.

A Combined Thesis of:

EPIGENETIC ANALYSIS OF SMALL CELL NUMBERS USING C-CHIP IN CONJUNCTION WITH HIGH THROUGHPUT SEQUENCING

CHARACTERISATION AND MODELLING OF MUTANT PALB2 PROTEINS

Submitted for the MRes in Molecular and Cellular Biology,
University of Birmingham

By Charlotte Eaton, 2nd September 2014

Contents

Characterisation and Modelling of Mutant PALB2 Proteins.

UNIVERSITY OF BIRMINGHAM

<u>Epigenetic Analysis of Small Cell Numbers Using C-ChIP in Conjunction with High Throughput Sequencing</u>

Abstract

Epigenetics and post-translational modifications (PTMs) are processes which contribute to the activation and deactivation of tumour development and cancer cell proliferation. ENCODE is a database incorporating all of the found modifications lying on the genome. Currently, a huge gap in the literature exists which proves that histone modifications are significantly underrepresented within cancer research. This gap is due to the lack of methods available for analysis of clinical samples. Chromatin immunoprecipitation (ChIP) is a vastly used technique however, it holds many limitations. This thesis describes a protocol which can be used to successfully identify the presence of PTMs of proteins from samples containing as little as 1,000-10,000 cells. By using carrier cells from *d.melanogaster,* target cells can be input into the solution and immunoprecipitated side by side. Results showed that by treating carriers with UV light in specific conditions, it is possible to inhibit the PCR amplification of the carrier cells 200 fold while still retaining carrier qualities. The treated cells were tested for their carrier and inhibition abilities where, it was proved that a successful ChIP can be carried out after treatment. Carrier cell inhibition provides an opportunity for subsequent analysis via high throughput sequencing, in turn producing more reliable and in depth results. This protocol could aid in the identification of new drug targets during pharmaceutical drug development.

1. Introduction

1.1 Epigenetics

1.1.1 The Nucleosome

The human genome project launched in 1990, 13 years later researchers had sequenced the 3 billion base pairs in all 23 diploid chromosomes. These 3 billion base pairs are found within all of the 50 trillion cells which form the human body. Each base pair is known to be 0.34 nanometers long, therefore each cell contains about 2 meters of ravelled DNA. With 50 trillion cells each encompassing around 2 meters of DNA, the human DNA component is long enough to stretch to the Sun and back more than 300 times. An intelligent packaging system, known as the nucleosome, was first imaged by Klug et al. in 1980, it comprises 2 copies each of the core histone proteins; H2A, H2B, H3 and H4. Together, these proteins constitute the histone octamer by which 147 base pairs of DNA is wrapped (Lugar et al. 1997). DNA is wrapped 1 ¾ times around the octamer in a left-handed turn, kept in place by histone H1, also known as the 'linker' histone. The idea of a 'linker' histone acting as the locking complex between the core octamer and DNA was recognized initially by Varshavshy et al. (1976). To add to the complexity of the packaging system, an arginine and lysine rich tail is located at the N-terminus of the core histones. While these tails may not largely impact the nucleosome structure they play a significant role in signalling, this role is interchangeable depending on the type and location of modification.

Not only is the nucleosome considered a packaging molecule accommodating 100 trillion meters of DNA within the human body, it also bares the capacity to covalently bind with chemical groups, in turn resulting in reversible changes to gene expression. By supporting or suppressing chromatin compaction the nucleosome can control promoter and enhancer access and moderate regulator binding. In

6

addition, the nucleosome can be used as a regulator with the accumulation of certain proteins which can also dictate the fate of regulator binding and gene expression. Histone modifications are molecular changes to the histone, the most notable being methylation, acetylation and phosphorylation. The 'epigenetic code' is a continuation of post translation modifications through mitosis to daughter cells (Turner 2000). This epigenetic code has many layers, initially the modification is put into place via enzymes, most of which are specific to a modification, i.e. KATs are enzymes which target lysine residues to acetylate, the acetylation domain is known as a 'writer'. Since these modifications are dynamic, 'erasers' exist which remove this modification, i.e. HDACs remove the acetyl group on an acetylated histone residue. And finally, the readers, are domains on the effectors which provide an accessible area to accommodate this modified residue, i.e. the methylation of lysine 3 on histone 9 provides a surface area specific for the binding of the protein HP1 (Yun *et al.* 2011)

1.1.2 Acetylation

Histones have been thought to play a significant role in signalling since before the nucleosome was first imaged (Klug et al. 1980). Pogo *et al*. (1966) recognised the positive correlation between histone acetylation and RNA synthesis by using phytohemaglutinin (PHA), a protein isolated from the red kidney bean. They concluded that acetylation alters the fine structure of chromatin, therefore loosening the grip of the negatively charged DNA allowing for more fluent binding of transcription factors (Jiang *et al.* 2011). Di-acetylation is the double acetylation of a residue, the addition of each acetyl group will incur a resulting increase in negative charge on the nucleosome structure making the DNA more accessible.

Histone acetylation is a dynamic and reversible reaction which can be catalysed by enzymes known as histone acetyltransferases (HATs). There are 2 main HAT groups, GNAT and MYST. The MYST family of proteins are highly conserved throughout human and yeast entities, they comprise of 5 main proteins: Moz, Qkf, Mof, TIP60 and Hbo (Thomas and Voss 2007, Zhang *et al.* 2014). The GNAT enzyme family is a huge group of enzymes named after the founding GCN5 enzyme (GCN5-related N-acetyltransferase), over 10,000 have been identified throughout all kingdoms (Vetting *et al*. 2004).

Histone acetylation can also be reversed by histone deacetylases (HDACs), the deacetylation of histones is just as important for gene regulation as acetylation. There are currently 18 known HDACs in mammalian cells which have been grouped in to class I, II, III and IV, characterised this way because of their similarities to various yeast proteins (summarised in Delcuve *et al.* 2012). Class I, II and IV are all part of the classical HDAC family and are zinc (Zn^{2+}) dependant. Class III however, consists of seven Sir2 (sirtuin) proteins with NAD^+ dependant reactions. HATs and HDACs work in partnership to control the levels of histone acetylation in the cell nucleus.

The second line of regulation is in what is known to be the 'readers', they are domains within proteins which specifically identify and bind to modified residues. Although the HAT and HDAC enzymes catalyse the acetyl transfer reactions, another domain can be found in most HATs which recognises the residue which needs to be acetylated, thereby ensuring a secure acetylation pattern.

Generally, but not exclusively, acetylation targets lysine residues, however Tweedie-Cullen *et al*. (2012) identified 58 never-before-described modifications including acetylation and di-acetylation of serine's and threonine's. In a publication entitled

"Histone acetylation and epigenetic code" (Turner 2000) lysine acetylation was named as being the most common, well understood and significant amongst all the post translational modifications.

1.1.3 Methylation
Histone methylation was first described 40 years ago by Murray (1964) and, along with acetylation, first identified as a post translation modification by Allfrey *et al.* later in 1964. Since its discovery vast research has elucidated numerous biological functions, mechanisms and molecular interactions by which histone methylation can occur. Histone methylation, unlike acetylation, does not have a direct effect on chromatin structure, i.e. DNA-histone affinity is left unchanged because methyl groups are neutrally charged. Another difference between methylation and acetylation is that lysine residues are capable of being methylated multiple times, in a mono-, di- and tri-methylated state. The main two residues affected by methylation are lysine and arginine; histidine methylation also occurs but is less common and not as well understood. Methylation is a broadly studied histone modification, despite this there is no definitive outcome which can summarise its effects. For example, tri-methylation of lysine 27 on histone 3 (H3K27me3) is associated with gene silencing, H3K4me3 however, is known to be linked with active transcription (Cao *et al.* 2002, Bernstein *et al.* 2005).

Methylation has two lines of regulation: firstly, the histone tail residue must be methylated. There are numerous enzymes involved in the methylation and demethylation of both arginine and lysine, each with a different purpose. Methyltransferases are enzymes which catalyse the relocation of one methyl group from the donor, usually S-adenosyl-L-methionine (SAM), to a substrate. There are currently thought to be 208 proteins with the potential to act as a catalyst for this

reaction (Petrossian and Clarke 2011). KMTs or lysine methyltransferases are lysine specific transfer enzymes, most KMTs carry a SET domain which is the fundamental factor when signalling to the lysine on the histone tail (Dillon *et al.* 2005). 8 families of KMT have been identified, all except one contain a SET domain (Feng *et al.* 2002).

Until the identification of histone demethylases by Shi *et al.* (2004), histone methylation was considered to be a permanent post-translational modification where, in order to remove the modification the histone tails were cut or the entire histone was replaced. Shi *et al.* (2004) identified the first demethylase known as Lysine-specific Demethylase 1 (KDM1), a flavin adenine dinucleotide (FAD)-dependant amine oxide reaction (Cloos *et al.* 2008). However, this enzyme family require a protonated methyl ε-ammonium group to catalyse this reaction, since trimethylated lysines do not hold this additional hydrogen atom this family is only suitable for the demethylation of mono- and di- methylated lysines. Further research discovered a subfamily of enzymes from 2-oxoglutarate (2OG) enzymes are capable of demethylating trimethylated histone residues via a hydroxylation reaction. This subgroup of the 2OG enzymes contains a JmjC catalytic domain and produces the demethylation action with Fe (II) and α-ketoglutarate (αKG) as cofactors (Klose *et al*. 2006). The third demethylation family, Petidylarginine deiminase 4, PADI4, doesn't demethylate arginine in the typical manor, but deimminates the methyl-arginine to citrulline thereby removing the methyl group (Cuthbert *et al.* 2004).

The second line of regulation are the recognition domains, or 'readers' found within the methyl transfer enzymes. The outcomes associated with histone methylation are pre-determined according the domain associated with that region. Once the proteins are bound they produce a signal calling for enhancers, promoters and other

effectors, thereby regulating transcription. The domains required for lysine or arginine interaction differ, methylated lysine residues are recognised by with Chromo-, Plant Homeo-, Malignant Brain Tumour- (MBT) and Tudor-domains. Similarly the domain family associated with arginine binding are the bromodomains, whereas phosphorylationis recognised by the 14-3-3 containing proteins (Izzo and Scneider 2011).

1.2 Examples of Well Studied Histone Modifications

Many histone modifications have been identified throughout the genome, each having a resulting effect on gene regulation and expression. In brief, the methylation of lysine 9 on histone 3 (H3K9me) is a heterochromatin mark (Bannister *et al*. 2001). In contrast to H3K9me and H3K9me3, acetylation of lysine 9-histone 3 (H3K9ac) and lysine 4 (H3K4ac) will provide a hallmark note of gene activation (Tjeertes *et al.* 2009). The methylation of lysine 4 on histone 3 (H3K4me) is uniformly known as an active chromatin marker (Flanagan *et al.* 2005). Acetylation of lysine 56 on histone 3 (H3K56ac) has also been suggested to be required for repair of DNA breaks during replication (Yuan *et al.* 2009).

Heterochromatin Protein 1 (HP1) selectively binds to H3K9me, it is the protein responsible for mediating gene silencing at heterochromatin (Fischle *et al.* 2005). The chromo domain located in HP1 recognises the methyl group on K9 as a 'marker' for binding. Methylation is catalysed by the SET domain within the Histone-lysine N-methyltransferase, SUV39H1 (Aagaard *et al.* 2000). Once SUV39H1 methylates the residue it signals for HP1 to bind where it is used as a transcriptional repressor.

<u>1.2.1 Histone Crosstalk</u>
Histone crosstalk is a mechanism by which histone modifications are enzymatically put into place and in turn have impact upon other potential modifications. An example of how this can occur is by substrate competition, catalyst inhibition, and by interfering with the space around it (Turner 2000). Histone crosstalk plays a fundamental part in gene regulation for example, cross talk occurs between H3S10ph and H3K9me1/2/3 during M phase. As previously mentioned, H3K9me1/2/3 are all uniform marks of heterochromatin, during M phase of the cell cycle certain heterochromatin marks have to be removed, therefore displacing the HP1 protein. Since the protein is already bound to the methyl marker a reaction must occur in order to dissociate this complex. Phosphorylation of Serine 10 on histone 3 (H3S10ph) is the mechanism by which HP1 can be removed. The kinase enzyme Aurora B is activated by factors within the chromosomal passenger complex (CPC), once active Aurora B has the catalytic capacity to phosphorylate H3S10. A study by Fischle *et al.* (2005) identified that phosphorylation of H3S10 inhibited the binding capacity of HP1 to H3K9me3. By mutating 3 aromatic amino acids in the chromodomain of HP1 Fischle *et al.* produced evidence that H3S10ph is the sole inhibitor of the HP1 chromodomain-H3K9me3 complex. Furthermore, they also found that H3S10 is not present during interphase of the cell cycle explaining how the HP1-H3K9me3 complex is restored during this time. By initiating H3S10ph, chromatin can keep its form and such modification can be passed on to daughter cells, if the mechanism were to merely remove the methylation from site it would mean a high risk methylation reformation when the cell comes through the other side of M phase, therefore leaving chromatin open for replication errors, and consequently, detrimental effects upon the organism.

1.3 Chromatin Immunopreciptation

In 1978 Vaughn Jackson (Jackson 1978) published a paper identifying what would become a ground breaking technique in molecular biology. Chromatin immunoprecipitation is a comparative technique using antibodies to bind to proteins on specific locations in the genome, it assesses where in the genome histone modifications occur and where non-histone proteins bind. The chromatin immunoprecipitation (ChIP) protocol has since been used for the mapping of numerous post-translational modifications. Modification and manipulation of the ChIP protocol have enabled more effective nucleosome digestion, use with small cell numbers and with or without protein cross-linking (Jackson 1978, Hebbes 1988, Gilmour and Lis 1985, O'Neil *et al*. 2006); it has also been used in conjunction with other techniques, e.g. Microarray (ChIP-chip); this partnership creates a more global view of DNA- protein binding (Lee *et al.* 2006).

Firstly, chromatin must be extracted, there are various ways in which this can be achieved. The native ChIP (NChIP) chromatin preparation involves the breakdown of cells via homogenisation and nucleosome digestion with micrococcal nucleases (O'Neil and Turner 2003). X-ChIP chromatin preparation is a method using formaldehyde as a cross-linking agent prior to nucleosome fragmentation via sonication (Jackson 1978); cross-linking can also be achieved by exposure to a UV source (Gilmour and Lis 1985). Other successful methods are worth mentioning to demonstrate the versatility of ChIP, Bauer *et al.* (2002) used a combination of formaldehyde cross linking followed by micrococcal nuclease digestion. Secondly, by binding antibodies to proteins found on the genome it is possible to isolate protein-bound DNA regions. Once again, there are various methods for immunoprecipitation depending on the chromatin preparation protocol used. Beads are bound to antibodies, introduction of the antibody-bound-beads to the chromatin gives the

13

opportunity for the antibodies to bind to their complimentary modifications. The output chromatin, or 'pull-down', can be taken on for further analysis with q-PCR, microarray and high throughput sequencing.
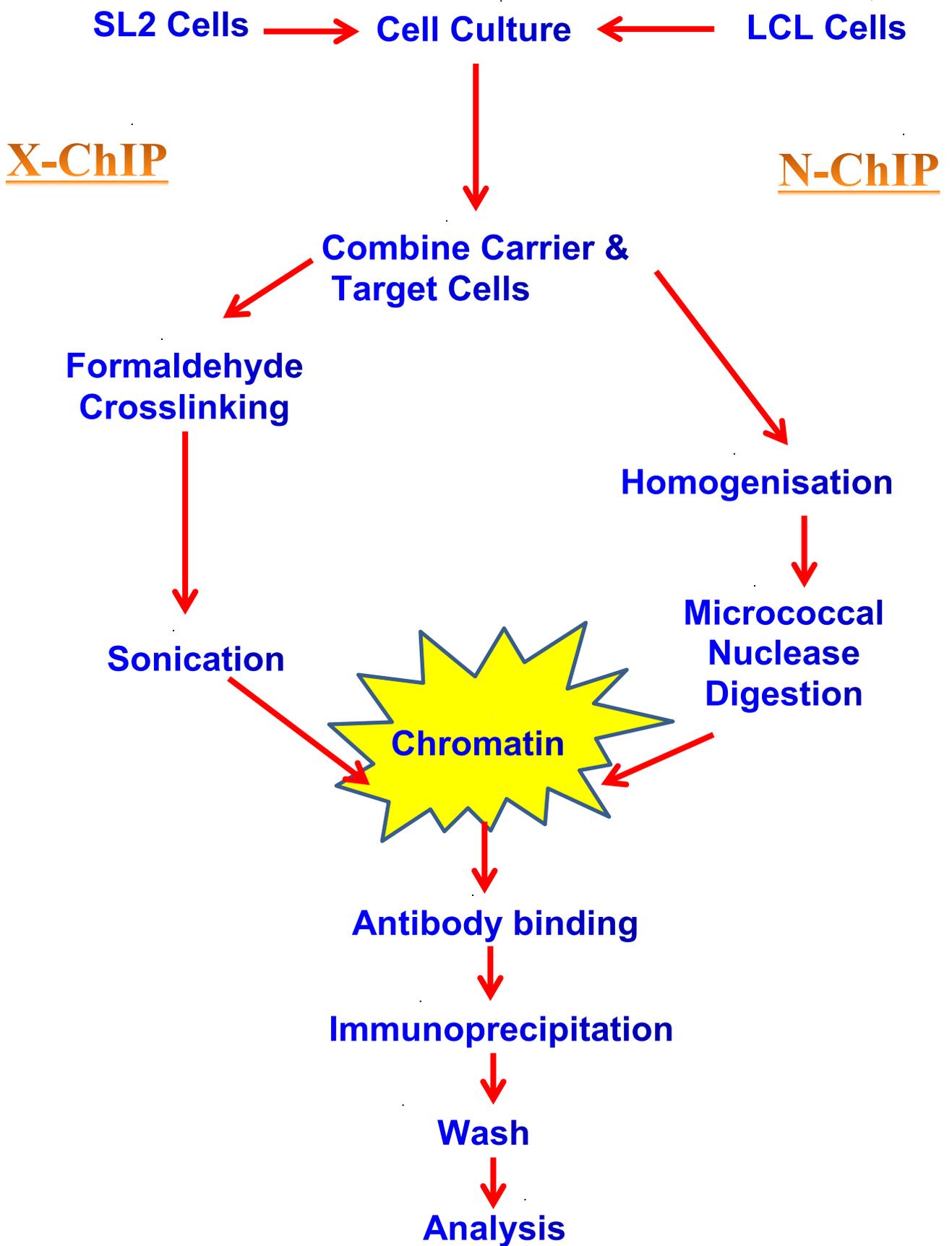
Figure 1.1 A figure to illustrate the main steps involved in carrier-ChIP for both native and formaldehyde crosslinked chromatin preparation

1.3.1 Native Chromatin Immunoprecipitation (N-ChIP)

Unlike X-ChIP, the native chromatin immunoprecipitation (N-ChIP) technique as first published by Hebbes *et al.* (1988) does not require any cross-linking. It is therefore only effective on histone proteins or other proteins that bind DNA with extremely high affinity. Nucleosome digestion or fragmentation is achieved using the enzyme micrococcal nuclease (MNase), polynucleosomal strands become mono-, di- and tri-nucleosomal segments or variably (yet reproducibly) bigger (concentration and duration dependant). The nuclease family are a group of enzymes which cut the phosphodeister bonds between nucleotides in a DNA or RNA sequence. MNase specifically is a ribo-, deribo-, endo-, exonuclease enzyme isolated from *Staphylococcus aureus*, which cuts chromatin at the linker strands between nucleosomes (Heins *et al.* 1966).

N-ChIP holds many advantages in identifying histone modifications compared to other ChIP techniques. N-ChIP provides a higher antibody specificity as micrococcal nuclease digestion is not as harsh as sonication (the fragmentation method used in X-ChIP) and does not chemically modify the histone tail as fixation does. Enzymatic fragmentation will leave the histone tails intact allowing for more efficient antibody binding, whereas when sonication is used to fragment the nucleosome, histone tails are damaged leading to fewer tails to bind to and in turn, decreased pulldown efficiency. In addition, DNA and protein output can be analysed leading to a more efficient precipitation; moreover, output DNA can be analysed without the use of PCR amplification (O'Neil and Turner 2003). However it has a major disadvantage, due to the nature of the protocol it is not possible to analyse non-histone proteins because non-histone proteins do not exert the same affinity to the post-translation modification as histones, therefore another technique is needed for these non-histone proteins. Additionally, the nucleosomes have the ability to rearrange during

nuclease digestion which can cause detrimental effects on further analysis. Furthermore, nuclease digestion is known to be somewhat selective potentially causing bias in the input chromatin.

<u>1.3.2 Cross-linked Chromatin Immunoprecipitation</u>
Cross-linked chromatin immunoprecipitation (X-ChIP) was the first method of ChIP to be developed, Jackson (1978) initially identified the use of formaldehyde as the optimal cross-linking agent. Due to the small size (just four atoms) and simple molecular structure it is easy for the formaldehyde (CH_2O, otherwise known as methanal) to infiltrate the cell, following absorption covalent bonds are formed between proteins, DNA, RNA etc (Sunderland *et al.,* 2008). The small size of the formaldehyde molecule means that only residues that are in close proximity to one another can be cross-linked, hence why it is the chosen chemical for protein-protein interactions, fig. 1.1. This theory has been utilised and broadly studied in order to bring about a more refined X-ChIP protocol. Cross-linking can also be achieved by using UV radiation, Gilmour and Lis (1985) used this technology as an alternative to the carcinogen formaldehyde, their results showed a cross-linking productivity of one protein molecule per 60kb. Following this, nucleosome fragmentation is achieved using sonication, a process which uses high energy sound waves to break down the nucleosomes into smaller sequences. However, sonication is not a reproducible way of obtaining fragmented chromatin, due to the erratic and intermittent use of sounds waves chromatin is sheered at unpredictable locations.

It is commonly known that there are a number of transcription regulatory factors (TRFs) that work in combinations to regulate gene expression and regulation. Miura *et al.* (2009) identified a set of transcription regulatory factors (TRFs) and various DNA regions which could be involved in combinatorial regulation. More specifically,

17

with the use of X-ChIP they were able to locate the regulatory region downstream from the CEBPA gene and confirm its functionality. These studies show that X-ChIP by formaldehyde, UV and other cross-linking methods is a largely versatile and reliable technique; particularly the use of non-histone proteins and those who's binding affinity to DNA is not effective enough for the use of N-ChIP. In organisms where native chromatin preparation is not easy, X-ChIP is the ideal technique due to the complexity of native chromatin preparation. X-ChIP can also be advantageous during instances where there is not enough cells available for N-ChIP, as the number of cells required for X-ChIP is generally less ($5 \times 10^7 - 1 \times 10^8$ cells/ N-ChIP vs $>1 \times 10^7$ cells/ X-ChIP) (O'Neil and Turner 2003, Abcam 2011). Moreover, the use of cross-linking reduces the likelihood of nucleosome rearrangement during sonication compared to N-ChIP. Never-the-less, cross-linking still has its limitations, the most prominent being the irregular and unreproducible way in which the nucleosomes are fragmented during sonication; cross-linking may inadvertently fix temporary interactions therefore not giving a true image of permanent modifications. Finally, despite the reduced number of cells required for X-ChIP compared to N-ChIP, it is still not possible to perform X-ChIP with small cell numbers therefore ruling out clinical application.

1.3.3 Carrier Chromatin Immunoprecipitation (C-ChIP)

Carrier ChIP is a development of the N-ChIP protocol to allow for the use of small cell numbers. As identified in the X-ChIP and N-ChIP sections, a major disadvantage of both techniques is that they do not have a clinical application due to the high number of cells required to start the experiment. When you consider that a typical mouse embryo's inner cell mass (ICM) consists of ~20 cells (O'Neill *et al.* 2006) it is possible to understand that obtaining the required 10^7-10^8 cells is unrealistic. By

using cultured SL2 cells from the *Drosophila melanogaster* (fruit fly) and combining them with as little as 100 mammalian cells, the SL2 cells will protect against the loss of the mammalian cells through chromatin preparation and act as a carrier throughout.

O'Neill *et al*. (2006) established this idea of using cultured *D.melanogaster* cells alongside mammalian cells in order to determine epigenetic changes in the early embryo. They used this carrier technology to identify and confirm the existence of modificated genes in the inner cell mass and embryonic stem cells, a project that until the introduction of C-ChIP would not have been possible.

