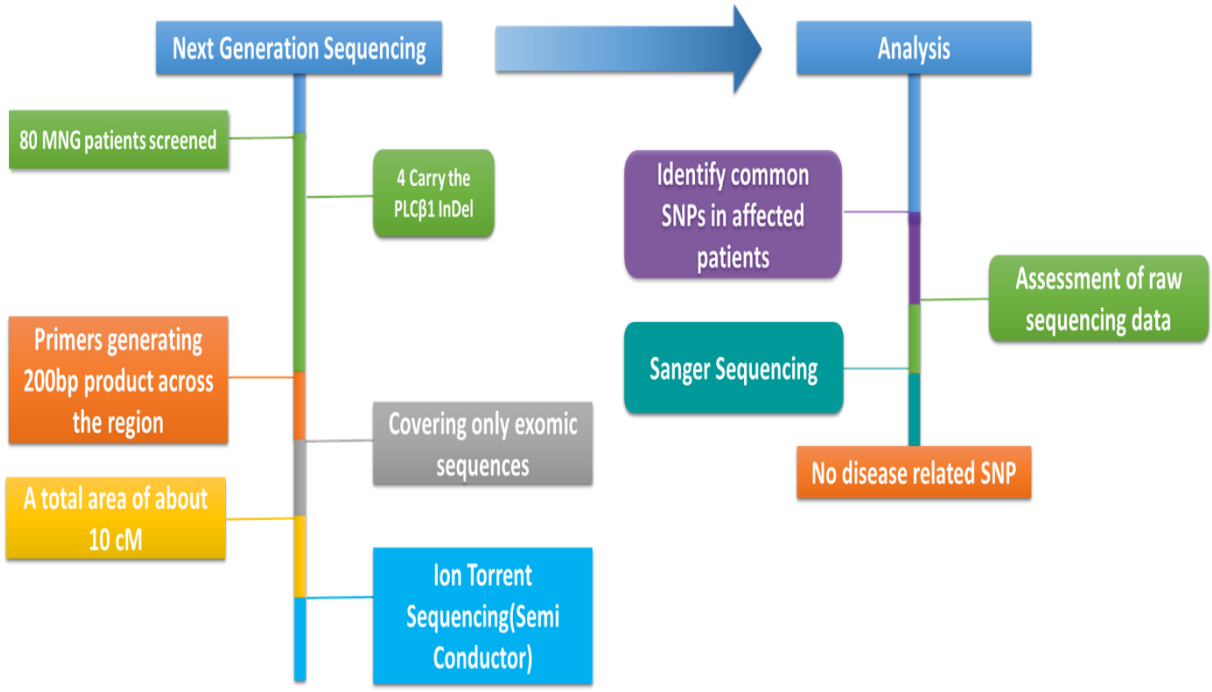
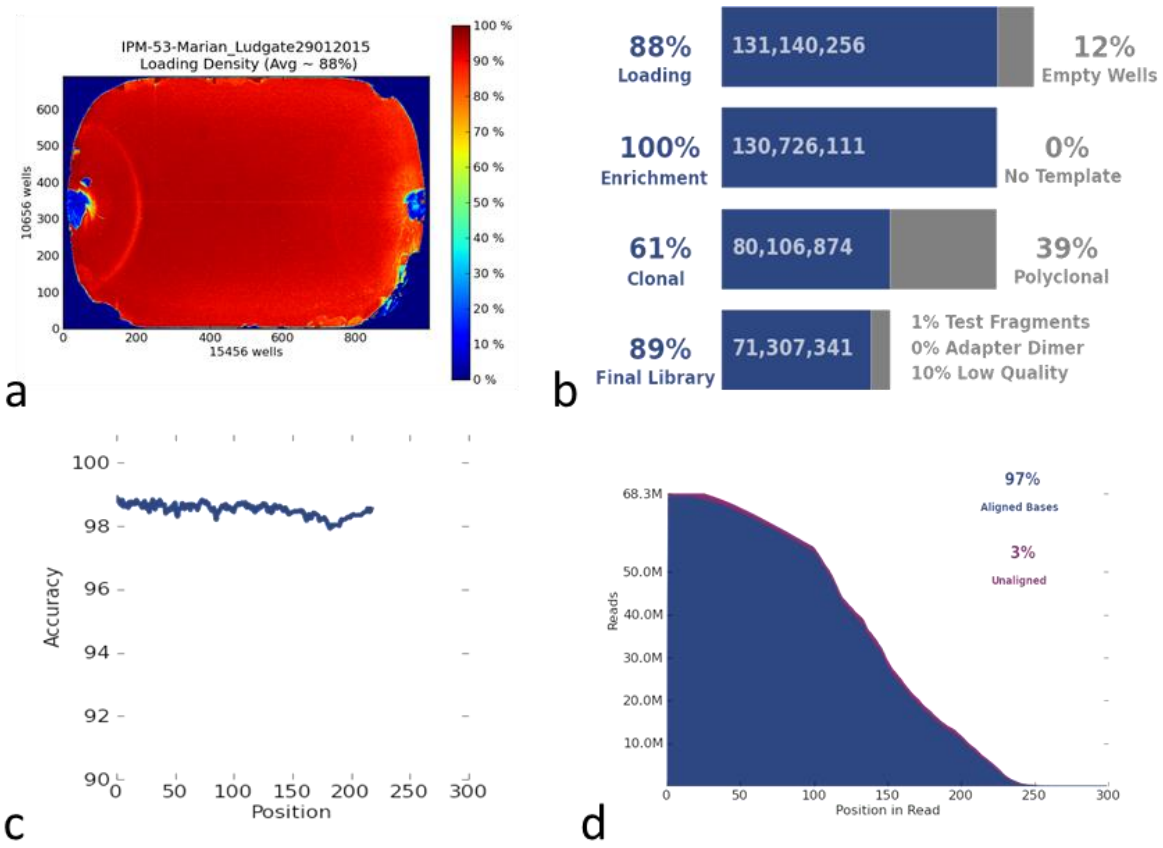


Figure 6.25 Summary of the next generation sequencing of the 80 MNG patients

6.3.8 Quality assessment of sequenced data

6.3.8.1 Sequencing with Ion Torrent Proton Sequencer

The Proton sequencer has generated a total of 9.9Gbp of data, achieving 98% accurately mapped sequences, and over 88% of the percentage of target bases covered by at least 0.2 times the average base read depth (Figure 6.26).



	AQ17	AQ20	Perfect
Total Number of Bases [Mbp]	8.1 G	6.9 G	5.2 G
Mean Length [bp]	133	120	95
Longest Alignment [bp]	299	286	275
Mean Coverage Depth	2.6	2.2	1.7

e

Figure 6.26 Quality assessment of the generated data. a) Chip loading with approximately 88% occupied wells , b) breakdown of the final number of libraries screened, c) Average raw (1x) accuracy across each individual base position in a read, d) Total number of bases aligned to the reference sequence, e) Alignment quality (AQ17, mega bases of DNA with one mismatch in the first 50 bases relative to the reference strain, AQ20, mega bases of DNA with one mismatch in the first 100 bases relative to the reference strain).

6.3.9 SNP assessment

The 80 MNG patients were unrelated to each and therefore, non-parametric linkage analysis could not be performed. The results had to be processed by direct comparison between the SNPs present in affected and absent in unaffected-control patients (Healthy patients from the affected family). Interestingly, 4 out 85 patients carried the PLCβ1 InDel with prevalence of 5% which is significantly higher compared to 1% in the general population. A list with the total number of SNPs was prepared similar to the pedigree file for non-parametric linkage analysis and comparison has been performed individually in each of the 308 different variants. The screening has been completed in two different levels, a first selection which were variants present in the 4 MNG patients carrying the PLCβ1 InDel and the same time absent in unaffected-control patients (Table 6.10). The second selection included variants absent in unaffected-control patients but not necessarily present in all 4 MNG patients carrying the PLCβ1 InDel. The first selection identified 1 deletion at antisense SNAP 25 gene and second selection identified 4 variants (Figure 6.27).

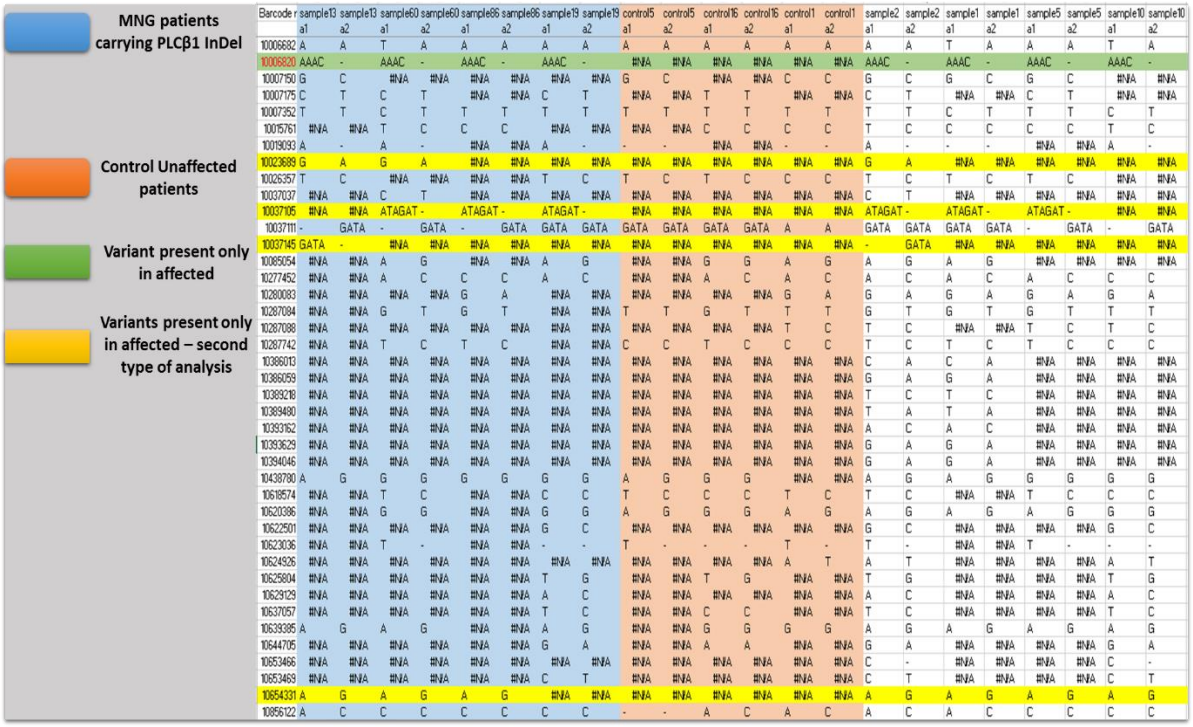


Figure 6.27 Variant comparison was completed in two levels of selection. The first selection which were variants present in the 4 MNG patients carrying the PLCβ1 InDel and the same time absent in unaffected-control patients identified one deletion highlighted with **green** and the second type of selection 4 variants highlighted with **yellow** (The second selection included variants absent in unaffected-control patients but not necessarily present in all 4 MNG patients carrying the PLCβ1 InDel)**.**

Table 6.10 Summary of the variant identified by the two types of screening

Region	Variant	Gene	Prevalence	Position
10006820	AAAC > -	Antisense SNAP-25	74/80	Intron
10023689	G > A	ANKRD5	42/80	Intron
10037105	ATAGATGATAAGATAG > -	ANKRD5	51/80	3'prime UTR
10037145	GATA > -	ANKRD5	22/80	3'prime UTR
10654331	A >G	Jag1	66/80	3'prime UTR

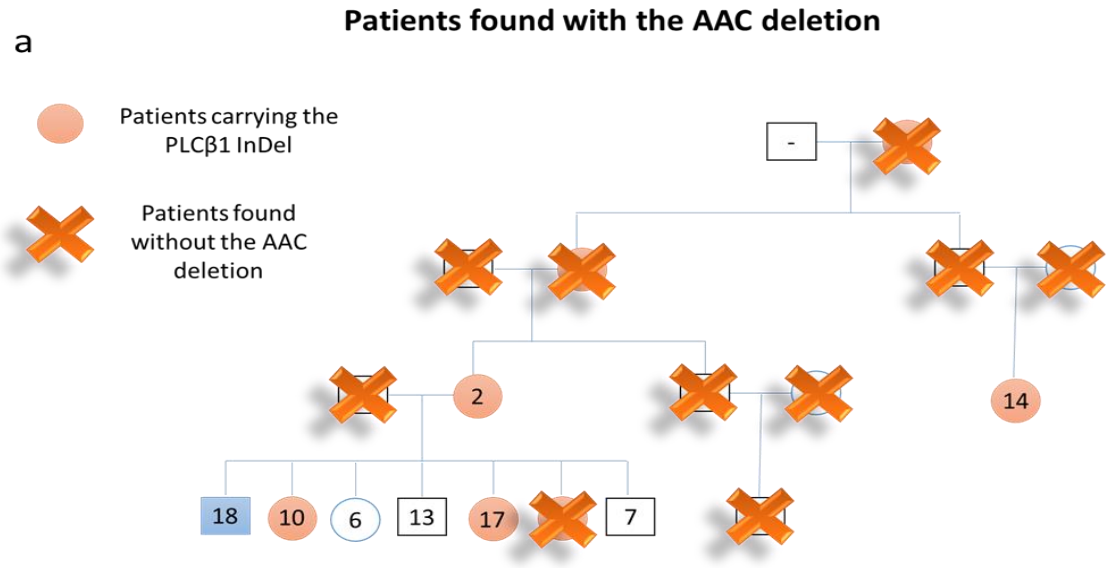
6.3.10 Filtering of identified variants

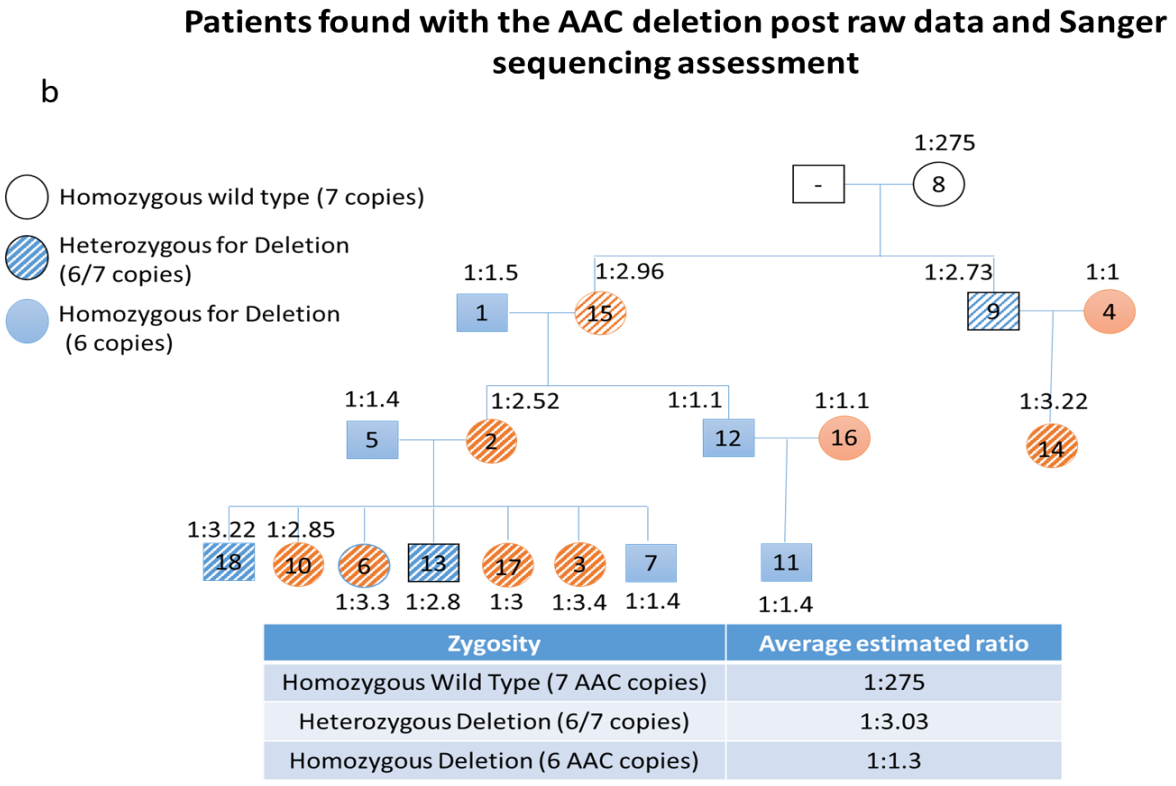
Alignment and direct comparison of the variants of the different patients has revealed 5 variants present in affected individuals and the same time absent from the unaffected patients. The comparison included two types of selection. The first included variants that are present in all four MNG patient with the PLCβ1 InDel and the same time absence in unaffected-control patients. On the other hand, the second type included variants that are mostly present but not necessarily in all of the four MNG patients with the PLCβ1 InDel and the rest of the MNG patients and the same time absent in unaffected-control patients.

Screening with the first selection found one deletion chr20:10006820-10006824 (AAAC) present in 74 out of 80 MNG patients. Screening for this deletion in the originally affected family was found in patients but with incompatible mode of inheritance indicating *de novo* mutagenesis. Thus further investigation was conducted for the identification of potential false negative patients for the deletion within the family. The results confirmed that the deletion was present in patients that appear to be false negative (indicated as absent) including healthy individuals and thus it was not important for the disease (Figure 6.28). Further investigation for identification of potential false negatives within the investigate family (Chapter 4) was performed by assessment of the raw data of 86 different variants indicated as negative-absent. The analysis has not found any false negative variant with potential important role in the disease.

Assessment of the raw data for the four variants found by the second selection has revealed that the ATAGATGATAGATAG deletion was an artefact of sequencing misalignment, the 100023689 and 10037145 SNPs were found to be present also in unaffected patients and therefore were not important variants. Variant 10654331 was found in 66 out of 80 MNG patients however it was absent in family members. The 66 patients were sequenced using Ion

Torrent Proton sequencer with the 14 MNG patients and the family members sequenced with the Personal Genome Machine. Therefore, investigation for potential presence of the SNP within the family was investigated by Sanger sequencing of the family members and of patients with the SNP from the Proton sequencing. The results have shown that none of the patients was carrying the SNP, indicating that it was an artefact of Proton sequencer (Table 6.11).





Zygosity	Average estimated ratio
Homozygous Wild Type (7 AAC copies)	1:275
Heterozygous Deletion (6/7 copies)	1:3.03
Homozygous Deletion (6 AAC copies)	1:1.3

Figure 6.28 Summary of the false positive patients carrying the deletion found with the first type of selection (described above) for screening. a) Patients appeared to carry the

AAC deletion in the family indicating the presence of false negatives. b) Raw data analysis and Sanger sequencing revealed the presence of false negative for the deletion in the family. The blue colour in the family tree represent males and red colour females.

Table 6.11 Summary of the results following raw data analysis and Sanger sequencing.

Region	**Variant**	**Gene**	**Prevalence**	**Position**	**Post-raw data and Sanger Sequencing assessment**
10006820	AAAC > -	Antisense SNAP-25	74/80	Intron	Not Important (present in Controls)
10023689	G > A	ANKRD5	42/80	Intron	Not Important (present in Controls)
10037105	ATAGATGATAGATAG > -	ANKRD5	51/80	3'prime UTR	Artefact
10037145	GATA > -	ANKRD5	22/80	3'prime UTR	Not important (present in Controls)
10654331	A > G	Jag1	66/80	3'prime UTR	Artefact

6.4 Discussion

The PLCβ1 InDel has been identified in a family where many members were suffering from MNG and many of them progressed to papillary thyroid cancer (Chapter 4). Also, the PLCβ1 InDel was identified to have increased prevalence in MNG patients with 5% compared to the general population with 1%. Also, genome wide linkage and haplotype analysis of the family by Dr. Bakhsh has identified a 10 cM region with high LOD score >3. That region contained 10 genes (PLCβ1, PLCβ4, LAMP5, PAK7, ANKEF1, SNAP25, MKK5, SLX4IP, JAG1, BTBD3), therefore a thorough investigation of that area for the identification of potential SNPs that may have role in the disease had to be conducted.

The exome of family members from this 10cM region was sequenced by Personal Genome Machine and semi conducting sequencing technology and primers were designed generating approximately 200bp fragments. The quality of sequencing data was satisfactory identifying more than 180 different variants. Non parametric linkage analysis was performed to investigate potential recombinations within that region that may indicate the presence of disease related variants. The analysis has found a recombination at the middle of the region indicating that one of the parts was disease related. Interestingly, even though the target region was a high LOD score region in Genome Wide Linkage Analysis as it was mentioned above, the LOD score following non parametric Linkage analysis was not significant. That was possibly due to the reduction of the size of the investigated region (10cM) compared to the initial analysis by Dr. Bakhsh where non-parametric analysis was performed in all chromosomes and the entire chromosome 20 where the high LOD score (20cM) was found. The reduction of the investigated region as only exomic sequences of the high LOD score region were included, had subsequently reduced the number of variants in the study thus the LOD score could not show significant linkage. Therefore, manual haplotyping had to be performed to identify the risk haplotype. Manual haplotyping of all the different variants from parent to offspring allowed the identification of the risk allele which is inherited from one generation to the other. Isolation of the risk alleles for each variant and direct comparison of the affected with unaffected- control patients, revealed that there is a haplotype of 12 variants segregating within the family. Filtering steps reduced the region of interest as only approximately 500kb from about 3.8Mb were found to be associated with the disease and segregated within the family. Filtering of these variants regarding its allele frequency found that only one variant was rare <2%. However, genotyping of 986 patients from the SHIP study containing nodules and/or goitres revealed that this SNP was unrelated with the disease as the prevalence was lower in affected compared to unaffected-control population.

PLCβ1 InDel was found in increased prevalence in MNG population. Therefore 80 MNG patients including the 4 found with the PLCβ1 InDel were screened for the same region as the family for potential presence of risk variants. The patients were sequenced with the use of proton sequencer and PGM with semi conducting sequencing technology. The quality of the sequencing data was satisfactory identifying more than 300 different variants across the 80 patients. A list with the total number of different variants present in all 80 patients was generated. The 80 patients were unrelated to each other therefore non parametric linkage analysis could not be performed. The variant assessment has to been done manually by direct comparison for their presence in affected and unaffected patients. The screening has been performed in two types of screening. The first included the presence of the variants in all 4 patients carrying the PLCβ1 InDel and the same time absent in unaffected-control patients. The second screening included the absence of the variant in unaffected patients but not necessary present in all 4 InDel carriers. The analysis identified 5 variants which had be investigated individually. Assessment included investigation of the raw data for potential identification of poor read depth misalignments and false negatives and Sanger sequencing for confirmation of presence or absence of the variant of interest.

Raw data assessment revealed that 2 variants were sequencing artefacts and therefore not real, also in 1 variant unaffected patients of the family were found as false negatives and therefore, it was not important for the disease. The other two variant were absent in most of the family members and only found in unaffected siblings thus were considered as not important

The presence of false negatives and the use of two different sequencing platforms, PGM and Proton indicated that a thorough screening of the variants illustrated as negative (absent) in the output files had to be performed for the identification of a potential important false negative. Therefore the raw data from 86 different variants were screened identifying 3 not important false negatives, confirming that there is no other important variant within that high LOD score region apart from the PLCβ1 InDel. The data following the sequencing of more than 100 patients revealed that there was no other disease related variant present with the high LOD score region suggesting that PLCβ1 InDel might be the solely responsible genetic cause for the disease pathogenesis (Figure 6.29).