Carrier ChIP was able to revolutionise the way in which we can look at histone modifications, other techniques have been recorded as identified in table 1.1. Moreover, researchers are now able to judge the validity of their research which has relied solely on cultured cells, made possible by using carrier ChIP as a comparative analysis to provide a realistic picture from non-cultured cells. Despite the wide application of this technique to many experiments it leaves one fundamental weakness. Due to the presence of a large excess of the *D.melanogaster* cells it is not possible to complete the analysis with high throughput sequencing. Analysis of C-ChIP chromatin has thus far been limited to PCR analysis and the global distribution of the modifications of samples prepared by C-ChIP is not possible.

1.3.4 Alternative Techniques for Small Cell Number ChIP

Other ChIP techniques have been published which are specific to the sequencing analysis of small cell number samples. Nano-ChIP-Seq uses the same concept as X-ChIP by initially crosslinking the modifications with formaldehyde, lysing and subsequently sonicating to sheer the nucleosomes. IP is achieved using sepharose A beads and subsequent incubation with the chosen antibody isolates the target

chromatin. Losses during the wash stage are reduced by binding the bead/antibody complex to columns. After IP libraries are prepared in three major steps, firstly, IP DNA is primed with custom primers, containing a hair pin loop, restriction site (BsiVI) and random nonamer sequence for DNA binding, step two involves the amplification of the target DNA plus the added restriction site. BciVI, when cut, leaves a 3' adenosine overhang which allows adaptor binding during the final step of library preparation followed by quantification by RT-PCR. Although Aldi and Bernstein (2011) were able to reduce the number of cells needed for ChIP to as little as 10,000 this method still has some limitations. Firstly, the method described has used formaldehyde crosslinking and sonication to isolate DNA, similar to that of X-ChIP, it's adaptation to NChIP is yet to be proven. In addition, this protocol can only be utilised for 10,000 cells when using embryonic stem cells, when using other cell types including FACS sorted cells 25,000 cell or more were needed. Indeed, the authors quote only a two to three orders of magnitude improvement in terms of the numbers of cells required compared to standard ChiP methodology. Furthermore, the time taken to perform the whole ChIP protocol including library preparation is upwards of five days. Thus, although at the time of publication this method was a breakthrough in the field, it can be concluded that a better, more efficient protocol could be adapted to eliminate many of these limitations.

iChIP (index-first ChIP) is a recently developed ChIP technique used for analysis of chromatin samples from small cell numbers by Lara-Astiaso *et al.* (2014). In iChIP, the chromatin is isolated by chemical fixation before using a carefully selected number of cycles for sonication. Using magnetic beads coated in H3 antibody, the chromatin is immobilised allowing for barcoding of individual samples which are then pooled together for IP. Sequencing analysis of the pooled samples is made possible

as the barcodes are unique to each chromatin sample and can therefore be separated upon analysis. This is a very new technique which is yet to be fully explored by other researchers. The authors of iChIP have gone into great depth on how to optimise various steps to ensure the integrity of the resulting DNA without the use of a carrier. The analysis of such small cell numbers without the use of a carrier increase loss of DNA to tubes and pipette tips. Using a carrier during chromatin preparation significantly reduces these losses, furthermore carrier cells are able to increase the reliability and reproducibility of each assay as the amount of chromatin in each reaction is careful controlled.

The number of cells recovered from a tumour biopsy can vary depending on the biopsy procedure used to recover the sample, type and location of tumour (Rajer and Kmet 2005). Although the number of cells retrieved from biopsies in some cases can be relatively useable, after the cells have been separated into their relevant groups and proportions have been sent for diagnostic purposes, the number of cells left to use for research purposes is limited. Although the genetic sequence is mostly a set sequence (with the exception of single nucleotide polymorphisms, SNPs, and artefacts), epigenetics and PTMs of histones provide a code which regulate factors on top of the genome, i.e. transcription factor binding sites. The same is true for non-histone proteins where PTMs can regulate binding capacity and modify their activation motifs etc. To provide an insight into the role of post-translational modifications on the mechanisms of tumour growth and survival, including cell apoptosis and proliferation, it is important to develop a reliable technique which identifies the presence of PTMs. Current techniques, mostly involving chromatin immunoprecipitation, hold many limitations and fundamental weaknesses which

raise issues with analysis. Since the number of cells required to initiate the first steps in ChIP are too large to enable use with primary clinical samples, current research is relying on the generation of new cell lines by cell culture which, by nature, has limitations growing exponentially as the passage line number increases. Analysis of primary cell PTMs was made possible by the introduction of a carrier, increasing the reliability of results and expanding the applications of ChIP to tumour cells after FACS separation, but introduces new barriers in analysis by high throughput sequencing. This project aimed to provide a new, versatile analysis for the identification of PTMs for both histone and non-histone proteins by addressing the limitations of the current techniques, thereby expanding the depth of current research in PTMs to primary cells from real, clinical samples.

Table 1.1 A comparison of the techniques available to identify and compare post-translational modifications of proteins outlining the advantages, limitations and sources for each. Table 1.1 also includes a comparision methods used for ChIP analysis.

Method	Advantages	Disadvantages	Protocol/Source
Mass Spectrometry (MS) Botton Up	o No prior knowledge of the modification site or type is needed.	o Required expensive software and use of complicated algorithms. o Success rate varies. o Short peptides are incompatible with MS o Trypsine digests K and R, histones are K and R rich therefore creates short peptides. o Trypsine cleaves mono-, but not di- and tri- methylated lysines.	Arnaudo and Garcia 2013
Mass Sepctrometry (MS) Middle Up	o Use with PTMap software increases reliability and reduces false positives.	o Misidentification of modifications due to: isobaric mass shift, sample preparation, incorrect database assignments and false positives. o Can cause methyltransfer within a peptide incurring false positives. o Loss of modification during sample processing. o Histone phosphorylation is lost during sample. processesing due to being acid labile. o Heavy isotope labelling required for reassurance of correct identification.	Arnaudo and Garcia 2013
µ-ChIP (Micro-ChIP)	o Uses samples with as little as 10,000 cells. o No carrier interference.	o Requires an amplification step. o Expensive equipment and kits. o Not enough sample available to use one	Acevedo *et al*. 2007

	o Can be used with microarray.	as an 'input'. o Reduced signal from less cells leads to underestimation of protein-DNA interaction.	
Fast ChIP	o Quicker due to shortened antibody binding times. o Can use multiple antibodies to simultaneously process many samples.	o Required Chelex-100 to facilitate DNA extraction. o Only suitable for large cell samples.	Nelson *et al*. 2006
Quick and Quantitative ChIP(Q^{2}ChIP)	o Faster protocol by reducing timing.	o Formaldehyde can cause disruptions to modifications creating inefficient epitope precipitation. o Only suitable for samples of a large cell number	Dahl and Collas 2007
Matrix ChIP	o Several samples in one plate. o Can be automated. o Quicker due to shortening of steps.	o Requires Chlenex-100 to facilitate DNA extraction. o Volume of cells required.	Flanagin *et al*. 2008
Native ChIP (N-ChIP)	o No crosslinking required. o Nucleosomes are cut at reproducible lengths between the nucleosomes. o High antibody	o Not suitable for non-histone proteins. o Nucelosome rearrangement . o Nuclease digestion can be selective. o Volume of cells required.	O'Neil and Turner 2003 Hebbes *et al*. 1988

	o specificity. o DNA and protein output can be analysed, therefore more efficient precipitation. o Analysis can be achieved without PCR.		
Formaldehyde Crosslinked ChIP	o Applicable to non-histone proteins. o Good for organisms where chromatin is difficult to prepare.	o Histone tails can be damaged during sonication causing issues with antibody binding. o Fragmentation by sonication is not reproducible. o Volume of cells required. o Formaldehyde can cause disruptions to modifications creating inefficient epitope precipitation.	Jackson 1978
UV Crosslinked ChIP	o Applicable to non-histone proteins. Good for organisms where chromatin is difficult to prepare.	o Histone tails can be damaged during sonication causing issues with antibody binding. o Fragmentation by sonication is not reproducible. o Volume of cells required.	Gilmour and Lis 1985
Carrier ChIP (C-ChIP)	o Can be used with samples of a small cell number. o Applicable with both crosslinked and native chromatin therefore,	o Carrier cells cause background 'noise' during high throughput sequencing.	O'Neil *et al*. 2006

	histone and non-histone proteins.		
ChIP-chip (ChIP with microarray analysis)	o Provides a whole genome analysis. o Broad applications.	o High quantity whole genome micro-array platforms are expensive. o Experimental process is long. o Challenging to troubleshoot. o Requires the use of replicate samples to account for experimental 'noise'. o Use of algorithms.	Lee *et al*. 2006
ChIP by sequencing analysis (ChIP-Seq)	o Higher resolution than ChIP-chip. o Fewer artefacts. o Large coverage of analysis.	o Analyses all input sample so not to be used as analysis for C-ChIP. o N-ChIP and X-ChIP analysis require large cell numbers. o Doesn't view whole genome.	Park 2009
Nano-ChIP-Seq	o Can analyse as few as 10,000 cells. o High throughput sequencing is possible.	o Duration can be up to 5 days o 10,000 is only applicable to ESC, other cell types require higher quantities.	Adli and Bernstein 2011
iChIP	o Can reduce the number of cells required to just 500 cells. o Can be analysed with high throughput sequencing.	o Requires the use of a specialised barcoding system. o Cells can cause detrimental effects to the quality of the chromatin.	Lara-Astiaso *et al*. 2014

1.4 Aims and Objectives

1.4.1 Primary Aims
The main focus of this experiment was to identify a novel way of coupling carrier chromatin immunoprecipitation (C-ChIP), a procedure for immunoprecipitating chromatin from small numbers of cells, to next generations sequencing (C-ChIP-seq). This requires the development of techniques that will suppress the amplification of carrier DNA when constructing sequencing libraries.

1.4.2 Objectives
In order to achieve this aim 3 objectives were identified.

1. To develop a protocol by which the DNA of carrier cells can be fragmented by UV light so as to supress its amplification, while still allowing the preparation of chromatin suitable for ChIP. Ensuring that the number of output cells (number of cells recovered after treatment) is not reduced so much that it hinders further experimental

2. Identify parameters for the use of crosslinked ChIP with carrier cells (a technique not previously attempted).

3. Perform chromatin immunoprecipitation assay on both crosslinked and native chromatin assessing the rate of inhibition induced by UV treated cells before and after the creation of DNA libraries.

2. Materials and Methods

2.1 Materials

Primer sets used (See appendix for position within the sequence)

Name	Sequence	Company
DM GapDH Forward	AGCGAACTGAAACTGAACGA	Invitrogen
DM GapDH R142	ACGGAGGCGCCCTTAT	Invitrogen
DM GapDH R221	GAAACGACCGTGAGTCGAG	Invitrogen
DM GapDH R319	CTGGCCCAGTTGATGTTG	Invitrogen
DM GapDH R419	CGAGATGATGACCTTCTTGG	Invitrogen
HS CD74 end F	TGCTGCTTCTCTCTCCAGTC	Invitrogen
HS CD74 end R	TGTCCAAGGGTGACGAAAGA	Invitrogen
HS CD74 -500F	TGCCCCTTCTCTCTGTCTTG	Invitrogen
HS CD74 -500R	GCACTAGAGGCCTCAGTTGA	Invitrogen
HS CD74 -1000F	CTTGCCAACACAGTCCCATC	Invitrogen
HS CD74 -1000R	GTGATCTTTACGGAGGCCCT	Invitrogen
Chromosome 18 desert F	ACTCCCCTTTCATGCTTCTG	Invitrogen
Chromosome 18 desert R	AGGTCCCAGGACATATCCATT	Invitrogen
HS CD74 F103	AGTGGGCGGAGTGGCCTTCT	Invitrogen
HS CD74 R172	CCCGGCTCGCCTCTTAAAGTCGGTGCT	Invitrogen
HS CD74 R221	CTTCCCGACAGCTCCTGCTTCTCCTCC	Invitrogen
HS CD74 R344	TCGGTGTGCCCTTTCCTGGTGCTCA	Invitrogen
HS CD74 R492	GAGAGGGAGCCGGAGCCCAGAGCA	Invitrogen
DM = *Drosophila Melanogaster* HS = *Homo Sapien*		

Buffers used

Name	Components	Concentration	Experiment(s)
1 x TAE	Tris Acetic Acid EDTA	40 mM 20 mM 1 mM	Agarose gel electrophoresis
0.1 % TE	Tris EDTA	1 mM 0.1 mM	
Buffer NB	Tris-HCl $MgCl_2$ NaCl KCl EGTA 2-mercaptoethanol	15 mM 5 mM 15 mM 60 mM 0.1mM 0.5 mM	N-ChIP

	Na Butyrate	5 mM	
	PMSF	0.1 mM	
Digestion buffer	Sucrose	0.32 M	N-ChIP
	Tris-HCl	50 mM	
	MgCl$_2$	4 mM	
	CaCl$_2$	1 mM	
	Na Butyrate	5 mM	
	PMSF	0.1 mM	
Lysis buffer (N-ChIP)	Tris-HCl	1 mM	N-ChIP
	EDTA	0.2 mM	
	Na Butyrate	5 mM	
	PMSF	0.2 mM	
Incubation buffer	NaCl	50 mM	N-ChIP
	Tris-HCl	20 mM	
	Na Butyrate	20 mM	
	EDTA	5 mM	
	PMSF	0.1 mM	
Wash Buffer A	Tric-HCl	50 mM	N-ChIP
	EDTA	10 mM	
	Na Butyrate	5 mM	
	NaCl	50 mM	
	PMSF	0.1 mM	
Wash Buffer B	Tric-HCl	50 mM	N-ChIP
	EDTA	10 mM	
	Na Butyrate	5 mM	
	NaCl	100 mM	
	PMSF	0.1 mM	
Wash Buffer C	Tric-HCl	50 mM	N-ChIP
	EDTA	10 mM	
	Na Butyrate	5 mM	
	NaCl	150 mM	
	PMSF	0.1 mM	
Dilution buffer	Triton X-100	1 %	X-ChIP
	EDTA	2 mM	
	NaCl	150 mM	
	Tris-HCl, pH 8	20 mM	
	Protease inhibitor cocktails		
0.1 M Citrate-phosphate buffer, pH 5	Citric Acid		X-ChIP
	Na$_2$HPO$_4$		
Low salt buffer	Triton X-100	1 %	X-ChIP

	SDS	0.1 %	
	EDTA	2 mM	
	NaCl	150 mM	
	Tris-HCl, pH 8	20 mM	
High salt buffer	Triton X-100	1 %	X-ChIP
	SDS	0.1 %	
	EDTA	2 mM	
	NaCl	500 mM	
	Tris-HCl	20 mM	
Elution buffer	NaHCO3	100 mM	X-ChIP
	SDS	1 %	

2.2.1 Cell Culture

Drosophila melanogaster SL2 cells (Schneider 1972) were cultured at 24 °C, 5 %

CO_2 and routinely split twice weekly in a 1:10 ratio with growth media [500 ml Insect-

Xpress + 5 % Fetal Bovine Serum (FBS) + 1 % Penicillin Streptomycin (Pen Strep)]

until required when the cells were split 1:3 and allowed to grow overnight. To

dissociate weakly adherent cells, flask was 'banged'. Once dislodged the cells were

washed with PBS (phosphate buffered saline) and resuspended in buffer NB to be

counted using a haemocytometer. Cells were resuspended at 2 x 10^7 cells ml^{-1}.

Frozen plates were made by pouring water into the contours of the plate between

each of the wells on both sides and leaving to freeze overnight. Cells to the

concentration of ~2 x 10^7 cells/ ml at a volume of 500 µl were transferred into various

wells on frozen and non-frozen 12-well plates ready for ultra violet treatment. The

plates were sealed with optical film and kept cool by using frozen plates, storing in

ice during movement and placing an ice block on one side during UV treatment. The

cells were treated by placing on a UV transilluminator for 30 minutes; 10 minutes in

UV conditions, 5 minute submerged in ice followed by 10 minutes in treatment and a

final 5 minutes submerged in ice. The cells were counted again to assess the level of

UV-induced damage.

2.2.2 DNA extraction

DNA was extracted according to the *QIAgen* DNA extraction from cultured cells

protocol.

The cultured cells were pelleted by centrifuging at 300 x g for 5 minutes. Proteins

were digested and removed by resuspending the pellet in 200 µl of PBS and 20 µl

Proteinase K. Cell lysis was achieved using 200 µl Buffer AL, and incubated at 56˚

for 10 minutes. To purify the DNA 200 µl of 100% ethanol was added to the solution,

mixed, transferred into a mini spin column and centrifuged (6,000 x g, 1 minute). Flow through was discarded and washed again with 500 µl of AW1 and AW2 at 6,000 x g for 1 minute and 20,000 x g for 3 minutes, respectively. DNA was eluted by placing the spin column into a new 1.5 ml Eppendorf, pipetting 200 µl of buffer AE directly onto the column membrane and centrifuging at 6,000 x g for 1 minute.

2.2.3 Chromatin immunoprecipitation (ChIP)
Native Chromatin Immunoprecipitation(N-ChIP)

The native chromatin immunoprecipitation (N-ChIP) protocol as described by O'Neil and Turner (2003) was followed. The cells were washed and resuspended in buffer NB to ~2 x 10^7 cells/ ml. An equal volume of 1 % tween in NB buffer was added and cells incubated on ice for 15 minutes with gentle mixing before homogenisation. Homogenisation was carried out using a dounce homogeniser with a tight pestle, in order to break down the cell membrane, leaving the nucleus exposed. Cells were checked under a microscope after every 3-5 strokes to avoid over- homogenisation, in turn creating sub-nucleosomal DNA.

To eliminate debris created by homogenisation the broken cells were centrifuged and washed sequentially (~1 x 10^7/ ml) in 5 % sucrose/buffer NB and digestion buffer. A nanospectrometer was used to assess the ƛ 260/280 values and DNA concentrations (µg/ µl) of the solutions. In preparation for digestion the pellet was resuspended in digestion buffer (0.5 mg/ ml) and divided into 1 ml aliquots. Micrococcal nuclease was added at 50 U/ml and incubated at 37 °C for 5 mins. 0.5 M EDTA was added to the solution (10 µl/ ml) and placed on ice for a further 5 minutes to stop the reaction. After centrifugation the supernatant was retained and kept chilled for further analysis (S1). Finally, the pellet was resuspended in 1 ml lysis buffer, and dialysed overnight against 2 L lysis buffer (4 °C).

Following dialysis, samples were centrifuged and the supernatant retained (S2). The resulting pellet was resuspended into 200 µl lysis buffer (P) and quantified, taking particular note of the ƛ 260/280 values and µg/ µl concentrations. During optimisation of chromatin preparation the S1, S2 and P samples were visualised on a 1.2 % agarose gel in 1 x TAE buffer containing GelRed dye. The purpose of this gel is to check chromatin integrity from the size and abundance of fragmented nucleosomes after micrococcal nuclease digestion. For each sample, 1 µg DNA was loaded in a final volume of 25 µl containing 0.1 % SDS and loading dye. S1 represents the supernatant after micrococcal nuclease digestion and will therefore contain mostly mono nucleosomes, S2 is the supernatant collected after overnight dialysis and therefore will have mostly di-, and tri-nucleosomes whereas P, or pellet, was the soluble portion of the sample resuspended after night dialysis.

N ChIP- Immunoprecipitation

For immunoprecipitation (IP), S1 and S2 were combined and the P sample was discarded. An initial pre-clear reaction was set up to eliminate unspecific binding from both the beads and the chromatin. Per IP, 30 µg – 100 µg of chromatin was used (depending on the amount available after chromatin preparation), 50 – 100 µl Protein A Sepharose beads, 100 µl 10 x incubation buffer, up to 900 µl with u.p. H_2O. After rotation at 4 °C for 1 hour an 'input' sample was retained as an unbound control. The antibody was bound to the chromatin by adding 4 µg to the chromatin solution and incubated overnight at 4 °C. Protein A sepharose beads are used to immunoprecipitate the antibody bound chromatin overnight at room temperature. Chromatin is washed on the beads using gradient salt washes buffer A, buffer B and buffer C and finally eluted using 500 µl of 10 x incubation buffer/SDS solution on rotation.

Crosslinking Chromatin Preparation (X-ChIP)

The cells were pelleted (centrifuge at 300 x g, 5 minutes) and resuspended in PBS (10^{6-7} cells/ ml) at room temperature. The crosslinking was achieved by dropping formaldehyde (1 % v/v) into the solution and leaving to rotate for 10 minutes. To quench formaldehyde activity 1/10 volume of 2 M glycine was added to the solution and centrifuged at 300 x g, 5 minutes, 4 ˚C. Non-target excess compounds were removed by washing in PBS. The supernatant was discarded and the pellet resuspended in lysis buffer at 5 x 10^7 cells/ ml. In clean eppendorfs the cells were sonicated; 30 seconds on, 30 seconds off for 5-30 cycles. The samples were pelleted by centrifuging at 16,000 x g for 10 minutes, 4 ˚C. The resulting pellet was discarded and the supernatant was retained overnight at 4 ᵒC.

X ChIP - Immunoprecipitation

One chromatin sample should be left un-immunoprecipitated (input sample) for further comparison on pull-down and relevant quantification. The other samples were diluted with ½ volume of dilution buffer, using a maximum of 1ml for each immunoprecipitation (IP).

To bind the anti-bodies with the magnetic Dyna-beads, the beads were washed using citrate-phosphate buffer and resuspended in citrate-phospate buffer + 0.5 % Bovine Serum Albumin (BSA), the protein in the BSA blocks non-specific binding of protein to the beads. Antibodies were bound using 4 µg of antibody per 15 µl Dynabeads, the solution was rotated for 2 hours at 4 ᵒC.

Any surplus antibodies were washed away with citrate-phosphate buffer, Dynabead-antibody complex was resuspended in 15 µl citrate-phosphate + 0.5 % BSA per immunoprecipitation. 1 ml of chromatin and the Dynabead-antibody complex were mixed and left to rotate for 4 hours at 4 ᵒC.

The antibody bound chromatin was washed twice each with low and high salt buffer. The chromatin was eluted twice using freshly prepared Elution buffer (50 µl) with 15 minute incubation and rotation at room temperature between each elution.

Using the antibody-bound chromatin and the input sample, the cross-linking was reversed by incubation with 20 mg/ ml of proteinase K at 65 °C, 300rpm overnight.

2.2.4 DNA Purification
DNA was extracted from the chromatin using the QIAgen PCR Purification Kit: Five volumes Buffer PB was added to one volume DNA in a spin column ready for removal via centrifugation. At this point the DNA is bound to the membrane of the spin column, and subsequently washed with Buffer PE to ensure all surplus fragments are removed. Finally, the DNA can be eluted with Buffer EB.

2.2.5 Polymerase Chain Reaction
Polymerase chain reaction (PCR) is a widely used DNA amplification method first developed by Nobel Prize winner Dr Kary Mullis in 1984. The primers used were specifically designed to create 132 bp, 211 bp, 309 bp and 409 bp amplicons (see appendix 1-4). It was anticipated that there would be considerably more DNA amplified in the smaller strands, and less amplification as the target size grows. This decline is due to the increased accumulation of UV-induced damage in larger. In order to significantly reduce amplification of the carrier DNA during library preparation, it was important to induce sufficient UV damage within the average size of a sequencing library. *Drosphila melanogaster* primers were designed to the housekeeping gene Glyceraldehyde 3-phosphate dehydrogenase (GapDH) encoding the enzyme which catalyses the sixth step of glycolysis and therefore highly expressed in nearly all animals (Greer *et al.* 2010). Human primers corresponded to the CD74 gene, a gene which is highly expressed in lymphoblastoid cells and therefore easy to detect. All primer sequences have been mapped within their gene

sequences, see appendix. The optimal PCR settings for these primers were 95°C, 30 seconds for denaturation; 60 °C, 30 seconds for primer annealing; and 72 °C, 30 seconds for elongation. A no template negative control is also produced for comparison. For evaluation, a 2 % agarose gel electrophoresis was run and observed stained with GelRed under a UV light.

Real-time PCR (RT-PCR)/ q-PCR

To be able to run the chromatin on the RT-PCR the chromatin must be purified into DNA. RT-PCR was set up using 10 µl reaction volume: 8.8 ng DNA, 3 pM reverse primer, 3 pM forward primer, 2×10^{-5} M SYBR.

Real Time-PCR (RT-PCR) is a quantitative technique for the detection of DNA molecules using either a probe or SYBR. In this project the SYBR detection technique was utilised. SYBR is a double stranded DNA binding dye such that fluorescence signal at each cycle is directly proportional to DNA concentration. In this experiment 7 point standard curves were constructed from a series of 1:5 (DNA: NF H_2O) standard DNA. A blank reaction series was also created as a negative control for possible contamination.

<u>2.2.6 Library preparation of ChIP samples</u>
DNA libraries were prepared using the New England Biolabs NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina kit according to the manufacturer's instructions.

Library preparation is the process of taking the immunoprecipitated DNA sample and manipulating it ready for injection into the DNA sequencer. Firstly, the DNA 'overhangs' need to be repaired; when the DNA is fragmented it is left with 'overhang' where one strand of the DNA is longer than the other, during the first step of library preparation this 'overhang' is fixed and blunt ends are produced. This is

made possible with the introduction of a repair enzyme buffer and a repair enzyme during an incubation period of 20 °C for 30 minutes. Following this an 'A' tail is added to the newly repaired double stranded DNA which provides a station for the docking of adaptors in the subsequent steps. dATPs are added to the repaired ends with the Klenow fragment (DNA polymerase I extracted from *E.coli*) (Klenow and Henningsen 1970). The DNA, dA-Tailing Reaction Buffer and Klenow fragment are incubated for 30 minutes at 37 °C. Adaptors are added to the 5' end and the 3' end of the A-tailed DNA, The double stranded adaptors contain the primer-binding sites for library amplification and paired end sequencing. DNA ligase is also used to reform the phosphodiester bonds between double stranded DNA. The DNA, ligation reaction buffer, adaptor and T4 DNA ligase are incubated for 15 minutes at 20 °C. After each incubation period the DNA is bound to magnetic Agencourt® AMPure® beads, washed with 80 % ethanol and eluted in 0.1 % TAE.

PCR amplification increases the abundance of the adapted DNA, this ensures there is enough fragmented sample to analyse. Real-time PCR can also be used as well as sequencing to assess the relative abundance of the different strands as a comparative study.

3. Results

The aim of this project was to establish whether ultraviolet damaged cells could be used as a viable carrier in order to create a modified version of the carrier - chromatin immunoprecipitation protocol to be used in conjuction with high throughput DNA sequencing (ChIP-Seq). To be successful, UV damage must inhibit PCR amplification of carrier DNA sufficiently enough that during PCR amplification of the ChIP-seq libraries target DNA is amplified and carrier DNA suppressed to such an extent as to overcome the initial excess of carrier chromatin. At the same time, UV damage must leave carrier cells and chromatin intact sufficiently to act as a viable carrier through all steps of the ChIP procedure.

During optimisation two experimental methods were used to assess the impact of UV treatment upon the cells and their DNA. To view the inflicted damage upon DNA the cells were taken through a chromatin preparation protocol after UV treatment. An agarose gel electrophoresis was performed after micrococcal nuclease digestion to view the productivity of the enzymatic reaction and to assess the level of UV damage
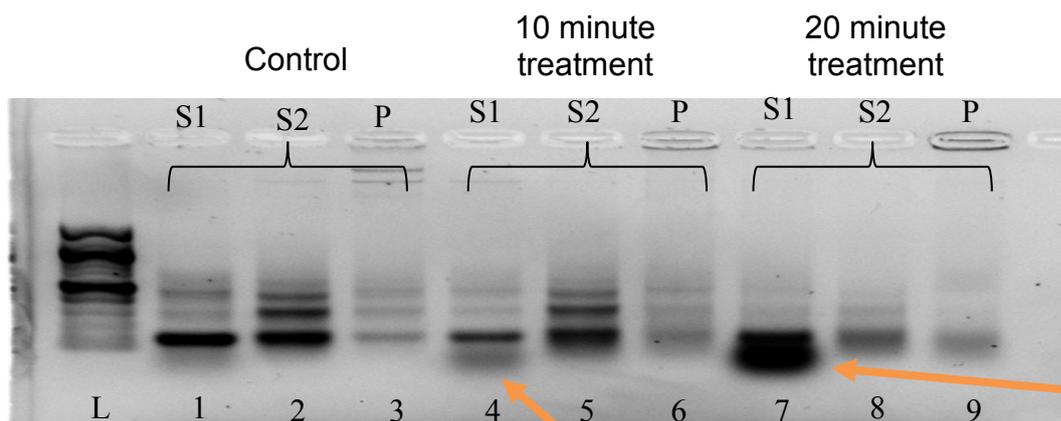
Figure 3.1 Initial identification of UV limits after micrococcal nuclease digestions with lanes 1-3 showing the control sample, 4-6 showing the 10 minute treated sample and 7-9 showing the 20 minute treatment sample.

Sub-nucleosomal DNA damage induced by UV radiation

inflicted upon the DNA. Secondly, four primer pairs were developed for use with PCR to assess the level of inhibition. Using the same forward primer, the reverse primers increased in amplicon length from 132 bp, 211 bp, 309 bp and 409 bp. This allowed a visual representation of how successful the UV treatment had been in inhibiting amplification.

3.1 Duration of Exposure

In order to identify optimal parameters for the UV treatment of *Drosophila melanogaster* SL2 cells the UV treatment conditions had to be adjusted. Firstly, the *D.melanogaster* cells were counted to the concentration of 2 x 10^7 cells/ ml and aliquoted into 0.5 ml samples in a 12 well plate and subsequent chromatin preparation undertaken using native chromatin methods. The first optimal condition we established was the duration of treatment. Here, assessment via electrophoresis following micrococcal nuclease digestion was used. From this we expected to see an
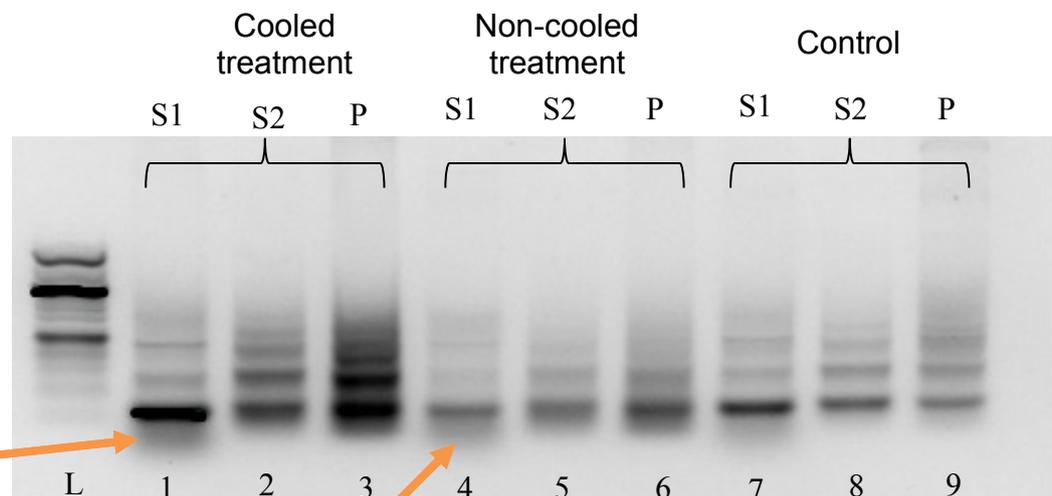
Figure 3.2 Picture after micrococcal nuclease digestion when using the frozen (lanes 1-3), non-frozen (lanes 4.6) and control sample (lanes 7-9).

Sub-nucleosomal DNA damage induced by UV radiation

increase in the level of damage in proportion to the duration of treatment. The initial tests of 10 and 20 minutes (figure 3.1) showed a large subnucleosomal band beneath the mononucleosome at the 20 minute treatment sample and a light band at the 10 minute treated sample compared with no band for the control (no UV treatment), in keeping with anticipated results. These results indicate that 20 minutes is too harsh on the cells, resulting in a loss of intact chromatin and an increase in subnucleosomal DNA, whereas 10 minutes is not damaging enough, therefore another treatment schedule had to be identified.

3.2 Temperature control

	Input number of cells	Output number of cells	% destroyed cells
Cooled plate	12 x 10^7 / ml	4.14 x 10^7 / ml	65.5
Non-cooled plate	12 x 10^7 / ml	1.375 x 10^7 / ml	86.25

Table 3.1
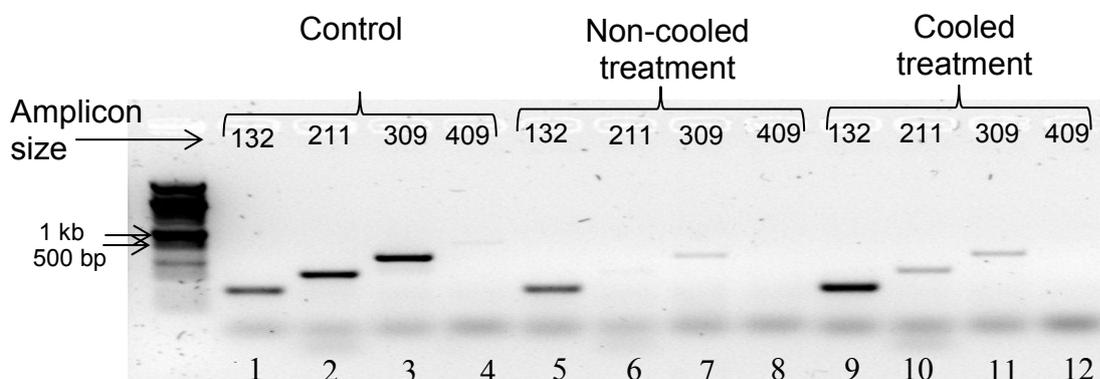
Figure 3.3. An electrophoresshowing the difference in amplification inhibition between a cooled and non-cooled plate. Lanes 1-4 identify PCR amplification of the control using primer pairs to generate 132 bp, 211 bp, 309 bp and 409 bp amplicons, respectively. Lanes 5-8 show 132 bp, 211 bp, 309 bp and 409 bp, amplicons, respectively from the non-cooled plate. Lanes 9-12 from the cooled plate shows amplicons 132 bp, 211 bp, 309 bp and 409 bp, respectively.

I investigated methods for reducing the heat of the cells during UV treatment to maintain chromatin and cellular integrity as well as inhibiting ongoing DNA repair. A 20 minute treatment time was used. Temperature of the cells was maintained by

freezing water in the contours between the wells of the plate before treatment. We predicted that temperatures wouldn't become as high as they previously had and, in turn would result in more cells surviving UV treatment. As expected, this added temperature control meant a higher number of output cells (table 3.1). There was also less DNA damage compared to the previous experiment, as seen in figure 3.2, and slightly less PCR inhibition (figure 3.3).

3.3 Increasing Sample Volume Increases Survival Rate but Reduces DNA Damage
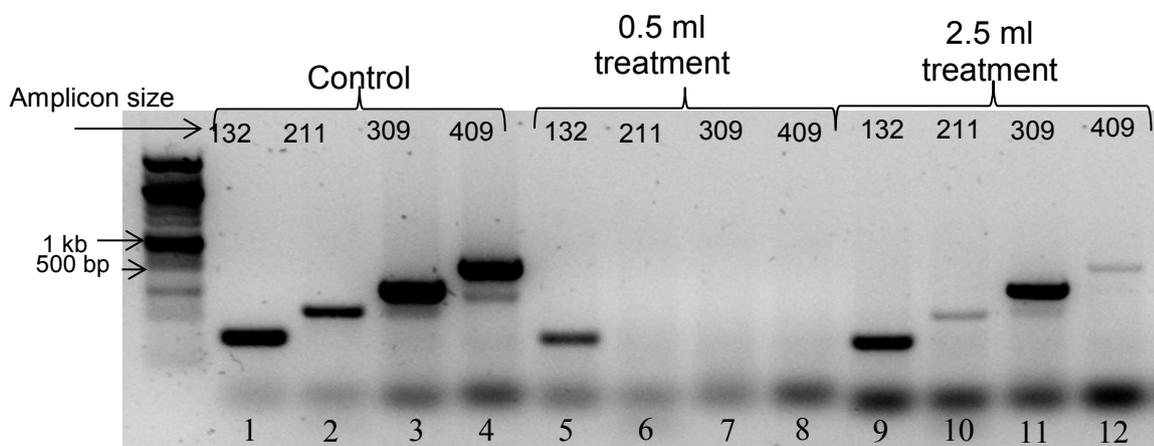
Figure 3.4 A figure showing the level of protection provided by the increased UV treated volume. Lanes 1-4 show PCR amplicons 132 bp, 211 bp, 309 bp and 409 bp, respectively from the control sample. Lanes 5-8 are amplicons 132 bp, 211 bp, 309 bp and 409 bp, respectively, from the 0.5ml treated sample. Lanes 9-12 from the 2.5ml treated sample show 132 bp, 211 bp, 309 bp and 409 bp amplicons, respectively.

It was initially forecast that by increasing the volume of sample in each aliquot from 0.5 ml (lanes 5-8) to 2.5 ml (lanes 9-12) the heat would distribute across the sample more efficiently in turn protecting the cells. With 77.5 % of cells surviving treatment, the initial analysis proved to be promising. Furthermore, PCR analysis following chromatin preparation (figure 3.4) showed that the 2.5 ml treatment sample had protected the DNA and amplification inhibition had been reduced to the point where amplicons were almost as abundant as those in the control (lanes 1-4) sample for

41

some amplicons, therefore concluding that the increased sample volume does not encourage enough DNA damage.

3.4 Alternating UV treatment with ice proved to be the best temperature control.

Given these results it was clear we had to provide an environment which could keep the sample cool while still inflicting optimal damage. A cooler environment for the cells during the UV damage period was provided by adopting a new treatment schedule. The cells were treated for 10 minutes under the UV lamp, taken off and submerged in ice for 5 minutes, followed by a further 10 minutes under a UV lamp

Sub-nucleosomal DNA damage induced by UV radiation
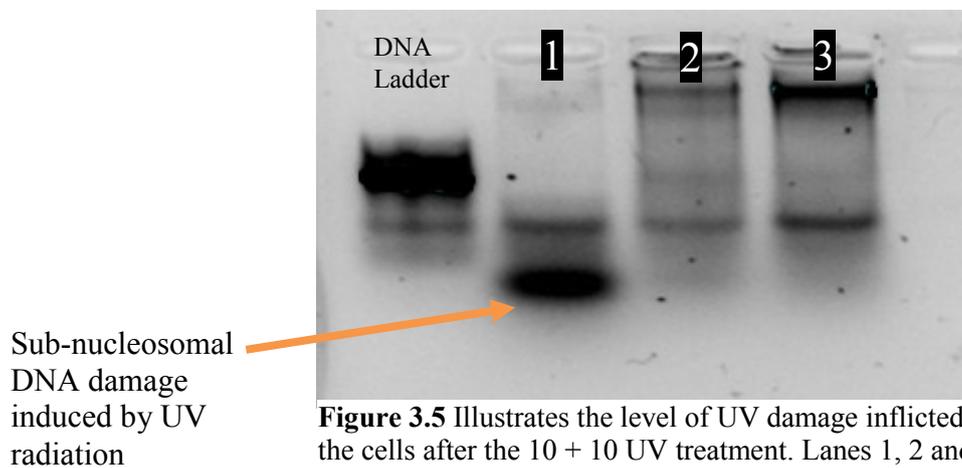
Figure 3.5 Illustrates the level of UV damage inflicted upon the cells after the 10 + 10 UV treatment. Lanes 1, 2 and 3 represent the S1, S2 and pellet, respectively.
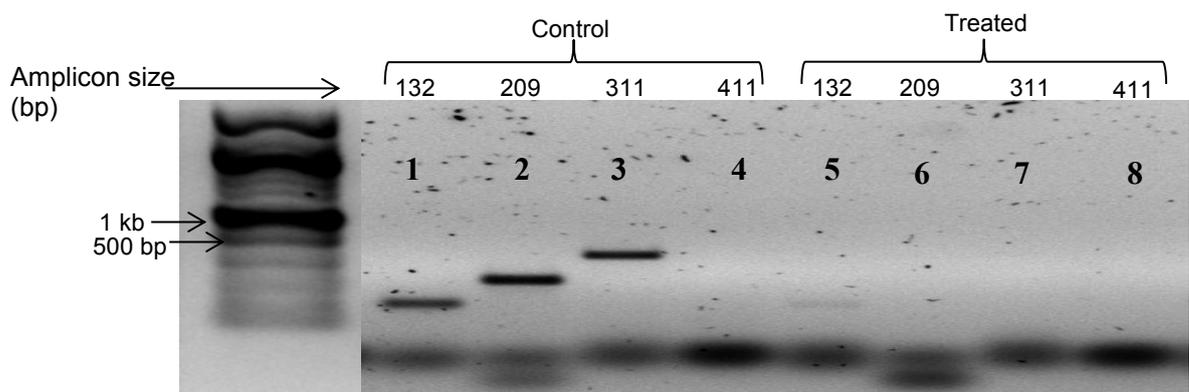
Figure 3.6 A figure showing the range of inhibition given by the duration of treatment for 10 minutes under a UV lamp, 5 minutes submerged in ice, 10 minutes back under the UV lamp and finally, 5 minutes in ice. Lanes 1-4 display PCR amplification of the control sample, lanes 5-8 show the UV treated sample.

and finally, 5 minutes in ice. The first electrophoresis after micrococcal nuclease digestion showed a band beneath the mononucleosome demonstrating UV damage had been inflicted on to the DNA, consistent with our aim to find an intermediate treatment (figure 3.5). Further to this we had to also ensure that the cooling time between treatments was not allowing for too much repair time and therefore affecting the level of inhibition provided by the UV. Figure 3.6 shows a control sample and a UV treated sample after PCR amplification of DNA following chromatin preparation. The first four lanes show the control sample, lanes five to eight represent the UV treated sample. A comparison of the 2 samples clearly identifies a high level of inhibition by UV treatment at all sample sizes with only a faint band for amplicons of 132 bp.

Real-time PCR was used to quantify the rate of inhibition and clarify the results of standard PCR. Figure 3.7 demonstrates the success of amplification inhibition relative to the control samples. Inhibition for the 132 bp long gene sequence proved to be 94 % successful compared with the untreated control sample, the 211 bp sequence saw 98.5 % inhibition, with 309 bp it was 99.3 % and 99.5 % inhibition was achieved for amplicons 409 bp long. These results show an exponential increase in success with the longer amplicon.

From these results it can be concluded that the required level of inhibition had been identified using the following parameters:

- ✓ Duration: 10 minutes under treatment conditions, 5 minutes in ice, 10 minutes in treatment and 5 minutes in ice.

- ✓ Temperature: using a pre-frozen 12-well plate with ice between each of the wells.

✓ Volume and concentration: cells were pipetted into 0.5 ml aliquots in each well at 2×10^7 cells/ ml.
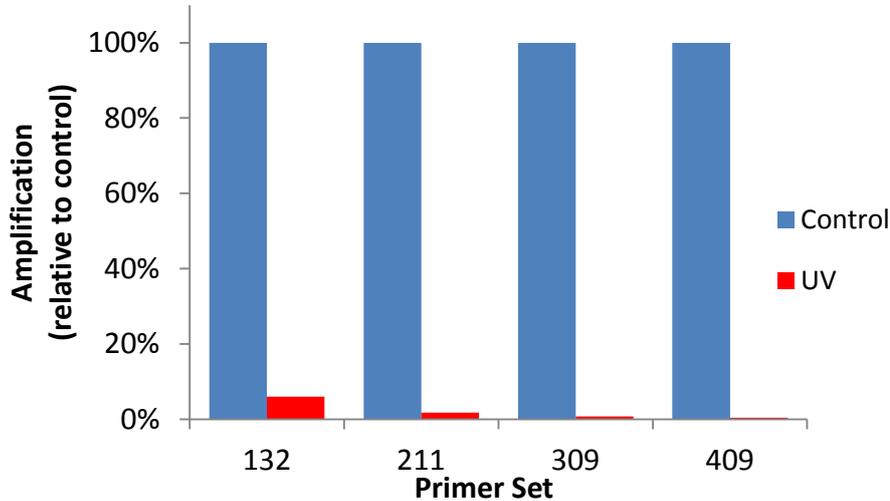
Figure 3.7 A graph to show the level of inhibition delivered by the UV treated SL2 cells during real-time PCR where the blue bars demonstrate comparison to the control DNA.

3.5 Reducing the stability of SL2 cells aids in optimising the sonication step.

Once the ideal treatment conditions had been ascertained the rest of the protocol had to be optimised. During the formaldehyde crosslinking variation of ChIP (X-ChIP) the chromatin is fragmented by sonication. Initial experiments using SL2 cells showed that after UV treatment the *d.melanogaster* cells fragmented to the required band length only after 30 cycles (30 seconds on, 30 seconds off), figure 3.8. However, experiments showed that human lymphoblastoid cells are not as robust,
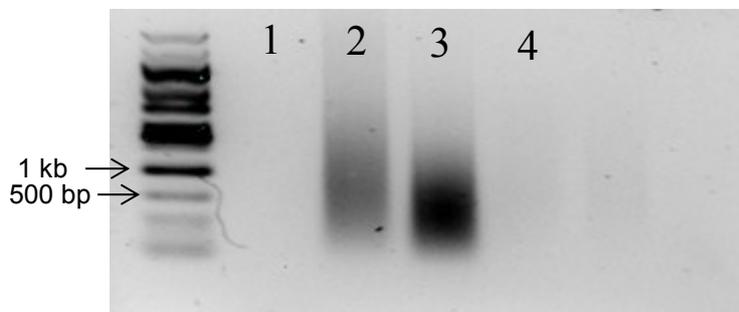
Figure 3.8 Chromatin sizes after sonication cycles. Lanes 1 and 3 show 20 cycles and 2 and 4 are 30 cycles. Where a cycle is 30 seconds on, 30 seconds off. Lanes 1-2 show the control sample and 3-4 identify the UV treated sample.

requiring only around 10 cycles of sonication. I attempted to address this by devising measures to destabilise the SL2 cells. An equilibrium in cell stability between the two cell lines was established by freezing the SL2 cells post-UV treatment, pre-chromatin preparation. Two electrophoresis photos are shown in figure 3.9, the areas highlighted by the red box are the required fragment length ascertained from samples which were sonicated for 10 cycles. This demonstrates that after freezing the SL2 cells for 1 hour, 10 sonication cycles was optimum for both cells lines.

3.6 UV treated cells for use as a carrier during chromatin immunoprecipitation

Table 3.2	To demonstrate the concentration and relative amount of pull down of target DNA from samples with treated and untreated carrier cells.		
		Concentration (ng/µl)	**% pull down**
	Control Input	177	
	Control Pol II	15	8.48
X-ChIP	Control H3K4me3	6	3.39
	UV Input	101	
	UV Pol II	24	23.76
	UV H3K4me3	22	21.78

After optimisation of chromatin preparation the immunoprecipitation steps were followed. Firstly, it was important to prove that the UV treated cells could still be used as a viable carrier. A full X-ChIP experiment using the antibodies complimentary to RNA polymerase II and H3K4me3 was performed using only SL2 cells. An input sample was retained prior to immunoprecipitation to quantify the level of pull down by comparison. The results showed good pull down from both the treated and untreated cells; furthermore these results also indicated that the UV treated cells showed increased pull down compared to the control when using crosslinked chromatin (table 3.2).

Next we optimised carrier ChIP. Experiments using the treated SL2 *D.melanogaster* cells as carrier, holding 10,000 and 1,000 human cells were set up for both crosslinked chromatin and native chromatin. The data was quantified using real-time PCR (figure 3.10) which showed enrichment of the crosslinked chromatin was higher in the UV treated cells than that of the untreated sample, in keeping with the rate of pull down in table 3.2. However, it should be noted that because the primers used are specific for human DNA, the results demonstrate that pull down has been achieved from the target LCL cells. This result was seen consistently throughout each crosslinked carrier ChIP. Native chromatin immunoprecipitation was also used to assess whether there was a difference in pull down between the two types of ChIP. N-ChIP demonstrated good powers of antibody affinity providing pull down percentages from UV treated cells of 6.47 and 5 % for 10,000 and 1,000 LCL cells, respectively. The results from these experiments have proven that UV treated cells can be used as a viable carrier for use in C-ChIP.

Creating DNA libraries is essential to enable next generation sequencing of the resulting DNA. Libraries were prepared following the New England BioLabs® Inc protocol: Illumina® Library Preparation for the creation of ChIP-Seq libraries. Finally, to check the enrichment of the target DNA subsequent analysis by Real Time PCR was conducted using primers: *H.sapien* CD74 -1000, HS CD74 (TSS) (see appendix 1).