The two platforms that were used in this project are using the power of semi-conducting technology for sequencing. Even though they provide robust and high-throughput data, this technology can identify only SNPs and very small deletions. Also, it is susceptible false positive calls in homopolymer repeat regions. The investigated area included a number of non-coding regions where great number of microsatellite repeats was identified and further

investigation by Sanger sequencing had to be performed as the variant call for these repeats was often not very clear. Also, the two platforms PGM, and proton had different sequencing potential as Proton uses a larger chip with more than 165 million wells and therefore is more sensitive in variant calling generating data with more than 1000 time read depth. Thus, comparison between the data acquired from the two different platforms had to be performed identifying that PGM illustrated higher false negative rate in variant calls than Proton sequencer.

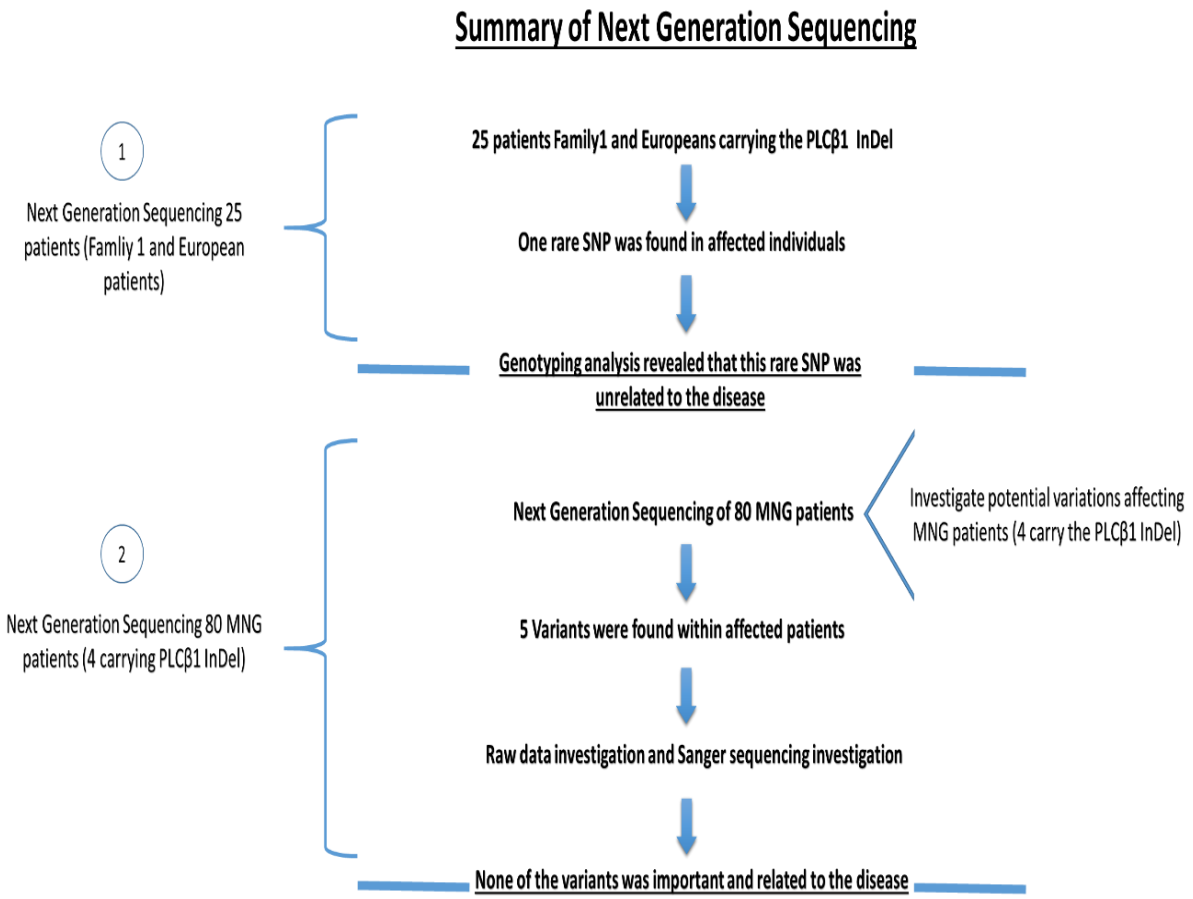


Figure 6.29. A summary of the next generation sequencing experiments.

Chapter 7. General Discussion

7.1 The role of phospholipase C enzymes in health and disease

The aim of my thesis was to investigate the role of PLCs in health and disease and more specifically the use of PLCζ as a potential therapeutic tool for male infertility and the role of a PLCβ1 InDel in patients with MNG likely to develop PTC.

The PLCβ1 project aims included:

1. Investigation of the prevalence of PLCβ1 InDel in different groups of patients
2. Investigate whether a potential ERα promoter binding site within the PLCβ1 InDel is an active promoter or contains silencing elements.
3. Evaluation of the role of PLCβ1 in cell cycle progression and cell proliferation.
4. Investigate whether the PLCβ1 InDel is solely responsible for the disease.

The PLCζ project aims included:

1. Expression and enrichment of recombinant human PLCζ
2. Activity assessment *in vivo* of enriched recombinant human PLCζ

In Chapter 4, a genotyping tool was designed for screening large patient cohorts for the identification of the PLCβ1 InDel. The presence of an active ERα promoter or silencing elements within the PLCβ1 InDel were also investigated by luciferase reporter gene assay.

The prevalence of the PLCβ1 InDel within the population varied as described in Chapter 4. Patients suffering with MNG illustrated an increased prevalence of 5% compared to 1% in the general population. On the other hand none of the PTC patients screened were found to carry the InDel. The prevalence of the PLCβ1 InDel has suggested that the InDel is not causing PTC but benign proliferation. The family that was found to carry the InDel included many patients with MNG developing PTC, suggesting that the increased proliferation in MNG patients carrying the InDel make them more susceptible to thyroid cancer. Thus, a qPCR based genotyping tool was designed to gain a better insight of the prevalence of the InDel. The technique provided a cost effective and fast method for screening simultaneously up to 40 patients at a time. The qPCR output provided a Ct value based comparison with a clear distinction between patients with the InDel and healthy individuals with a difference of 10 cycles as described in (Figure 4.11b in Chapter 4). The primers that were found to be ideal for the genotyping screening were the ones flanking the InDel as Ct difference between affected and healthy individuals was approximately 10 Ct values, in contrast primers within the deletion the Ct value difference was close to 1 making the distinction extremely difficult.

The 208 subclinical hypothyroid patients genotyped, illustrated a PLCβ1 InDel prevalence of approximately 1% similar to the prevalence in the general population. This cohort was expected to be similar with the general population (euthyroid) as their peripheral thyroid hormone levels were normal with only their thyroid stimulating hormone being slightly elevated. Their thyroid function is normal with an increased likelihood of becoming hypothyroid. The original family affected by the PLCβ1 InDel was euthyroid with an increased prevalence of MNG developing to PTC. Also, the 5% prevalence of the PLCβ1 InDel within MNG patients suggested that patients with MNG would be the ideal cohort to test. Furthermore, none of the 70 PTC patients tested was found to carry the InDel suggesting that it is more likely to be present in PTC patients initially suffering from MNG, rather than PTC presenting as a single nodule. The patients from the study of health in Pomerania (SHIP) were good candidate for genotyping as they had increased prevalence of both thyroid goitres and nodules. However, it was not possible to genotype these patients for the InDel as the PLCβ1 region does not include any SNPs represented in existing genotyping data from the cohort. Thus we would have needed to obtain genomic DNA for qPCR genotyping of the InDel and such samples were not available to us.

In silico analysis had identified that within the PLCβ1 InDel there is oestrogen receptor (ERα) α binding site. This lead to the hypothesis that the InDel might be an alternative promoter for PLCβ1 or for another transcript within the long intron. Patients carrying the InDel had also illustrated increased thyroid PLCβ1 transcript levels that could indicate even the presence of potential silencing elements within the InDel. It could even exert effects at considerable distance, as in the case of the FTO gene (Smemo *et al*. 2014). To investigate the potential promoter activity of the InDel I used a luciferase reporter gene assay in MCF-7 cells which are an oestrogen receptor positive cell line. The cells were treated with oestrogen and luciferase response would indicate presence of an active promoter. The assay has confirmed that there is no active promoter within the InDel as there was no significant luciferase light emission. However, even though empty PGL3 basic vector contained a luciferase gene and emitted light in luciferase assay as expected, there was a reduction of luciferase emission in cells which contained PLCβ1 InDel-PGL3 basic vector which could indicate the presence of silencing elements within the InDel. The presence of luciferase in these plasmids was confirmed by PCR therefore, luciferase emission of similar levels as the positive control (empty PGL3 basic vector) was expected. The successful transfection and validation of a functional assay was achieved with the use of positive control plasmids which were co-transfected into cells along with the plasmid of interest. Renilla plasmid and α-Luciferase plasmids emitted Renilla luciferase and Firefly luciferase respectively confirming the successful cell transfection.

In Chapter 5 the role of the PLCβ1 gene in cell proliferation was investigated by knockdown and overexpression of PLCβ1 by SiRNA and lentiviral transduction respectively. SiRNA knockdown was shown to delay cell proliferation up to 20% in comparison with scrambled SiRNA (Control) treated cells. Studies by (Faenza *et al.* 2000) have shown that overexpression of PLCβ1 in mouse erythroleukemia (MEL) cells led to cell cycle progression by affecting the MAPK/ERK pathway inducing cell cycle progression from G1 to S phase. The siRNA knockdown on rat thyroid FRTL-5 cells has supported these findings indicating a clear role of PLCβ1 in cell proliferation. The FRTL-5 cells provide a suitable cell model for thyroid cell studies as they share common characteristics with human thyroid epithelial cells. However, the doubling time of these cells is very high requiring approximately 3 days for cell division which is a drawback for short lasting experiments such SiRNA treatment as SiRNA clears from the cells relatively fast. A human thyroid cancer cell line such as thyroid papillary carcinoma (TPC-1) would be the optimal choice for such study however none was available at the time of the project. Also, a complete investigation of the effect of PLCβ1 on cell cycle would require a flow cytometry assay to confirm the stage of the cell cycle the cells are held in. It is hypothesised that PLCβ1 knockdown cells would be held at G1 phase and that there would be an increased percentage of apoptotic cells as studies by (Lee *et al.* 2000) have revealed that overexpression of PLCβ1 suppresses apoptotic genes such as c-fos.

Overexpression, of PLCβ1 was attempted with lentiviral expression vectors for PLCβ1 provided by a collaborator Dr. Roberta Fiume from the University of Bologna. Supernatant for GFP positive control was also provided by Dr. Darley from Cardiff University. Even though active GFP lentiviral vector from the (Cardiff) samples was successfully generated. It was not possible to generate an active lentiviral vector for GFP or PLCβ1 using the components provided by Dr. Fiume, even after several attempts. Direct transfection into HEK293T cells and FRTL-5 cells was attempted with no positive result. The patients from the family affected by the PLCβ1 InDel were shown to have increased PLCβ1 transcript levels in their thyroid tissue. Therefore, it was particularly important to mimic the disease condition by overexpression of PLCβ1.

7.1.1 The cell signalling of PLCβ1

A further insight regarding the molecular pathway of the PLCβ1 reveals its vital role in cell proliferation and carcinogenesis. PLCβ1b is activated following the binding of a stimulus (e.g. a hormone or light) to its cognate GPCR which releases Gαq making it available to bind to PLCβ1, as previously described. However it has also been found to be activated in response

to different stimuli such as, interleukin-2, insulin and insulin-like growth factor-1 (Cocco *et al.* 2002) which function via tyrosine kinase receptors. Interleukin-2, activates nuclear PLCβ1 in human natural killer cells (NK) following treatment for 60 minutes. The activation is thought to occur via the ERK/MAPK pathway as IL-2 can activate PI3 kinase via IL-2Rβ chain which in turn activates PKC isoforms (Vitale *et al.* 2001). The second activation stimulus is insulin and studies on NIH 3T3 cells treated with insulin, reported enhanced cell proliferation. Following insulin treatment, increased DAG quantity was measured and PLC activity assays supported the indication that insulin activates nuclear PLCβ1 (Martelli *et al.* 2000b). A number of studies had initially identified the nuclear presence of PLCβ1, however it was not clarified until the correlation between IGF-1 treatment and nuclear activity of PLCβ1 was found (Martelli *et al.* 1992). A later study (Manzoli *et al.* 1997) had revealed the mitogenic role of PLCβ1, where overexpression of PLCβ1 followed by IGF-I stimulation resulted in significant enhancement of DNA synthesis. However, the missing part of the puzzle was how nuclear PLCβ1 is activated as Gαq and Gβγ do not localize to the nucleus. It was found that upon stimulation with IGF-1, MAP kinase translocates to the nucleus and phosphorylates PLCβ1 on serine 982 present within a MAP kinase consensus sequence pro- ser- ser- pro at its characteristic C terminal region. This phosphorylation site is unique for PLCβ1 as it is not present in any other PLCβ isoforms (Martelli *et al.* 1999; Martelli *et al.* 2001). Mutagenesis of this site by replacing serine with alanine or treating the cells with a specific MEK inhibitor PD98059 had the same result in failing to activate nuclear PLCβ1. These results were a strong indication that nuclear activation is regulated by the MAP kinase translocation in the nucleus(Xu *et al.* 2001a). In addition to this finding another piece of the nuclear PLCB1 signalling puzzle was put in place when a negative feedback was identified between PKCa and nuclear PLCβ1. Stimulation with IGF-1 and translocation of MAP- kinase leads to activation of PLCβ1b. The increased concentration of DAG by PIP2 hydrolysis within the nucleus attracts the PKCa from the cytoplasm which in turn translocates to the nucleus and phosphorylates PLCβ1b at serine 887. The phosphorylation of PLCβ1b by PKCa terminates PLCβ1b activity. This hypothesis has been supported by the overexpression of mutated PLCβ1b on the PKCa phosphorylated site with alanine which resulted in sustained activation of PLCβ1b. Suggesting that PLCβ1a is activated by GPCR route and PLCβ1b by Tyrosine kinase receptors (Taniguchi *et al.* 1995; Martelli *et al.* 2000a; Xu *et al.* 2001b)(Xu *et al*. 2001).

7.1.2 Overexpression of PLCβ1 promotes cell cycle progression

The correlation of PLCβ1 expression and the cell cycle regulation through a PKC mediated pathway has also been suggested by a very recent study where a decrease of PKCa following

overexpression of PLCβ1 was observed in K562 cells. Further analysis of this finding was carried out by silencing PKCa which resulted in upregulation of cyclin D3, similarly with PLCβ1 effect, indicating that PLCβ1 leads to cell cycle progression by targeting to cyclin D3 via the PKCa pathway (Poli *et al.* 2013). A greater insight regarding the role of nuclear PLCβ1 has been given by (Faenza *et al*. 2000). PLCβ1b has been found to have a key role in cell cycle regulation, as its overexpression in mouse erythroleukemia (MEL) cells, lead to overexpression of cyclin D3 and cdk4. The complex of D3/cdk4 is known to phosphorylate the Retinoblastoma protein (pRb). The phosphorylation of this tumour suppressor resulted in the release of proteins from the E2F family which in turn induced the expression of genes that are required for cell cycle progression from G1 to S phase (Faenza *et al*. 2000). The overexpression of nuclear PLCβ1 was found to be correlated with the enzymatic activity which was 3 fold higher than in wild type cells. It is noteworthy that increased PLC activity has been reported in thyroid neoplasms (Kobayashi *et al.* 1993). Unpublished work by Dr. Lei Zhang and Dr. Fiona Grennan Jones by our group has indicated significantly higher transcript levels of isoform a in all thyroids tested ($p<0.01$) but no significant difference in the ratio of PLCβ1-a to PLCβ1-b in PLCβ1 InDel-affected and non-affected thyroids. The increased enzymatic activity which could follow-on from the higher transcript levels is considered to be vital as IP_3 and DAG are generated, with the latter attracting PKCa into the nucleus and triggering a nuclear signalling cascade which leads to cell cycle progression (Faenza *et al*. 2000). A few years later, a study on serum-stimulated HL-60 cells found that there are two waves of PLCβ1b activity throughout the cell cycle. The first occurs at 30 minutes in serum stimulated HL 60 cells during the G1 phase and the second at 11.5 hours. Both PLCβ1 activity peaks were associated with serine phosphorylation and they were inhibited following the use of PI-PLC and Erk kinase (MEK) inhibitors, such as ET-18-OCH_3 and PD98059 respectively. Interestingly it has been suggested that the two peaks are equally important for the transition of G1 to S phase. Inhibition of only the second peak resulted in the same mitotic index profile as the cells which were initially treated with inhibitors, with both cells failing to enter S phase underlining the importance of both PLC activity peaks for cell cycle progression from G1 to S phase (Lukinovic-Skudar *et al.* 2007). Initial evidence regarding the role of nuclear PLCβ1 in transition of G2/M phase has been provided by (Lukinovic-Skudar *et al.* 2005). This suggestion was supported by the finding that PLCβ1b is associated with laminB1. The study has revealed that both co-localise to the nuclear speckles and that the loss of laminB1 causes structural changes to the speckles, suggesting a potential role in RNA splicing as they are rich in pre messenger RNA splicing factors. The study has revealed that once JNK is activated by an extracellular signal can activate PLCβ1b which in turn leads to generation of DAG. The increased concentration of DAG chemo–attracts PKCa from the cytoplasm to the nucleus resulting in phosphorylation of laminB1. The depolymerisation of laminB1 caused by its

phosphorylation induces the cell cycle progression from G2 to M phase and is considered to have a key role in the breakdown of the nuclear envelope (Fiume *et al.* 2009). Lamins and more specifically the B type are present in most somatic cells and are considered to be vital for cell survival as they regulate apoptosis, gene expression, and nuclear architecture and are often involved in cancer and tumour progression. Alteration of the expression of the B type lamins is essential for cell survival as laminB proteins are always present in cancer cells (Foster *et al.* 2010). Another hypothesis suggests that overexpression of PLCβ1 may prevent cell death caused by oxidative stress by suppressing specific apoptotic genes such as c-fos. Overexpression of PLCβ1 caused 5 fold increase of PLCβ1 transcript levels in NIH3T3 cells, followed by approximately 2 fold increase in the PLCβ1 enzymatic activity (Lee *et al*. 2000). The cellular survival due to overexpression of PLCβ1 is suggested to be followed by cell cycle progression to S phase. PLCβ1 was speculated to inhibit via Translin associated factor (TRAX) the downregulation of lactate dehydrogenase (LDH) and Glyceraldehyde 3-phosphate dehydrogenase (GAPD) which are known to be involved in histone synthesis as they are part of the octamer-binding factor 1 coactivator (OCA-S) which has a vital role in histone expression by binding to H2B promoter in S phase at the time of histone synthesis(Philip *et al.* 2012). Therefore, the overexpression of PLCβ1 is suggested to prevent cell death by oxidative stress and by regulating histone synthesis, to promote cell cycle progression (Lee *et al*. 2000; Philip *et al*. 2012). The key role of PLCβ1 on the cell cycle was supported by another study where copy number alteration of PLCβ1 gene was investigated in a number of breast cancers. The study identified that the genetic alterations of PLCβ1 had an effect on the mitotic index of the cells (Molinari *et al.* 2012). A summary of the studies described above is illustrated in (Figure 7.1).

7.1.3 Downregulation of PLCβ1

On the other hand, downregulation of the PLCβ1b in patients with myelodysplastic syndrome (MDS) has been associated with high risk of developing acute myeloid leukaemia (AML). A mono-allelic deletion on PLCβ1 gene has been shown to have more severe clinical outcome in high risk MDS patients compared with patients having both alleles. In addition, the high risk MDS patients carrying the deletion were found to have decreased expression of PLCβ1b, indicating that a potential decrease of the nuclear PLCβ1 could affect the cell cycle in MDS pathogenesis (Follo *et al.* 2009). Interestingly, PLCβ1 has been found to affect the Akt signalling pathway, in MDS patients. A recent study where MDS patients were treated with azacitidine revealed that untreated patients had elevated activated Akt in contrast with the treated which had elevated levels of PLCβ1 expression and dowregulated activated Akt. The

study revealed direct correlation between the PLCβ1 and activated Akt levels, as the PIP_2 levels fluctuated according to the changes of PLCβ1 and activated Akt having an inverse and direct relationship respectively (Follo *et al.* 2008).

The PLCβ1 transcript levels were also found to be altered in another recent study by (Guerrero-Preston *et al.* 2014) where promoter methylation differences of PLCβ1 were found in head and neck squamous cell carcinoma (HNSCC) leading to dowregulation of the gene. This important finding suggested that somatic mutations along with promoter methylation may be an important factor for the downregulation of tumour suppressor genes in HNSCC.

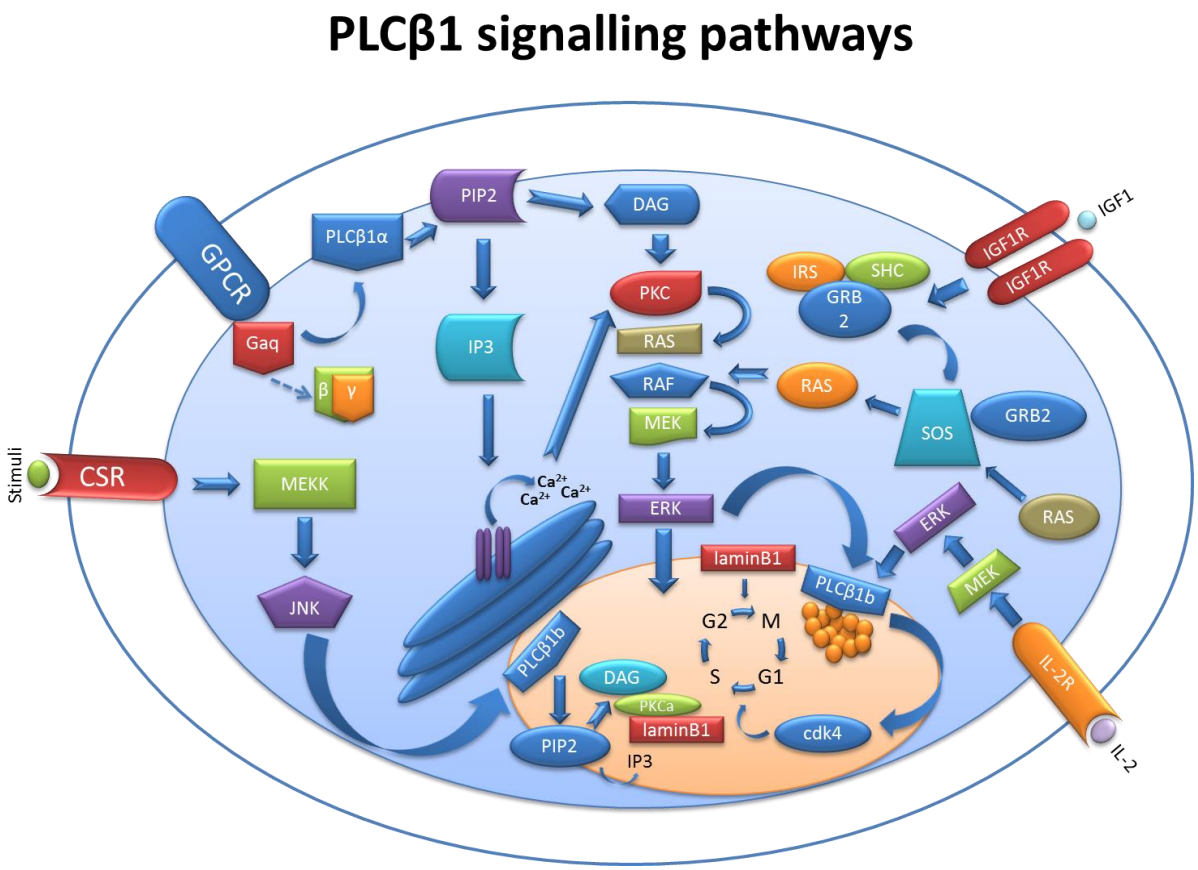


Figure 7.1 Cell signalling of PLCβ1. Overexpression of PLCβ1 is found to be correlated to cell cycle progression and involved in a number of different cancers. PLCβ1b is activated via phosphorylation by MAPK (ERK) and promotes cell cycle progression from G1 to S and G2 to M phase due to up-regulation of cdk4 and laminB1 respectively.

7.1.4 The Correlation of PLCβ1 with the main somatic and germline mutations in human cancer

The studies mentioned above have indicated a number of genes which seem to be directly affected by the overexpression of PLCβ1 transcript levels and activity. Databases such as the Catalogue of Somatic Mutations In Cancer (COSMIC) and the International Cancer Genome Consortium (ICGC) were used to gain a better understanding regarding the importance of PLCβ1 and its correlation with the main somatic and germline mutations occurring in the four main cancers in which PLCβ1 is involved, (Figure 7.2). The main PLCβ1 signalling cascade, contains several downstream components implicated in a range of cancers, e.g. BRAF and RB1. RB1 is phosphorylated from cyclin D3 following PLCβ1 overexpression, which disables its ability to bind to E2F, allowing the latter to be released and promote cell cycle progression from G1 to S phase (Faenza *et al*. 2000). In addition, BRAF gene is one of the most common genes to be altered in PTC which in turn is most frequent type of thyroid cancer. The B-type Raf kinase (BRAF V600E) mutation accounts for 90% of the BRAF mutations (Frasca *et al.* 2008). Patients carrying the BRAF mutation have illustrated approximately 5 times higher mortality rate in comparison with the mutation negative patients, thus it is used as biomarker for PTC prognosis (Haq *et al.* 2013; Xing 2013).

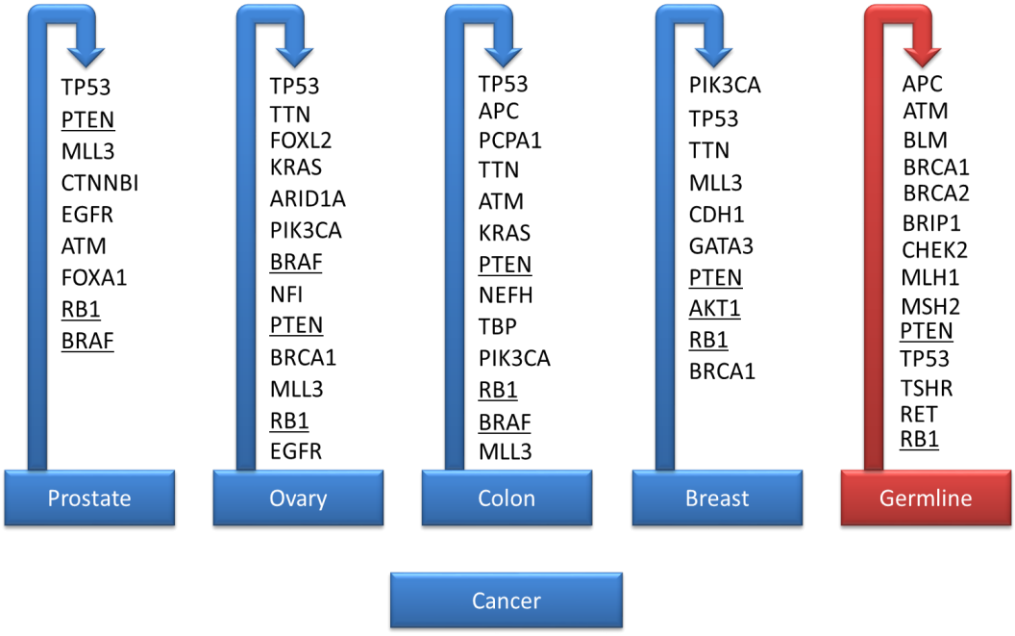


Figure 7.2 Somatic germline mutations in PLCβ1 found in patients with Prostate, Ovary, Colorectal and Breast cancer. The data were acquired from Catalogue of Somatic Mutations In Cancer (COSMIC) and International cancer genome consortium (ICGC). The genes correlating directly and indirectly with the PLCβ1 pathway are underlined. The genes underlined are present within PLCβ1 cell signalling cascades.

7.1.6 PLCβ1 InDel and goitre formation hypothesis

The family patients found with the InDel (as described in Chapter 4) have illustrated increased levels of PLCβ1 transcripts. Furthermore, Akt in a recent study was found to be associated with PLCβ1 in patients with myelodysplastic syndrome (MDS) with high risk of developing acute myeloid leukemia (AML) (Follo *et al*. 2008). The study has shown that PLCβ1 levels were correlated with the levels of activated Akt, possibly due to the PIP_2 levels as the latter is the common substrate for both PLCβ1 and Akt. Phosphoinositide 3-kinase (PI_3K) is known to generate Phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) via dephosphorylation of PIP_2. Then PIP_3 activates Akt following its binding at Akt-PH domain. Activated Akt then phosphorylates tuberous sclerosis complex 2 (TSC2) which prevents the TSC1/TSC2 complex to inhibit Ras homolog enriched in brain (RHEB) activity by GTP hydrolysis. Active RHEB-GTP in turn activates mTOR by a mechanism not well understood. It is hypothesized that the lack of PIP_2 substrate due to the increased PLCβ1 transcript levels in the patients carrying the PLCβ1 InDel may downregulate Akt activity and therefore leading to dowregulation of mTOR activity (Hay and Sonenberg 2004). A study by (de Souza *et al.* 2010) has illustrated that inhibition of mTOR in TSH-stimulated PCCL3 thyroid cell line has increased iodide uptake suggesting that mTOR has a key role in regulating thyroid function. A study from (Zimmermann *et al.* 2005) which has investigated the consequence of a chronic high iodine intake in children from Hokkaido Island in Japan has shown that excess Iodine intake is associated with goitre formation and increased thyroid volume. Also in a study for chronic iodine excess in American workers in West Africa (Pearce *et al.* 2002) illustrated increased goitre prevalence suggesting the goitrogenic effect of excess iodine intake. As described in Chapter 1, the iodide imported by NIS is oxidised in the follicular lumen by TPO via the activity of H_2O_2. Studies by (Corvilain *et al.* 1994; Moreno-Reyes *et al.* 2007) have supported the hypothesis that H_2O_2 is generated predominantly by PIP_2-DAG-Ca^{2+} cascade. Therefore, the excess iodine uptake by the thyroid which may lead to goitre formation and the increased PLCβ1 transcript levels found in patients carrying the PLCβ1 InDel leading to overactivation of PI-PLC-DAG-Ca^{2+} could increase H_2O_2 generation. As described above overexpression of PLCβ1 can lead to cell proliferation and cell cycle progression. Thus we suggest that the InDel can be goitrogenic due to potential increased iodide uptake, can cause benign proliferation as overexpression of PLCβ1 leads to cell cycle progression and the subsequent overactivation of the cascade can lead to excess H_2O_2 generation which in turn can be mutagenic and cause oxidative DNA damage and cancer (Song *et al.* 2007). However, the MNG patients in the original family affected by the PLCβ1 InDel were euthyroid. Thus further investigation will be required to identify the role of the PLCβ1 InDel in thyroid hormone synthesis and whether it

affects iodine uptake with the resulting increased H_2O_2 generation possibly leading to cancer (Figure 7.3).

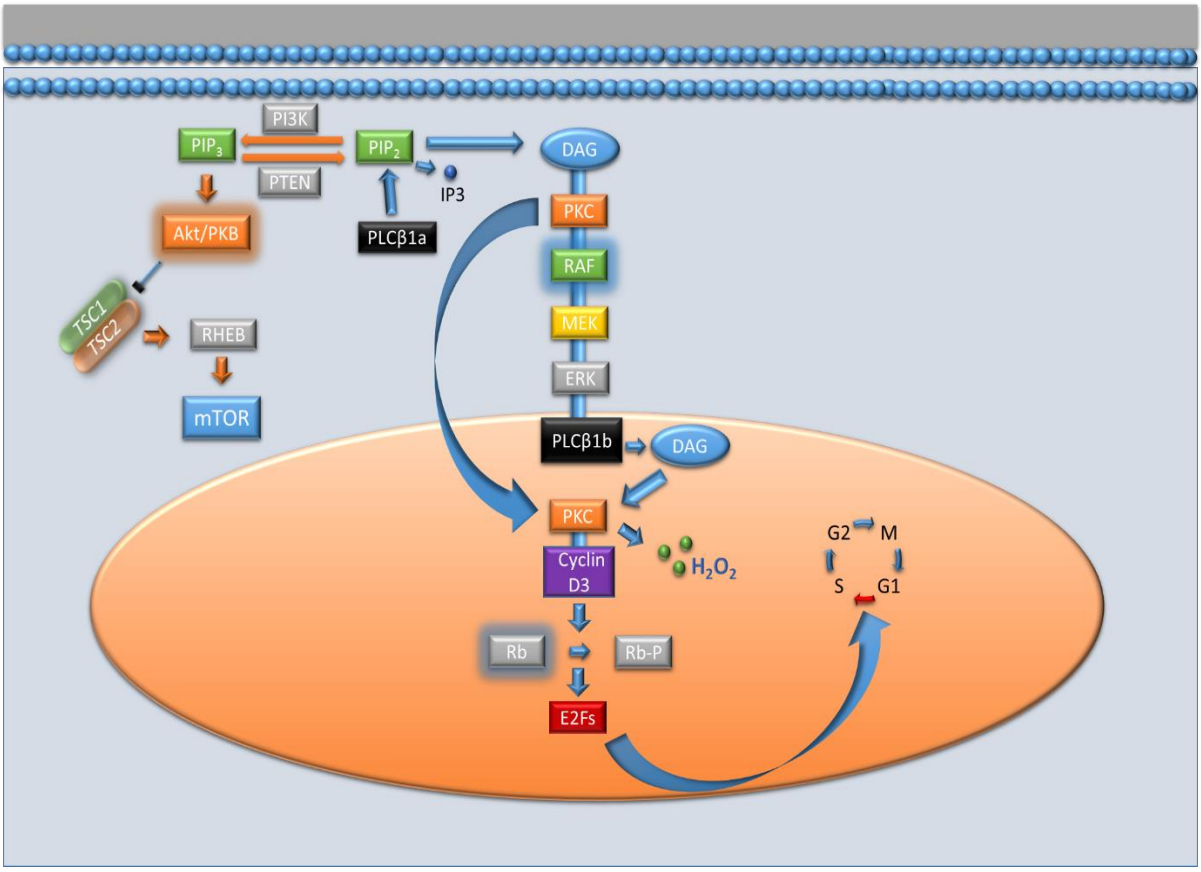


Figure 7.3. The associated cell signalling cascades in the disease pathogenesis. Overexpression of PLCβ1 downregulates mTOR cascade causing in turn increased iodide uptake which is goitrogenic. The overactivation of the PIP_2-DAG-Ca^{2+} has also been suggested to increase the generation of H_2O_2.

In Chapter 6, the high LOD score region on chromosome 20 identified by Dr. Bakhsh has been investigated by next generation sequencing for the potential presence of other variants that may be responsible for the disease pathogenesis. Semi conducting sequencing technology by Ion Torrent was chosen for the screening of the exomic regions of that Chr.20 20cM. The target patients were the member of the family found to carry the InDel (18), the 4 patients found to carry the InDel in chapter 4 by the genotyping tool and 80 MNG patients 4 of whom were known to carry the InDel. The screening has not identified additional variants that could be disease related confirming that PLCβ1 InDel is solely responsible for the disease. The Ion Torrent sequencing technology is high throughput generating a vast amount of data. Although

it can only detect SNPs and small deletion therefore large deletions or InDels could be missed, although previous studies using analysis of copy number variation identified only the InDel on Chr 20.

PLCζ and its potential use in biomedical applications

In Chapter 2, different protein constructs were designed to allow optimal protein expression and solubility of recombinant human PLCζ. The protein was expressed in a bacterial expression system therefore, the expression of a mammalian protein such as human PLCζ was very challenging. Thus a range of different protein solubility partners were used to enhance protein solubility.

Dr. Nomikos had successfully expressed recombinant hPLCζ with NusA solubility tag (NusA-hPLCζ) (Nomikos *et al.* 2013). In this study, hPLCζ was subcloned into a plasmid of the petMM series and generated NusA- hPLCζ which in turn exhibited a high yield in protein expression and solubility. Also the protein without the presence of a solubility tag illustrates rapid degradation. Thus NusA and other solubility tags were used to retain protein solubility. The major drawback of this strategy is that the protein is destined to be injected into human eggs to trigger egg activation in cases of severe male infertility. The bacterial origin of such solubility tags would be an impediment for future biomedical applications of hPLCζ in human eggs. Thus the human thioredoxin (hTrx), a protein that occurs naturally in human eggs, was employed as a potential solubility tag. The bacterial Trx (nTrx) was found to retain hPLCζ solubility thus hTrx was subcloned as second solubility tag at the N terminal of hPLCζ in between hPLCζ and the bacterial solubility tag.

In Chapter 3, the protein constructs with the highest yield of soluble recombinant hPLCζ were selected and assessed for their *in vivo* PLCζ activity following recombinant hPLCζ microinjection into mouse eggs. The expressed protein was enriched with affinity chromatography using the His_6 protein tag which binds with high affinity to Ni^{2+} beads. Activity assay illustrated that different active forms of hPLCζ were expressed which could potentially be used in biomedical applications. However, the level of purity of the enriched protein was low therefore, multiple protein purification steps should be performed.

PLCζ, in common with many proteins produced in the testis, is temperature sensitive. Bacterial cells were lysed using sonication, which unfortunately produces heat, even when performed on ice. An alternative would be the use of a French press to break open the bacteria but it can be used only for larger volumes of bacterial lysates >300ml.

As mentioned above, hPLCζ has to be partnered with a solubility tag of human origin and the same time has to be purified free from bacterial contaminants in order to be ethically approved for biomedical applications in human oocytes. Thus hTrx was subcloned at the N-Terminus and C -Terminus of hPLCζ. Bacterial solubility tags were removed by proteolytic cleavage and hTrx-hPLCζ was shown to be soluble and active *in vivo* a vital step for the potential use of hPLCζ in biomedical applications.

Final Conclusion

In conclusion, the data from chapters 2 and 3 have illustrated that active recombinant hPLCζ was successfully expressed and enriched. However, even though the protein was found to be active *in vivo* a number of bacterial protein contaminants were present in the enriched protein. Thus further optimization of protein production and additional purification steps are required to generate a product suitable for clinical use.

1. Assessment of the protein solubility levels was performed by direct comparison of each protein construct following western blot. The culture conditions and volumes were equal and experiments were conducted at the same time. This technique is not quantitative but can provide clear visual indication of the proportion of protein solubility. A potential quantitative analysis that could be performed is by enzyme-linked immunosorbent assay (ELISA).
2. Protein contaminant in affinity chromatography elution could be eliminated by introducing a double His_6 tag in the protein construct to maximize the potential binding of the protein to the Ni^{2+} beads and the same time including more washing steps with more stringent washing buffer to remove remaining contaminants. Also, affinity chromatography could be combined with either ion exchange chromatography or gel filtration to maximize purification levels of hPLCζ.
3. hPLCζ as described in the general introduction, has a positively charge X-Y linker which binds strongly to the negatively charged PIP_2. It is hypothesized that these proteins are attached to hPLCζ via this positively charged X-Y linker as western blot assay with anti-MBP and anti-HIS_6 antibodies were unable to detect the protein suggesting that is a bacterial origin protein and not a degraded part of recombinant hPLCζ. Therefore, the potential use of a buffer with increased salt concentration would be likely to prevent the protein interactions and a combination of two chromatography methods as mentioned above could achieve higher levels of protein purity.

Several conclusions can be drawn from the data presented in chapters 4, 5 & 6. The frequency of the PLCβ1 InDel is about 1% in healthy individuals or patients with SH but 5% in those with MNG. It could contain silencing elements that could influence PLCβ1 transcript levels. Reducing PLCβ1 transcripts (the opposite of the situation in the family) using siRNA also decreased proliferation, supporting the idea that it contributes to benign thyroid proliferation as occurs in MNG. Finally NGS analysis of the chr 20 high LOD score region did not identify other variants which are associated with disease. Thus the PLCβ1 InDel could provide a biomarker for MNG and identify patients most likely to develop PTC.

Future investigations could include:

1. The genotyping of patient cohorts suffering from MNG developing PTC as it is the population more likely to be affected by the PLCβ1 InDel.
2. Developing assays that can investigate the effect of the overexpression of hPLCβ1 on the cell cycle, mimicking the physiological effect of the PLCβ1 InDel in PLCβ1 transcript levels.
3. Investigate the correlation of PLCβ1 overexpression with BRAF down regulation and goitre formation.

Appendices

(Appendix A)

(Appendix A1)

pETMM vector series

hPLCζ was cloned into the pETMM vector series which contain a range of different protein solubility tags.

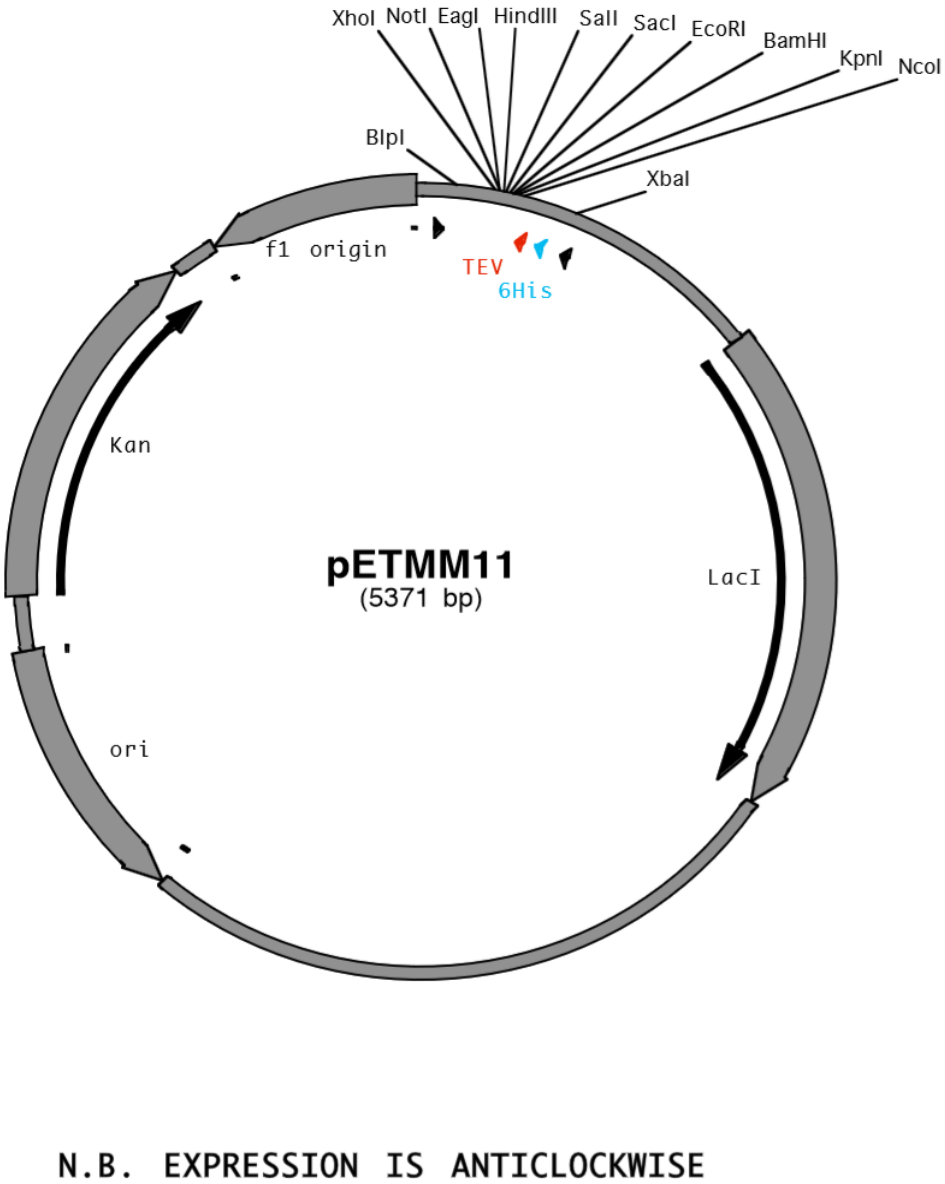


Figure App 8.1 pETMM 11. Contains a His_6 region for protein purification with Ni^{2+} resin

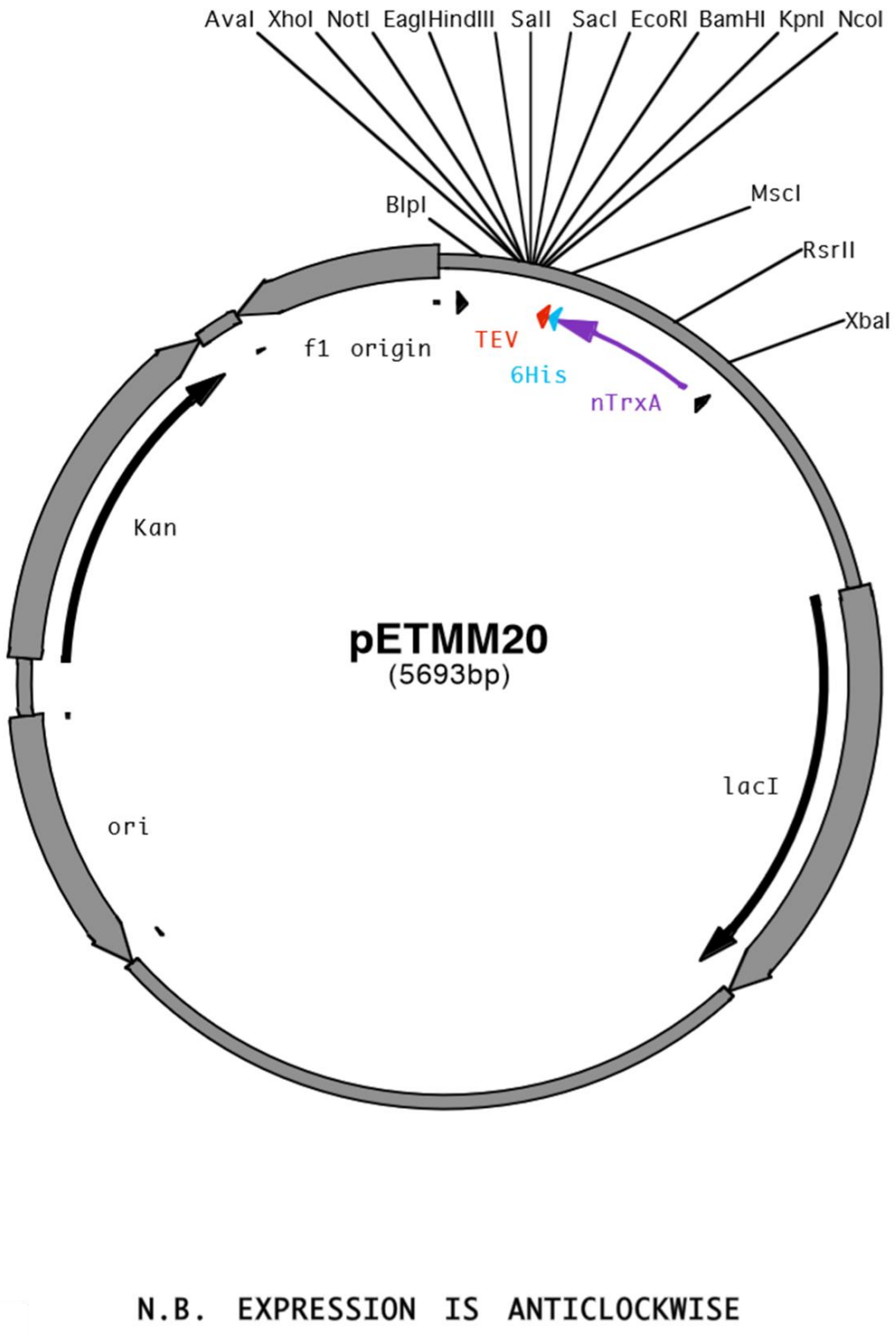


Figure App 8.2 pETMM 20. Contains a His_6 and an nTrx region allowing protein purification and enhancing protein solubility respectively.