3.7 Analysis of library prepared samples

The DNA samples were prepared for DNA sequencing using the New England Biolabs NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina. Figure 3.11 shows the relative enrichment of the genes at 1,000 bp upstream the CD74 transcription start site (TSS) and at the transcription start site relative to the input sample (unbound chromatin). Since H3K4me3 is associated with gene activation it would be expected to see higher fold enrichment of genes at the transcription start site of an active gene as that is where the H3K4me3 binds. In addition to this, judging by the results from previous experiments we would also expect to see an increased enrichment of these genes in ChIP with a UV-treated carrier in comparison to the control. Figure 3.11 visualises the enrichment of the two gene sequences at -1000 bp and the TSS. As expected the gene sets in the UV treated sample are more enriched.
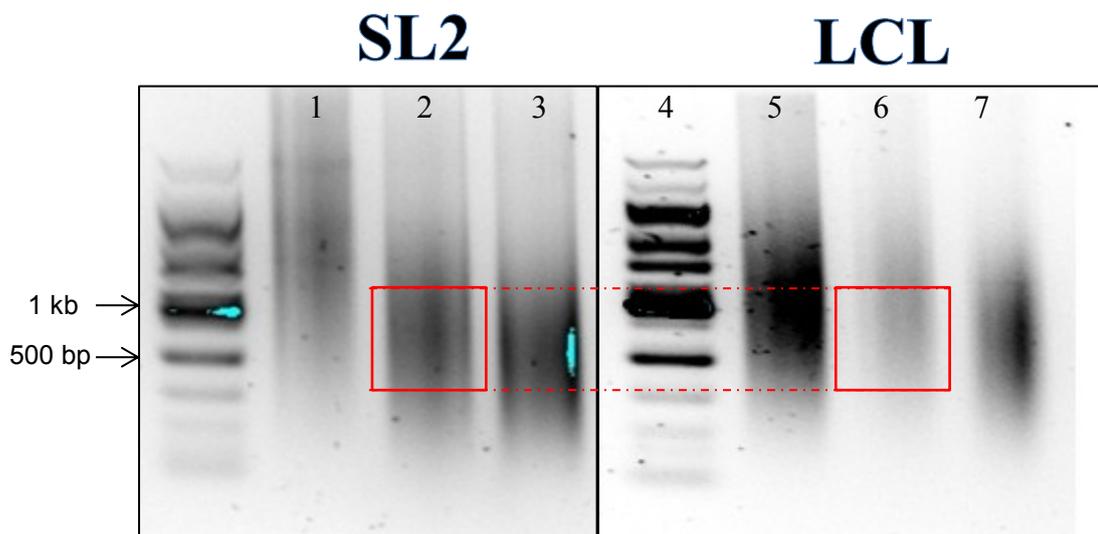
Figure 3.9 Electrophoresis showing the optimisation of sonication cycles between LCL cells and SL2 cells after freezing the SL2 cells post UV treatment. Lanes 1 - 3 show size of fragmented SL2 DNA, lanes 5-7 show fragmented LCL DNA after sonication. Samples in lanes 1 and 5 were sonicated for 5 cycles, lanes 2 and 6 for 10 cycles and lanes 3 and 7 for 15 cycles.
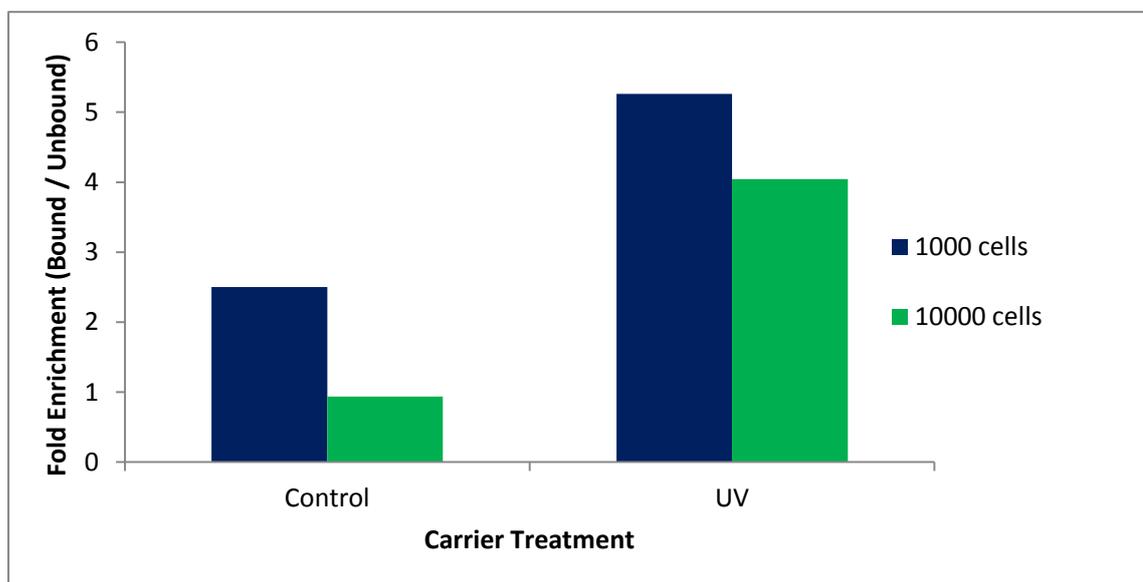
Figure 3.10 Graph to show the fold enrichment of target DNA during real-time PCR amplification using primer sets from the human CD74 gene (see appendix). Shows that quantifiable target DNA within each sample is present.
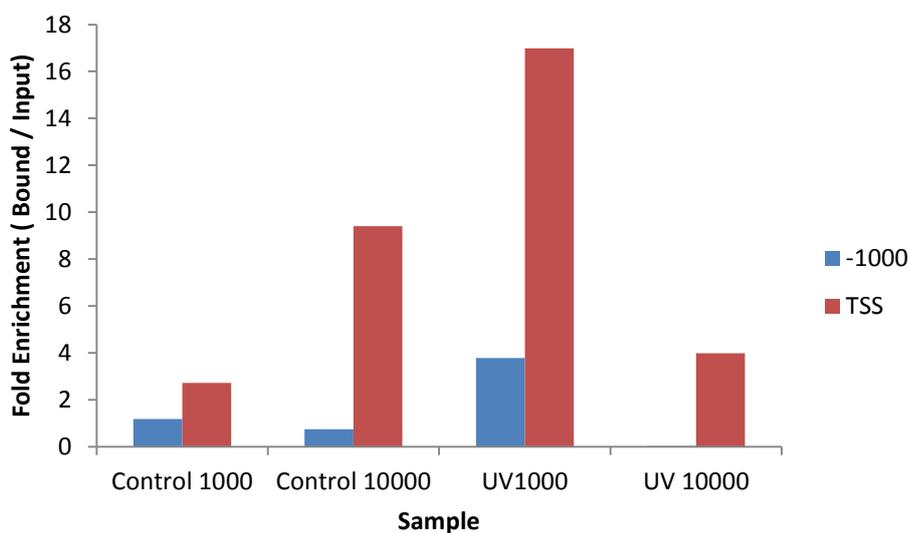
Figure 3.11 Real-time PCR analysis of Library Prepared Samples. The graph exhibits the relative enrichment genes around the transcription start site (TSS) (blue) and 1000 bp before the TSS (red). The graph shows that there is human DNA available for amplification therefore human DNA has been pulled down after immunopecipitation. It also shows that after pull down using the antibody with an epitope for H3K4me3, more DNA is available from the transcription start site than further up in the gene therefore, validating our results.

4. Discussion

The ENCODE project is a follow up of the human genome project, established in 2003 upon the completion of the human genome project it set out to provide a better understanding of the human body's responses to health and disease (The ENCODE Project Consortium 2004). In terms of epigenetics, ENCODE had provided a substantial insight into the impact of DNA modifications and post-translational histone modifications. Here however is where problems arise, ENCODE data is generated from cell lines where large amounts of starting material are readily available. However, the size of clinical samples generally available to study means there are significant gaps in the research on histone modifications in cancers, tumours and disease research. The National Centre of Biotechnology Information 'PubMed' database (www.NCBI.nlm.nih.gov/pudmed), on 23rd February 2014 provides 19,969 results in relation to DNA methylation and cancer, histone modifications however are largely underrepresented at 1,791 results when searched in relation to cancer. Without a reputable procedure to analyse histone modification of cancer cells there is little hope of increasing the current understanding of the roles of post-translational modifications of histone and other proteins in cancers and tumour cells. Chromatin immunoprecipitation is one of the main sources of experimental data for ENCODE protein post-translational modification investigations (ENCODE Project Consortium 2011). ChIP is a widely used technique providing a tool for identification and interpretation of histone post-translational modifications and other DNA-protein interactions. The original and ultimately, most reliable ChIP protocols - NChIP and X-ChIP utilise the affinity of the histone to the DNA or crosslink the proteins to the DNA prior to chromatin preparation, respectively. The large number of cells required to initiate the first steps of these two protocols is the

main drawback which have hindered progression of understanding epigenetic processes for mammalian cancers, tumour and other immune stresses. Given that clinical samples produce significantly less cells than required it is evident that a technique is needed to utilise the analysis properties of ChIP-Seq without the need for such large quantities of input cells. As summarised in table 1.1 many variations of this technique exist, all of these however carry fundamental weaknesses which inhibit and sometimes carry the potential to misidentify analysis products. Other ChIP techniques which enable the sequencing of ChIP samples from small cell numbers have been published. Nano-ChIP-Seq is a technique which has begun to address this issue by providing a technique which requires only 10,000 cells as a lower limit of detection however, this lower limit applied to embryonic stem cells and was higher for other cell types (Adli and Berntstein 2011). We therefore proposed that by modifying the carrier-ChIP technique by O'Neil *et al*. (2006) it would be possible to sequence an exponentially smaller volume of immunoprecipitated target cells using a carrier. The impact of this new, adapted protocol could see a breakthrough in the way in which we are able to view post-translation modifications from a clinical perspective giving greater understanding and insight into the epigenetic mechanisms of real cancer and tumour cells by being able to analyse a broader range of clinical sample.

4.1 Explanation of the Results

4.1.1 Optimisation of exposure time induces appropriate damage upon DNA.
Previous experiments had shown that UV treatment induced more effective damage to the cells than exposure to chemicals. Upon this discovery the UV treatment conditions had to be optimised in keeping with the project objective of causing enough damage to the DNA in order to inhibit PCR amplification, while not inducing

so much that the cells can no longer be used as viable carriers, but also to ensure the loss of cells after treatment is not so drastic as to hinder further experimental use. The most important guideline to establish first was the duration of treatment; the initial experiments saw 2 treatments of 10 and 20 minutes with very different results. While 10 minutes induced minor damage on the cells with only a faint band beneath the mononucleosomes representing DNA damage, 20 minutes proved to induce significantly more damage, closer to the range of damage required (Fig 3.1). In a comparison to check the effect of UV light on cell survival, cells were counted pre- and post-treatment. Although the severity of damage occurring to the 20 minute cells was in keeping with the level of damage we anticipated, the percentage of cells recovered (31 %) was less than ideal. The natural progression from here would have been to moderate the duration of treatment to 15 minutes, indeed this would increase the number of cells returning from treatment however, it would have also seen a decrease in the level of DNA damage induced to the cells which in turn, would decrease the rate of inhibition during PCR amplification. Using the 20 minute conditions another idea was pitched which involved better temperature control during UV exposure.

4.1.2 Temperature control using a coolled 12-well plate increases cell volume output.
During UV treatment the UV light inflicts temperatures onto the cells which are high enough to cause a percentage to burst open. From this, an idea emerged which gave greater temperature control of the solution containing the cells during UV exposure. A 12-well plate containing crevices in which ice was allowed to form around the wells was created to assess the difference temperature control could have on cell destruction. The sample was then treated using these plates and a non-cooled control to compare the protection offered by the temperature. As expected,

51

the number of cells returning after treatment increased to a 65.5% loss, which although was higher than the previous experiment it was still not within the range that was aimed for (table 3.1). Amplification inhibition was seen as in fig. 3.2, inevitably the reduction in temperature had a negative impact on the UV's ability to inhibit PCR amplification however the reduction in inhibition was well within acceptable limits for this project in comparison to the control sample. This finding suggests that although more output cells are still required, the cooler temperature helps to protect cells during treatment producing a higher volume of cells returning from treatment without abolishing the desired level of inhibition.

4.1.3 Increasing sample volume absorbs too much UV and reduced DNA damage

Further experimentation on temperature control to reduce cell death during treatment lead to the variation of sample volume (concentration was kept to 2×10^7 cells/ ml at all times). It was proposed that the larger sample size would encourage the distribution of heat more effectively throughout the sample discouraging cell death and therefore bringing back more output cells from treatment. Previous experiments had used 0.5 ml aliquots of sample in each well of the plate during treatment, for the purpose of this investigation some 2.5 ml samples were also treated. As hypothesised the number of cells coming back after treatment had increased to 77.5 % in the 2.5 ml sample. This huge increase was also reflected in the amount of inhibition seen during PCR. It was thought that the buffer (buffer NB) used to treat the cells created a protective screen between the cells and UV light, or that the increased volume distributed the heat too effectively. The electrophoresis in fig. 3.3 demonstrates that the level of amplification inhibition given by the 2.5 ml sample was on par with that provided by the non-UV treated control, the 0.5 ml sample however, continued to provide more desirable results. From this data it is evident that another

method of temperature control is required which is in keeping with the objective of optimising UV treatment which does not protect the cells as much as the increase in sample volume.

4.1.4 Alternating UV treatment with ice proved to be the best temperature control.

As demonstrated by these results temperature control is key to the optimisation of the UV treatment. With this, a treatment schedule was applied to prevent the cells reaching temperatures that would be detrimental to their survival but in keeping with the already established ideal inhibition by the 20 minute treated sample. Given that the 20 minute schedule was too harsh a revised schedule was adopted which involved 10 minutes in UV treatment, 5 minutes submerged in ice, 10 further minutes under the UV light with a final 5 minute ice treatment (abbreviated to 10 + 10). Figure 3.4 shows faint bands at only the 132bp amplicon size for the sample following the 10 + 10 treatment protocol in comparison to the solid bands given from the control for all amplicon sizes. Furthermore, the number of cells coming back from UV treatment increased to 63 %, double that of the original experiment. From this it is possible to determine that the inhibition provided by these conditions were on point for the desired effects as stated in the objectives section. In summary, the optimal UV treatment conditions were sample concentration of 2 x 10^7 cells/ ml in buffer NB, separated into 0.5 ml aliquots in a frozen, 12-well plate at a treatment schedule of 10 + 10.

4.1.5 Reducing the stability of SL2 cells aids in optimising the sonication step.

Sonication is the mechanism used which produces ultrasonic sound waves at high intensity to a solution, as a result powerful vibrations are inflicted onto the cells causing them to break and expose the contents of the nucleus. The sound waves continue to work by fragmenting the chromatin into variable sizes which are used for immunoprecipitation. Preliminary experiments using both *D.melanogaster* SL2 and

H.sapien LCL cell types showed significantly different fragmented sizes when sonicated at 5-30 cycles with SL2 cells showing much larger DNA fragments. This difference in size suggests that the SL2 cells are more robust than the LCL cells, therefore an additional step was implemented resulting in cells of an equal stability. A follow up experiment which involved freezing the SL2 cells for 1 hour, post-UV-treatment and prior to chomatin preparation meant the SL2 cells were destabilised enough to create fragments of the same size as the LCL cells during sonication. As shown in figure 3.5, the lanes showing 10 cycles for both SL2s and LCLs demonstrated an abundance of chromatin between the desired lengths of 250-550 bp. In conclusion, this data shows that the optimal parameters for using SL2 cells as carriers during X-ChIP are freezing the sample post-UV treatment at aliquots of 5 x 10^7 cells/ ml per sonication sample.

4.1.6 UV Treated Cells are a Viable Carrier during Cross Linked Chromatin Immunoprecipitation.

Despite the measures taken to create optimal conditions for variables in the protocol so far, it had not yet been proven that the UV treated cells were a viable option for use as carriers during ChIP. The first step in achieving objective three, to assess the practicality of using UV treated cells as carriers during the ChIP experiments for both native and crosslinked chromatin, was to carry out the full protocol using only UV treated and control *D.melanogaster* cells. This was done for the formaldehyde crosslinked ChIP experiment where all optimised conditions were implemented. Table 3.2 identifies the resulting concentration of chromatin after each ChIP and the percentage pull-down relative to the input samples. From this table it is justifiable to conclude that the UV treated cells could be used as sustainable carriers. Interestingly, table 3.2 shows an increase in the percentage LCL pull-down for the sample using UV treated carriers compared with the control sample after

formaldehyde crosslinking. Initially, the explanation that UV radiation was inflicting enough damage to the cells to initiate DNA repair responses resulting in added stress markers on the chromatin was suggested however, there was one vital issue with this idea; histone markers H3K4me3 and Pol II have already been identified as a markers for active transcription, not DNA damage responses. To further investigate thesis theories another C-ChIP experiment could be performed using a completely unrelated marker, for example H3K9me.

4.1.7 Immunoprecipitation of as little as 1,000 Human LCL cells using UV Treated Carriers.

The introduction of C-ChIP in 2006 has led to broader applications of ChIP to samples of a small cell number, e.g. clinical samples. O'Neil *et al.* (2006) reported successful pull down and real-time PCR amplification when using carrier ChIP with as little as 100 mouse cells. Detection of such small quantities required the use of radioactive RT-PCR which uses x-ray light to detect amplicons. However, radioactive RT-PCR is only semi-quantitative and therefore it was decided that real-time PCR would be a more effective use of resources. Recently, Lara-Astiaso and colleagues developed a technique known as iChIP, this was published after the work on this thesis had been completed, their method reduced the number of cells required for ChIP-Seq by five orders of magnitude to just 500 cells (Lara-Astiaso *et al.* 2014). iChIP is a technique which has become of great interest within our laboratory however, working through the chromatin isolation procedure by chemical fixation and sonication with such small numbers of cells could prove to be detrimental to the quality of the resulting DNA. By using a carrier during chromatin preparation the target chromatin is less vulnerable to loss during the fixing and sheering steps. C-ChIP using 1,000 and 10,000 target cells were set up and analysed for both native and crosslinked experiments. Figure 3.6 shows the percentage pull down relative to

the input for the UV and control samples using 1,000 and 10,000 target cells during X-ChIP. In addition to the previous results which showed a higher concentration of pull down of target cells after UV treatment compared to the control, similar results were observed here showing that this data was not accidental or a 'one off'. The enrichment of target cells in the UV treated samples, as quantified by RT PCR, is hypothesised to be a result of damaged DNA. During UV treatment the radiation will damage the DNA therefore reducing the amount available for PCR amplification, thereby resulting in more reagents available to amplify the target DNA. Furthermore, figure 3.7, indicates that the pull down percentage from samples containing 1,000 target cells was greater than that of the 10,000 cells. It should be noted that these results need be addressed with caution, although it is possible for the 1,000 sample to result in higher levels of pull down, it was concluded that amplification assessment of just 1,000 cells was edging close to the detection limits for the real-time PCR machine.

4.1.8 Analysis of library prepared samples

For this project to be termed a success the samples would be sequenced after library preparation. The produced DNA should correlate with the known genomic distribution of RNA PolII and H3K4me3 in LCLs, available form the ENCODE project. Before this however, gene-specific enrichment was checked using real-time PCR. It is important that the library prepared sample is amplified using both *D.melanogaster* and human primers. In the case of the *Drosphila* primers amplification would be expected to be considerably low in the UV treated sample in comparison to the control. Testing against the human primers however, it would be expected that the human genes would amplify more readily in comparison to the UV treated *D.melanogaster* amplicons. Furthermore, in the case of amplifying human DNA in

the UV treated carrier, lessons from previous results indicate that the LCL DNA would amplify better than in the non-UV treated carrier. In addition, given that the primer sets used corresponded to different locations on the gene, one of which being the TSS (methyl mark H3K4me3), it would be expected to find an abundance of DNA associated with this locus in comparison to the other sequences. The purpose of this experiment was to use human specific primers for real-time PCR amplification to confirm the presence of human DNA in the sample.

Firstly, the results as shown in figure 3.11 confirm the presence of human DNA and therefore pull down of target LCL cells using 1,000 and 10,000 target cells. Secondly, when comparing the abundance of DNA in the control to the UV treated carrier sample, it is clear that the enrichment of the pull down DNA is greater in the treated sample when comparing to the input sample, correlating with the prediction. As mentioned above, it was predicted that this graph should show an abundance of material at the TSS. Figure 3.11 shows that human sequences are successfully amplified in both samples, proving the presence of human DNA in all experiments. To summarise, figure 3.11 shows human DNA is present within all samples and that the UV treated carriers provide a consistently higher enrichment of human DNA compared to the non-UV treated. To further investigate the success of a potential new carrier ChIP technique next generation sequencing is required.

4.2 Troubleshooting

Since the first paper on chromatin immunoprecipitation was published in 1978 (Jackson 1978) a huge number of modifications have been applied to the ChIP protocol in order to establish optimal running conditions for many variables. It was by reading and combining information collected from these variations of ChIP that the advantages, limitation and applications of these techinques were collaborated to

provide a better, more versatile technique which is more applicable to real, clinical samples (see table 1.1).

Most of the technical issues that occurred have been addressed for the purpose of optimising the protocol. However, there were various aspects of the protocol which were technically difficult and required adjustment between each experiment. Firstly, during native chromatin preparation the cell walls are broken down using a homogeniser (O'Neil *et al.* 2006). The number strokes required for each sample varied from 1-15, in order to assess the rate of cell breakdown during homogenisation the sample was viewed underneath a microscope between strokes. It was found that the UV treated samples were somewhat easier to breakdown than the control samples which proved to be problematic. Over homogenisation of a sample can cause the premature destruction of chromatin and in turn, sample is lost; under homogenisation leads to a lack of available chromatin kept as it is within the unbroken cell, neither outcome will provide good chromatin to pursue with further experimentation. The complex nature of the homogenisation step caused issues with a number of experiments that were attempted and even meant that some experiments had to be abandoned. The only way to overcome this was practicing being able to spot the difference between a free nucleus, cell debris and unbroken cells, this practice came from the volume of experiments that were attempted during this project.

4.3 Further Work and Applications to the Field

The successful development of cChIP-seq would allow the analysis of smaller populations of cells than is currently possible. All cell types contain different cell surface markers, a factor which can aid with therapeutic drug targeting, cancer diagnostics and cell sorting (Andrews *et al.* 2013). When tumour cells are recovered

from clinical samples a number of cell types are incorporated with in the sample. Flow cytometry is a cell sorting technique which has the ability to discriminate cells according to their cell surface markers, by doing so tumour cells and non-tumour cells can be separated. Further cell sorting can be achieved by separating cells according to their position in the cell cycle. Since clinical samples sent for research purposes are limited this project could provide a useful tool in analysis of primary cells lines. The use of these sample types can have major implications on pharmaceutical study, including drug discovery and target identification.

DNA sequencing is an invaluable tool for the identification of genes, gene sequences and as identified here, for use as an analysis of ChIP data. Although the samples were prepared and libraries were analysed by real-time PCR, DNA sequencing was left incomplete due to time restraints. This project could be continued by the sequencing analysis of the library samples. From here, a definitive conclusion could be made as to whether or not UV treatment can be used as a valid regime to inhibit carrier DNA amplification during sequencing. If, as predicted, the damage inflicted upon the cells is enough to create reliable target DNA sequences, the improvement could provide a breakthrough in the way research on clinical samples is conducted, incorporating how post-translation modifications of histones and other proteins impact gene regulation.

Although the results displayed reason to believe that the 1,000 cell sample contained pull down from the LCL cells, the results are unconvincing when addressing the relative amounts in the TSS and -500 ad -1,000 bp sequences. Additional time to continue with this research could mean the optimisation of the protocol with 1,000 LCL cells. Since the real-time PCR analysis after immunoprecipitation showed enrichment of the target genes it can be concluded that DNA was lost during library

preparation. Further work could include the preparation of more ChIP samples, in turn more library DNA to take for analysis. Real-time PCR analysis of these libraries could provide evidence as to whether or not using 1,000 cells as a target for ChIP is beyond the detection limits for. Additionally, further experiments on the inhibition of the *D.melanogaster* DNA would provide a detailed insight into the rate of inhibition, especially if carried out with LCL analysis on the same plate where a comparision could be made between the enrichment of LCL cells and *D.melanogaster*.

Ultimately, next generation sequencing is the next step for this project, since next generation sequencer can determine exactly what proportion of the read throughs are human, and what are *Drosophilia* final quantification cannot be made until after sequencing.

5. Conclusion

Here, I have described the modification of a well-established technique which can be used to aid in the discovery of gene regulatory mechanisms for clinical samples e.g. cancers, tumours, diseases etc. Although other versions of ChIP for small cell numbers exist, their applications are limited. Despite efforts by researchers the analysis of post-translation modifications for clinical samples has proved technically difficult leading to gaps in the literature. Introduction of a valid and reliable technique which incorporates all of the advantages of using the original chromatin immunoprecipitation technique and discarding the disadvantage would have huge impacts and potentially revolutionise the way biopsies and clinical samples are researched, providing insight into epigenetics of tumour development which have not been possible before. The method described holds the benefits of C-ChIP for its application with a small cell number for both N-ChIP and X-ChIP experiments, while discarding the limitations of background interference of *D.melanogaster* by C-ChIP. The results and figures shown provide solid evidence that by UV treating the *Drosohila melanogaster* carrier cells prior to combination with target cells it is possible to inhibit PCR amplification to a substantial degree. It was not possible to complete DNA sequencing of the produced libraries within the timeframe of the project and this will clearly provide the definitive evidence of the success of the technique. Without the results of DNA sequencing, providing a solid conclusion to the aim of this project would be naïve and presumptuous. It is possible however, to reflect on the proposed objectives.