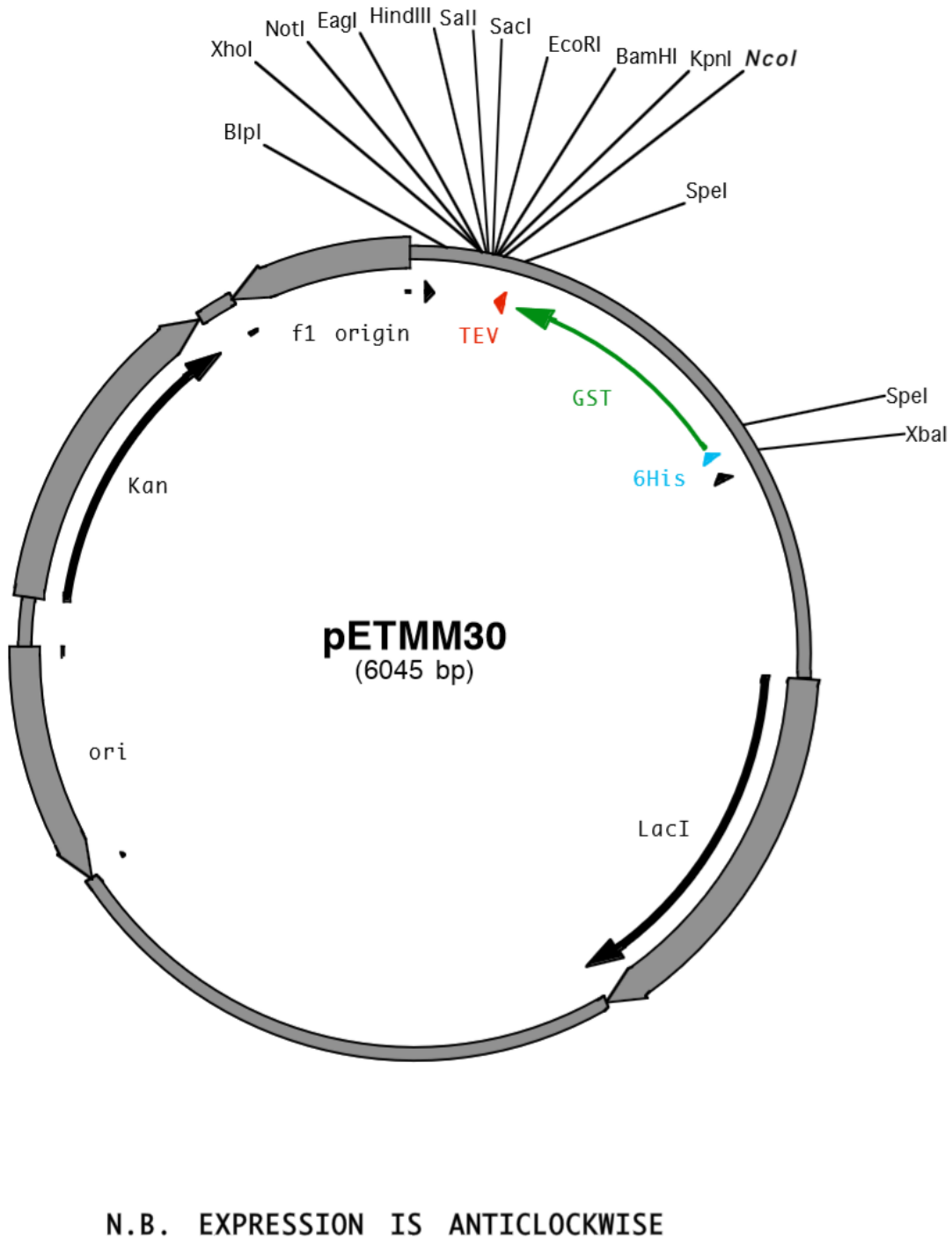


Figure App 8.3 pETMM 30. Contains a His_6 and a GST region allowing protein purification and enhancing protein solubility respectively.

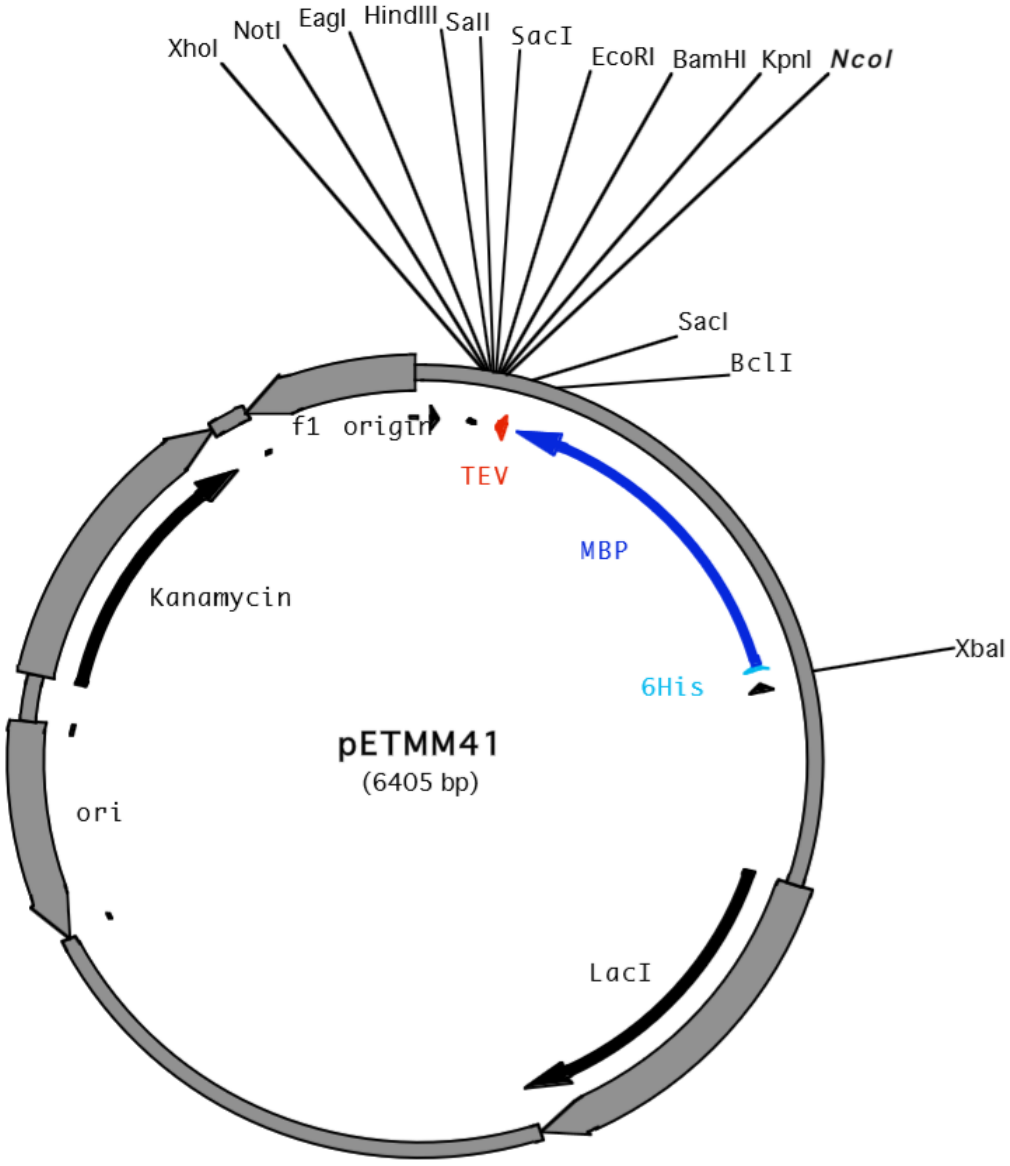


Figure App 8.4 Petmm 41. Contains a His_6 and a MBP region allowing protein purification and enhancing protein solubility respectively.

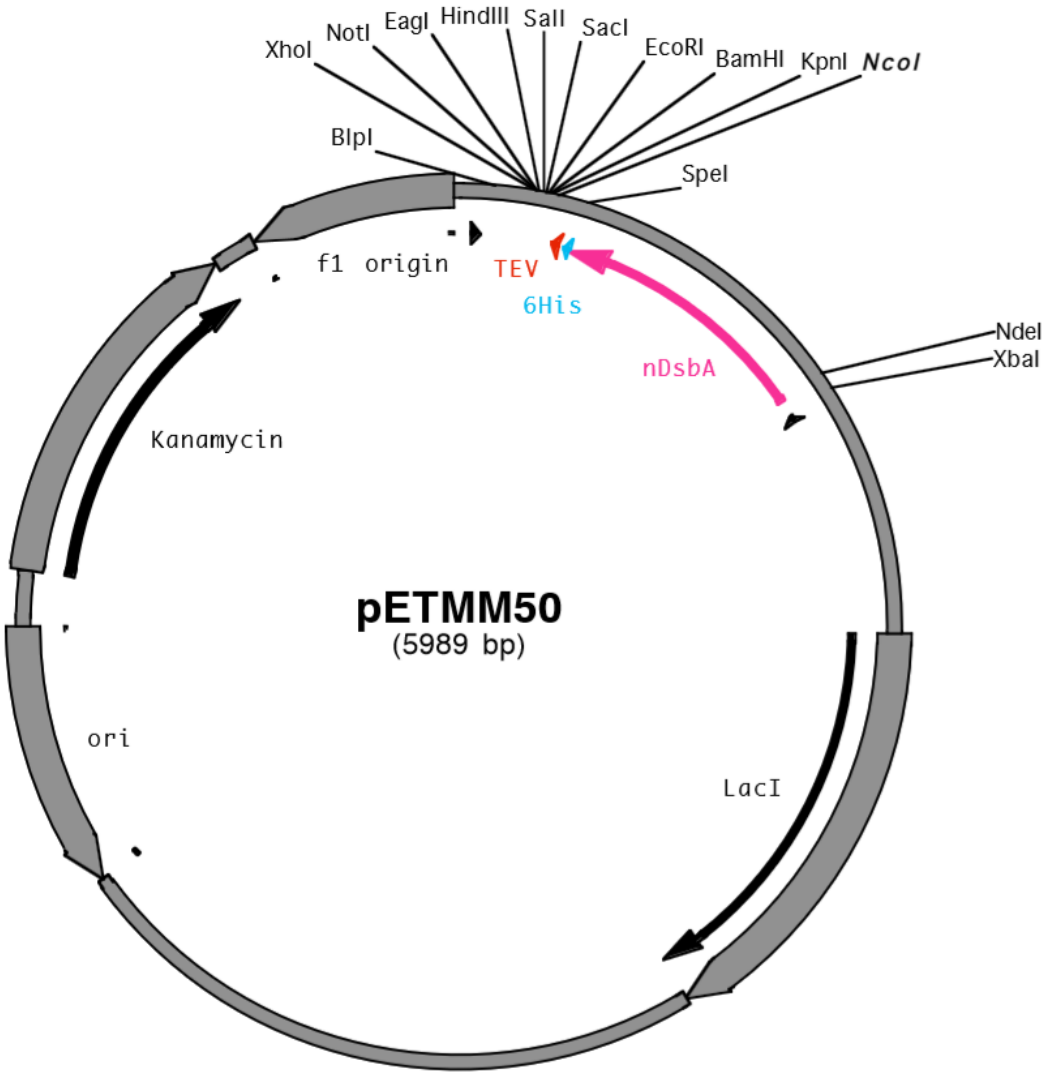


Figure App 8.5 pETMM 50. Includes nDsbA as a solubility tag and a His_6 for purification with Ni^{2+} resin.

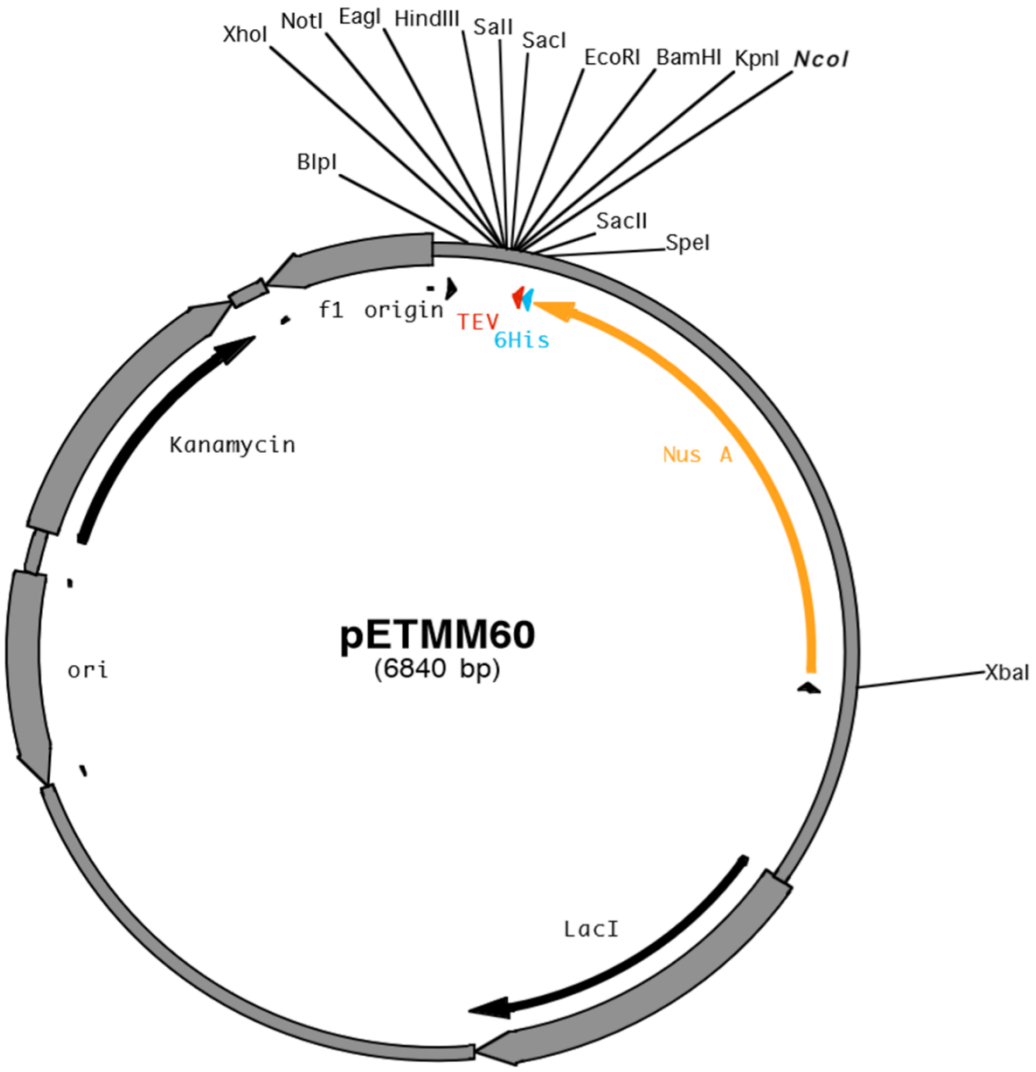


Figure App 8.6 pETMM 60. Includes NusA as a solubility tag and a His_6 for purification with Ni^{2+} resin.

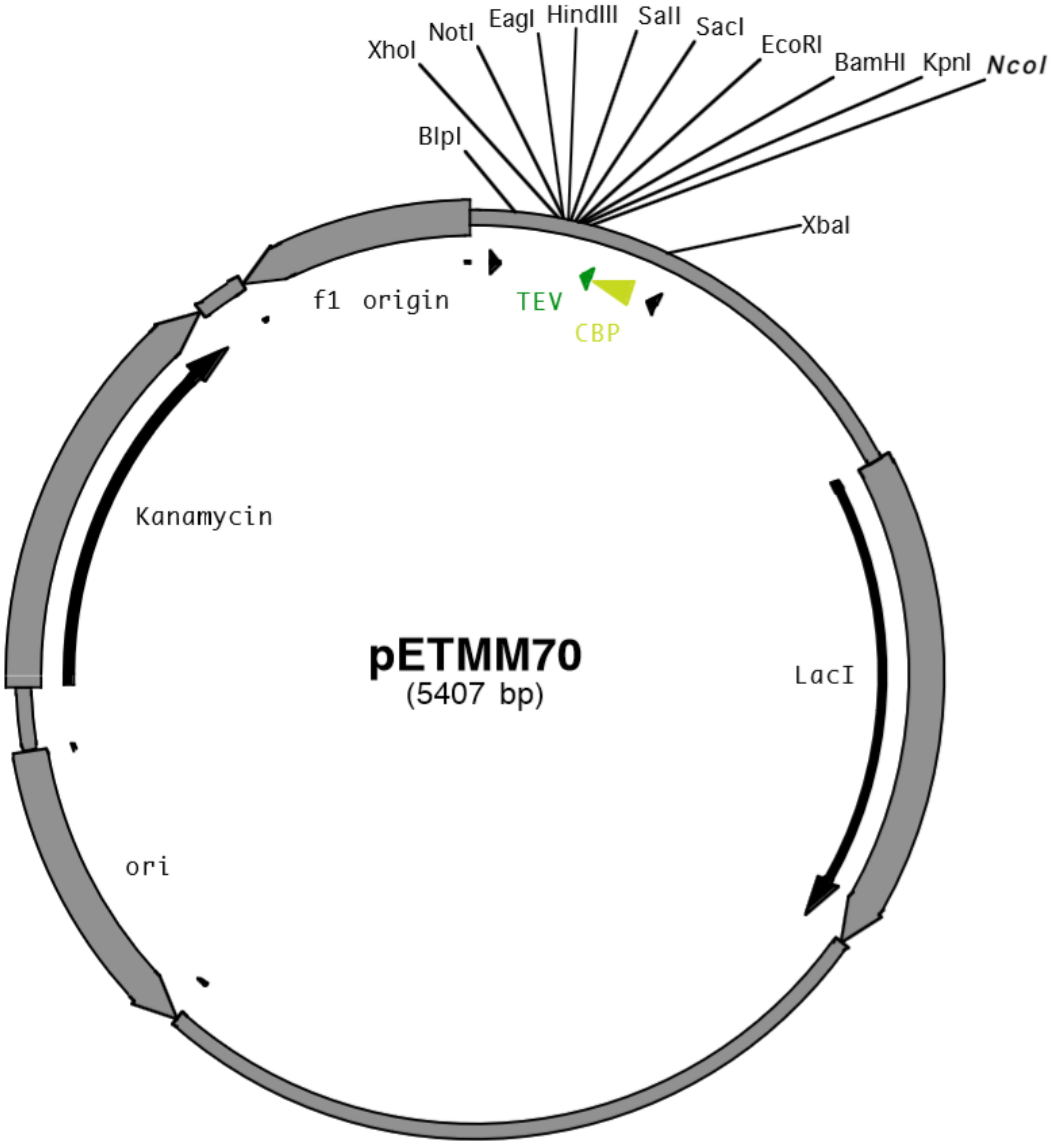


Figure App 8.7 pETMM 70. Includes only CBP as a solubility tag which allows protein purification with Calmodulin resin.

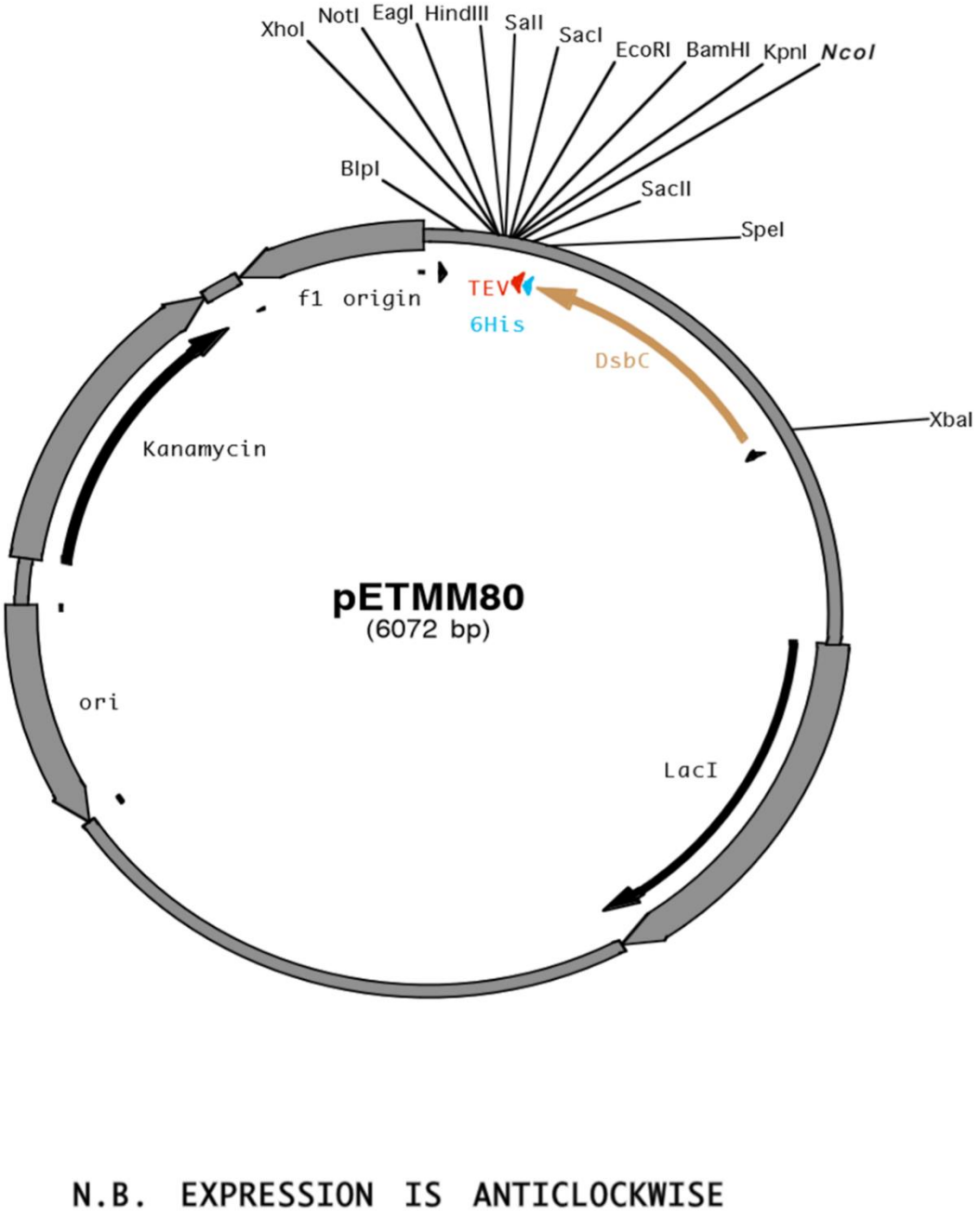


Figure App. 8.8 pETMM 80. Includes DsbC as a solubility tag and a His_6 for purification with Ni^{2+} resin

(Appendix B)

(Appendix B1)

Ampliseq primer pools

Primers were designed with the use of Ampliseq 3.0.1 software (https://ampliseq.com/) according to the manufacturer's protocol. A total number of 429 primers were designed generating 100-300bp amplification products.