Ultraviolet conditions were optimised by variation of the duration, volume and temperature of cells during treatment. Troubleshooting with the various conditions meant that, in keeping with objective one, it was possible to inhibit PCR amplification

without abolishing carrier qualities of the *D.melanogaster* cells. Unexpected issues arose during X-ChIP optimisation leading to an added step for use with a carrier after UV treatment. Objective two was achieved during this troubleshooting stage. Finally, immunoprecipitation was achieved for both native and crosslinked chromatin with further quantification and real-time analysis on the library prepared samples, demonstrating the success of objective three. In addition to these findings, ultraviolet treatment of the carrier cells proved to hold more benefits than initially expected. Firstly, with the use of real-time PCR, significant amplification inhibition was observed for both the native chromatin and formaldehyde crosslinked chromatin using carriers. Finally, the data consistently represented highly enriched pull down of target DNA when carried by UV treated cells compared to the LCL cells carried by non-UV treated *D.melanogaster,* as quantified by RT-PCR. To conclude, although solid conclusions cannot be made without results from DNA sequencing, the data represented here tells a story which leaves little doubt that UV treatment of carrier cells inflicts enough damage upon DNA to use C-ChIP as a viable and reproducible tool in conjunction with high throughput sequencing.

Bibliography

Abcam. (2011). **A Beginners Guide to ChIP.** [Online]. Available at: http://docs.abcam.com/pdf/chromatin/A-beginners-guide-to-ChIP.pdf [Accessed 10th February 2014].

Acevedo, L. Iniguez, A. Holster, H. Zhang, X. Green, R. Farnham, P. (2007). Genome-scale ChIP-chip analysis using 10,000 human cells. **Biotechniques**, 43: (6): 791 - 797.

Adli, M. Bernstein, B. (2011). Whole-genome chromatin profiling from limited number of cells using nano-ChIP-Seq. **Nature Protocols**, (10): 1656-1668.

Allfrey, V. Faulkner, R. Mirsky, A. (1964). Acetylation and Methylation of histones and Their Possible Role in the Regulation of RNA Synthesis. **PNAS,** 51: (5): 786 - 794.

Andrews, T. Wang, D. Harki, D. (2013). Cell Surface Markers of Cancer Stem Cells: Diagnostic Macromolecules and Targets for Drug Delivery. **Drug Delivery and Translational Research**, 3: (2): 121 - 142.

Arnaudo, A. Garcia, B. (2013). Proteomic characterisation of novel histone post-translational modifications. **Epigenetics and Chromatin** [Online]. 6: (1). Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737111/pdf/1756-8935-6-24.pdf. [Accessed: 10th February 2014].

Bannister, A. Zegerman, P. Partridge, J. Miska, E. Thomas, J. Allshire, R. Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. **Nature**, 410: (6824): 120 – 124.

Bauer, U. Daujat, S. Nielsen, S. Nightingale, K. Kouzarides, T. (2002). Methylation at arginine 17 of histone H3 is linked to gene activation. **EMBO Reports**, 3: (1): 39 - 44.

Bernstein, B. Kamal, M. Lindblad-Toh, K. Bekiranov, S. Bailey, S. Huebert, D. McMahon, S. Karlsson, E. Kulbokas III, E. Gingeras, T. Schreiber, S. Lander, E. (2005). Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse. **Cell**, 120: (2): 169 - 181.

Cao, R. Wang, L. Wang, H. Xia, L. Erdjument-Bromage, H. Tempst, P. Jones, R. Zhang, Y. (2002). Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. **Science**, 298: (5595): 1039 - 1043.

Cloos, P. Christensen, J. Agger, K. Helin, K. (2008). Erasing the methyl mark: histone demethylases at the centre of cellular differentiation and disease. **Genes and Development,** 22: (9): 1115 - 1140.

Cuthbert, G. Daujat, S. Snowden, A. Erdjument-Bromage, H. Yamada M. Schenider, R. Gregory, P. Tempst, P. Bannister, A. Kouzarides, T. (2004). Histone deimintation antagonizes arginine methylation. **Cell,** 118: (5): 545 – 553.

Dahl, J. Collas, P. (2007). Q^2ChIP, a Quick and Qantitative Chromatin Immunoprecipitation Assay, Unravels Epigenetic Dynamics of Developmentally Regulated Genes in Human Carcinoma Cells. **Stem Cells**, 25: (4): 1037 - 1046.

Deluve, G. Khan, D. Davie, J. (2012). Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors. **Clinical Epigenetics**. [Online], 4: (5). Available at: http://www.clinicalepigeneticsjournal.com/content/pdf/1868-7083-4-5.pdf Accessed: 20th February 2014.

Dillon, S. Zhang, X. Trievel, R. Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methylatransferases. **Genome Biology**, 6: (8): 227.

ENCODE Project Consortium. (2004). The ENCODE (ENCyclopedia of DNA Elements) Project. **Science,** 306: (5696): 636-640.

ENCODE Project Consortium. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). **PLoS Biol** [online], 9: (4). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079585/. [Accessed: 27th February 2014].

Feng, Q. Wang, H. Hui Ng, H. Erdjument-Bromage, H. Tempst, P. Struhl, K. Zhang, Y. (2002). Methylation of H3-lysine 79 is Mediated by a New Family of HMTases without a SET domain. **Current Biology**, 12: (12): 1052 - 1058.

Fischle, W. Tseng, B. Dormann, H. Ueberheide, B. Garcia, B. Shabanowitz, J. Hunt, D. Funabiki, H. Allis, C. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. **Nature**, 438: (7071): 1116 – 1122.

Flanagan, J. Mi, L. Chruszcz, M. Cymborowski, M. Clines, K. Kim, Y. Minor, W. Rastinejad, F. Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognise the methylated histone tail. **Nature**, 438: (7071): 1181 – 1185.

Flanagin, S.,Nelson, J.,Castner, D. Denisenko, O. Bomsztyk, K. (2008). Micro-plate based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signalling of a complex genomic events. **Nuceic Acids Research**. [Online], 36: (3): Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2241906/pdf/gkn001.pdf [Accessed: 1st March 2014].

Gilmour, D. Lis, J. (1985). *In Vivo* interactions of RNA Polymerase II with genes of *Drosophila Melanogaster*. **Molecular and Cellular Biology**, 5: (8): 2009 - 2018.

Greer, S. Honeywell, R. Geletu, M. Arulanandam, R. Raptis, L. (2010). Housekeeping genes; expression levels may change with density of cultured cells. **Journal of Immunological Methods,** 355: (1-2): 76 - 79.

Hebbes, T. Thorne, A. Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. **The EMBO Journal**, 7: (5): 1395 - 1402.

Heins, J. Suriano, J. Taniuchi, H. Anfinsen, C. (1966). Characterisation of Nuclease Produced by *Staphylocuccus aureus*. **Journal of Biological Chemistry**, 242: (5): 1016 - 1020.

Izzo, A. Schneider, R. (2011). Chatting Histone Modifications in Mammals. **Briefings in functional genomics**, 9: (6): 429 – 443.

Jackson, V. (1978). Studies on Histone Organisation in the Nucleosome using Formaldehyde as a Reverse Cross-linking Agent. **Cell,** 15: (3): 945 - 954.

Jiang, J. Lu, J. Lu, D. Liang, Z. Li, L. Ouyang, S. Kong, X. Jiang, H. Shen, Luo, C. (2011). Investigation of the Acetylation Mechanism by GCN5 Histone Acetyltransferase**. PLOS One,** [Online]. 7 (5). Available at: http://www.plosone.org/article/fetchObject.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0036660&representation=PDF [Accessed: 17th February 2014].

Klenow, H and Henningsen, I. (1970). Selective elimination of the Exonuclease Activity of the -Deoxyribonucleic Acid Polymerase from *Escherichia coli* B by limited proteolysis. **PNAS,** 65: (1): 168 – 175.

Klose, R. Kallin, E. Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. **Nature Reviews Genetics**, 7: (9): 715 - 727.

Klug, A. Rhodes, D. Smith, J. Finch, J.T. Thomas, J.O. (1980). A low resolution structure for the histone core of the nucleosome**. Nature,** 287: (5782): 509-516.

Lara-Astiaso, D. Weiner, A. Lorenzo-Vivas, E. Zaretsky, I. Jaitin D. David, E. Keren-Shaul, H. Mildner, A. Winter, D. Jung, S. Friedman, N. Amit, I. Chromatin state dynamics during blood formation. **Science**. 345: (6199): 943-949.

Luger, K. Mader, A. Richmond, R. Sargent, D. Richmond, T. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. **Nature**, 389: (6648): 251 - 260.

Lee, T. Johnstone, S. Young, R. (2006). Chromatin Immunoprecipitation and microarray-based analysis of protein location. *Nat. Proc*, 1: (2): 729 - 748.

Miura, H. Tomaru, Y. Nakanishi, M. Kondo, S. Hayashizaki, Y. Suzuki, M. (2009). Identification of DNA regions and a set of transcriptional regulatory factors involved in transcriptional regulation of several human liver-enriched transcription factor genes. **Nucleic Acids Research,** [Online]. 37 (3). Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2647325/pdf/gkn978.pdf. [Accessed: 14th November 2013].

Murray, K. (1964). The occurrence of Epsilon-N-Methyl Lysine in Histones. **Biochemistry**, 3: 10 - 15.

Nelson, J. Denisenko, O. Sova, P. Bomsztyk, K. (2006). Fast Chromatin immunoprecipitation assay. **Nucleic Acids Research,** [Online]. 34: (1). Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1325209/pdf/gnj004.pdf.   [Accessed: 10th February 2014]

O'Neil, L and Turner, B. (2003). Immunoprecipitation of native chromatin: NChIP. **Methods**, 31: (1): 76 - 82.

O'Neil, L. VerMilyea, M. Turner, B. (2006). Epigenetic Characterisation of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. **Nature Genetics,** 38: (7): 835 – 841.

Park, J. (2009). ChIP-seq: advantages and challenges of a maturing technology. **Nature Reviews Genetics,** 10: (10): 669 - 680.

Petrossian, T. Clarke, S. (2011). Uncovering the human methyltransferosome. **Molecular and Cellular Proteomics** [online], 10: (1). Available at: http://www.mcponline.org/content/10/1/M110.000976.full.pdf+html. [Accessed: 14th February 2014].

Pogo, B. Allfrey, V. Mirksy, A. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. **Proc Natl Acad Sci USA**, 55: (4): 805 - 812.

Rager, M. Kmet, M. (2005). Quantitative Analysis of Fine Needle Aspiration Biopsy Samples. **Radiol Oncol,** 39: (4): 269 - 272.

Schneider, I. (1972). Cell lines derived from the late embryonic stages of *Drosphila melanogaster*. **Journal of Embryology and Experimental Morphology,** 27: (2): 353 - 365.

Shi, Y. Lan, F. Matson, C. Mulligan, P. Whetstine, J. Cole, P. Casero, R. Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. **Cell**, 119: (7): 941 – 953.

Sutherland, B. Toews, J. Kast, J. (2008). Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. **Journal of Mass Spectrometry**, 43: (6): 699 – 715.

Thomas, T. Voss, A. (2007). The Diverse Biological Roles of MYST Histone Acetyltransferase Family Proteins. **Cell Cycle**, 6: (6): 696 - 704.

Tjeertes, J. Miller, K. Jackson, S. (2009). Screen for DNA Damage responsive histone modifications identifies H3K9ac and H3K56ac in human cells. **EMBO Journal**, 28: (13): 1878 - 1889.

Turner, B. (2000). Histone acetyltation and epigenetic code. **BioEssays**, 22: (9): 836 - 845.

Tweedie-Cullen, R. Brunner, A. Grossman, J., er al. (2012). Identification of combinatorial patterns of post-translational modifications on individual histones in the mouse brain. **PloS one** [online], 7: (5): Available at: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0036980. [Accessed: 14th November 2013].

Varshovsky, A. Bakayev, V. Georgiev, G. (1976). Heterogeneity of chromatin subunits in vitro and location of histone H1. **Nucleic Acids Research**, 3: (2): 477 – 492.

Vetting, M. Magnet, S. Nieves, E. Roderick, S. Blanchard, J. (2004). A bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones. **Chemistry and Biology**, 11: (4): 565 - 573.

Zhang, X. Ouyang, S. Kong, X. Liang, Z. Lu, J. Zhu, K. Zhao, D. Zheng, M. Jiang, H. Liu, X. Marmorstein, R. Luo, C. (2014). Catalytic Mechanism of Histone Acetyltransferase p300: From the Proton Transfer to Acetylation Reaction**. Journal of Physical Chemistry** [Online], Available at: http://pubs.acs.org/doi/pdf/10.1021/jp409778e. [Accessed: 17th February 2014].

Yun, M. Wu, J. Workman, J. Li, B. (2011). Readers of histone modification. **Cell Research,** 21: (4): 564 – 578.

Yuan, J. Pu, M. Zhang, Z. Lou, Z. (2009). Histone H3-K56 acetylation is important for genomic stability in mammals. **Cell Cycle**, 8: (11): 1747 – 1753.

Appendices

Appendix 1: Human CD74 gene primers for -1,000, -500 and TSS.

gagtgcagtggcgtgatcttggctcactgcaagctccgcctcccgggttcacgccattctcctgcctcagcctccagagtatctgg

gactacaggcgcctaccaccacgcccggttaatttttttgtattttttttagtagagatggggtttttgccgtgttagccaggatggtct

cgatctcctgacctcgtgatccgcccgcctcggcctcccaaagtgctgggattacaggcgtgagccactgcccccagccataag

ctaggattttaaaggcaggggtaaatctgaggaaagcagaagttataggcaaaattctaaatcaatacatgaaggttatatatt

ggtttgacctaaaaaggtgggacatcttgaagtggggggtccacaggtcatgggtggatttaaagattttctcatttgtcattggtt

aaggacacgaacgtttgtctaaaaatctggggtcagaaaaaaggaatgttagctctggcctgtgggcatgacctcctccaggcc

cctcagaaagaaatttatttattttttaatgtttaaaatgtattatttattgagacagggggctcgctctgtcacccagactggagttta

gtgggacaatcagtgctcactgtagcctcaaactcctgggctcaagtgatccttctgcctcagcctcccatctagctgggactac

agatatgtgctgccacacctggctaattaaattttgttttgttttgttttgttttgtagaaatggggggggtcttgctttgttgcccag

gctggtcttgaacttctggcctcaagcagtccttccaccttggcgtcccaaagtgctgagattacaggtgtgaaaggaagaaatt

tagagcaaagattggtggttagagtttagtcctctgctcccccttgtctgagatctatgagccagtggatctgtttggtgggggtct

gggtttatgaaaagcaatgcagggacataagttaagatgttatctttagttttataggagccaaataccccatgattctaactt

ttttggctattgtcttaagttattactatattttttgtttatcagcttgctcatttacttctcaaggctagctaggtgcctgaaatactttt

gaagaaacgtaagtttcttttttcttttcttttcttttttctctttagagatggtcttactctgtcacccaagctgcagtacagtggcatc

atcacagctcactgcagcctggacactcaagcctccggggatcctcctgcctcagcctcctgagtagctgggactacaggcacg

tgccaccacgcctggctaattttagaaaaattttgtggagacggggtctcactgtgttgcccaggctggtctcgaactcctggc

ctcaagcgatcctcccaccttggacctgctaaagtgttggaattacaggtgtgagccaccatttctggcctagattttctctctttt

tatgttggtggggtgctcagctggcccctaagaagggtccctgctctgtctcacagctatttgaaactgtatgttattgttaactat

agtcatcccacaggggtatagaaaatgagaacttagtcctatctaggtgcaattttgtatccttttgcatccccgtccctcccttcc

ccctacccttcccagcctctagtatcctctgttctacttttcacttctatgagttgaactttttttagcttccacatatgagttagaaca

ttctgtgtttaactttctgttcctggcttatttcacttacaaaatgtctccagttccatccagtttcatatcttcagcatcttctatggtg

cgaggctctggctaaatgcttatttagtctaacagctctgcaaggctagtattaccatccctgcattacaaagcagaaagcaaaa

tcagaaaggttgaggggccaggcgtggtggctcacgcctgtaatcccagcactttgggaggccaaggcaggcggatcacgag

71

gtcaggagatcgagaccatcctggctaacacagtgaaacaccatctctactaaaaaaatacaaaaaattagctgggtgtggtg

gcgggcacctgtagtcccagctacttgggaggctgaggcaggagaatggcatgaacccgggaggcggagcttgcagtgagcc

gagatcgtgccactgcactccagcctgggcaacagagcaagactctgtctcaaaaaaaaaaaaaaaaagtaaagaaaggttg

agggactggttcagggactcacatgcctaggatgtttctgacttgggatttgaactcaggtctggccacatcaaaggccatgtct

gctcttcgctctcctcccaaaatacatgtactatctcatttgactcttgaaatccaaaactcattgaaccttcgaaccacaagtcta

aatttattaaaaaaaatttattaatgcaaattgtaaagcacacacaatacataaagattaataaaacaaatatggtatataaaa

ggaaacttgtttttttacactcacagcacttctgacaccaaatgtgtggattttttctacaccaacaacaaatgttctaactctctgg

acaccaactgggcatcctacaattcaattctgttcagacactatctacctgggggttagcatctgaacccataggttgagggctca

gtcagacaagactcatccacttcagatgccaatcctaagtcccgggactccagtacttctgatgagttggctacaaattgggagt

tcccacgactccctcttcatacttgacaaatagctacaatggcccacagaactgagaagcactttacttactattaccagtttatta

taaaggatacagatgagcagctgatgaagaggtctatagcgtacggcatgtgggaggggtcctgaacacagcagcctctgtcc

ctgtccttgtggagttggggtgcaccaccctcctggcacatggatgtgttcgtctactcgaaagctctccaaacctagttgtttgg

gattttatgcaagttccattatgtagacatggttggttaaatcactgattgaactcaacctccaaccccctctctcctctctggaagt

cggggtggggcgggtgggggtgaggtctgaaagtcccgaccctctcctcacatggcacccagcctccatctgaagctctctagg

ggctttcagccaccagcatttcctgaatatataaaaaaagacattcttctcactctggagatactttttttttttgagatggagtctca

ctctgttgcccaggctgaagtgcagtggtgcgatctcgcacctccgcctccagggttcaagcgattttcctgcctcagcctctcga

gtagctgcgattacaggcacacgtcaacatgtccggctaattttttgtgtttttagtagcggggtttcactatgttggtcaggcttgt

ctcaaactcctgacctcatgatacacccacctcgacctcccaaagtgctgggattacaagcgtgagccaccgtgcctggtcttttt

ttctttttttttcttgtttttttttttttagacagagtcttgctctgtcacccaggctggagtgcagtggtgcgttcttggctcactgcagc

ctctgcctcctgggttcaagcaattcccctgcctcagactcccaagtagctgggattacaagtgtgtgccaccatgcctggctaat

ttttaaattttttgtagagacagggtttcaccatgttggccaggcaagtctcgaactcctgacctcaactgatcctcccgcctcagc

ctcccaaagtgctggaattacaggtctgagccaccacacccaggcttactctggagactctttggttcttagtggtccttgtgtca

ggaaccagggacagagaccagagacgtttttcactatgtcacagcacaggagc<mark>cttgccaacacagtcccatc</mark>ctcccaggca

ccctttcccaggctgtcctcttccctcccagcacaggcccccattctcctggatctgaaccacacaaggaggcagaggggtgga

gtggttaaaaatgcaggcgctactactttctggaggtgtgatcctggacaagagcccctgtgtccctatctgtaaatggagatga

taaaacactcctcaa**agggcctccgtaaagatcac**aggaaacaaagcatggacaacgtgcctggcacgaaggagagcttggt

aagtggaggcaagtctgagtgccagagccttaaaaccatggactcttaggtttgaaggaaatcaaaatattcctctctaaaaac

ctgcggatttttagttaaaggtaaaaggttaaaacccaggggcaccctctgccccttctctctgtcttgatcagctcagaagcaga

gctcagaggatcctggatcagacttgactgttcccatacattcaccttcccacatttccagccctttggaaccctgaagatatcctc

ttctctgtcctgtcactctatgggacttacggctctttgttaaaatactatttaagcaaggcccctaagccactgccttgagagaga

aatacttctcaactgaggcctctactgcataatgggtacagcatgtgttaatcaactttggctgattttttcttttgttaagttgacttt

tgtttccaggggagcgtgtcaactaagaacctaaacagggaaagaaaagaaatgttttcacccctgcacgtcgtccaaacttcc

tcctgcgtgtttgagacgacataacggaggccagagaagaggacttgtctgagctcaccaagggtgcacagcccctttctgcac

ccactcatctgtggatgagaaggggagacaaacaaaggtgtcttctgtttcaaagtgctttcctgtctagggagtggacatttgc

ctgtttcttgaaacattcaaagagccttatgaatccaaaggcctgcccagaaacaagtgatgagggccctgggcagccaatgg

gatcgtgctggcctttctacctgcctggggagcccccccgccccacatcctgccccgcaaaaggcagcttcaccaaagtggggt

atttccagcctttgtagctttcacttccacatctaccaagtgggcggagtggccttctgtggacgaatcagattcctctccagcacc

gactttaagaggcgagccggggggtcagggtcccag**atgcacaggaggaga**agcaggagctgtcgggaagatcagaagcc

agtcatggatgaccagcgcgaccttatctccaacaatgagcaactgcccatgctgggccggcgccctggggccccggagaggt

atgtgtgagcaccaggaaagggcacaccgatcctggactgcagagccgtgcgcatgtgctgggaagaagcaccggccagga

gtcacaggaaagagggattcaggctctgactcta**acattgactttttgggt**gattttgagcaagtctttggcctgctctgggctcc

ggctccctctcttgtcaaatgaggggggttcctttaagagcaggagtttgtaactttccttgcatcacaccgcctccttctgaaaagc

agaagagagcatgggttctctccctagagtttcacatgctatgcaggggctttgagatcccactaaagctgggctgtgaatgcta

ggttaaaggccttggaacaggctgtgcctggtggctcccgcctgtaatcccagcactttgggaggctgaggcgggcagatcagc

tgaggccaggagttcgagaccagcctggccaacatgttgaaaccccatctctactaaaaatacaaaaattagccagatgtggt

ggtgtacacctgtaatcccacctacttgggaggttgagacatgagaatcacttcaacctgggaggcagaggttgcagtgagcca

agattgtgcaactgtaccccaccctgggtgacacagtgacactctgtctcaaaaaaaataaataaataaataaagaccttgcac

cagattttgaaaatgttcaagatcccagtggagaaagagtttctaaaagaggatggtagaagcagttcttcacagaggagttcc

ctcctaattaccacggatcaagggaatgatagttgcaggtgtctgtagagtagatggagacatctgccatcacaggagggattg

tgagatgttgggggcaggcaagaaaggagggacaggctctggttaaatttctggaggggggctcagggttgacggggggtcatgt

ccttgcctgtgcccctctcccttgcagccagaggccagcacagccatccaggggacaccaggcccccacttcccagccccaaa

atagcctcacccatgtgtttccctcaaaaagggctagaaggctattaccccaatgcccctgctaccccagtctccaaaaagtaa

tttgcgatctacagggacttacagctaggtgtgatgtcagctgttgccaggcagaaaggggtttgggagaaccctgcagagatg

ttatgacaagctatgttctggggaacaaggtaacctgctgcagaccattagagctgtggcctgtgaaaaccccagggaatccca

tggcaaggagggagactgttttgtgaggaatggattcaggtctaggccacgcgtgaccttggataaatcactgcctctccctga

gtctcagtaataatataatagccataatagttctcatccactgtgtacctgtcaccaggcctcagccactgttagttagtactgtga

ttatttacatacgtaaaggctaaactcaattttaattcttacagtcaccctatgaatcaggtactttgttgtcccttttacagaag

aggaaattgaggtcatattggtaagtgacttacccaagtaatcatagctagtgagtggttggaggcagaatttgaaccaggact

agtctgacatcaaagtgttagctcactccaccatttggcccccctttcagaggtgcattttgtgggaagggggaggtgagacccctc

tgagctcagaccactcagtgatttacagacccaggtgggctcagtgttcagagtgcatgtcccccttgtctaatccttgtctagtttt

ccctttattcagttactgcttctaaagctggctatgtgtgcttaggccattccagactggagcctgttttggaggctttgtgaatgc

ttctgcctggcaacatctgtgtcacgattcctgctagctctggttggctcctgactttggctttttttttttttttttttaattgcagaaga

aggtagtgtaaatgaggaaataaagagggggggtgacaagagttggcactctaccagaaccaggggacaggcgttcaagatg

cataaagaggaagaaaagcagatgctaaacttaactacgaactcccctgagaattgcactgagtgtaatttatttatagagtgc

taacaatgtgctaggtgtggagccaagagctttaccctcatcatcttatttaatgtaacagcccttggagcagatgtttttacccga

ttttactcacacagagggttaatcaacttagatggtcccagaactaaccgactccagagccaggattccaatccaggttatgtgg

tctcgtcttcagggtgaatctctaggttcttaggcctcctgaggaactgtggtgcagagaggggaaggtactgtcccatgtcgtgt

gtgggaggtagagctggggctagaattgggtgtggggttgcagaagcactgaacaggtggaggatgaggctggagagcctga

acacaggctaggccacccgtgttccttgagaaggccaagaagtagaatttgttttcccttcctgtctggtctccccacttcccccc

cgccccccgcccaccgtcgggttttttgctgtttccttcccgaataactctcccttgggctgggccttatcccagaaagcacagact

ggcacgcagcaagccaatgagaatgagccagtctgtgatgtcatgagttaccaggcacaagggactcctggagcgtcctaaac

catccttgctggttgctggagtaggcagggtgcagcctcatccttgcagttcccaatgggtctgttttagtccgttttgtgccgcttt

aacagaataccacagactgcgtaatatacaaacaaacaaacaaacaaacaaaaaacacccagaaatttatttctcacagttct

ggaggggctttcttgttttttttttttttttttttgttatgttttttgagacagagtctcactctatcccccaggatggaatgcagtggtgt

gatcttggctcacagcaacctctgcctcccaggttcaagtgattctcctgcctcagcctcctgagtagctgggactacaggtgccc