Table App. 8.1 List of Ampliseq primers

AmpliSeq_Version=3.0.1 Workflow=DNA" db=hg19

chr20	8113337	8113476	GENE_ID=PLCB1
chr20	8130861	8130992	GENE_ID=PLCB1
chr20	8130992	8131094	GENE_ID=PLCB1
chr20	8351850	8351982	GENE_ID=PLCB1
chr20	8351982	8352196	GENE_ID=PLCB1
chr20	8608886	8609043	GENE_ID=PLCB1
chr20	8609043	8609185	GENE_ID=PLCB1
chr20	8626686	8626907	GENE_ID=PLCB1
chr20	8628444	8628673	GENE_ID=PLCB1
chr20	8629954	8630163	GENE_ID=PLCB1
chr20	8637681	8637808	GENE_ID=PLCB1
chr20	8637808	8637988	GENE_ID=PLCB1
chr20	8639180	8639401	GENE_ID=PLCB1
chr20	8665485	8665583	GENE_ID=PLCB1
chr20	8665583	8665795	GENE_ID=PLCB1
chr20	8678187	8678395	GENE_ID=PLCB1
chr20	8678393	8678494	GENE_ID=PLCB1
chr20	8689158	8689308	GENE_ID=PLCB1
chr20	8689308	8689479	GENE_ID=PLCB1
chr20	8696853	8697048	GENE_ID=PLCB1
chr20	8698233	8698425	GENE_ID=PLCB1
chr20	8698425	8698576	GENE_ID=PLCB1
chr20	8702907	8703048	GENE_ID=PLCB1
chr20	8703048	8703148	GENE_ID=PLCB1
chr20	8705239	8705461	GENE_ID=PLCB1

chr20	8707795	8707961	GENE_ID=PLCB1
chr20	8707961	8708119	GENE_ID=PLCB1
chr20	8709583	8709712	GENE_ID=PLCB1
chr20	8709712	8709897	GENE_ID=PLCB1
chr20	8713789	8713908	GENE_ID=PLCB1
chr20	8713908	8714089	GENE_ID=PLCB1
chr20	8717591	8717744	GENE_ID=PLCB1
chr20	8717744	8717883	GENE_ID=PLCB1
chr20	8719831	8719932	GENE_ID=PLCB1
chr20	8719932	8720103	GENE_ID=PLCB1
chr20	8720920	8721060	GENE_ID=PLCB1
chr20	8721060	8721161	GENE_ID=PLCB1
chr20	8722013	8722124	GENE_ID=PLCB1
chr20	8722124	8722270	GENE_ID=PLCB1
chr20	8737642	8737845	GENE_ID=PLCB1
chr20	8737842	8737943	GENE_ID=PLCB1
chr20	8740957	8741157	GENE_ID=PLCB1
chr20	8745697	8745877	GENE_ID=PLCB1
chr20	8745877	8746074	GENE_ID=PLCB1
chr20	8755108	8755268	GENE_ID=PLCB1
chr20	8755268	8755428	GENE_ID=PLCB1
chr20	8768935	8769066	GENE_ID=PLCB1
chr20	8769066	8769222	GENE_ID=PLCB1
chr20	8769212	8769435	GENE_ID=PLCB1
chr20	8770086	8770295	GENE_ID=PLCB1
chr20	8770755	8770974	GENE_ID=PLCB1
chr20	8782570	8782726	GENE_ID=PLCB1
chr20	8782726	8782899	GENE_ID=PLCB1
chr20	8862215	8862360	GENE_ID=PLCB1
chr20	8862360	8862557	GENE_ID=PLCB1
chr20	8862557	8862774	GENE_ID=PLCB1
chr20	8862771	8862976	GENE_ID=PLCB1
chr20	8862976	8863169	GENE_ID=PLCB1
chr20	8863169	8863380	GENE_ID=PLCB1
chr20	8863379	8863592	GENE_ID=PLCB1
chr20	8863592	8863793	GENE_ID=PLCB1
chr20	8864011	8864115	GENE_ID=PLCB1

chr20	8864115	8864298	GENE_ID=PLCB1
chr20	8864234	8864396	GENE_ID=PLCB1
chr20	8864396	8864518	GENE_ID=PLCB1
chr20	8864518	8864684	GENE_ID=PLCB1
chr20	8864674	8864834	GENE_ID=PLCB1
chr20	8864832	8864960	GENE_ID=PLCB1
chr20	8864960	8865068	GENE_ID=PLCB1
chr20	8865068	8865264	GENE_ID=PLCB1
chr20	8865264	8865481	GENE_ID=PLCB1
chr20	8865481	8865610	GENE_ID=PLCB1
chr20	9049674	9049860	GENE_ID=PLCB4
chr20	9197970	9198165	GENE_ID=PLCB4
chr20	9288297	9288473	GENE_ID=PLCB4
chr20	9288473	9288649	GENE_ID=PLCB4
chr20	9317638	9317750	GENE_ID=PLCB4
chr20	9317750	9317903	GENE_ID=PLCB4
chr20	9318572	9318795	GENE_ID=PLCB4
chr20	9319483	9319580	GENE_ID=PLCB4
chr20	9319580	9319781	GENE_ID=PLCB4
chr20	9343409	9343525	GENE_ID=PLCB4
chr20	9343525	9343672	GENE_ID=PLCB4
chr20	9346015	9346231	GENE_ID=PLCB4
chr20	9351710	9351931	GENE_ID=PLCB4
chr20	9352836	9352986	GENE_ID=PLCB4
chr20	9352986	9353139	GENE_ID=PLCB4
chr20	9353593	9353817	GENE_ID=PLCB4
chr20	9360563	9360686	GENE_ID=PLCB4
chr20	9360686	9360879	GENE_ID=PLCB4
chr20	9364782	9364923	GENE_ID=PLCB4
chr20	9364923	9365116	GENE_ID=PLCB4
chr20	9367955	9368138	GENE_ID=PLCB4
chr20	9368138	9368253	GENE_ID=PLCB4
chr20	9370428	9370658	GENE_ID=PLCB4
chr20	9371103	9371321	GENE_ID=PLCB4
chr20	9374077	9374213	GENE_ID=PLCB4
chr20	9374213	9374438	GENE_ID=PLCB4
chr20	9376044	9376191	GENE_ID=PLCB4

chr20	9376191	9376417	GENE_ID=PLCB4
chr20	9382073	9382295	GENE_ID=PLCB4
chr20	9385834	9386056	GENE_ID=PLCB4
chr20	9388509	9388610	GENE_ID=PLCB4
chr20	9388610	9388758	GENE_ID=PLCB4
chr20	9389220	9389429	GENE_ID=PLCB4
chr20	9389554	9389765	GENE_ID=PLCB4
chr20	9389765	9389934	GENE_ID=PLCB4
chr20	9391604	9391821	GENE_ID=PLCB4
chr20	9400323	9400460	GENE_ID=PLCB4
chr20	9400460	9400626	GENE_ID=PLCB4
chr20	9401863	9401994	GENE_ID=PLCB4
chr20	9401994	9402158	GENE_ID=PLCB4
chr20	9404311	9404426	GENE_ID=PLCB4
chr20	9404426	9404619	GENE_ID=PLCB4
chr20	9416076	9416240	GENE_ID=PLCB4
chr20	9416240	9416385	GENE_ID=PLCB4
chr20	9417554	9417686	GENE_ID=PLCB4
chr20	9417686	9417881	GENE_ID=PLCB4
chr20	9424534	9424747	GENE_ID=PLCB4
chr20	9424755	9424959	GENE_ID=PLCB4
chr20	9433938	9434166	GENE_ID=PLCB4
chr20	9438002	9438208	GENE_ID=PLCB4
chr20	9440199	9440308	GENE_ID=PLCB4
chr20	9440308	9440507	GENE_ID=PLCB4
chr20	9449161	9449375	GENE_ID=PLCB4
chr20	9453356	9453554	GENE_ID=PLCB4
chr20	9453872	9454083	GENE_ID=PLCB4
chr20	9457272	9457489	GENE_ID=PLCB4
chr20	9459480	9459634	GENE_ID=PLCB4
chr20	9459634	9459833	GENE_ID=PLCB4
chr20	9459833	9460046	GENE_ID=PLCB4
chr20	9460046	9460258	GENE_ID=PLCB4
chr20	9460258	9460460	GENE_ID=PLCB4
chr20	9460460	9460667	GENE_ID=PLCB4
chr20	9460658	9460843	GENE_ID=PLCB4
chr20	9460843	9461016	GENE_ID=PLCB4

chr20	9461016	9461237	GENE_ID=PLCB4
chr20	9461237	9461384	GENE_ID=PLCB4
chr20	9461384	9461512	GENE_ID=PLCB4
chr20	9485757	9485917	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9485917	9486016	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9486016	9486232	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9486232	9486416	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9486416	9486639	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9486634	9486779	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9486779	9486996	SUBMITTED_REGION=RP5-1119D9.4_EX1
chr20	9487522	9487623	SUBMITTED_REGION=RP5-1119D9.4_EX2
chr20	9487623	9487838	SUBMITTED_REGION=RP5-1119D9.4_EX2
chr20	9487801	9487907	SUBMITTED_REGION=RP5-1119D9.4_EX2
chr20	9489497	9489641	SUBMITTED_REGION=RP5-1119D9.4_EX3
chr20	9489604	9489805	SUBMITTED_REGION=RP5-1119D9.4_EX3
chr20	9494981	9495119	SUBMITTED_REGION=LAMP5_EX1
chr20	9495119	9495344	SUBMITTED_REGION=LAMP5_EX1
chr20	9495242	9495358	SUBMITTED_REGION=LAMP5_EX1
chr20	9495451	9495651	SUBMITTED_REGION=LAMP5_EX1
chr20	9496007	9496116	SUBMITTED_REGION=LAMP5_EX2
chr20	9496116	9496330	SUBMITTED_REGION=LAMP5_EX2
chr20	9496651	9496764	SUBMITTED_REGION=LAMP5_EX3
chr20	9496764	9496985	SUBMITTED_REGION=LAMP5_EX4
chr20	9496985	9497120	SUBMITTED_REGION=LAMP5_EX4
chr20	9498617	9498758	SUBMITTED_REGION=LAMP5_EX5
chr20	9498758	9498904	SUBMITTED_REGION=LAMP5_EX5
chr20	9510300	9510526	SUBMITTED_REGION=LAMP5_EX6
chr20	9510526	9510754	SUBMITTED_REGION=LAMP5_EX6
chr20	9510754	9510889	SUBMITTED_REGION=LAMP5_EX6
chr20	9510889	9510992	SUBMITTED_REGION=LAMP5_EX6
chr20	9510992	9511204	SUBMITTED_REGION=LAMP5_EX6
chr20	9517953	9518106	GENE_ID=PAK7
chr20	9518106	9518207	GENE_ID=PAK7
chr20	9518207	9518330	GENE_ID=PAK7
chr20	9518330	9518437	GENE_ID=PAK7
chr20	9518437	9518636	GENE_ID=PAK7
chr20	9518622	9518745	GENE_ID=PAK7

chr20	9518745	9518882	GENE_ID=PAK7
chr20	9518882	9519055	GENE_ID=PAK7
chr20	9519055	9519257	GENE_ID=PAK7
chr20	9519257	9519468	GENE_ID=PAK7
chr20	9519462	9519649	GENE_ID=PAK7
chr20	9519633	9519813	GENE_ID=PAK7
chr20	9519813	9519945	GENE_ID=PAK7
chr20	9519945	9520170	GENE_ID=PAK7
chr20	9520170	9520334	GENE_ID=PAK7
chr20	9523119	9523226	GENE_ID=PAK7
chr20	9523226	9523417	GENE_ID=PAK7
chr20	9524890	9525082	GENE_ID=PAK7
chr20	9525082	9525224	GENE_ID=PAK7
chr20	9538137	9538287	GENE_ID=PAK7
chr20	9538287	9538431	GENE_ID=PAK7
chr20	9543434	9543534	GENE_ID=PAK7
chr20	9543534	9543722	GENE_ID=PAK7
chr20	9546464	9546591	GENE_ID=PAK7
chr20	9546591	9546724	GENE_ID=PAK7
chr20	9546686	9546902	GENE_ID=PAK7
chr20	9546894	9547117	GENE_ID=PAK7
chr20	9560711	9560882	GENE_ID=PAK7
chr20	9560882	9561075	GENE_ID=PAK7
chr20	9561075	9561177	GENE_ID=PAK7
chr20	9561177	9561395	GENE_ID=PAK7
chr20	9561395	9561627	GENE_ID=PAK7
chr20	9624699	9624896	GENE_ID=PAK7
chr20	9624896	9625042	GENE_ID=PAK7
chr20	9691833	9691954	GENE_ID=PAK7
chr20	9691954	9692150	GENE_ID=PAK7
chr20	9799181	9799356	GENE_ID=PAK7
chr20	9799356	9799511	GENE_ID=PAK7
chr20	9819330	9819513	GENE_ID=PAK7
chr20	9819530	9819755	GENE_ID=PAK7
chr20	9966966	9967184	SUBMITTED_REGION=LOC101929371_EX1
chr20	9987130	9987332	SUBMITTED_REGION=LOC101929371_EX2
chr20	9987332	9987554	SUBMITTED_REGION=LOC101929371_EX2

chr20	9987554	9987780	SUBMITTED_REGION=LOC101929371_EX2
chr20	10004418	10004633	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10004532	10004753	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10004753	10004915	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10004915	10005073	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005062	10005203	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005203	10005372	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005372	10005500	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005464	10005644	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005644	10005853	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005853	10005953	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10005953	10006123	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10006123	10006320	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10006320	10006502	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10006502	10006692	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10006706	10006872	SUBMITTED_REGION=SNAP25-AS1_EX1
chr20	10007128	10007353	SUBMITTED_REGION=SNAP25-AS1_EX2
chr20	10015672	10015809	GENE_ID=ANKRD5
chr20	10015809	10016031	GENE_ID=ANKRD5
chr20	10016144	10016372	GENE_ID=ANKRD5
chr20	10018917	10019125	GENE_ID=ANKRD5
chr20	10019125	10019348	GENE_ID=ANKRD5
chr20	10023685	10023848	GENE_ID=ANKRD5
chr20	10023848	10024041	GENE_ID=ANKRD5
chr20	10024987	10025106	GENE_ID=ANKRD5
chr20	10025106	10025259	GENE_ID=ANKRD5
chr20	10026169	10026395	GENE_ID=ANKRD5
chr20	10029973	10030144	GENE_ID=ANKRD5
chr20	10030144	10030288	GENE_ID=ANKRD5
chr20	10030288	10030483	GENE_ID=ANKRD5
chr20	10030483	10030696	GENE_ID=ANKRD5
chr20	10030693	10030902	GENE_ID=ANKRD5
chr20	10032233	10032383	GENE_ID=ANKRD5
chr20	10032383	10032600	GENE_ID=ANKRD5
chr20	10033674	10033812	GENE_ID=ANKRD5
chr20	10033812	10033981	GENE_ID=ANKRD5
chr20	10035022	10035122	GENE_ID=ANKRD5

chr20	10035122	10035315	GENE_ID=ANKRD5
chr20	10036093	10036226	GENE_ID=ANKRD5
chr20	10036226	10036368	GENE_ID=ANKRD5
chr20	10036368	10036573	GENE_ID=ANKRD5
chr20	10036573	10036786	GENE_ID=ANKRD5
chr20	10036786	10036906	GENE_ID=ANKRD5
chr20	10036906	10037072	GENE_ID=ANKRD5
chr20	10037072	10037252	GENE_ID=ANKRD5
chr20	10037252	10037459	GENE_ID=ANKRD5
chr20	10084882	10085087	SUBMITTED_REGION=SNAP25-AS1_EX3
chr20	10177413	10177626	SUBMITTED_REGION=SNAP25-AS1_EX4
chr20	10199370	10199491	GENE_ID=SNAP25
chr20	10199490	10199710	GENE_ID=SNAP25
chr20	10199843	10200012	SUBMITTED_REGION=SNAP25-AS1_EX5
chr20	10255961	10256063	GENE_ID=SNAP25
chr20	10256049	10256261	GENE_ID=SNAP25
chr20	10258243	10258468	GENE_ID=SNAP25
chr20	10265288	10265503	GENE_ID=SNAP25
chr20	10273403	10273517	GENE_ID=SNAP25
chr20	10273517	10273714	GENE_ID=SNAP25
chr20	10273596	10273824	GENE_ID=SNAP25
chr20	10273824	10273976	GENE_ID=SNAP25
chr20	10277451	10277548	GENE_ID=SNAP25
chr20	10277529	10277748	GENE_ID=SNAP25
chr20	10279837	10279938	GENE_ID=SNAP25
chr20	10279938	10280112	GENE_ID=SNAP25
chr20	10286692	10286902	GENE_ID=SNAP25
chr20	10286902	10287125	GENE_ID=SNAP25
chr20	10287125	10287343	GENE_ID=SNAP25
chr20	10287343	10287511	GENE_ID=SNAP25
chr20	10287511	10287657	GENE_ID=SNAP25
chr20	10287657	10287785	GENE_ID=SNAP25
chr20	10287785	10288001	GENE_ID=SNAP25
chr20	10288001	10288116	GENE_ID=SNAP25
chr20	10385802	10385976	GENE_ID=MKKS
chr20	10385976	10386168	GENE_ID=MKKS
chr20	10386168	10386347	GENE_ID=MKKS

chr20	10388130	10388224	GENE_ID=MKKS
chr20	10388224	10388430	GENE_ID=MKKS
chr20	10389194	10389309	GENE_ID=MKKS
chr20	10389309	10389520	GENE_ID=MKKS
chr20	10393092	10393294	GENE_ID=MKKS
chr20	10393294	10393504	GENE_ID=MKKS
chr20	10393504	10393720	GENE_ID=MKKS
chr20	10393720	10393937	GENE_ID=MKKS
chr20	10393937	10394157	GENE_ID=MKKS
chr20	10394157	10394379	GENE_ID=MKKS
chr20	10394296	10394433	GENE_ID=MKKS
chr20	10394433	10394648	GENE_ID=MKKS
chr20	10401106	10401248	GENE_ID=MKKS
chr20	10401248	10401466	GENE_ID=MKKS
chr20	10412256	10412401	GENE_ID=MKKS
chr20	10412401	10412623	GENE_ID=MKKS
chr20	10414664	10414802	GENE_ID=MKKS
chr20	10414741	10414917	GENE_ID=MKKS
chr20	10415889	10416061	GENE_ID=C20orf94
chr20	10416061	10416192	GENE_ID=C20orf94
chr20	10438771	10438960	GENE_ID=C20orf94
chr20	10536810	10537034	GENE_ID=C20orf94
chr20	10541222	10541394	GENE_ID=C20orf94
chr20	10541394	10541521	GENE_ID=C20orf94
chr20	10579215	10579315	GENE_ID=C20orf94
chr20	10579315	10579453	GENE_ID=C20orf94
chr20	10582326	10582528	GENE_ID=C20orf94
chr20	10601907	10602118	GENE_ID=C20orf94
chr20	10603237	10603463	GENE_ID=C20orf94
chr20	10603463	10603646	GENE_ID=C20orf94
chr20	10603646	10603859	GENE_ID=C20orf94
chr20	10603859	10604077	GENE_ID=C20orf94
chr20	10618256	10618358	GENE_ID=JAG1
chr20	10618358	10618534	GENE_ID=JAG1
chr20	10618534	10618752	GENE_ID=JAG1
chr20	10618752	10618849	GENE_ID=JAG1
chr20	10618849	10619056	GENE_ID=JAG1

chr20	10619052	10619236	GENE_ID=JAG1
chr20	10619236	10619404	GENE_ID=JAG1
chr20	10619404	10619588	GENE_ID=JAG1
chr20	10619498	10619709	GENE_ID=JAG1
chr20	10619707	10619838	GENE_ID=JAG1
chr20	10619836	10619936	GENE_ID=JAG1
chr20	10619936	10620158	GENE_ID=JAG1
chr20	10620158	10620345	GENE_ID=JAG1
chr20	10620345	10620555	GENE_ID=JAG1
chr20	10620555	10620656	GENE_ID=JAG1
chr20	10621350	10621519	GENE_ID=JAG1
chr20	10621511	10621732	GENE_ID=JAG1
chr20	10621732	10621942	GENE_ID=JAG1
chr20	10622057	10622187	GENE_ID=JAG1
chr20	10622187	10622414	GENE_ID=JAG1
chr20	10622370	10622593	GENE_ID=JAG1
chr20	10623008	10623233	GENE_ID=JAG1
chr20	10623233	10623340	GENE_ID=JAG1
chr20	10624353	10624579	GENE_ID=JAG1
chr20	10624907	10625126	GENE_ID=JAG1
chr20	10625405	10625632	GENE_ID=JAG1
chr20	10625632	10625821	GENE_ID=JAG1
chr20	10625803	10626021	GENE_ID=JAG1
chr20	10626021	10626183	GENE_ID=JAG1
chr20	10626561	10626787	GENE_ID=JAG1
chr20	10627510	10627616	GENE_ID=JAG1
chr20	10627616	10627832	GENE_ID=JAG1
chr20	10628520	10628746	GENE_ID=JAG1
chr20	10628746	10628849	GENE_ID=JAG1
chr20	10629113	10629212	GENE_ID=JAG1
chr20	10629212	10629435	GENE_ID=JAG1
chr20	10629589	10629818	GENE_ID=JAG1
chr20	10629922	10630138	GENE_ID=JAG1
chr20	10630138	10630348	GENE_ID=JAG1
chr20	10630840	10631063	GENE_ID=JAG1
chr20	10632105	10632208	GENE_ID=JAG1
chr20	10632208	10632430	GENE_ID=JAG1

chr20	10632656	10632760	GENE_ID=JAG1
chr20	10632760	10632971	GENE_ID=JAG1
chr20	10632939	10633167	GENE_ID=JAG1
chr20	10633128	10633322	GENE_ID=JAG1
chr20	10636959	10637181	GENE_ID=JAG1
chr20	10639036	10639250	GENE_ID=JAG1
chr20	10639250	10639448	GENE_ID=JAG1
chr20	10644546	10644764	GENE_ID=JAG1
chr20	10653284	10653509	GENE_ID=JAG1
chr20	10653509	10653705	GENE_ID=JAG1
chr20	10654099	10654330	GENE_ID=JAG1
chr20	10654330	10654499	GENE_ID=JAG1
chr20	10654499	10654617	GENE_ID=JAG1
chr20	10654604	10654777	GENE_ID=JAG1
chr20	10855961	10856161	SUBMITTED_REGION=RP4-697P8.3_EX1
chr20	10856161	10856311	SUBMITTED_REGION=RP4-697P8.3_EX1
chr20	10856311	10856467	SUBMITTED_REGION=RP4-697P8.3_EX1
chr20	10856467	10856677	SUBMITTED_REGION=RP4-697P8.3_EX1
chr20	10856677	10856893	SUBMITTED_REGION=RP4-697P8.3_EX1
chr20	10858898	10859088	SUBMITTED_REGION=RP4-697P8.3_EX2
chr20	10874509	10874725	SUBMITTED_REGION=RP4-697P8.3_EX3
chr20	10875201	10875300	SUBMITTED_REGION=RP4-697P8.3_EX4
chr20	10875300	10875521	SUBMITTED_REGION=RP4-697P8.3_EX4
chr20	10876610	10876794	SUBMITTED_REGION=RP4-697P8.3_EX5
chr20	10876766	10876920	SUBMITTED_REGION=RP4-697P8.3_EX5
chr20	10876920	10877071	SUBMITTED_REGION=RP4-697P8.3_EX5
chr20	10877677	10877865	SUBMITTED_REGION=RP4-697P8.3_EX6
chr20	10889548	10889726	SUBMITTED_REGION=RP4-697P8.3_EX7
chr20	10889726	10889946	SUBMITTED_REGION=RP4-697P8.3_EX7
chr20	11247250	11247359	SUBMITTED_REGION=LOC339593
chr20	11247359	11247582	SUBMITTED_REGION=LOC339593
chr20	11247582	11247750	SUBMITTED_REGION=LOC339593
chr20	11247750	11247844	SUBMITTED_REGION=LOC339593
chr20	11247844	11248037	SUBMITTED_REGION=LOC339593
chr20	11248037	11248184	SUBMITTED_REGION=LOC339593
chr20	11248184	11248403	SUBMITTED_REGION=LOC339593
chr20	11248403	11248608	SUBMITTED_REGION=LOC339593

chr20	11248608	11248799	SUBMITTED_REGION=LOC339593
chr20	11248799	11249006	SUBMITTED_REGION=LOC339593
chr20	11249006	11249118	SUBMITTED_REGION=LOC339593
chr20	11249118	11249344	SUBMITTED_REGION=LOC339593
chr20	11249318	11249483	SUBMITTED_REGION=LOC339593
chr20	11898477	11898683	GENE_ID=BTBD3
chr20	11898679	11898893	GENE_ID=BTBD3
chr20	11898893	11899085	GENE_ID=BTBD3
chr20	11899085	11899299	GENE_ID=BTBD3
chr20	11899652	11899883	GENE_ID=BTBD3
chr20	11900233	11900407	GENE_ID=BTBD3
chr20	11900407	11900537	GENE_ID=BTBD3
chr20	11903199	11903349	GENE_ID=BTBD3
chr20	11903349	11903565	GENE_ID=BTBD3
chr20	11903565	11903785	GENE_ID=BTBD3
chr20	11903785	11903961	GENE_ID=BTBD3
chr20	11903961	11904186	GENE_ID=BTBD3
chr20	11904186	11904413	GENE_ID=BTBD3
chr20	11904413	11904629	GENE_ID=BTBD3
chr20	11904629	11904824	GENE_ID=BTBD3
chr20	11904824	11904967	GENE_ID=BTBD3
chr20	11904967	11905181	GENE_ID=BTBD3
chr20	11905181	11905294	GENE_ID=BTBD3
chr20	11905292	11905504	GENE_ID=BTBD3
chr20	11905504	11905694	GENE_ID=BTBD3
chr20	11905694	11905922	GENE_ID=BTBD3
chr20	11905922	11906127	GENE_ID=BTBD3
chr20	11906127	11906332	GENE_ID=BTBD3
chr20	11906332	11906521	GENE_ID=BTBD3
chr20	11906520	11906703	GENE_ID=BTBD3
chr20	11906703	11906860	GENE_ID=BTBD3
chr20	11906855	11907060	GENE_ID=BTBD3
chr20	11907055	11907240	GENE_ID=BTBD3
chr20	11907208	11907302	GENE_ID=BTBD3

(Appendix B2)

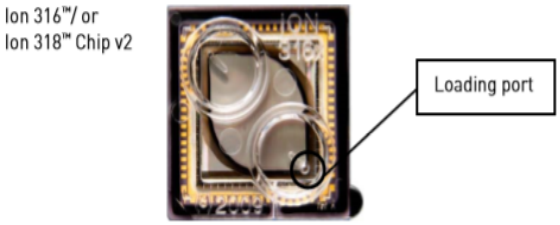
Loading of Ion 316 Chip v2 for PGM

According to the manufacturers protocol Adapted from Life Technologies.

Bind Sequencing Polymerase to the ISPs

1. After annealing the Sequencing Primer, remove the ISPs from the thermal cycler and add 3 µL of Ion PGM™ Sequencing 200 v2 Polymerase to the ISPs, for a total final volume of 30 µL.
2. Pipet the sample up and down to mix, and incubate at room temperature for 5 minutes.

Load the chip



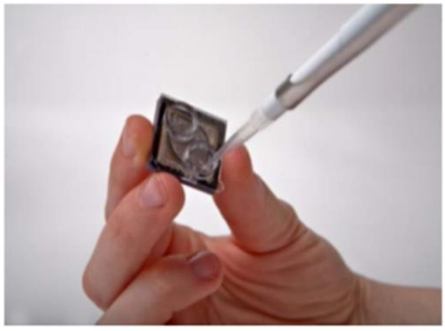
Alternate chip loading method: Ion PGM™ Chip loading with the Ion PGM™ Weighted Chip Bucket

For an alternate chip loading method with fewer handling steps, see the *Ion PGM™ Chip Loading with the Ion PGM™ Weighted Chip Bucket User Bulletin* (Pub. no. MAN0007517), which is available for download from the Ion Community at ioncommunity.lifetechnologies.com.

This protocol requires the use of the Ion PGM™ Weighted Chip Bucket (Cat. no. 4480283), which can be ordered by contacting Life Technologies customer service as described on page 70.

Remove liquid from the chip

1. Tilt the chip 45 degrees so that the loading port is the lower port, as shown below.



2. Insert the pipette tip firmly into the loading port and remove as much liquid as possible from the loading port. Discard the liquid.

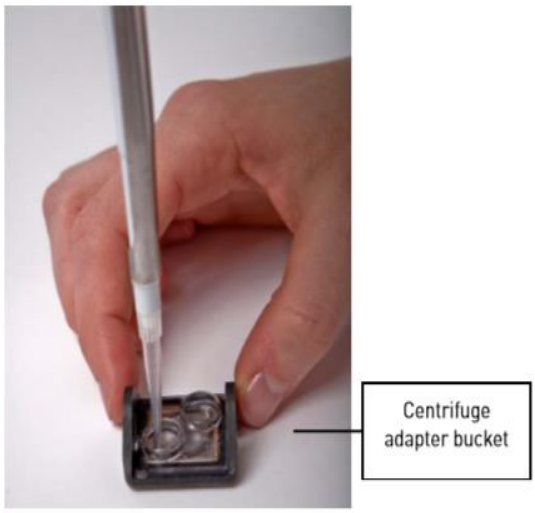
IMPORTANT! For the next steps, balance the centrifuge adapter with a used chip of the same chip type and orientation if you are preparing one Ion PGM™ Chip at a time. Be careful to balance an upside-down chip with another upside-down chip. Mark the used chip with a laboratory marker to differentiate it from the new chip containing the sample.

3. Place the chip **upside-down** in the centrifuge adapter bucket and transfer the bucket to the MiniFuge **with the chip tab pointing in** (toward the center of the MiniFuge), as shown below.

4. Centrifuge for 5 seconds to completely empty the chip. Remove the chip from the bucket and wipe off any liquid on the bucket.

Load the sample on the chip

IMPORTANT! When loading liquid into the chip, keep the pipette tip at a 90° angle to the chip, press the tip firmly into the circular loading port, and apply gentle pressure between the pipette tip and chip.



1. Place the Ion PGM™ Chip back in the centrifuge adapter bucket and place the bucket on a flat, stable surface such as a benchtop.
2. Following polymerase incubation, collect the entire sample (~30 µL) into a Rainin® SR-L200F pipette tip and insert the tip firmly into the loading port of the chip.
3. Dial down the pipette as shown below to gently and slowly deposit the ISPs at a rate of ~1 µL per second. To avoid introducing bubbles into the chip, leave a small amount of sample in the pipette tip (~0.5 µL).

4. Remove and discard any displaced liquid from the other port of the chip.

5. Transfer the chip to the MiniFuge **with the chip tab pointing in** (toward the center of the MiniFuge).

6. Centrifuge for 30 seconds, then remove the chip from the centrifuge bucket.
7. Mix the sample in the chip:
   a. Set the pipette volume to 30 µL.
   b. Tilt the chip 45 degrees so that the loading port is the lower port, and insert the pipette tip into the loading port.
   c. Without removing the tip, slowly pipet the sample in and out of the chip three times. **Pipet slowly to avoid creating bubbles.**
8. Centrifuge the chip for 30 seconds **with the chip tab pointing out** (away from the center of the MiniFuge).

9. Repeat the chip mixing in step 7 one more time, then spin for 30 seconds **with the chip tab pointing in** (toward the center of the MiniFuge).
10. Tilt the chip at a 45-degree angle and slowly remove as much liquid as possible from the loading port by dialing the pipette. Discard the liquid.
11. If some liquid remains in the chip, perform a 5-second quick spin with the chip tab pointing out and remove and discard any additional liquid. **Do not spin the chip upside down.**
12. If some liquid remains in the chip after the quick spin, lightly and rapidly tap the point of the chip tab against the benchtop a few times, and remove and discard any collected liquid. Do not flush the chip.
13. When chip loading is complete, press **Next** on the touchscreen and proceed immediately to **Select the Planned Run and perform the run**.

(Appendix B2)

Loading of Ion Pi Chip v2 for Proton Sequencer

According to the manufacturers protocol Adapted from Life Technologies.

Prepare and calibrate the Ion PI™ Chip v2 for loading

IMPORTANT! To avoid possible damage to the chip due to electrostatic discharge, ground yourself before picking up a chip or placing a chip on a surface such as a lab bench. For example, touch the metal trim on the chip compartment before inserting or removing a chip from the chip clamp.

Prepare the chip for loading

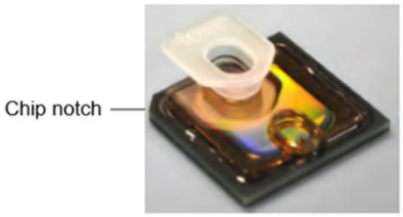
IMPORTANT! Wear gloves for the following steps. The Ion PI™ Chip Preparation Solution is a skin irritant.

1. Label a new Ion PI™ Chip v2 to identify the experiment, then place the chip on a stable surface such as a benchtop.

   Note: If the chip was not used to initialize the Ion Proton™ Sequencer, skip the next step.

2. If the chip was used to initialize the Ion Proton™ Sequencer, inject 100 µL isopropanol into the chip loading port, then inject 100 µL of nuclease-free water into the loading port and remove the expelled liquid from the opposite port. Repeat the water flush one more time and remove the expelled liquid.

3. Attach an Ion Proton™ Chip Adapter to the chip exit well (located on the side of the chip with the notched corner, opposite the loading port; see figure below). Press firmly to ensure the adapter is tightly engaged around the well.



4. Inject 200 µL of 100% isopropanol into the chip loading port, then remove the expelled liquid from the exit well fitted with the adapter.

5. Attach a P200 pipette tip to a vacuum line, insert the pipette tip into the chip loading port, and aspirate the isopropanol from the chip flow cell for 5–10 seconds. Inspect the chip to make sure it is dry.

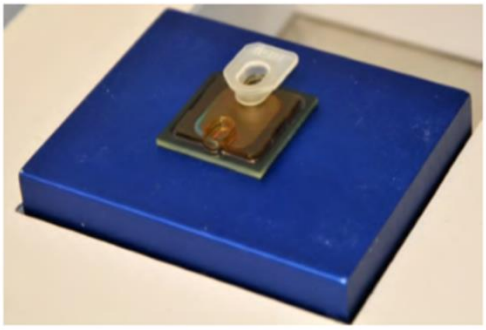
   Note: Make sure that the Ion PI™ Chip Preparation Solution has completely equilibrated to room temperature before proceeding to the next step. Immediately re-cap the solution tube after use to prevent evaporation and exposure to air.

6. Inject 100 µL of Ion PI™ Chip Preparation Solution into the chip loading port, then remove the expelled liquid from the exit well.

   Note: Be careful not to introduce bubbles in the chip. If bubbles appear, repeat this wash step.

7. Place the chip on the flat surface of an aluminum heat block pre-heated to 50°C and incubate for 2 minutes.

Note: Place the chip on the side of the heat block that does not contain wells, as shown in the figure below.



8. Remove the chip from the heat block and place it on a stable surface such as a benchtop.

9. Inject 200 µL of 100% isopropanol into the chip loading port, then remove the expelled liquid from the exit well. Repeat the isopropanol flush one more time.

10. Inject 200 µL of nuclease-free water into the chip loading port, then remove the expelled liquid from the exit well.

11. Inject 200 µL of 0.1 M NaOH into the chip loading port, then remove the expelled liquid from the exit well.

12. Wait 1 minute, then inject 200 µL of nuclease-free water into the chip loading port and remove the expelled liquid from the exit well.

13. Repeat steps 4–12 two more times, for a total of three chip washes.

14. Inject 200 µL of 100% isopropanol into the chip loading port, then remove the expelled liquid from the exit well.

15. Pipet 100 µL of 100% isopropanol into the chip loading well (do not inject directly into the port), then remove the liquid from the same well. This step cleans any residue from the surface of the loading well.

16. Remove the Ion Proton™ Chip Adapter from the chip.

Optional: Store the prepared chip

The prepared chip can be stored for up to 24 hours before calibration and loading. To store the chip:

1. Attach a P200 pipette tip to a vacuum line, insert the pipette tip into the chip loading port, and aspirate the isopropanol from the chip flow cell for 5–10 seconds. Inspect the chip to make sure it is dry.

2. Place the dry chip back in its original anti-static bag and store at room temperature in the dark (for example, in the box that the chip came in).

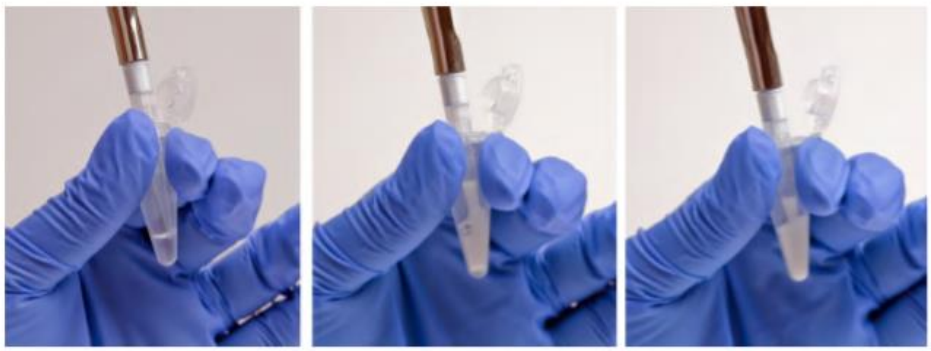
3. When you are ready to calibrate and load the chip, inject 100 µL of 100% isopropanol into the chip flow cell, remove the expelled liquid from the opposite port, then proceed to the next section.

3. Centrifuge the chip for 10 minutes in the Ion Chip™ Minifuge.

4. In a 1.5-mL tube, combine 45 µL of 50% Annealing Buffer with 5 µL of 2% Triton® X-100.

5. Create foam by injecting air into the 50-µL mixture from the previous step using a Rainin® SR-L200F pipette set to dispense 100 µL. Next, break the large bubbles into smaller bubbles by rapidly pipetting for ~5 seconds. Repeat this step one more time.

   Note: Do not over-inject the air; the final volume of foam should be approximately 250 µL.



6. Place the chip on a stable surface such as a benchtop, then inject 100 µL of foam into the chip loading port. Remove the expelled liquid from the opposite port.

7. Place the chip back in the centrifuge bucket with the chip notch pointing out, then dispense 55 µL of 50% Annealing buffer into the chip loading well (not the chip loading port).

8. Centrifuge the chip for 30 seconds in the Ion Chip™ Minifuge.

9. Place the chip on a stable surface such as a benchtop. Remove the liquid that has accumulated in both of the chip loading wells.

10. Briefly "re-foam" the foam sample by pipetting rapidly for ~5 seconds, then inject 100 µL of foam into the chip loading port. Remove the expelled liquid from the opposite port.

11. Place the chip back in the centrifuge bucket with the chip notch pointing out, then dispense 55 µL of 50% Annealing buffer into the chip loading well (not the chip loading port).

12. Centrifuge the chip for 30 seconds in the Ion Chip™ Minifuge, then proceed to flushing the chip.

Calibrate the chip

1. Secure the prepared chip in the chip clamp.
2. Press **Run** in the Main Menu, then press **Next** and confirm that "Cleaning fluid lines" displays.
   The instrument begins cleaning the fluid lines, then begins chip calibration.
3. During calibration, observe the chip for leaks. If there is a leak, press the **Abort** button immediately to stop the flow to the chip, then see "Liquid in drip pan below chip clamp" on page 42.

   IMPORTANT! Never open the chip clamp during calibration.

4. When the chip passes calibration, remove the chip, and replace it with a cleaning chip to prevent backflow in the fluid lines.

   Note: If chip calibration fails, see "Error message: Failed: Reseat chip, then press Next to recalibrate" on page 45.

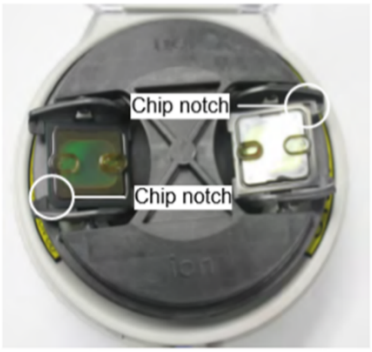
5. Inject 100 µL of nuclease-free water two times into the chip loading port. After each injection, remove the expelled liquid from the opposite port.
6. Inject 100 µL of isopropanol two times into the chip loading port. After each injection, remove the expelled liquid from the opposite port.
7. Aspirate the remaining isopropanol from the chip flow cell for 5–10 seconds. Confirm that the chip is dry.

   Note: To aspirate the isopropanol, attach a P200 pipette tip to a vacuum line, then place the pipette tip in the chip loading port.

Load the Ion PI™ Chip v2

Load the sample on the chip

1. Place the calibrated chip in the centrifuge bucket with the chip notch pointing out. Place a used chip in the opposite bucket with the chip notch pointing out.



2. Dispense the entire prepared sample (55 µL) into the chip loading well (not the chip loading port) of the calibrated chip.

   Note: Some sample enters the flow cell at this point by capillary action; the remaining sample is loaded into the flow cell during centrifugation.

Flush the chip and load the Ion PI™ Sequencing Polymerase

1. Inject 100 µL of the Flushing solution into the chip loading port two times. After each injection, discard the solution that is expelled from the opposite port.

2. Inject 100 µL of 50% Annealing Buffer into the chip loading port three times. Do not introduce air bubbles. After each injection, remove the expelled liquid from the opposite port.

IMPORTANT! For Ion AmpliSeq™ Exome libraries, use Ion PI™ Sequencing Polymerase v3, included in the Ion AmpliSeq™ Exome Kit (Cat. no. 4487084), instead of Ion PI™ Sequencing Polymerase in the next step.

---

3. Combine 6 µL of Ion PI™ Sequencing Polymerase with 60 µL of 50% Annealing buffer, then inject 65 µL of this solution into the chip loading port. Avoid introducing air bubbles. Remove the expelled liquid from the opposite port.

4. Allow the chip to incubate for 5 minutes, then immediately proceed to "Select the planned run and start the sequencing run" on page 37.

Select the planned run and start the sequencing run

Complete "Prepare and calibrate the Ion PI Chip v2 for loading" on page 33 and "Load the Ion PI Chip v2" on page 35 before performing this procedure.

1. Secure the chip loaded with template-positive Ion PI™ ISPs in the chip clamp, close the chip compartment lid, then press **Next**.

2. In the drop-down list, select a planned run that you created in the Torrent Suite Software, then press **Next**.

   Note: You can also select **Planned Run (none)**, then enter your run information on the following screen, but selecting a predefined planned run is recommended.

3. Confirm the pre-populated settings are correct, or make changes using the buttons and drop-down lists if necessary.

   Note: If an error message appears, see "Error message: Not enough disk space for the necessary number of flows" on page 41.

4. Confirm that the reagent compartment door is closed, then press **Next** to begin the sequencing run.

(Appendix B3)

PedCheck software

The PedFile without the disease column was used and genotyping format was converted to numerical format (A,B >1,2). Also, MapFile included only the SNP ID as the output is expected to be SNPs with Mendelian errors.

The Commands used for PedCheck analysis:

-p PedFile_2.txt –n MapFile_2.txt -4

Merlin for npl Analysis

Merlin requires 3 files in order to operate as described in Chapter 6, Data File, Ped File and Map File.

The commands used for Non- Parametric Linkage Analysis:

-d DataFile.txt –p PedFile.txt –m MapFile.txt –npl –exp

--swap –smalSwap –quiet

--markerNames –pdf –prefix npl_1 >output_npl_1

References

Aagaard, L. and Rossi, J. J. (2007). RNAi therapeutics: principles, prospects and challenges. *Adv Drug Deliv Rev* **59:**75-86.

Abou-haila, A. and Tulsiani, D. R. (2009). Signal transduction pathways that regulate sperm capacitation and the acrosome reaction. *Arch Biochem Biophys* **485:**72-81.

Adriana, G. (2015). Epidimiological aspects of euthyroid diffuse goiter in a group of adults with thyroid diseases and diabetes mellitus and other changes in glycemic balance. European Scientific Journal pp. 28-39.

Aitken, R. J., Nixon, B., Lin, M., Koppers, A. J., Lee, Y. H. and Baker, M. A. (2007). Proteomic changes in mammalian spermatozoa during epididymal maturation. *Asian J Androl* **9:**554-564.

Allard, S. T. M. (2008). Bioluminescent Reporter Genes [Online]. http://www.promega.co.uk/resources/pubhub/enotes/bioluminescent-reporter-genes. Available at: [Accessed.

Amballi, A. A. (2007). Thyrotoxicosis-A Review. Middle-East Journal of Scientific Research: pp. 98-103.

Amdani, S. N., Jones, C. and Coward, K. (2013). Phospholipase C zeta (PLCζ): oocyte activation and clinical links to male factor infertility. *Adv Biol Regul* **53:**292-308.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnol* **25:**195-203.

Bakhsh, A., Kirov, G., Gregory, J. W., Williams, E. D. and Ludgate, M. (2006). A new form of familial multi-nodular goitre with progression to differentiated thyroid cancer. *Endocr Relat Cancer* **13:**475-483.

Berridge, M. J. (2007). Inositol trisphosphate and calcium oscillations. *Biochem Soc Symp* 1-7.