74

accactacgcccggctaattttttcttttttttttgagatggagcctcactctgttgcccaggctggagtgcaatggcatgatctcag

ctcactgcaacctccaccttccaggttcaagggattctcctgcctcagcctcccgagtagctgggattacaggtgctcgccacca

cgtccggctaattttttgcatttttagtagagacagggtttcaccacgttggccaggctggtctcgaactcctgacctcaggtgatct

acctgtcttggcttcctaaagtgctgggattacaggtgtgagccaccgcgcccggcctagcttgctttcttttttttttctgagacag

tctccttctgtcacccaggctggagtgcagtgacgtgatctcggctcactgcaacctccgcctcccgggttcaagtgattctccca

cctcagcctcccaagtagctgggactagaggtgcacgccaccatgcctggctaagttttttgtatttttttgtagagacggggtttc

gccatgttggccagactggtctcgaactcctggcctcaagtgatccacctgccttggcctcccaaagtgctgggattacaggcgt

gagtcaccacacccagcctggtgaggggcttccttgatgtgtcctcatatggtggaaggcagaagggcaagagagaaccaact

ccttccatcaagccgtcttacagggcacctaatcccattcgtgagggagaaggccttgtggcccgatcacctcccaacaccttca

cattggcaacacctggactagggtcttagtagtcccacccctggggtcgtcaaagaacagatcagaaactcctggggtgtagg

ggcaggatccttacatcccctgtactgtaacatatgcatgcaccccccagtcgcatccaagtcccacagatacaagggcttgctg

aggtccgagggtctggatggtgagtccagggagcatgagaaattcatccaggccacacaggctcaggcacctatactcctgcc

cacccctgcaggatggatattaccatttgaaatactttggcgcttgcagtgagatcttgcttaaaagaagaaacccaccaaatt

ctagaatcgcagaggtgggggcacctatgtgacaaatgagggaacagaggcacagagattggcagggacatgctcaaagac

atggagagcctgggcagaggatggccccagagtgggtggtgcctggccctgcttcagcctgacattgtccaagcctcccccag

cccactcctaaaatggccaggcccccacttcttcccaggtgggctacattttctgctgtcccggtcccacttccttacctttctcctt

aaatctgatttcactgtgagggtgggaggttgggcagaggggtgaaaaccctcttccccttacttggtaagcacatacaatttttt

ttggttttgtcaattaaaaaaaataacatacctggttgatcctgccagcagaaaaaaaaaacaaaacaaaaggctgggcgcgg

tgactcacgcttgtaatcccagcactttgggaggctgaggtgggtggatcacgagatcaggagatcgagaccacggtgaaacc

ccgtctctatcaaaaatacaaaataattagccaggcgtggtggcgggtgcctgtagtcccagctactcggagaggctgaggaa

ggagaatggcataaacccaggaggtggaggttgcagtgagccgagatcgcgccactgcactccagcctgggcgacagagca

agactccgtctcaaaaaacaaaacaaaacaaaacaaaacaaaaaccaaacctgttccccttgagggcagccctgatcatcag

gagatccgcacggcaaagggagcctgggaagacccccccacctcaaccccctcatcgcacaggcaggaccctaaggtccaga

gaggacaaagagcatgcctctttgctcttctctgttcaactcaggcccgtggcagctgcagggtctctgccagggaagacttcag

gcacatcgatggttggaaggagggggtggggtgaggtcttcaagctgtgtttccatagccagcttactgtttgtgcttcaaatgcat

tgtttattaggggtggcttaggggtctaatggagatgggggtatacatttccccattcctagccagcccttggggcagccaccact

gtggatcctcctctgacctatcctcccacctcccacagcaagtgcagccgcggagccctgtacacaggcttttccatcctggtg

actctgctcctcgctggccaggccaccaccgcctacttcctgtaccagcagcagggccggctggacaaactgacagtcacctcc

cagaacctgcagctggagaacctgcgcatgaagcttcccaagcgtgcgtgcacccctacatcctgatacccccacctcccacc

atccctcaactcagagacccgcatccctgcacccagctgggcccactgtcctctccctccggtttggaattccagcccttcctcat

ctgggtctgataccctcctccctgggcaccggggccacacttaccctcgttcctgtccccacagctcccaagcctgtgagcaaga

tgcgcatggccaccccgctgctgatgcaggcgctgcccatgggagccctgccccaggggggtaaggacagccccagggtggtg

ggaggggcaaggttatcccgcctggatggaggacagtgccaaggggaggggcagggaagagagcccacctggggagggt

cctgactgctgcgggagggacagtgcctgcctcaggaagaatcgggctccccaggtgtggagggcacaggtgaagagtctctt

ggtgccatccctgggaggaaggctcagccctctacagtttacaaagtgcttctcatttcctatagcatctcactgtcctctcccatt

ctcaaagaccttgctatcatgcattgacaatatttatattcacaatactgtgctgtggacaaaaccctgggcaggaaagcttatgc

cagtttgaccaatgaggacattgaggcaggaagctaaagtgacttgctcgagctctcatgtttggaggtggcagagatggaac

cattgaccaagtgcctgcgattccagtctcttacctagatcccagcaaccggctcctgctccataccccctgctccagggaccag

ctctggtaaccttctgttacttcctcccacagcccatgcagaatgccaccaagtatggcaacatgacagaggaccatgtgatgca

cctgctccaggtgagtgcagggagctagctgggtggtcctgcctgcccacccaggaccctggccgggccaagctccaaggcct

gtatacctggctcatggcagactttcaacatgtgctctctggattactgaatacaagggtgacccttaaatgttatatgtagtctgg

cctctgcatttttgagataaacaggctcggctgggtgcagtggctcatgcctgtaatctcagcactttgggaggcctaggcaggc

agatcacctgaggccaggagttcgagaccagcctggccaacatagtgaaagtctctactaaaaatacaaaaattagccggatg

tggtggcagtcacctgtaatcccagctactcatgaggttgaggcaggagaatcgctttaacccgggaagtggaggttgcagtga

gccgagatcatcacgccactgcactccagcctgggtgacagagtgagactccatttcaaaaaataaaataaaataaaataaa

aaaagaaactgaggctcaaagtttggaagtgacttgctcaagacacccttttttctccccaaatttgtactgcgtgtctgtcctat

gccagcactgtgcaaggctctgggcccacaatggggagcaattggacccagctcctgaagccatggggatcccaggctagtgt

gggagacagacaagtgaccaggtgatgacagaagtgcaggggctttgtgaagaaagaggcagaggactcaacctagctgg

gggcagtcaaggttgggtcctcacctgctggagcaggtaagacctgaaagctgagcagccatccacagggaggacactgggg

cttgttggtagagctggcctggaatctggaactcgactcgcaacgctgcagtctcgtagctgaacagctggactacccttacccg

taccctcccctccctcccatcatggctgaagtccagggtttaggcttggtgtagctctggtccattcctcagaatggaagaccaa

gggaacgtgggctacacatttcacactgcctgccagggagctcttcttgaaccatccttcctgccctgctaccctgtagaatgctg

accccctgaaggtgtacccgccactgaaggggagcttcccggagaacctgagacaccttaagaacaccatggagaccataga

ctggaaggtcagcaggtttccctgcatgggaactctctctttcctctggtgtctagggcagggctaggagaggtggggtgagggt

gggctggggaagccattctcaggaagctgaaggggtttaccagcacttccaaaacctggatgctgcagagtatgtggggttca

gccccaggggtctttaagagggaaccaggctgcagctggacccgggtgttggggccctatttggccctgcatgtttctgtcccca

ggggcaaagccaggcagtgtaggggcttggtggtggccatcgaacctgacctccacctctatccgtattaggtctttgagagct

ggatgcaccattggctcctgtttgaaatgagcaggcactccttggagcaaaagcccactgacgctccaccgaaaggtacaggg

agtgggagctttagcgtgccagggcttctggaccctcggggctctcctgaagctgctgaggccggggcctccagcactccctgg

tcccagcaccgcggaatctcccatcctctcagctctcacttctttctcacttctctctctcctgtctgttccttcttggtttggctgtc

cccctcccctgacccacccccatcttgtctaagggttcctaagggcccacagaggcctgtcaccacagggtacaggtgacctctc

tcatagaggatggtacagcacagggtgcctggggtagaacctgcccgaaacactgcagaaaggaatccttgtaatgaccttgc

cccagtgctgcccgcaatccagtgagggcgccaaggtcacagctgctgccaagagagccttgggcgtttcccacctcatggac

aatgcagactaggatgttttttagacccaaagacagagtgctgtttctatcccggtgctgtctctaatttgctatgtaactttggaca

agtcccccttccctctaggattcagggtcctgaagtagaaggtcaaagggccaccctgcctggggcctcagtttctgcatcagatt

catagaaggcaccttacatgctatctccaactcctagctgatgcttaaaccccctctaagacatctccaacaaacggtaaaccccca

tttctacttgagaactcccagtaacaggaagcttagacttaccaaggtgcccattgcttttctctagaatcagaatcgcaagtgca

attccaaactgtaatggtgtttttttgtttgattgtttgttttgagatagagtctggctctgtcgcccaggctggagtgcagtggtgc

gatctcagctcactgcaacctcagcctcccgggttcaagcaattctcctgcctcggcctcccgagtagctgggattataggcatgt

gccaccacgcccagctaattttgtattttagtagagacagggtttcaccatgttggccaggctggtcttgaactcctgacctcaa

gtgatccgcccacctcggcctcccaaagtgctaagattacaggcatgagccactgtgcctggcccaaaactgtaatgggctttg

agtgtcagaagaaaccatcaacttctgaggtgaattggacacatggccattcacttccttttttgatctcagaccttgttggtctag

gcctcagtttcccatccgtgtgatggctggagtgagtaaagccacttgggaagaaggcattaagcccacagcagtggtgtgtg

ggtctttagctctgctcagaccctggttcagagctcactcactcactgtgtcctcatcatgcctgtcgcttcagtactgaccaagtg

ccaggaagaggtcagccacatccctgctgtccacccgggttcattcaggcccaagtgcgacgagaacggcaactatctgccac

tccagtgctatgggagcatcggctactgctggtgtgtcttccccaacggcacggaggtccccaacaccagaagccgcgggcac

cataactgcagtggtaagcagtggcactgtgccagtgtcagaggaccaggaaggactaggaaggttgaggggcaagaggtcc

cctctgaagcacatgggaccaggacacccaggatggcagctcctgggggcagtgacgtagtcatgcgtccagcccctttatccat

ccacccacctgtgcattcctatttgtccatttatcatctctcctcttacatgcactcatatttgttcacttatccgtcagtcctttcatct

gtgaatttcatccatccacgcgccatcctctagcatccaggagtcctacagacccacccatctcatttcatcacccccttctcactc

gagcccccattacacctctttgttgctgcttgcagctgtccttccctgggtacctgctttcccaggcactaagcctgtggctaggtg

ggaagcactgccctcaggcatcttgggtcaggtaggctgcacttcaagtgacaaacggacttgctgctcctttgcagagtcactg

gaactggaggacccgtcttctgggctgggtgtgaccaagcaggatctgggcccaggtaagggccttgcagaggggcatctggt

caccagcagctcatccccagcagggccagctcctttgtgggcaggtgaaggagtgtgacgctgggccactctccaacattcctg

ggatgtccatttcacagacgaaggaacagggttggggggctgtggggagttacacaaatccatgatggtcattattggacctga

gcaggggcaggggagggtggacagtccttaactgctctgcaggtccaggatgttagaaaggggcagggacaacaaatggg

tgaccccaacctcaacctgctgcttctctctccagtccccatgtgagagcagcagaggcggtcttcaacatcctgccagccccac

acagctacagctttcttgctcccttcagcccccagcccctcccccatctcccaccctgtacctcatcccatgagaccctggtgcctg

gctctttcgtcacccttggacaagacaaaccaagtcggaacagcagataacaatgcagcaaggccctgctgcccaatctccatc

tgtcaacaggggcgtgaggtcccaggaagtggccaaaagctagacagatccccgttcctgacatcacagcagcctccaacac

aaggctccaagacctaggctcatggacgagatgggaaggcacagggagaagggataaccctacacccagaccccaggctgg

acatgctgactgtcctctcccctccagcctttggccttggctttttctagcctatttacctgcaggctgagccactctcttccctttccc

cagcatcactccccaaggaagagccaatgttttccacccataatcctttctgccgacccctagttccctctgctcagccaagcttg

ttatcagctttcagggccatggttcacattagaataaaaggtagtaattagaacactctggctcctggctcttctgttgagatcca

gttcatttatcctaaagaggtaactaaagacgctaaatccagtcactgaggtgggagaacgcagcccaagtaacccacagaga

tgcctgcgggggccaggcaggcctgggggggggcttgtagtctttggacacagacatacttgattacatgtttaacacacacgaa

cactctacttccaaaaagaaataggctcttgcctagtttaagcacaagctcttcttggtacagctttcattttcctatgtccccaatt

tcatttctatactaagaacagtgtgaatgtgatgacagatggcaaccgacatagggcctccgtgtcaggatggacaggaagtag

gggactgtggagaactgggtgcccatgccttacctaaggggctgactgctactcagctctatctggttgctgccaggaggaagc

acagcccccgttgccagatgttatcttttaagaggagccagaaatcaagaactgcagtgaaatcaccccattttttaagtttaaa

gtcttgctgtgggtcaacaaggcccaatgaggacactactctgtagttctgttctagagcaacatgcttgagaggagagaggcc

agatggtggagggtctggcccccgggccccgagtcaagggtgcagctggatgccaacctgggtctccgacacctccagccatgc

attccttttctgcactgcctctggctgtttccagcatgaggctaaaagctctggagaccatctccaggacccctggtgttccacatg

gctctgggcccagggaagtgggcttctgcatttccaggctggagcagcagaccacctcctttgctatccctcccctccctcagtcc

acacccaagacatcatcctgtctttcatgttcatccacttcttgccctcaccaccacagaaccatggtgagctcgatttgaggggc

aactcacttccgattttcctagacctgctgtgcctgccacagctgttcaccaggcacctccagggagcctgcccctagctcgagg

cctaactcgaggattctcagcccccttccatttcttgtctttgggtatttacatgacttactctggcaacaggaaatcaaccatgacc

atcctttagttccatcctttagtttagatgccacctcatccaggaagccttccccaccaattcggcactcaggcacttcactgctca

ctgcctgtcccactaaccacggtcacccatagcctcaacatgcaggcaagagggaatcccgagtgtcggcgagacagtcccctt

ccaagaaacatgtgccccagatgcactgggagcttgcacccttcctccaagtcttttatttttttttttgagttattaaaaaaaaaac

aaaaaacacaaccaaaagaactgacgagaggaagacaaagcctcgggatcagccagagaccgagcggcacgggatgagg

cacaggtcaggcctagctggatctgtggagcaaggtggtgggaagtcccgaggcctttgggcatgtggggtctgtggaaggag

aggaggtcacttaaaaaaatacatatactgtacacatctatatgaagcgtcctgtgttggggagtggggtgaggtcatgggag

ctggtccactttctccttggaccagcctcccccgcaggctcactggcttggccagcacagggacccctgaagctggctgtcttcttc

tgtggagcgagggaaggtaacagtttagttccaacccatggtgggcagaggtcggtggaggtcagaggcatggggatggcca

ctgctcggctaggaaatccctaggggaaggaaacacatgcagatgtcaaatttcctgacccttttcccctgcaaggaaaatgaa

atctggcatctccaatgcaggggggaggcattaaccacagtggacattttttagcacttggaatgttctgagaactgggctgagt

cctttagtgtttagtttagcttc

-1000 Primers   F – C T T G C C A A C A C A G T C C C A T C

                     R – G T G A T G T T T A C G G A G G C C C T

Transcription Start Site   F - A T G C A C A G G A G G A G A

                     R – A C C C A A A A A G T C A A T G T

Appendix 2: *Drosophila melanogaster* GapDH gene primers for 132 bp amplicon.

tctagaacttaaagaactaacttggcaacaatgtaaagaaaaacttttagtgcaagatgaaggaattttatacttgacaatgcta

atttttgtataaacttcattacgtaaagggtcctacaacccatatatatcattagttgttagaggctcttttttaaggcgcttatataa

tcaaaccctttgtaaaaattaaagttttaaatggaattctaatcgatttatttcacattagctttatttaagtgtgacctacgcagaa

agctagcgaaatactcatcaaccctcccccgccatcgccagcgccattctcctaatttgcgaaaaaagctccgggaaaaggaa

aaagcggcagtcgtaatagcgaactgaaactgaacgagagtaaaagttaaaagacagcaggaactcagccatgtcgaagat

cggaatcaacggatttggccgcatcggccgcttggtgctccgcgccgccatcgataagggcgcctccgtggtggccgtcaacga

tcccttcatcgatgtcaactacatggtttacctgtttaaattcgactcgactcacggtcgtttcaagggcaccgttgcggctgagg

gcggattcctggtggtgaacggccagaagatcaccgtgttcagcgagcgcgacccggccaacatcaactgggccagtgctgga

gccgagtatgtggtggagtccaccggagtgttcaccaccatcgacaaggcgtccacccacttgaagggcggtgccaagaaggt

catcatctcggccccatccgccgatgcgcccatgttcgtgtgcggcgttaacctggacgcctacagccccgacatgaaggtggt

ctccaacgcctcgtgcaccaccaactgcctggctcccctggccaaggtcatcaatgacaacttcgagatcgtcgagggtctgat

gaccaccgtgcacgccaccactgccacccagaagaccgtcgacggtccctctggcaaactgtggcgcgatggacgtggcgcc

gcccagaacatcatcccggccgccaccggagccgccaaggctgtgggcaaggtcatccccgccctgaacggcaagctgaccg

gcatggctttccgcgtgcccacgcccaatgtctccgttgtggatcttaccgtccgcctgggcaagggagccacctatgacgaaat

caaggctaaggtcgaggaggcctccaagggacccctgaagggaatcctgggctacaccgacgaggaggtggtctccaccgac

ttcctcagcgacacccattcgtctgtgttcgacgccaaggctggcatttcgctgaacgataagttcgtcaagctaatctcgtggta

cgacaacgagttcggttactccaaccgcgtcatcgacctgatcaagtatatgcagagcaaggactaaactagccacaactatcg

tacaaacccggccagcagctggtcgggaatcactgttgcataatccgcaaggggcgcaattgaggatgctttttttttcaaatgc

acaaaatagccataattacgcaaagtacatatttattgttgttgaataaagtatcgccctgttactgtatttaaatgccagctattt

gaaaaataaaacttgtgcatcgaatgaagaacgttgcttgtgtaagcccaaggcgtgtgaaggttatatcgggcggacttgctg

agatatgcagtgaatacacaaataagtatctgtgcttcgcctgggggagtcatcccaatcccaaaggaacgatgtttctatttat

gtcatcatatgtatgtatgtacgtttgttgttgccaactctggcagtctcgaaattcgaagataacgaacaaatatattctcttacc

gctgcactttcaaacctgtgcggatgctgtttgtttgtctcggatttttgattccttcggcaaaacaacataaatacacataacgacc

aattgaactcagaccatataaacatatatgcttatatatactttatacttttttggtggataaagggtttgttttcatatatcgacctg

cgttcctatacaaagaggctttcttatgagcataaccaaggtcacgaaaggtcatacatatacttgctgcttggtacataacttat

cggcattacagaattgaatcgat

Forward primer: AGCGAACTGAAACTGAACGA

Reverse primer: TATTCCCGCGGAGGCA

Total amplified sequence: 133bp

Appendix 3: *Drosophila melanogaster* GapDH gene primers for 211 bp amplicon.

tctagaacttaaagaactaacttggcaacaatgtaaagaaaaactttagtgcaagatgaaggaattttatacttgacaatgcta

atttttgtataaacttcattacgtaaagggtcctacaacccatatatatcattagttgttagaggctcttttttaaggcgcttatataa

tcaaacccttttgtaaaaattaaagtttttaaatggaattctaatcgatttatttcacattagctttatttaagtgtgacctacgcagaa

agctagcgaaatactcatcaaccctcccccgccatcgccagcgccattctcctaatttgcgaaaaaagctccgggaaaaggaa

aaagcggcagtcgtaat<mark>agcgaactgaaactgaacga</mark>gagtaaaagttaaaagacagcaggaactcagccatgtcgaagat

<mark>cggaatcaacggatttggccgcatcggccgcttggtgctccgcgccgccatcgataagggcgcctccgtggtggccgtcaacga</mark>

<mark>tcccttcatcgatgtcaactacatggtttacctgtttaaattcga</mark>ctcgactcacggtcgtttcaagggcaccgttgcggctgagg

gcggattcctggtggtgaacggccagaagatcaccgtgttcagcgagcgcgacccggccaacatcaactgggccagtgctgga

gccgagtatgtggtggagtccaccggagtgttcaccaccatcgacaaggcgtccacccacttgaagggcggtgccaagaaggt

catcatctcggccccatccgccgatgcgcccatgttcgtgtgcggcgttaacctggacgcctacagccccgacatgaaggtggt

ctccaacgcctcgtgcaccaccaactgcctggctcccctggccaaggtcatcaatgacaacttcgagatcgtcgagggtctgat

gaccaccgtgcacgccaccactgccacccagaagaccgtcgacggtccctctggcaaactgtggcgcgatggacgtggcgcc

gcccagaacatcatcccggccgccaccggagccgccaaggctgtgggcaaggtcatccccgccctgaacggcaagctgaccg

gcatggctttccgcgtgcccacgcccaatgtctccgttgtggatcttaccgtccgcctgggcaagggagccacctatgacgaaat

caaggctaaggtcgaggaggcctccaagggacccctgaagggaatcctgggctacaccgacgaggaggtggtctccaccgac

ttcctcagcgacacccattcgtctgtgttcgacgccaaggctggcatttcgctgaacgataagttcgtcaagctaatctcgtggta

cgacaacgagttcggttactccaaccgcgtcatcgacctgatcaagtatatgcagagcaaggactaaactagccacaactatcg

tacaaacccggccagcagctggtcgggaatcactgttgcataatccgcaaggggcgcaattgaggatgctttttttttcaaatgc

acaaaatagccataattacgcaaagtacatatttattgttgttgaataaagtatcgccctgttactgtatttaaatgccagctattt

gaaaaataaaacttgtgcatcgaatgaagaacgttgcttgtgtaagcccaaggcgtgtgaaggttatatcgggcggacttgctg

agatatgcagtgaatacacaaataagtatctgtgcttcgcctgggggagtcatcccaatcccaaaggaacgatgtttctatttat

gtcatcatatgtatgtatgtacgtttgttgttgccaactctggcagtctcgaaattcgaagataacgaacaaatatattctcttacc

gctgcactttcaaacctgtgcggatgctgtttgtttgtctcggattttgattccttcggcaaaacaacataaatacacataacgacc

aattgaactcagaccatataaacatatatgcttatatatactttatacttttttggtggataaagggtttgttttcatatatcgacctg

cgttcctatacaaagaggctttcttatgagcataaccaaggtcacgaaaggtcatacatatacttgctgcttggtacataacttat

cggcattacagaattgaatcgat

Forward primer: AGCGAACTGAAACTGAACGA

Reverse primer: GAAACGACCGTGAGTCGAG

Total amplified sequence: 212bp

Appendix 4: *Drosophila melanogaster* GapDH gene primers for 310 bp amplicon.

tctagaacttaaagaactaacttggcaacaatgtaaagaaaaacttttagtgcaagatgaaggaattttatacttgacaatgcta

atttttgtataaacttcattacgtaaagggtcctacaacccatatatatcattagttgttagaggctctttttaaggcgcttatataa

tcaaacccttgtaaaaattaaagtttttaaatggaattctaatcgatttatttcacattagctttatttaagtgtgacctacgcagaa

agctagcgaaatactcatcaaccctcccccgccatcgccagcgccattctcctaatttgcgaaaaaagctccgggaaaaggaa

aaagcggcagtcgtaat<mark>agcgaactgaaactgaacga</mark>gagtaaaagttaaaagacagcaggaactcagccatgtcgaagat

cggaatcaacggatttggccgcatcggccgcttggtgctccgcgccgccatcgataagggcgcctccgtggtggccgtcaacga

tcccttcatcgatgtcaactacatggtttacctgtttaaattcgactcgactcacggtcgtttcaagggcaccgttgcggctgagg

gcggattcctggtggtgaacggccagaagatcaccgtgttcagcgagcgcgacccggc<mark>caacatcaactgggccag</mark>tgctgga

gccgagtatgtggtggagtccaccggagtgttcaccaccatcgacaaggcgtccacccacttgaagggcggtgccaagaaggt

catcatctcggccccatccgccgatgcgcccatgttcgtgtgcggcgttaacctggacgcctacagccccgacatgaaggtggt

ctccaacgcctcgtgcaccaccaactgcctggctcccctggccaaggtcatcaatgacaacttcgagatcgtcgagggtctgat

gaccaccgtgcacgccaccactgccacccagaagaccgtcgacggtccctctggcaaactgtggcgcgatggacgtggcgcc

gcccagaacatcatcccggccgccaccggagccgccaaggctgtgggcaaggtcatccccgccctgaacggcaagctgaccg

gcatggctttccgcgtgcccacgcccaatgtctccgttgtggatcttaccgtccgcctgggcaagggagccacctatgacgaaat

caaggctaaggtcgaggaggcctccaagggacccctgaagggaatcctgggctacaccgacgaggaggtggtctccaccgac

ttcctcagcgacacccattcgtctgtgttcgacgccaaggctggcatttcgctgaacgataagttcgtcaagctaatctcgtggta

cgacaacgagttcggttactccaaccgcgtcatcgacctgatcaagtatatgcagagcaaggactaaactagccacaactatcg

tacaaacccggccagcagctggtcgggaatcactgttgcataatccgcaaggggcgcaattgaggatgctttttttttcaaatgc

acaaaatagccataattacgcaaagtacatatttattgttgttgaataaagtatcgccctgttactgtatttaaatgccagctattt

gaaaaataaaacttgtgcatcgaatgaagaacgttgcttgtgtaagcccaaggcgtgtgaaggttatatcgggcggacttgctg

agatatgcagtgaatacacaaataagtatctgtgcttcgcctgggggagtcatcccaatcccaaaggaacgatgtttctatttat

gtcatcatatgtatgtatgtacgtttgttgttgccaactctggcagtctcgaaattcgaagataacgaacaaatatattctcttacc

gctgcactttcaaacctgtgcggatgctgtttgtttgtctcggattttgattccttcggcaaaacaacataaatacacataacgacc

aattgaactcagaccatataaacatatatgcttatatatactttatactttttggtggataaagggtttgttttcatatatcgacctg

cgttcctatacaaagaggctttcttatgagcataaccaaggtcacgaaaggtcatacatatacttgctgcttggtacataacttat

cggcattacagaattgaatcgat

Forward primer: AGCGAACTGAAACTGAACGA

Reverse primer: CTGGCCCAGTTGATGTTG

Total amplified sequence: 310bp

Appendix 5: *Drosophila melanogaster* GapDH gene primers for 410 bp amplicon.

tctagaacttaaagaactaacttggcaacaatgtaaagaaaaactttagtgcaagatgaaggaattttatacttgacaatgcta

atttttgtataaacttcattacgtaaagggtcctacaacccatatatatcattagttgttagaggctcttttttaaggcgcttatataa

tcaaaccctttgtaaaaattaaagttttaaatggaattctaatcgatttatttcacattagctttatttaagtgtgacctacgcagaa

agctagcgaaatactcatcaaccctcccccgccatcgccagcgccattctcctaatttgcgaaaaaagctccgggaaaaggaa

aaagcggcagtcgtaat<mark>agcgaactgaaactgaacga</mark>gagtaaaagttaaaagacagcaggaactcagccatgtcgaagat

<mark>cggaatcaacggatttggccgcatcggccgcttggtgctccgcgccgccatcgataagggcgcctccgtggtggccgtcaacga</mark>

<mark>tcccttcatcgatgtcaactacatggtttacctgtttaaattcgactcgactcacggtcgtttcaagggcaccgttgcggctgagg</mark>

<mark>gcggattcctggtggtgaacggccagaagatcaccgtgttcagcgagcgcgacccggccaacatcaactgggccagtgctgga</mark>

<mark>gccgagtatgtggtggagtccaccggagtgttcaccaccatcgacaaggcgtccacccacttgaagggcggtg</mark><mark>ccaagaaggt</mark>

<mark>catcatctcg</mark>gccccatccgccgatgcgcccatgttcgtgtgcggcgttaacctggacgcctacagccccgacatgaaggtggt

ctccaacgcctcgtgcaccaccaactgcctggctcccctggccaaggtcatcaatgacaacttcgagatcgtcgagggtctgat

gaccaccgtgcacgccaccactgccacccagaagaccgtcgacggtccctctggcaaactgtggcgcgatggacgtggcgcc

gcccagaacatcatcccggccgccaccggagccgccaaggctgtgggcaaggtcatccccgccctgaacggcaagctgaccg

gcatggctttccgcgtgcccacgcccaatgtctccgttgtggatcttaccgtccgcctgggcaagggagccacctatgacgaaat

caaggctaaggtcgaggaggcctccaagggacccctgaagggaatcctgggctacaccgacgaggaggtggtctccaccgac

ttcctcagcgacacccattcgtctgtgttcgacgccaaggctggcatttcgctgaacgataagttcgtcaagctaatctcgtggta

cgacaacgagttcggttactccaaccgcgtcatcgacctgatcaagtatatgcagagcaaggactaaactagccacaactatcg

tacaaacccggccagcagctggtcgggaatcactgttgcataatccgcaagggggcgcaattgaggatgctttttttttcaaatgc

acaaaatagccataattacgcaaagtacatatttattgttgttgaataaagtatcgccctgttactgtatttaaatgccagctattt

gaaaaataaaacttgtgcatcgaatgaagaacgttgcttgtgtaagcccaaggcgtgtgaaggttatatcgggcggacttgctg

agatatgcagtgaatacacaaataagtatctgtgcttcgcctggggggagtcatcccaatcccaaaggaacgatgtttctatttat

gtcatcatatgtatgtatgtacgtttgttgttgccaactctggcagtctcgaaattcgaagataacgaacaaatatattctcttacc

gctgcactttcaaacctgtgcggatgctgtttgtttgtctcggattttgattccttcggcaaaacaacataaatacacataacgacc

aattgaactcagaccatataaacatatatgcttatatatactttatactttttggtggataaagggtttgttttcatatatcgacctg

cgttcctatacaaagaggctttcttatgagcataaccaaggtcacgaaaggtcatacatatacttgctgcttggtacataacttat

cggcattacagaattgaatcgat

Forward primer: AGCGAACTGAAACTGAACGA

Reverse primer: CGAGATGATGACCTTCTTGG

Total amplified sequence: 410bp

Characterisation and Modelling of Mutant PALB2 Proteins.

Abstract

Fanconi anaemia (FA) is a rare, recessive, inherited disease characterised by the biallelic instability of one of 16 proteins associated with the DNA double-stranded break (DSB) repair pathway – FANCA, FANCB, FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ, FANCL, FANCM, FANCN, FANCO, FANCP and FANCQ. Affected individuals with gene deficiency affecting the FA core complex have a similar clinical appearance, while patients whose defect (eg in FANCD1, FANCN) is in the pathway downstream of this show a more severe presentation including the occurrence of embryonal tumours and may not survive infancy. The *FANCN* gene otherwise known as PALB2 (Partner and Localiser of BRCA2), was originally described as a FA associated gene in 2006 and most of the laboratory work reported to date has been carried out on cells from a patient, EUFA1341. I describe here work towards an investigation of the role of two mutant *PALB2* proteins identified in an unusual FANCN family, with a milder presentation of Fanconi anaemia. I was able to confirm the presence of the correct *PALB2* mutant DNA sequence following transfection of separate constructs into U2OS cells. I was also able to show PALB2 protein expression from each *PALB2* mutant construct in these cells as the first step to examining the functions of these separate mutant PALB2 proteins.

1. Introduction

1.1 Fanconi Anaemia

Fanconi anaemia (FA) was originally identified as a familial panmyelopathy by Dr Fanconi (Lobitz 2006) and later confirmed as inherited in an autosomal recessive manner. The typical clinical presentations in FA patients include marked growth retardation, congenital malformations (most famously absent thumbs and radial aplasia), progressive bone marrow failure (the origin of the name), and tumour predisposition, particularly, myeloid leukaemia and also some epithelial tumours.

It was not until 1964 however, that spontaneous chromosome instability was identified as being associated with FA (Schroeder *et al.* 1964, Lobitz 2006). At about the same time it was discovered that FA patient's cells were chromosomally unusually sensitive to DNA crosslinking agents like Mitomycin C and subsequently, diepoxybutane (DEB). It was not until 1991, however, that the first Fanconi anaemia defective gene was identified, FA-A, on chromosome 20q (Mann *et al.* 1991). Since then, 15 additional genes have been identified which, when mutated, cause FA. These are now named *FANCB*, *FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ, FANCL, FANCM, FANCN, FANCO, FANCP* and *FANCQ* (Strathdee *et al.* 1992, Joenje *et al.* 1995, Levitus *et al*. 2004, Reid *et al.* 2007, Xia *et al.* 2006a, Vaz *et al.* 2010, Cybulski and Howlett 2011, Osorio *et al.* 2013). Fanconi anaemia occurs when the offspring of two carrier parents inherit both FANC mutations (biallelic mutation). Heterozygous carriers of some *FANC* genes are also known to have a predisposition to some cancers (Reid *et al.* 2007, Tischkowitz and Xia 2010).

1.2 The Fanconi anaemia signalling pathway

Over the past 20 years the proteins associated with the many *FANC* genes have been studied in detail and it is now clear that there are many protein complexes

90

which form during the repair pathway. The first to be defined was the FA core complex. The core complex was initially identified when Kupfer *et al.* (1997) identified an interaction between FANCA and FANCC, enabling their localisation to the nucleus which could not be achieved as a single protein. Further work identified FANCG as being part of the same complex (Garcia-Higuera *et al.* 1999). Eight proteins, FANCA, FANCB (FAAP95), FANCC, FANCE, FANCF, FANCG, FANCL (PHF9/ FAAP43) and FAAP100 (Kupfer *et al.* 1997, Garcia-Higuera *et al.* 1999, Garcia-Higuera *et al.* 2001, de Winter *et al*. 2000, Meetei *et al.* 2003a, Meetei *et al.* 2003b, Ling *et al*. 2007) have been identified as the proteins which function as the Fanconi anaemia core complex. Following exposure of cells to a DNA damaging agent, the FANC core complex facilitates ubiquitination of FANCD2 via FANCL which allows an association with FANCI (the ID2 complex). The ID2 protein complex is targeted to chromatin where it interacts with BRCA1/BRCA2 and other proteins to allow repair of the cross link.

A chromatin bound DNA translocase protein, FANCM (FAAP250) in association with FAAP24 (Fanconi anaemia-associated protein 24), recognises DNA damage; in response FANCM is phosphorylated at Ser1045 in an ATR-dependant reaction (Kim *et al.* 2008, Singh *et al.* 2013). FANCM is able to recognise a stalled replication fork and stabilise it until downstream proteins are localised to damaged foci. Phosphorylation of FANCM activates the DNA damage response by assembling of the FA core complex. In turn FANCM is recruited to the site of DNA damage where the FANCM-FAAP24 complex acts as an anchor, stably binding the core complex in position. Meanwhile, FANCD2 and FANCI are phosphorylated in another ATR-dependent reaction following DNA-damage response activation (Ishiai *et al.* 2008). Monoubiquitination of FANCD2 (FANCD2ub) via the E3 ubiquitin ligase, FANCL of

91

the core complex, at K561 sequentially follows the assembly of the core complex; an action essential for the targeting and co-localisation of FANCD2 within the nuclear foci (Kim *et al.* 2008, Garcia-Higuera *et al.* 2001, Ling *et al*. 2007, Meetei *et al.* 2003b). FANCD2ub is consequently transported to chromatin at the site of DNA damage where it signals for the pre-formed complex between FANCN-FANCD1 (PALB2-BRCA2) to form FANC(CD1-N-CD2) (Zhang *et al* 2009a, Zhang *et al*. 2009b, Sy *et al*. 2009a), thereby initiating DNA repair via homologous recombination through interactions between proteins such as RAD51 (Xia *et al*. 2006a).The Fanconi anaemia pathway is complex, it is involved in repair of DNA crosslinks, recognition of stalled replication forks and in the repair of DNA double strand breaks. Phenotypically, Fanconi anaemia patients diagnosed with mutatons within proteins of the core complex (FANC-A, -B, -C, -E, -F, -G, -L, -M) display a less severe version of the disease compared to patients with mutated proteins downstream of the core complex, i.e. FANCN, FANCD1 and FANCJ. FA core complex patients on average, show onset of squamous cell carcinoma (SCC) at around 31 years (Kutler *et al.* 2003), in contrast Reid *et al*. (2007) published a collection of several FA-N patients and their phenotypes (extended version in table 1.1), the median onset of tumour growth started at 2 years, these patients also suffered with more severe and often multiple tumours. This difference in severity and age of growth suggests a more aggressive disease which can be easily differentiated from core complex related FA (Howlett 2007).

1.3 PALB2
Located on chromosome 16q12.2, PALB2 is a 131 kDa, 1,186 residue nucleoprotein involved in the DSB repair pathway (Xia *et al.* 2006a). The structure of PALB2 is complex and it is composed of several domains and motifs which are essential for

various protein-protein interactions required within the DSB pathway. Figure 1.1 shows the structure of PALB2. PALB2 N-terminal sequence from a.a 6-90 reveals a conserved coiled coil domain (see figure 1.3) (Sy *et al.* 2009a, Zhang *et al.* 2009a, Zhang *et al.* 2009b). Similar to that of the DNA double helix, a coiled coil domain is a protein domain in which a collection of heptad repeats crystallise to form an α-helix. When 2-4 α-helices are present they form a super-helix. The amino side chains perform in a 'knobs into hole' mechanism whereby one amino acid (the knob) will lodge into the surrounding area of four residues on an equivalent helix (see figure 1.2). The 'knob' residue is located at either the 'a' or 'd' position in an a-g heptad repeat structure where each repeat includes two coils (Lupas 1996, Liu *et al.* 2006). Residues 395-446 encompass a structural motif broadly conserved among vertebrates known as the Chomatin Association Motif, ChAM (figure 1.4). The 52 amino acid sequence mediates PALB2-chromatin interaction via nucleosome binding and deletion of this motif has been shown to cause disruption to chromatin localisation of PALB2 in damaged and normal cells (Bleuyard *et al*. 2012). The PALB2 C-terminus is comprised of 7 chain WD40 repeat sequence, whereby each repeat constitutes a symmetrical, 4 tier anti-parallel β-sheet; together it combines in a velcro-like locking mechanism where the N-terminal repeat links via hydrogen bonding to the C-terminus on the 7th sheet creating a 7 spear β-propeller (Fulop and Jones 1999, Oliver *et al.* 2009). This domain is essential for stability of the protein and has been implicated in many protein-protein interactions required for DSB repair such as BRCA2, RAD51, RAD51C (figure 1.5) (Park *et al*. 2013) and Polyermase η during DNA synthesis at blocked replication forks (Buisson *et al.* 2014).

PALB2:
Chromosome 16q12.2;
131 kDa; 1,186 a.a
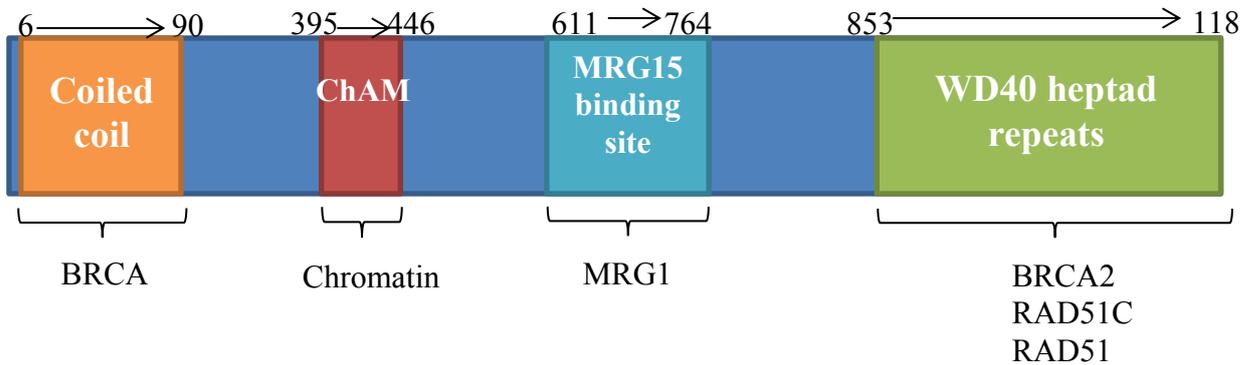
Figure 1.1 PALB2 basic structure identifying the locations of the coiled coil domain, ChAM motif and WD40 repeats. The figure also shows where BRCA1, chromatin, MRG15, BRCA2, RAD51 and RAD51C bind.
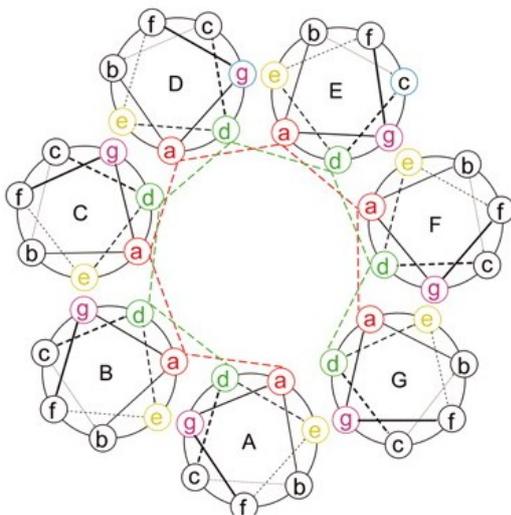
Figure 1.2 [Reproduced from Liu *et al.* 2006, figure 3F] A Standard Coiled Coil Motif. The diagram shows 7 heptad repeats where positions 'a' and 'd' form the side chains needed for the 'lock in hole' mechanism.
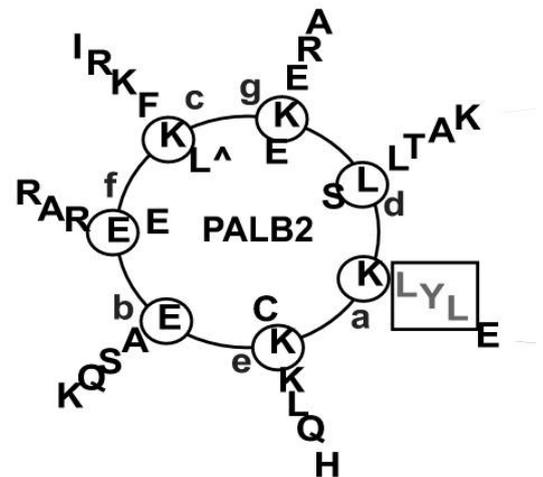
Figure 1.3. [Modified from Sy *et al*. 2009a, figure 2A] PALB2 coiled-coil domain. Located at the N-terminus of PALB2, the heptad repeat structure interacts with BRCA1 at the isolated LYL 'a' position.

94
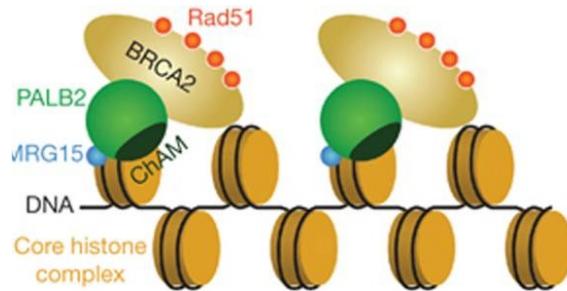
Figure 1.4 [modified from Bleuyard *et al.* 2012, figure 5] ChAM in association with PALB2 and chromatin. The diagram shows the location of ChAM on PALB2 with respect to chromatin, BRCA2, MRG15 and RAD51
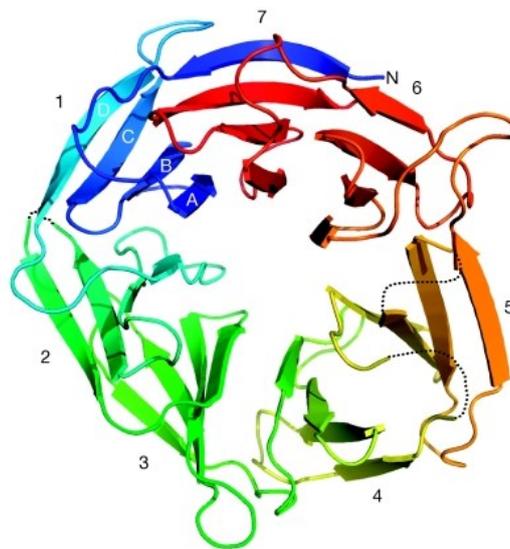
Figure 1.5 [Reproduced from Oliver *et al.* 2009, figure 1b] PALB2 WD40 domain. Each β-propeller (1 - 7) is made up of 4 tiers (A – D), the arrows indicate the direction N-terminal to C-terminal. The blue tier (7D) represents the 'velcro-like' locking mechanism.

1.4 BRCA1/BRCA2 Complex

A major breakthrough came when BRCA2 was identified as FANCD1, a discovery which confirmed a long debate about the association between the BRCA and FA proteins. Howlett *et al.* (2002) sequenced both BRCA1 and BRCA2 in cells of FA-B and FA-D1 patients. Normal sequences were found in the BRCA1 gene but biallelic mutations were discovered in the BRCA2 gene leading to the finding that FA-D1 is characterised by defects in BRCA2.

PALB2 functions in homologous recombination as one part of a multi-subunit complex with BRCA1 and BRCA2 (FANCD1). Shortly after the discovery of PALB2, Reid *et al.* (2007) and Xia *et al*. (2006b) identified PALB2 as the gene for Fanconi anaemia subtype-N. This discovery allowed for diagnosis of FA-N in many patients who were of a previously unclassified FA subtype.

Previous studies have shown that interaction between PALB2, BRCA1 and BRCA2 proteins is an essential component of the DSB pathway. Using co-IP these studies have shown that BRCA1 binds to N terminus of PALB2 (Xia *et al.* 2006b, Reid *et al*. 2007, Sy *et al.* 2009a, Zhang *et al.* 2009a, Zhang *et al.* 2009b) in the coiled coil motif. BRCA1 also contains a C-terminal, double helical, coiled coil domain between exons 11 and 13 at residues 1364-1437 (Clark *et al.* 2012). Coiled coil domains are known for their role in protein-protein interactions and the BRCA1-PALB2 interaction is no exception, refinement experiments showed that the second α-helix in BRCA1 (a.a. 1397-1424) is responsible for formation of a BRCA1-PALB2 heterodimer complex (Sy *et al*. 2009a).