Berridge, M. J., Bootman, M. D. and Roderick, H. L. (2003). Calcium signalling: dynamics, homeostasis and remodelling. *Nat Rev Mol Cell Biol* **4:**517-529.

Bhatkar, S. V., Rajan, M. G., Velumani, A. and Samuel, A. M. (2004). Thyroid hormone binding protein abnormalities in patients referred for thyroid disorders. *Indian J Med Res* **120:**160-165.

Boehning, D. and Joseph, S. K. (2000). Functional properties of recombinant type I and type III inositol 1, 4,5-trisphosphate receptor isoforms expressed in COS-7 cells. *J Biol Chem* **275:**21492-21499.

Bornhorst, B. J. and Falke, J. J. (2011). Reprint of: Purification of Proteins Using Polyhistidine Affinity Tags. *Protein Expr Purif*.

Braunstein, G. D. (2012). Thyroid cancer. *Endocrine updates,*. New York: Springer,. pp. 1 online resource (x, 322 pages).

Brown, R. L., de Souza, J. A. and Cohen, E. E. (2011). Thyroid cancer: burden of illness and management of disease. *J Cancer* **2:**193-199.

Bullock, M., Duncan, E. L., O'Neill, C., Tacon, L., Sywak, M., Sidhu, S., Delbridge, L. *et al.* (2012). Association of FOXE1 polyalanine repeat region with papillary thyroid cancer. *J Clin Endocrinol Metab* **97:**E1814-1819.

Bunney, T. D. and Katan, M. (2011). PLC regulation: emerging pictures for molecular mechanisms. *Trends Biochem Sci* **36:**88-96.

Burgess, G. M., McKinney, J. S., Fabiato, A., Leslie, B. A. and Putney, J. W. (1983). Calcium pools in saponin-permeabilized guinea pig hepatocytes. *J Biol Chem* **258:**15336-15345.

Capon, F., Tacconelli, A., Giardina, E., Sciacchitano, S., Bruno, R., Tassi, V., Trischitta, V. *et al.* (2000). Mapping a dominant form of multinodular goiter to chromosome Xp22. *Am J Hum Genet* **67:**1004-1007.

Carmell, M. A., Xuan, Z., Zhang, M. Q. and Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16:**2733-2742.

Chatterjee, V. K., Madison, L. D., Mayo, S. and Jameson, J. L. (1991). Repression of the human glycoprotein hormone alpha-subunit gene by glucocorticoids: evidence for receptor interactions with limiting transcriptional activators. *Mol Endocrinol* **5:**100-110.

Chi, H. C., Chen, C. Y., Tsai, M. M., Tsai, C. Y. and Lin, K. H. (2013). Molecular functions of thyroid hormones and their clinical significance in liver-related diseases. *Biomed Res Int* **2013:**601361.

Cocco, L., Martelli, A. M., Vitale, M., Falconi, M., Barnabei, O., Stewart Gilmour, R. and Manzoli, F. A. (2002). Inositides in the nucleus: regulation of nuclear PI-PLCbeta1. *Adv Enzyme Regul* **42:**181-193.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:**57-74.

Corvilain, B., Laurent, E., Lecomte, M., Vansande, J. and Dumont, J. E. (1994). Role of the cyclic adenosine 3',5'-monophosphate and the phosphatidylinositol-Ca2+ cascades in mediating the effects of thyrotropin and iodide on hormone synthesis and secretion in human thyroid slices. *J Clin Endocrinol Metab* **79:**152-159.

Cossu, A., Budroni, M., Paliogiannis, P., Palmieri, G., Scognamillo, F., Cesaraccio, R., Attene, F. *et al.* (2013). Epidemiology of thyroid cancer in an area of epidemic thyroid goiter. *J Cancer Epidemiol* **2013:**584768.

Costa, S., Almeida, A., Castro, A. and Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front Microbiol* **5:**63.

Coughlin, C. R., Scharer, G. H. and Shaikh, T. H. (2012). Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Med* **4:**80.

Coward, K. and Wells, D. (2013). Textbook of Clinical Embryology. Cambridge: Cambridge University Press,. p. 1 online resource (402 p.).

D'haene, B., Vandesompele, J. and Hellemans, J. (2010). Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50:**262-270.

Dacheux, J. L., Gatti, J. L. and Dacheux, F. (2003). Contribution of epididymal secretory proteins for spermatozoa maturation. *Microsc Res Tech* **61:**7-17.

Dale, B., Elder, K., Harper, J., Huntriss, J. and Ménézo, Y. (2010). In-Vitro Fertilization. Cambridge: Cambridge University Press,. p. 1 online resource (288 p.).

Dam, A. H., Feenstra, I., Westphal, J. R., Ramos, L., van Golde, R. J. and Kremer, J. A. (2007). Globozoospermia revisited. *Hum Reprod Update* **13:**63-75.

Davies, T. F., Ando, T., Lin, R. Y., Tomer, Y. and Latif, R. (2005). Thyrotropin receptor-associated diseases: from adenomata to Graves disease. *J Clin Invest* **115:**1972-1983.

Dawn Teare, M. and Barrett, J. H. (2005). Genetic linkage studies. *Lancet* **366:**1036-1044.

de Souza, E. C., Padrón, A. S., Braga, W. M., de Andrade, B. M., Vaisman, M., Nasciutti, L. E., Ferreira, A. C. *et al.* (2010). MTOR downregulates iodide uptake in thyrocytes. *J Endocrinol* **206:**113-120.

Dowal, L., Provitera, P. and Scarlata, S. (2006). Stable association between G alpha(q) and phospholipase C beta 1 in living cells. *J Biol Chem* **281:**23999-24014.

Dudbridge, F. (2003). A survey of current software for linkage analysis. *Hum Genomics* **1:**63-65.

Esposito, D. and Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol* **17:**353-358.

Essen, L. O., Perisic, O., Katan, M., Wu, Y., Roberts, M. F. and Williams, R. L. (1997). Structural mapping of the catalytic mechanism for a mammalian phosphoinositide-specific phospholipase C. *Biochemistry* **36:**1704-1718.

Evans, J. P. (2002). The molecular basis of sperm-oocyte membrane interactions during mammalian fertilization. *Hum Reprod Update* **8:**297-311.

Faenza, I., Matteucci, A., Manzoli, L., Billi, A. M., Aluigi, M., Peruzzi, D., Vitale, M. *et al.* (2000). A role for nuclear phospholipase Cbeta 1 in cell cycle control. *J Biol Chem* **275:**30520-30524.

Falcone, T. and Hurd, W. W. (2007). Clinical reproductive medicine and surgery. Philadelphia: Mosby/Elsevier, pp. xiv, 831 p.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:**806-811.

Fiume, R., Faenza, I., Matteucci, A., Astolfi, A., Vitale, M., Martelli, A. M. and Cocco, L. (2005). Nuclear phospholipase C beta1 (PLCbeta1) affects CD24 expression in murine erythroleukemia cells. *J Biol Chem* **280:**24221-24226.

Fiume, R., Ramazzotti, G., Teti, G., Chiarini, F., Faenza, I., Mazzotti, G., Billi, A. M. *et al.* (2009). Involvement of nuclear PLCbeta1 in lamin B1 phosphorylation and G2/M cell cycle progression. *FASEB J* **23:**957-966.

Florman, H. M., Jungnickel, M. K. and Sutton, K. A. (2008). Regulating the acrosome reaction. *Int J Dev Biol* **52:**503-510.

Follo, M. Y., Finelli, C., Bosi, C., Martinelli, G., Mongiorgi, S., Baccarani, M., Manzoli, L. *et al.* (2008). PI-PLCbeta-1 and activated Akt levels are linked to azacitidine responsiveness in high-risk myelodysplastic syndromes. *Leukemia* **22:**198-200.

Follo, M. Y., Finelli, C., Clissa, C., Mongiorgi, S., Bosi, C., Martinelli, G., Baccarani, M. *et al.* (2009). Phosphoinositide-phospholipase C beta1 mono-allelic deletion is associated with myelodysplastic syndromes evolution into acute myeloid leukemia. *J Clin Oncol* **27:**782-790.

Foster, C. R., Przyborski, S. A., Wilson, R. G. and Hutchison, C. J. (2010). Lamins as cancer biomarkers. *Biochem Soc Trans* **38:**297-300.

Frasca, F., Nucera, C., Pellegriti, G., Gangemi, P., Attard, M., Stella, M., Loda, M. *et al.* (2008). BRAF(V600E) mutation and the biology of papillary thyroid cancer. *Endocr Relat Cancer* **15:**191-205.

Fresno Vara, J. A., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C. and González-Barón, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treat Rev* **30:**193-204.

Friederichs, J., Zeller, Y., Hafezi-Moghadam, A., Gröne, H. J., Ley, K. and Altevogt, P. (2000). The CD24/P-selectin binding pathway initiates lung arrest of human A125 adenocarcinoma cells. *Cancer Res* **60:**6714-6722.

Fukami, K. (2002). Structure, regulation, and function of phospholipase C isozymes. *J Biochem* **131:**293-299.

Fulton, B. P. and Whittingham, D. G. (1978). Activation of mammalian oocytes by intracellular injection of calcium. *Nature* **273:**149-151.

Futatsugi, A., Nakamura, T., Yamada, M. K., Ebisui, E., Nakamura, K., Uchida, K., Kitaguchi, T. *et al.* (2005). IP3 receptor types 2 and 3 mediate exocrine secretion underlying energy metabolism. *Science* **309:**2232-2234.

Gabriel, J. (2007). The biology of cancer. 2nd ed. Chichester ; Hoboken, NJ: John Wiley & Sons, p. 204p.

Gadella, B. M., Tsai, P. S., Boerke, A. and Brewis, I. A. (2008). Sperm head membrane reorganisation during capacitation. *Int J Dev Biol* **52:**473-480.

Gasdaska, J. R., Berggren, M. and Powis, G. (1995). Cell growth stimulation by the redox protein thioredoxin occurs by a novel helper mechanism. *Cell Growth Differ* **6:**1643-1650.

Gifford, J. L., Walsh, M. P. and Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. *Biochem J* **405:**199-221.

Golebiewska, U. and Scarlata, S. (2010). The effect of membrane domains on the G protein-phospholipase Cbeta signaling pathway. *Crit Rev Biochem Mol Biol* **45:**97-105.

Goodman, H. M. (2003). Basic medical endocrinology. San Diego, Calif.: Academic Press,. pp. 1 online resource (xxi, 469 pages).

Grienberger, C. and Konnerth, A. (2012). Imaging calcium in neurons. *Neuron* **73:**862-885.

Grubb, D. R., Vasilevski, O., Huynh, H. and Woodcock, E. A. (2008). The extreme C-terminal region of phospholipase Cbeta1 determines subcellular localization and function; the "b" splice variant mediates alpha1-adrenergic receptor responses in cardiomyocytes. *FASEB J* **22:**2768-2774.

Gudmundsson, J., Sulem, P., Gudbjartsson, D. F., Jonasson, J. G., Sigurdsson, A., Bergthorsson, J. T., He, H. *et al.* (2009). Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet* **41:**460-464.

Guerrero-Preston, R., Michailidi, C., Marchionni, L., Pickering, C. R., Frederick, M. J., Myers, J. N., Yegnasubramanian, S. *et al.* (2014). Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer. *Epigenetics* **9**.

Gundem, G., Perez-Llamas, C., Jene-Sanz, A., Kedzierska, A., Islam, A., Deu-Pons, J., Furney, S. J. *et al.* (2010). IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat Methods* **7:**92-93.

Gupta, P. K. (2008). Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol* **26:**602-611.

Hall, L. L., Smith, K. P., Byron, M. and Lawrence, J. B. (2006). Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol* **288:**664-675.

Halpern, J. and Whittemore, A. S. (1999). Multipoint linkage analysis. A cautionary note. *Hum Hered* **49:**194-196.

Hanumanthappa. (2012). The Incidence of Malignancy in Multi-nodular Goitre: A Prospective Study at a Tertiary Academic Centre. In: Gopinathan S., R.S., Guruprasad Rai D. Gautham Shetty, Ashith Shetty, Naren Shetty ed. Journal of clinical and diagnostic research. pp. 267-270.

Haq, R., Shoag, J., Andreu-Perez, P., Yokoyama, S., Edelman, H., Rowe, G. C., Frederick, D. T. *et al.* (2013). Oncogenic BRAF regulates oxidative metabolism via PGC1α and MITF. *Cancer Cell* **23:**302-315.

Hay, N. and Sonenberg, N. (2004). Upstream and downstream of mTOR. *Genes Dev* **18:**1926-1945.

Hicks, S. N., Jezyk, M. R., Gershburg, S., Seifert, J. P., Harden, T. K. and Sondek, J. (2008). General and versatile autoinhibition of PLC isozymes. *Mol Cell* **31:**383-394.

Hiromatsu, Y., Satoh, H. and Amino, N. (2013). Hashimoto's thyroiditis: history and future outlook. *Hormones (Athens)* **12:**12-18.

Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H. and Iino, M. (1999). Spatiotemporal dynamics of inositol 1,4,5-trisphosphate that underlies complex Ca2+ mobilization patterns. *Science* **284:**1527-1530.

Hisatsune, C., Nakamura, K., Kuroda, Y., Nakamura, T. and Mikoshiba, K. (2005). Amplification of Ca2+ signaling by diacylglycerol-mediated inositol 1,4,5-trisphosphate production. *J Biol Chem* **280:**11723-11730.

Ho, C. W. (2008). Comparative evaluation of different cell disruption methods for the release of recombinant hepatitis B core antigen from *Escherichia Coli*. In: Wen Siang Tan, W.B.Y., Tau Chuan Ling, Beng Ti Tey ed. Biotechnology and Bioprocess Engineering. pp. 577-583.

Hokin, M. R. and Hokin, L. E. (1953). Enzyme secretion and the incorporation of P32 into phospholipides of pancreas slices. *J Biol Chem* **203:**967-977.

Holmgren, A. (1985). Thioredoxin. *Annu Rev Biochem* **54:**237-271.

Hu, Y., Kireev, I., Plutz, M., Ashourian, N. and Belmont, A. S. (2009). Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J Cell Biol* **185:**87-100.

Huang, W., Barrett, M., Hajicek, N., Hicks, S., Harden, T. K., Sondek, J. and Zhang, Q. (2013). Small molecule inhibitors of phospholipase C from a novel high-throughput screen. *J Biol Chem* **288:**5840-5848.

Huang, W., Hicks, S. N., Sondek, J. and Zhang, Q. (2011). A fluorogenic, small molecule reporter for mammalian phospholipase C isozymes. *ACS Chem Biol* **6:**223-228.

Iitaka, M., Kakinuma, S., Fujimaki, S., Oosuga, I., Fujita, T., Yamanaka, K., Wada, S. *et al.* (2001). Induction of apoptosis and necrosis by zinc in human thyroid cancer cell lines. *J Endocrinol* **169:**417-424.

Jaffe, L. F. (1983). Sources of calcium in egg activation: a review and hypothesis. *Dev Biol* **99:**265-276.

Jaffe, L. F. (1991). The path of calcium in cytosolic calcium oscillations: a unifying hypothesis. *Proc Natl Acad Sci U S A* **88:**9883-9887.

Jemal, A., Siegel, R., Xu, J. and Ward, E. (2010). Cancer statistics, 2010. *CA Cancer J Clin* **60:**277-300.

Jia, W. H., Zhang, B., Matsuo, K., Shin, A., Xiang, Y. B., Jee, S. H., Kim, D. H. *et al.* (2013). Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet* **45:**191-196.

Jones, K. T., Soeller, C. and Cannell, M. B. (1998). The passage of Ca2+ and fluorescent markers between the sperm and egg after fusion in the mouse. *Development* **125:**4627-4635.

Kaji, K. and Kudo, A. (2004). The mechanism of sperm-oocyte fusion in mammals. *Reproduction* **127:**423-429.

Kashir, J., Heindryckx, B., Jones, C., De Sutter, P., Parrington, J. and Coward, K. (2010). Oocyte activation, phospholipase C zeta and human infertility. *Hum Reprod Update* **16:**690-703.

Kashir, J., Nomikos, M., Swann, K. and Lai, F. A. (2015). PLCζ or PAWP: revisiting the putative mammalian sperm factor that triggers egg activation and embryogenesis. *Mol Hum Reprod* **21:**383-388.

Katan, M. and Williams, R. L. (1997). Phosphoinositide-specific phospholipase C: structural basis for catalysis and regulatory interactions. *Semin Cell Dev Biol* **8:**287-296.

Kay, M. A., Glorioso, J. C. and Naldini, L. (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med* **7:**33-40.

Kimura, S. (2011). Thyroid-specific transcription factors and their roles in thyroid cancer. *J Thyroid Res* **2011:**710213.

Kirchhoff, C. (1998). Molecular characterization of epididymal proteins. *Rev Reprod* **3:**86-95.

Knobil, E. and Neill, J. D. (1998). Encyclopedia of reproduction. San Diego ; London: Academic.

Knowles, M. A., Selby, P. and MyiLibrary. (2005). Introduction to the cellular and molecular biology of cancer. Oxford: Oxford University Press,. pp. xvii, 532 p.

Knudsen, N., Bülow, I., Laurberg, P., Perrild, H., Ovesen, L. and Jørgensen, T. (2002). High occurrence of thyroid multinodularity and low occurrence of subclinical hypothyroidism among tobacco smokers in a large population study. *J Endocrinol* **175:**571-576.

Kobayashi, K., Shaver, J. K., Liang, W., Siperstein, A. E., Duh, Q. Y. and Clark, O. H. (1993). Increased phospholipase C activity in neoplastic thyroid membrane. *Thyroid* **3:**25-29.

Kondo, T., Ezzat, S. and Asa, S. L. (2006). Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nat Rev Cancer* **6:**292-306.

Kong, A. and Cox, N. J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* **61:**1179-1188.

Kouchi, Z., Fukami, K., Shikano, T., Oda, S., Nakamura, Y., Takenawa, T. and Miyazaki, S. (2004). Recombinant phospholipase Czeta has high Ca2+ sensitivity and induces Ca2+ oscillations in mouse eggs. *J Biol Chem* **279:**10408-10412.

Kouchi, Z., Shikano, T., Nakamura, Y., Shirakawa, H., Fukami, K. and Miyazaki, S. (2005). The role of EF-hand domains and C2 domain in regulation of enzymatic activity of phospholipase Czeta. *J Biol Chem* **280:**21015-21021.

Kristiansen, G., Denkert, C., Schlüns, K., Dahl, E., Pilarsky, C. and Hauptmann, S. (2002). CD24 is expressed in ovarian cancer and is a new independent prognostic marker of patient survival. *Am J Pathol* **161:**1215-1221.

Krohn, K., Führer, D., Bayer, Y., Eszlinger, M., Brauer, V., Neumann, S. and Paschke, R. (2005). Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr Rev* **26:**504-524.

Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. and Lander, E. S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58:**1347-1363.

Kuroda, K., Ito, M., Shikano, T., Awaji, T., Yoda, A., Takeuchi, H., Kinoshita, K. *et al.* (2006). The role of X/Y linker region and N-terminal EF-hand domain in nuclear translocation and Ca2+ oscillation-inducing activities of phospholipase Czeta, a mammalian egg-activating factor. *J Biol Chem* **281:**27794-27805.

Lamond, A. I. and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4:**605-612.

Lasserre, A., Barrozo, R., Tezon, J. G., Miranda, P. V. and Vazquez-Levin, M. H. (2001). Human epididymal proteins and sperm function during fertilization: un update. *Biol Res* **34:**165-178.

Laurent, T. C., Moore, E. C. and Reichard, P. (1964). Enzymatic Synthesis of Deoxyribonucleotides. Iv. Isolation and Characterization of Thioredoxin, the Hydrogen Donor from Escherichia Coli B. *J Biol Chem* **239:**3436-3444.

Lawrence, Y., Whitaker, M. and Swann, K. (1997). Sperm-egg fusion is the prelude to the initial Ca2+ increase at fertilization in the mouse. *Development* **124:**233-241.

Lee, J. H. and Jeon, J. T. (2008). Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels. *Cytogenet Genome Res* **123:**333-342.

Lee, Y. H., Kim, S. Y., Kim, J. R., Yoh, K. T., Baek, S. H., Kim, M. J., Ryu, S. H. *et al.* (2000). Overexpression of phospholipase Cbeta-1 protects NIH3T3 cells from oxidative stress-induced cell death. *Life Sci* **67:**827-837.

Levenson, A. S. and Jordan, V. C. (1997). MCF-7: the first hormone-responsive breast cancer cell line. *Cancer Res* **57:**3071-3078.

Lewit-Bentley, A. and Réty, S. (2000). EF-hand calcium-binding proteins. *Curr Opin Struct Biol* **10:**637-643.

Lin, C. Y., Vega, V. B., Thomsen, J. S., Zhang, T., Kong, S. L., Xie, M., Chiu, K. P. *et al.* (2007). Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3:**e87.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D. *et al.* (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012:**251364.

Liu, Y. J., Qiang, W., Shi, J., Lv, S. Q., Ji, M. J. and Shi, B. Y. (2013). Expression and significance of IGF-1 and IGF-1R in thyroid nodules. *Endocrine* **44:**158-164.

Liò, P. and Morton, N. E. (1997). Comparison of parametric and nonparametric methods to map oligogenes by linkage. *Proc Natl Acad Sci U S A* **94:**5344-5348.

Lopata, A., Sibson, M. C., Enders, A. C., Bloomfield, K. L., Gregory, M. S., Trapani, G. D., Perkins, A. V. *et al.* (2001). Expression and localization of thioredoxin during early implantation in the marmoset monkey. *Mol Hum Reprod* **7:**1159-1165.

Lukinovic-Skudar, V., Donlagic, L., Banfíc, H. and Visnjic, D. (2005). Nuclear phospholipase C-beta1b activation during G2/M and late G1 phase in nocodazole-synchronized HL-60 cells. *Biochim Biophys Acta* **1733:**148-156.

Lukinovic-Skudar, V., Matkovic, K., Banfic, H. and Visnjic, D. (2007). Two waves of the nuclear phospholipase C activity in serum-stimulated HL-60 cells during G(1) phase of the cell cycle. *Biochim Biophys Acta* **1771:**514-521.

Lyon, A. N., Pineda, R. H., Hao, L. T., Kudryashova, E., Kudryashov, D. S. and Beattie, C. E. (2013). Calcium binding is essential for plastin 3 function in Smn-deficient motoneurons. *Hum Mol Genet*.

Makino, T., Skretas, G. and Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microb Cell Fact* **10:**32.

Manole, D., Schildknecht, B., Gosnell, B., Adams, E. and Derwahl, M. (2001). Estrogen promotes growth of human thyroid tumor cells by different molecular mechanisms. *J Clin Endocrinol Metab* **86:**1072-1077.

Manzoli, L., Billi, A. M., Rubbini, S., Bavelloni, A., Faenza, I., Gilmour, R. S., Rhee, S. G. *et al.* (1997). Essential role for nuclear phospholipase C beta1 in insulin-like growth factor I-induced mitogenesis. *Cancer Res* **57:**2137-2139.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9:**387-402.

Martelli, A. M., Billi, A. M., Manzoli, L., Faenza, I., Aluigi, M., Falconi, M., De Pol, A. *et al.* (2000a). Insulin selectively stimulates nuclear phosphoinositide-specific phospholipase C (PI-PLC) beta1 activity through a mitogen-activated protein (MAP) kinase-dependent serine phosphorylation. *FEBS Lett* **486:**230-236.

Martelli, A. M., Bortul, R., Tabellini, G., Aluigi, M., Peruzzi, D., Bareggi, R., Narducci, P. *et al.* (2001). Re-examination of the mechanisms regulating nuclear inositol lipid metabolism. *FEBS Lett* **505:**1-6.

Martelli, A. M., Cocco, L., Bareggi, R., Tabellini, G., Rizzoli, R., Ghibellini, M. D. and Narducci, P. (1999). Insulin-like growth factor-I-dependent stimulation of nuclear phospholipase C-beta1 activity in Swiss 3T3 cells requires an intact cytoskeleton and is paralleled by increased phosphorylation of the phospholipase. *J Cell Biochem* **72:**339-348.

Martelli, A. M., Gilmour, R. S., Bertagnolo, V., Neri, L. M., Manzoli, L. and Cocco, L. (1992). Nuclear localization and signalling activity of phosphoinositidase C beta in Swiss 3T3 cells. *Nature* **358:**242-245.

Martelli, A. M., Tabellini, G., Bortul, R., Manzoli, L., Bareggi, R., Baldini, G., Grill, V. *et al.* (2000b). Enhanced nuclear diacylglycerol kinase activity in response to a mitogenic stimulation of quiescent Swiss 3T3 cells with insulin-like growth factor I. *Cancer Res* **60:**815-821.

Matsui, M., Oshima, M., Oshima, H., Takaku, K., Maruyama, T., Yodoi, J. and Taketo, M. M. (1996). Early embryonic lethality caused by targeted disruption of the mouse thioredoxin gene. *Dev Biol* **178:**179-185.

McCudden, C. R., Willard, F. S., Kimple, R. J., Johnston, C. A., Hains, M. D., Jones, M. B. and Siderovski, D. P. (2005). G alpha selectivity and inhibitor function of the multiple GoLoco motif protein GPSM2/LGN. *Biochim Biophys Acta* **1745:**254-264.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **11:**31-46.

Miao, Y. L., Stein, P., Jefferson, W. N., Padilla-Banks, E. and Williams, C. J. (2012). Calcium influx-mediated signaling is required for complete mouse egg activation. *Proc Natl Acad Sci U S A* **109:**4169-4174.

Miyazaki, S. (1988). Inositol 1,4,5-trisphosphate-induced calcium release and guanine nucleotide-binding protein-mediated periodic calcium rises in golden hamster eggs. *J Cell Biol* **106:**345-353.

Miyazaki, S. and Ito, M. (2006). Calcium signals for egg activation in mammals. *J Pharmacol Sci* **100:**545-552.

Miyazaki, S., Shirakawa, H., Nakada, K. and Honda, Y. (1993). Essential role of the inositol 1,4,5-trisphosphate receptor/Ca2+ release channel in Ca2+ waves and Ca2+ oscillations at fertilization of mammalian eggs. *Dev Biol* **158:**62-78.

Molinari, C., Medri, L., Follo, M. Y., Piazzi, M., Mariani, G. A., Calistri, D. and Cocco, L. (2012). PI-PLCβ1 gene copy number alterations in breast cancer. *Oncol Rep* **27:**403-408.

Moreno-Reyes, R., Tang, B. N., Seret, A., Goldman, S., Daumerie, C. and Corvilain, B. (2007). Impaired iodide organification in autonomous thyroid nodules. *J Clin Endocrinol Metab* **92:**4719-4724.

Moretti, S., De Falco, V., Tamburrino, A., Barbi, F., Tavano, M., Avenia, N., Santeusanio, F. *et al.* (2009). Insights into the molecular function of the inactivating mutations of B-Raf involving the DFG motif. *Biochim Biophys Acta* **1793:**1634-1645.

Morozova, O. and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92:**255-264.

Moulick, N. D. (2008). Multinodular Thyroid Disease. *Medicine*.

Nada, A., Mohamed Ahmed, A., Vilallonga, R., Armengol, M. and Moustafa, I. (2011). A giant euthyroid endemic multinodular goiter with no obstructive or compressive symptoms. *Case Rep Med* **2011:**620480.

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C. F. *et al.* (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485:**242-245.

Nelson, M. R., Thulin, E., Fagan, P. A., Forsén, S. and Chazin, W. J. (2002). The EF-hand domain: a globally cooperative structural unit. *Protein Sci* **11:**198-205.

Neumann, S., Willgerodt, H., Ackermann, F., Reske, A., Jung, M., Reis, A. and Paschke, R. (1999). Linkage of familial euthyroid goiter to the multinodular goiter-1 locus and exclusion of the candidate genes thyroglobulin, thyroperoxidase, and Na+/I- symporter. *J Clin Endocrinol Metab* **84:**3750-3756.

Nomikos, M., Blayney, L. M., Larman, M. G., Campbell, K., Rossbach, A., Saunders, C. M., Swann, K. *et al.* (2005). Role of phospholipase C-zeta domains in Ca2+-dependent phosphatidylinositol 4,5-bisphosphate hydrolysis and cytoplasmic Ca2+ oscillations. *J Biol Chem* **280:**31011-31018.

Nomikos, M., Elgmati, K., Theodoridou, M., Calver, B. L., Cumbes, B., Nounesis, G., Swann, K. *et al.* (2011a). Male infertility-linked point mutation disrupts the Ca2+ oscillation-inducing and PIP(2) hydrolysis activity of sperm PLCzeta. *Biochem J* **434:**211-217.

Nomikos, M., Elgmati, K., Theodoridou, M., Georgilis, A., Gonzalez-Garcia, J. R., Nounesis, G., Swann, K. *et al.* (2011b). Novel regulation of PLCzeta activity via its XY-linker. *Biochem J* **438:**427-432.

Nomikos, M., Kashir, J., Swann, K. and Lai, F. A. (2013a). Sperm PLCζ: from structure to Ca2+ oscillations, egg activation and therapeutic potential. *FEBS Lett* **587:**3609-3616.

Nomikos, M., Mulgrew-Nesbitt, A., Pallavi, P., Mihalyne, G., Zaitseva, I., Swann, K., Lai, F. A. *et al.* (2007). Binding of phosphoinositide-specific phospholipase C-zeta (PLC-zeta) to phospholipid membranes: potential role of an unstructured cluster of basic residues. *J Biol Chem* **282:**16644-16653.

Nomikos, M., Theodoridou, M., Elgmati, K., Parthimos, D., Calver, B. L., Buntwal, L., Nounesis, G. *et al.* (2014). Human PLCζ exhibits superior fertilization potency over mouse PLCζ in triggering the Ca(2+) oscillations required for mammalian oocyte activation. *Mol Hum Reprod* **20:**489-498.

Nomikos, M., Yu, Y., Elgmati, K., Theodoridou, M., Campbell, K., Vassilakopoulou, V., Zikos, C. *et al.* (2013b). Phospholipase Cζ rescues failed oocyte activation in a prototype of male factor infertility. *Fertil Steril* **99:**76-85.

O'Carroll, S. J., Mitchell, M. D., Faenza, I., Cocco, L. and Gilmour, R. S. (2009). Nuclear PLCbeta1 is required for 3T3-L1 adipocyte differentiation and regulates expression of the cyclin D3-cdk4 complex. *Cell Signal* **21:**926-935.

Ock, S., Ahn, J., Lee, S. H., Kang, H., Offermanns, S., Ahn, H. Y., Jo, Y. S. *et al.* (2013). IGF-1 receptor deficiency in thyrocytes impairs thyroid hormone secretion and completely inhibits TSH-stimulated goiter. *FASEB J* **27:**4899-4908.

Okada, M., Fujii, M., Yamaga, M., Sugimoto, H., Sadano, H., Osumi, T., Kamata, H. *et al.* (2002). Carboxyl-terminal basic amino acids in the X domain are essential for the nuclear import of phospholipase C delta1. *Genes Cells* **7:**985-996.

Paredes, R. M., Etzler, J. C., Watts, L. T., Zheng, W. and Lechleiter, J. D. (2008). Chemical calcium indicators. *Methods* **46:**143-151.

Pearce, E. N., Gerber, A. R., Gootnick, D. B., Khan, L. K., Li, R., Pino, S. and Braverman, L. E. (2002). Effects of chronic iodine excess in a cohort of long-term American workers in West Africa. *J Clin Endocrinol Metab* **87:**5499-5502.

Philip, F., Guo, Y., Aisiku, O. and Scarlata, S. (2012). Phospholipase Cβ1 is linked to RNA interference of specific genes through translin-associated factor X. *FASEB J* **26:**4903-4913.

Piazzi, M., Blalock, W. L., Bavelloni, A., Faenza, I., D'Angelo, A., Maraldi, N. M. and Cocco, L. (2013). Phosphoinositide-specific phospholipase C β 1b (PI-PLCβ1b) interactome: affinity purification-mass spectrometry analysis of PI-PLCβ1b with nuclear protein. *Mol Cell Proteomics* **12:**2220-2235.

Placzkowski, K. A., Reddi, H. V., Grebe, S. K., Eberhardt, N. L. and McIver, B. (2008). The Role of the PAX8/PPARgamma Fusion Oncogene in Thyroid Cancer. *PPAR Res* **2008:**672829.

Poli, A., Faenza, I., Chiarini, F., Matteucci, A., McCubrey, J. A. and Cocco, L. (2013). K562 cell proliferation is modulated by PLCβ1 through a PKCα-mediated pathway. *Cell Cycle* **12:**1713-1721.

Primakoff, P. and Myles, D. G. (2007). Cell-cell membrane fusion during mammalian fertilization. *FEBS Lett* **581:**2174-2180.

Primrose, S. B., Twyman, R. M. and ebrary Inc. (2003). Principles of genome analysis and genomics. Malden, Mass.: Blackwell Pub.,. p. 1 electronic text.

Pulst, S. M. (1999). Genetic linkage analysis. *Arch Neurol* **56:**667-672.

Ramadan, W. M., Kashir, J., Jones, C. and Coward, K. (2012). Oocyte activation and phospholipase C zeta (PLCζ): diagnostic and therapeutic implications for assisted reproductive technology. *Cell Commun Signal* **10:**12.

Rastogi, M. V. and LaFranchi, S. H. (2010). Congenital hypothyroidism. *Orphanet J Rare Dis* **5:**17.

Razzini, G., Brancaccio, A., Lemmon, M. A., Guarnieri, S. and Falasca, M. (2000). The role of the pleckstrin homology domain in membrane targeting and activation of phospholipase Cbeta(1). *J Biol Chem* **275:**14873-14881.

Rhee, S. G. (2001). Regulation of phosphoinositide-specific phospholipase C. *Annu Rev Biochem* **70:**281-312.

Ribeiro, R., Monteiro, C., Catalán, V., Hu, P., Cunha, V., Rodríguez, A., Gómez-Ambrosi, J. *et al.* (2012). Obesity and prostate cancer: gene expression signature of human periprostatic adipose tissue. *BMC Med* **10:**108.

Rice, A., Parrington, J., Jones, K. T. and Swann, K. (2000). Mammalian sperm contain a Ca(2+)-sensitive phospholipase C activity that can generate InsP(3) from PIP(2) associated with intracellular organelles. *Dev Biol* **228:**125-135.

Ronaghi, M., Uhlén, M. and Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* **281:**363, 365.

Rosen, A., Lundman, P., Carlsson, M., Bhavani, K., Srinivasa, B. R., Kjellstrom, G., Nilsson, K. *et al.* (1995). A CD4+ T cell line-secreted factor, growth promoting for normal and leukemic B cells, identified as thioredoxin. *Int Immunol* **7:**625-633.

Ross, M. H. and Dobler, J. (1975). The Sertoli cell junctional specializations and their relationship to the germinal epithelium as observed after efferent ductule ligation. *Anat Rec* **183:**267-291.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H. *et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475:**348-352.

Rotin, D., Honegger, A. M., Margolis, B. L., Ullrich, A. and Schlessinger, J. (1992). Presence of SH2 domains of phospholipase C gamma 1 enhances substrate phosphorylation by increasing the affinity toward the epidermal growth factor receptor. *J Biol Chem* **267:**9678-9683.

Runft, L. L., Watras, J. and Jaffe, L. A. (1999). Calcium release at fertilization of Xenopus eggs requires type I IP(3) receptors, but not SH2 domain-mediated activation of PLCgamma or G(q)-mediated activation of PLCbeta. *Dev Biol* **214:**399-411.

Russell, P. J. (2002). iGenetics. San Francisco ; London: Benjamin Cummings, pp. xix, 828 pages.

Rusten, T. E. and Stenmark, H. (2006). Analyzing phosphoinositides and their interacting proteins. *Nat Methods* **3:**251-258.

Saunders, C. M., Larman, M. G., Parrington, J., Cox, L. J., Royse, J., Blayney, L. M., Swann, K. *et al.* (2002). PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development. *Development* **129:**3533-3544.

Schneider, D. F. and Chen, H. (2013). New developments in the diagnosis and treatment of thyroid cancer. *CA Cancer J Clin* **63:**374-394.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26:**1135-1145.

Shreve, D. P. and Carter, D. R. (2009). Non-coding RNAs: Bridging Biology and Therapy. *J RNAi Gene Silencing* **5:**351-353.

Sinha, R. A., You, S. H., Zhou, J., Siddique, M. M., Bay, B. H., Zhu, X., Privalsky, M. L. *et al.* (2012). Thyroid hormone stimulates hepatic lipid catabolism via activation of autophagy. *J Clin Invest* **122:**2428-2438.

Sledz, C. A. and Williams, B. R. (2005). RNA interference in biology and disease. *Blood* **106:**787-794.

Song, Y., Driessens, N., Costa, M., De Deken, X., Detours, V., Corvilain, B., Maenhaut, C. *et al.* (2007). Roles of hydrogen peroxide in thyroid physiology and disease. *J Clin Endocrinol Metab* **92:**3764-3773.

Steinhardt, R. A., Epel, D., Carroll, E. J. and Yanagimachi, R. (1974). Is calcium ionophore a universal activator for unfertilised eggs? *Nature* **252:**41-43.

Stiblar-Martincic, D., Cör, A. and Pajer, Z. (2002). Astereological analysis of FRTL-5 cells as an effect of thyrotropin and irradiation. *Folia Biol (Krakow)* **50:**173-177.

Stone, D., David, A., Bolognani, F., Lowenstein, P. R. and Castro, M. G. (2000). Viral vectors for gene delivery and gene therapy within the endocrine system. *J Endocrinol* **164:**103-118.

Strachan, T. and Read, A. P. (2011). Human molecular genetics. Fourth edition. ed. New York: Garland Science, pp. xxv, 781 pages.

Strauch, K., Fimmers, R., Kurz, T., Deichmann, K. A., Wienker, T. F. and Baur, M. P. (2000). Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. *Am J Hum Genet* **66:**1945-1957.

Suh, P. G., Park, J. I., Manzoli, L., Cocco, L., Peak, J. C., Katan, M., Fukami, K. *et al.* (2008). Multiple roles of phosphoinositide-specific phospholipase C isozymes. *BMB Rep* **41:**415-434.

Sullivan, P. F., Neale, B. M., Neale, M. C., van den Oord, E. and Kendler, K. S. (2003). Multipoint and single point non-parametric linkage analysis with imperfect data. *Am J Med Genet B Neuropsychiatr Genet* **121B:**89-94.

Swann, K. (1990). A cytosolic sperm factor stimulates repetitive calcium increases and mimics fertilization in hamster eggs. *Development* **110:**1295-1302.

Swann, K. and Lai, F. A. (2013). PLCζ and the initiation of Ca(2+) oscillations in fertilizing mammalian eggs. *Cell Calcium* **53:**55-62.

Swann, K., Saunders, C. M., Rogers, N. T. and Lai, F. A. (2006). PLCzeta(zeta): a sperm protein that triggers Ca2+ oscillations and egg activation in mammals. *Semin Cell Dev Biol* **17:**264-273.

Szkudlinski, M. W., Fremont, V., Ronin, C. and Weintraub, B. D. (2002). Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. *Physiol Rev* **82:**473-502.

Takahashi, T., Nozaki, J., Komatsu, M., Wada, Y., Utsunomiya, M., Inoue, K., Takada, G. *et al.* (2001). A new locus for a dominant form of multinodular goiter on 3q26.1-q26.3. *Biochem Biophys Res Commun* **284:**650-654.

Takenawa, T. and Nagai, Y. (1981). Purification of phosphatidylinositol-specific phospholipase C from rat liver. *J Biol Chem* **256:**6769-6775.

Taniguchi, T., Miyazaki, T., Minami, Y., Kawahara, A., Fujii, H., Nakagawa, Y., Hatakeyama, M. *et al.* (1995). IL-2 signaling involves recruitment and activation of multiple protein tyrosine kinases by the IL-2 receptor. *Ann N Y Acad Sci* **766:**235-244.

Taylor, C. W., da Fonseca, P. C. and Morris, E. P. (2004). IP(3) receptors: the search for structure. *Trends Biochem Sci* **29:**210-219.

Taylor, C. W. and Tovey, S. C. (2010). IP(3) receptors: toward understanding their activation. *Cold Spring Harb Perspect Biol* **2:**a004010.

Taylor, D. A., Sack, J. S., Maune, J. F., Beckingham, K. and Quiocho, F. A. (1991). Structure of a recombinant calmodulin from Drosophila melanogaster refined at 2.2-A resolution. *J Biol Chem* **266:**21375-21380.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60:**523-533.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72:**211-222.

Theodoridou, M., Nomikos, M., Parthimos, D., Gonzalez-Garcia, J. R., Elgmati, K., Calver, B. L., Sideratou, Z. *et al.* (2013). Chimeras of sperm PLCζ reveal disparate protein domain functions in the generation of intracellular Ca2+ oscillations in mammalian eggs at fertilization. *Mol Hum Reprod* **19:**852-864.

Thompson, J. F. and Steinmann, K. E. (2010). Single molecule sequencing with a HeliScope genetic analysis system. *Curr Protoc Mol Biol* **Chapter 7:**Unit7.10.

Tripathi, V., Song, D. Y., Zong, X., Shevtsov, S. P., Hearn, S., Fu, X. D., Dundr, M. *et al.* (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol Biol Cell* **23:**3694-3706.

Truter, I. (2011). Clinical review: hyper- and hypothyroidism. S Afr Pharm J.

Tsien, R. Y. (1980). New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures. *Biochemistry* **19:**2396-2404.

Tufano, R. P., Teixeira, G. V., Bishop, J., Carson, K. A. and Xing, M. (2012). BRAF mutation in papillary thyroid cancer and its value in tailoring initial treatment: a systematic review and meta-analysis. *Medicine (Baltimore)* **91:**274-286.

Umekawa, T., Sugiyama, T., Kihira, T., Murabayashi, N., Zhang, L., Nagao, K., Kamimoto, Y. *et al.* (2008). Overexpression of thioredoxin-1 reduces oxidative stress in the placenta of transgenic mice and promotes fetal growth via glucose metabolism. *Endocrinology* **149:**3980-3988.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet* **30:**418-426.

Vanderpump, M. P. (2011). The epidemiology of thyroid disease. *Br Med Bull* **99:**39-51.

Vanderpump, M. P., Tunbridge, W. M., French, J. M., Appleton, D., Bates, D., Clark, F., Grimley Evans, J. *et al.* (1995). The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)* **43:**55-68.

Vitale, M., Matteucci, A., Manzoli, L., Rodella, L., Mariani, A. R., Zauli, G., Falconi, M. *et al.* (2001). Interleukin 2 activates nuclear phospholipase Cbeta by mitogen-activated protein kinase-dependent phosphorylation in human natural killer cells. *FASEB J* **15:**1789-1791.

Völzke, H., Alte, D., Schmidt, C. O., Radke, D., Lorbeer, R., Friedrich, N., Aumann, N. *et al.* (2011). Cohort profile: the study of health in Pomerania. *Int J Epidemiol* **40:**294-307.

Wang, T., Dowal, L., El-Maghrabi, M. R., Rebecchi, M. and Scarlata, S. (2000). The pleckstrin homology domain of phospholipase C-beta(2) links the binding of gbetagamma to activation of the catalytic core. *J Biol Chem* **275:**7466-7469.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H. *et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453:**539-543.

Wen, W., Yan, J. and Zhang, M. (2006). Structural characterization of the split pleckstrin homology domain in phospholipase C-gamma1 and its interaction with TRPC3. *J Biol Chem* **281:**12060-12068.

Werner, S. C., Ingbar, S. H., Braverman, L. E., Utiger, R. D. and Ovid Technologies Inc. (2005). Werner & Ingbar's the thyroid

a fundamental and clinical text. *LWW Doody's all reviewed collection*. Philadelphia: Lippincott Williams & Wilkins,. pp. xvii, 1166 pages.

Williams, C. J., Mehlmann, L. M., Jaffe, L. A., Kopf, G. S. and Schultz, R. M. (1998). Evidence that Gq family G proteins do not function in mouse egg activation at fertilization. *Dev Biol* **198:**116-127.

Williams, G. R. and Bassett, J. H. (2011). Deiodinases: the balance of thyroid hormone: local control of thyroid hormone action: role of type 2 deiodinase. *J Endocrinol* **209:**261-272.

Wu, H., Smyth, J., Luzzi, V., Fukami, K., Takenawa, T., Black, S. L., Allbritton, N. L. *et al.* (2001). Sperm factor induces intracellular free calcium oscillations by stimulating the phosphoinositide pathway. *Biol Reprod* **64:**1338-1349.

Xing, M. (2013). Molecular pathogenesis and mechanisms of thyroid cancer. *Nat Rev Cancer* **13:**184-199.

Xing, M., Westra, W. H., Tufano, R. P., Cohen, Y., Rosenbaum, E., Rhoden, K. J., Carson, K. A. *et al.* (2005). BRAF mutation predicts a poorer clinical prognosis for papillary thyroid cancer. *J Clin Endocrinol Metab* **90:**6373-6379.

Xu, A., Suh, P. G., Marmy-Conus, N., Pearson, R. B., Seok, O. Y., Cocco, L. and Gilmour, R. S. (2001a). Phosphorylation of nuclear phospholipase C beta1 by extracellular signal-regulated kinase mediates the mitogenic action of insulin-like growth factor I. *Mol Cell Biol* **21:**2981-2990.

Xu, A., Wang, Y., Xu, L. Y. and Gilmour, R. S. (2001b). Protein kinase C alpha -mediated negative feedback regulation is responsible for the termination of insulin-like growth factor I-induced activation of nuclear phospholipase C beta1 in Swiss 3T3 cells. *J Biol Chem* **276:**14980-14986.

Yoon, S. Y., Jellerette, T., Salicioni, A. M., Lee, H. C., Yoo, M. S., Coward, K., Parrington, J. *et al.* (2008). Human sperm devoid of PLC, zeta 1 fail to induce Ca(2+) release and are unable to initiate the first step of embryo development. *J Clin Invest* **118:**3671-3681.

Zennou, V., Petit, C., Guetard, D., Nerhbass, U., Montagnier, L. and Charneau, P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101:**173-185.

Zhao, R., Bodnar, M. S. and Spector, D. L. (2009). Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev* **19:**172-179.

Zhou, X., Ren, L., Li, Y., Zhang, M., Yu, Y. and Yu, J. (2010). The next-generation sequencing technology: a technology review and future perspective. *Sci China Life Sci* **53:**44-57.

Zimmermann, M. B., Ito, Y., Hess, S. Y., Fujieda, K. and Molinari, L. (2005). High thyroid volume in children with excess dietary iodine intakes. *Am J Clin Nutr* **81:**840-844.

Zufferey, R., Donello, J. E., Trono, D. and Hope, T. J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* **73:**2886-2892.