BRCA2 binds to the PALB2 C-terminus. Zhang *et al.* (2009b) originally specified the point of interaction to the 3rd and 4th WD40 repeat, however more recent evidence located this interaction to the 4th and 5th β-sheet (Oliver *et al.* 2009). These residues

have since been identified as a hydrophobic 'pocket' found between the 4th and 5th spear in the β-propeller. Residues Val 1019, Met 1022, Ala 1025, Ile 1037, Leu 1046, Lys 1047, Leu 1070, Pro 1097 and Lys 1098 have so far been implicated in the functional formation of this pocket. Immunoprecipitation using mutant PALB2 proteins, A1025R, M1022A and M1022R indicated that the alanine at amino acid position 1025, located at the bottom of the hydrophobic pocket on PALB2 has a more pivotal role in the bond between BRCA2 and PALB2, abolishing the BRCA2 binding affinity in mutant cells (Oliver *et al.* 2009). Xia *et al.* (2006a) were first to locate the site of PALB2 binding within the first 40 residues of the amino tail on BRCA2, further evidence identified that a tryptophan at residue 31 located at the N-terminus of BRCA2 is responsible for binding to PALB2 alanine 1025, therefore indicating that this residue also holds a crucial role in the BRCA2-PALB2 interaction.

1.5 Other PALB2 interacting proteins
RAD51, a 339 amino acid recombinase protein, is the central protein involved in homologous recombination. It forms helical nucleoprotein filaments at the site of DNA damage in an ATP-dependent reaction (Ogawa *et al.* 1993, Galkin *et al*. 2006). BRCA2 plays a critical role in the localisation of RAD51 to sites of DSB; however PALB2 is required for enhanced RAD51 function (Buisson *et al.* 2010). PALB2 utilises the role of its N-terminal WD40 repeat domain to mediate the direct interaction with RAD51 and the location has been refined to residues 853-1186 on PALB2 (Park *et al*. 2013). Additionally, a RAD51 binding site has been observed at the C-terminal coiled coil located at residues 1-200. RAD51 was reported to bind to PALB2 at amino acids 184-257.

RAD51C, another FANC protein categorised as complementation group FA-O, is one of five paralogues of RAD51 which also include, RAD51B, RAD51D, XRCC2

and XRCC3 and all play a role in HR. It has been proposed that RAD51C, PALB2 and BRCA2 work in a complex as the central component controlling homologous recombination and in turn, genomic stability. The 376 aa polypeptide, RAD51C, has been identified as a direct binding partner of PALB2. Mutant PALB2 lacking blades 5-7 from the WD40 repeat domain was unable to co-immunoprecipitate Rad51C, therefore, indicating these residues (1098-1186) on PALB2 as essential for RAD51C binding affinity (Dosanjh *et al*. 1998, Park *et al.* 2013).

MRG15 has been indicated as another direct binding partner of PALB2 within the BRCA complex where residues 611-764 of PALB2 are required for MRG15 interaction (Sy *et al*. 2009b). Conflicting reports on the effects of MRG15 and its function have been published. In one study Sy *et al.* (2009b) reported that PALB2 mutant cells deficient in the MRG15 binding domain demonstrated no difference in BRCA2 localisation to the site of DNA damage after MMC treatment compared to unperturbed cells, and that these mutant cell lines conferred resistance to MMC. Conversely, Hayakawa *et al.* (2010) later suggested that MRG15 acted independently of BRCA1 in localising PALB2, BRCA2 and RAD51 to DNA damage foci and that depletion of MRG15 also lead to the decrease of BRCA2 to DNA damage site therefore inducing hypersensitivity to MMC.

1.6 Fanconi anaemia, caused by mutation of *FANCN*.
In 2006, Xia et al. (2006b) published the first case of Fanconi anaemia caused by mutation of PALB2 which they designated *FANCN*. Having ruled out all other *FANC* genes known at that time, they tested for mutation of *PALB2* as a plausible candidate gene, and showed that their patient, indeed had biallelic *PALB2* mutations. Since then, more studies have revealed more cases of the FA-N subtype with PALB2 mutations (see table 1.1), including an interesting mutation giving a truncated

protein just 3 residues before the original stop codon. This mutant is noteworthy as it represents the importance of the WD40 domain located at the C-terminus of PALB2. The locking mechanism found within the WD40 domain, where the N-terminus joins the C-terminus as the 4th tier on the 7th blade, is removed. In turn, the loss of this mechanism renders the protein unstable and PALB2 can no longer localise to the site of DSB on chromatin, leading to ineffective DSB repair and therefore, chromatin instability (Xia *et* al. 2006b, Reid *et al*. 2007, Oliver *et al*. 2009).

Analysis of one FA-N patient, EUFA1341, identified two novel mutations, the paternally inherited deletion of exons1-10 and the maternal mutation, a nonsense nucleotide mutation resulting in a truncated protein at residue Y551X (Xia *et al.* 2006b). In their study Xia *et al*., identified a revertant cell line in which the maternal mutation had undergone a further change (as a result of growth of the cultured cells in the presence of a cross linking agent) and had become resistant to the killing effect of this agent. The revertant protein occurred as a result of loss of exon 4 and retained both N- and C-terminal sequences of PALB2 and therefore potential function. These mutants (the germline Y551X and the revertant due to loss of exon 4 mutation) are significant here since they may have similar end products to the likely proteins expressed from the mutations in the family that I studied. Xia *et al.* concluded that, the PALB2 revertant with the large exon 4 deletion was functional. However, the N-terminal Y551X PALB2 mutant was unable to function. Moreover, in another publication by Reid *et al.* (2007) a truncated protein, Y1183X, depleted PALB2 activity due to the loss of just the final 3 residues in the C-terminal WD40 repeat locking mechanism destabilising the protein, and in turn PALB2 activity is lost.

1.7 Lymphoma

Lymphoma is a cancer affecting lymphoid cells of the immune system and can occur through a variety of mechanisms most notably through the inhibition of tumour suppressor genes or the activation of oncogenes. All lymphomas arise as a consequence of a chromosome translocation involving an immune system gene and an oncogene. Prior loss of a particular tumour suppressor gene may increase the frequency of such translocations. When the body is invaded by an infectious agent the immune response is activated, this initiates lymphocytes to recognise the infection and increase production of antibodies as a line of defence. T cells control the pathway by recognising the antigen and in turn, secreting other proteins which initiate the full immune response. B cells produce two types of immunoglobulins with antigen binding capacity, the B-cell receptor and antibodies, the first is an integral membrane protein which sits on the membrane and lies dormant until acted upon, the antibody is a secreted protein with duplicated antigen binding specificity. The production of both functional immunoglobulin and T cell receptor requires the rearrangement of both immunoglobulin and T cell receptor genes, which involves breakage and re-joining of the DNA in these regions. The same is true for class switching. DNA DSB repair pathways are required for this process. The cell may be vulnerable to chromosome translocation at these times and indeed many lymphoid tumours carry translocations where the breakpoint is within an immune system gene, indicating a role for this process in tumourigenesis (Bastard *et al*. 1992; Kerckaert *et al.* 1993; Miki *et al.* 1994). These translocations can either be the switch needed to activate oncogenes, or deactivate tumour suppressor genes (Ekstrom-Smedby 2006). In light of this, the importance of the DSB repair pathway becomes increasingly highlighted by stabilising genomic aberrations.

Dr Thomas Hodgkin first described 7 patients with what is now known as Hodgkin lymphoma in 1832 (Hodgkin 1832). The most recent universally accepted classification, as outlined by the World Health Organisation (WHO), organised lymphoma cases into a number of sub-types. The document denoted an adaptation of the original, REAL classification (Revised American-European Lymphoma Classification) (Campo *et al.* 2011). Lymphoma neoplasms are either derived from B or T cells, with B cell neoplasms being the most prominent in 2010 (94%) (Cancer Research UK 2013). Inevitably, since there are significantly more cases, B cell lymphoma is considerably better understood with the 5 most common subtypes including Diffuse Large B-cell Lymphoma (DLCL), Follicular lymphoma, Marginal Zone lymphoma, Mantle Cell lymphoma and Burkitt lymphoma. According to the American Cancer Society more than 350,000 new cases of NHL were diagnosed in 2008, globally (American Cancer Society 2008). In the UK, more recent data identified NHL as being the 6th most common cancer in 2010 with 12,180 new patients diagnosed (Cancer Research UK 2013).

1.8 Fanconi Anaemia and Lymphoma

Fanconi anaemia is associated with myeloid tumours. However, since the FANC pathway is involved in DNA DSB repair, lymphoid tumours (which arise as result of a defective repair in DNA DSB) might be expected to occur in at least a proportion of FA patients; those who have defects in DNA DSB repair. Indeed there are occasional reports of such tumours in FA patients but, they are unusual (Shen *et al.* 2006, Smedby *et al.* 2006, Hill *et al.* 2006, Xiao *et al.* 2008). Shen *et al.* sequenced a catalogue of 29 NHL cell lines and four genes associated with the FANC/DSB repair pathway were identified, WRN, XRCC3, BRCA1 and BRCA2/FANCD1. The endpoint examined was loss of protein by Western blotting. However, according to their

results only single nucleotide polymorphisms (SNPs) found in BRCA1 and WRN could be correlated to NHL with statistical confidence. More interestingly, light has been cast on to the expression of FANCN/PALB2 and its correlation with lymphoma (Xiao *et al.* 2008). Xiao *et al.* speculated a negative correlation between the level of FANCN/PALB2 expression and susceptibility to lymphoma from a SNP in exon 5 found in 2 patients. Since FANCN/PALB2 plays a major role in homologous recombination during DNA double strand break repair it would be valuable to look deeper into the epidemiology of lymphoma caused by genetic variation in PALB2/FANCN.

Table 1.1. Summary of published FA-N patients. Table extended from Reid *et al*. 2007, Table 1.

Patient	Location		Modification		Resulting in	Associated phenotype	Reference
	Nucleotide position	Amino acid position	Nucleotide change	Amino acid change			
1	1,653	551	T > A	Y > X	Truncated protein at residue 551X	Skin, thumb, heart and kidney abnormalities; growth retardation; endothelial cancer.	Xia *et al*. 2006b
	Exon 1 - 10		Deletion		Shortened protein		
2	3113	1105	2835_3113del 279/A946		Protein truncates at frame shifted residue 1,093X	Growth retardation, loss of right kidney, medullablastoma	Reid *et al*. 2007
	395	132	Del T	Frame shift	Protein truncates at frame shifted residue 527X		
3	757_758	253	Del CT	Frame shift	Protein truncates at frame shifted residue 254X	Growth retardation, left pelvic kidney, hypoplastic thumbs, Wilms tumor, AML, medulloblastoma	Reid *et al*. 2007
	3294_3298	1,098	Del GACGA	Frame shift	Protein truncates at frame shifted residue 1,119X		
4	2257	753	C > T	R > X	Truncated protein	Growth retardation, microphthalmis, microcephaly, bifurcated anus, Wilms Tumor	Reid *et al*. 2007
	3549	1183	C > A	Y > X	Truncated protein		
5	2393_2394	799	Ins CT	Frame shift	Protein truncates at frame shifted residue 850X	Growth retardation, microphthalmia, microcephaly,	Reid *et al*. 2007

	3350	1118	Ins GCAG	Frame shift	Protein truncates at frame shifted residue 1,122X	bifurcated anus	
6	2521	841	Del A	Frame shift	Protein truncates at frame shifted residue 849X	Growth retardation, microcephaly, horseshoe kidney, gonadal dysgenesis, Wilms tumor, medullablastoma	Reid *et al.* 2007
	3323	1108	Del A	Frame shift	Protein truncates at frame shifted residue 1,122X		
7	2962	988	C > T	Q > X	Truncated protein	Growth retardation, hypoplastic thumb, microcephaly, medullablastoma	Reid *et al.* 2007
	3549	1183	C > G	Y > X	Truncated protein		
8	3116	1039	Del A	Frame shift	Protein truncates at frame shifted residue 1,039X	Growth retardation, microcephaly, hand abnormalities, skin hyperpigmentation, Neuroblastoma	Reid *et al.* 2007
	3549	1183	C > G	Y > X	Truncated protein		
9	1676_1677 Homozygous	526	Del AA, Ins G	Frame shift	Protein truncates at frame shifted residue 559X	Abnormal thumbs, Horseshoe kidney, Neuroblastoma, Wilms tumor.	Serra *et al.* 2012
11	1176_1677	526	Dell AA, Ins G	Frame shift	Protein truncates at frame shifted residue 559X	Non-Hodgkin lymphoma	**This thesis**
	Exon 6		In frame deletion		Shortened protein		

1.9 Aims and Objectives

<u>1.9.1 Objective</u>
Using experimental procedures my aim was to clone the two *PALB2* mutations, in a particular PALB2 family, into different cell lines so that the consequences of each mutation could be examined independently at the cellular level. This involved isolating many clones, testing each one for PALB2 expression and potentially the activity of the two different proteins. This was to be done by determining whether these mutant proteins were able to directly interact with BRCA1 and BRCA2, and whether they localised BRCA2 to the site of DSBs. EUFA1341 mutant cells, Y551X, and the exon 4 revertant were used as controls.

<u>1.9.2 Aims</u>
1. Using Western blotting to confirm the size and quantity of PALB2 protein in cells from family members

2. Using site directed mutagenesis to create separate cell lines for each mutation from this new family and using DNA sequencing to confirm that the target mutations were present and also check protein expression in these lines.

Site directed mutagenesis cell lines to include:

- Y551X- EUFA1341

- Q559R fsX2 (Del/ins) – my patient

- Exon 4 deletion (revertant of Y551X) – EUFA1341

- Exon 6 deletion- my patient

- Wild type

3. Perform siRNA knockdown of the endogenous protein in order to remove endogenous PALB2 activity to allow visualisation of induced PALB2.

4. Immunoprecipitate the knockdown cells with antibodies for BRCA1, BRCA2 and

FLAG to determine whether mutant PALB2 interacted with these proteins.

2. Materials and Method

2.1 Creating U2OS/FRT/T-REX cell lines that express wild type or the mutant PALB2

PALB2 mutant cell lines were created for wild type (WT), revertant exon 4 deletion (X4), patient exon 6 deletion (X6), Y551X and Q559RfsX2 (1676_1677delAAinsG deletion_insertion (del/ins) for experimental purposes. To facilitate the integration of the PALB2 gene into the FRT site in the U2OS Flp-In T-REX cells, wild type and mutant cDNA copies of PALB2 had been cloned previously into the expression vector pcDNA5/FRT/TO that had been modified to include the FLAG peptide sequence N-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-**C (DYKDDDDK).** Induction of expression of the PALB2 cDNAs in the U2OS cells would produce PALB2 protein with the FLAG peptide fused in frame to the N-terminus of the PALB2 protein. Inclusion of the FLAG peptide would allow the PALB2 protein expressed from the integrated cDNA copies of PALB2 to be identified against the background of the PALB2 protein normally expressed by the U2OS cells. The FLAG sequence, DYKDDDDK, had been inserted into pcDNA5/FRT/TO plasmid from pcDNA3-NTAP (Gingras *et al*. 2005, Tasto *et al*. 2001) using the restriction endonucleases KpnI and PmeI (Chen *et al.* 2008). Parental U2OS/FRT/T-REX cells were maintained in DMEM medium containing 10% foetal bovine serum (FBS), 50μg/ml Zeocin and 5μg/ml blasticidin. Covalently closed circular plasmid (bulk prep) DNA was transfected into the U2OS/FRT/T-REX cells using lipofectamine® 3000 (Invitrogen). Adherent U2OS/FRT/T-REX cells were plated at 3x10^6 cells per 10cm dish in DMEM and serum, but without antibiotics, prior to transfection the following day. Plasmid DNA was diluted to 160ng/μl in Opti-MEM medium without serum. Diluted plasmid DNA (1.6μg in 10μl) was mixed with 14.4μg

of plasmid pOG44 and 32 μl P3000 Reagent in Opti-MEM without serum, made up to 1ml with the medium. Lipofectamine 3000 Reagent (40μl) was diluted to 1ml in Opti-MEM medium without serum. The 1ml aliquot of diluted DNA (PALB2 plasmid and pOG44) with P3000 Reagent was mixed with the 1ml Opti-MEM/Lipofectamine 300 Reagent and incubated at room temperature for 5 minutes. During the incubation period the U2OS cells, still adhering to the tissue culture plate, were washed in phosphate buffered saline (PBS) which was subsequently replaced with 8ml of Opti-MEM medium without serum. The 2ml of DNA/Lipofectamine, now containing DNA-lipid complexes that were formed during the incubation period, was added to the cells and incubated at 37°C in a humidified atmosphere of 5% CO_2 in air. After 4.5 hours the Opti-MEM containing the DNA and Lipofectamine was removed and replaced with DMEM containing 10% FBS but omitting antibiotics. The following day each dish of transfected cells was split to four $75cm^2$ tissue culture flasks and each culture incubated in 12ml DMEM with FBS but without antibiotics. The following day the medium was removed and replaced with DMEM 10% Tetracycline-free FBS containing 200 μg/ml hygromycin and 5 μg/ml blasticidin. The medium was removed and replaced with the same medium every five days. After approximately 14 days colonies of selected cells became apparent. The four flasks of selected colonies for each PALB2 plasmid construct were pooled together and aliquots taken for further growth and freezing down. Cells were frozen down in 1.0ml vials in freezing medium that contained 50% Tetracycline-free FBS, 40% DMEM and 10% dimethylsulfoxide (DMSO). U2OS/FRT/T-REX cells that had been selected for retention of PALB2 constructs were grown routinely in DMEM 10% Tetracycline-free FBS containing 200 μg/ml hygromycin and 5 μg/ml blasticidin. Confluent cultures of cells were

washed in PBS, detached from the plastic surface of the tissue culture plate/flask using small volumes of TrypLE Express (1x) stable trypsin replacement enzyme and diluted into fresh DMEM 10% FBS containing antibiotics.

An aliquot of the pooled selected colonies for each PALB2 construct was taken for DNA extraction to check that the U2OS/FRT/T-REX cells had been transfected with the correct plasmid. DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit for Purification of Total DNA from Animal Blood and Cells (Spin-Column Protocol) (Qiagen 2006). An initial PCR reaction was used as an indicator of whether or not the transfections had been successful. The PCR products were analysed on a 1 % agarose gel (containing Midori Green to aid visualization of the DNA) against a 1 kb ladder, actual amplicon sizes were compared with expected values.

Table 2.1 Primer sequences used during sequencing and futher PCR reactions

Primers	
77143F	G T T G T T G A A T T C G G C G C G C C A A T G G A C G A G C C T C C C G G G A A G C
1975R	G A C T C A G G A T T T G C G T A C C T A T G T
1513F	C T T G T C T A G G A A G G C A G T T G
1803R	T T T A A A C T C A G C A T T C C A T C
2143F	C T T T A A A T A C G G T T G C G C C T G A T G
2463R	C T A T C A A G G T C A C G G A C T A C A

2.2 DNA sequencing

Following PCR amplification the sample was cleaned using the ExoSAP-IT® protocol from Affymetrix removing excess primers and dNTPs. The Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit provides fluorescently tagged dNTPs for use during sequencing using PCR amplification of single stranded amplicons. Following the manufacturers guidelines a BigDye® reaction was set up.

After BigDye® PCR, precipitation of 50 ng of PCR product was achieved by incubation in a solution of ethanol and sodium acetate for 20 minutes. DNA and other moieties were separated by centrifuging for 22 minutes, 16,000 x g. The supernatant was removed and the pelleted DNA was washed with 70 % ethanol. DNA was air dried at room temperature for 10 minutes and resuspended in 12 µl of HiDi (highly deionised) formamide. DNA was denatured at 100 ºC for 5 minutes and quenched on ice to inhibit any re-annealing during sample loading. Samples were loaded on to a 96 well plate and sequenced using an Applied Biosystems 3500XL Genetic Analyser.
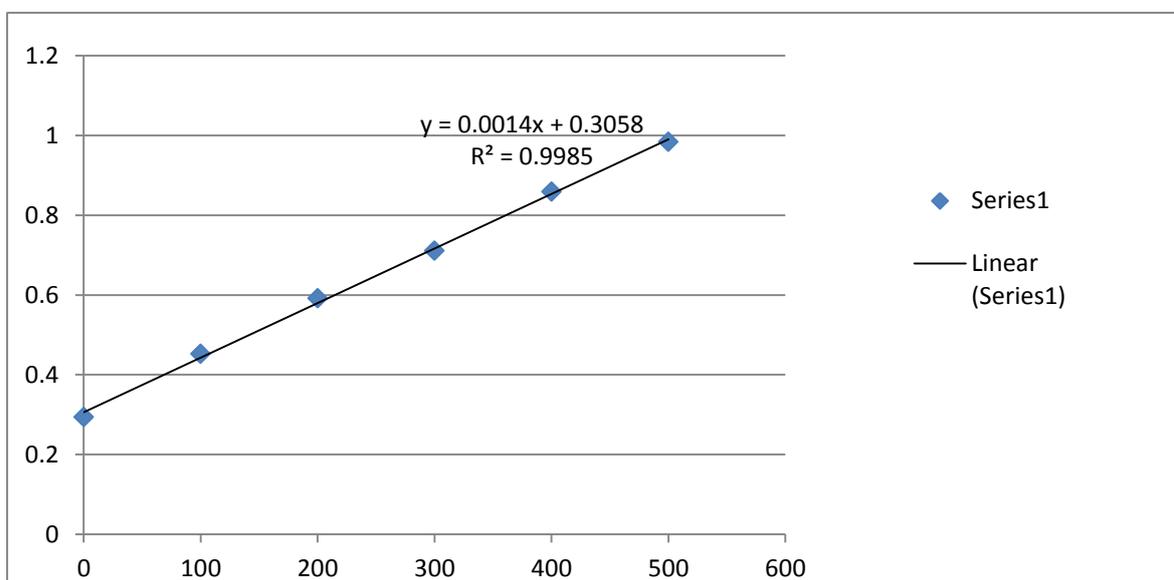
Figure 2.1 Example graph used for R^2 values to assess pipetting accuracy during the BioRad Bradford protein quantification assay.

2.3 Bradford Assay

Protein was quantified using the Bradford assay by calculating the light absorbance of the protein in a coloured dye at various wavelengths. In high protein conditions the dye turns blue, in low protein conditions the dye is red. A controlled protein gradient using diluted BSA at concentrations of 100, 200, 300, 400 and 500 µg /µl was used to

monitor accuracy of pipetting and potted on a graph for R^2 values, an example is shown in figure 2.1.

2.4 Western Blot

A Western blot is used to measure protein length and expression by blotting the protein against an antibody specific to the protein of interest. Firstly, the cell lysate was created by purification from other cellular debris and quantified using the BioRad Bradford protein quantification assay. A 6 % SDS polyacrylamide gel (0.1 M Tris, 0.1 M Bicine, 6 % Acrylamide, 0.1 % SDS, Temed, APS) was set, loaded and run for 4 hours 30 minutes to ensure clear separation of the larger proteins. Proteins were transferred onto nitrocellulose paper overnight by electrophoresis in transfer buffer (400 mM glycine, 50 mM tris, 20 % v/v methanol) and then cut to size ready for immunoblotting. To block the unspecific binding of antibodies the protein bound nitrocellulose paper was shaken in a 5 % milk in TBST solution (20 mM tris, 200 mM NaCl, 0.1 % tween 80) for 1-4 hours. Following unspecific blocking, the nitrocellulose paper was washed in TBST buffer and the primary antibody was added (See table 2.2). The primary antibody was blotted in a 5 % milk/ TBST solution, shaking overnight. The primary antibody is specific to the protein of interest and will therefore only bind where the designated protein is located. After washing, the secondary protein was blotted by shaking the nitrocellulose paper in diluted 5 % milk in TBST buffer. The secondary antibody contains a tag which binds to the primary antibody and reacts with luminescent reagents during the development of the autoradiography film. The nitrocellulose was washed a final time and treated with the imaging agents, ECL or illuminol.

Table 2.2 Antibodies used during experimental procedures

Antibody	Epitope binding site	Purchased from
PALB2 246A (Rabbit)	Residues 200-250	Bethyl Laboratories Inc.
PALB2 247A (Rabbit)	Residues 675-725	Bethyl Laboratories Inc.
Anti-FLAG M2 (Mouse)	FLAG sequence (DYKDDDDK)	Sigma Aldrich

2.5 Colony selection

Colonies of growing mutants with variable morphologies were selected by pipetting the target colony under a microscope and transferring to 1 ml of media in a 3 x 4 well plate. 48-96 colonies were chosen for each cell line.

2.6 Doxycycline induction

The plasmid vector, pcDNA5-FLAG contains a tetracycline inducible promoter region, in order to test the activity of the cloned PALB2 peptides the cells were induced using the tetracycline derivative, doxycycline. After 24 and 48 hours the adherent cells were treated with trypsin and harvested ready for protein quantification before Western blot analysis.

2.7 Knockdown of endogenous PALB2 using siRNA.

The siRNA used here was specific to the 3' UTR region of PALB2 mRNA (see table 2.3), since pcDNA5-PALB2 constructs did not include this site only endogenous PALB2 would be knocked down. Cell were split, counted and divided up into approximately 0.25×10^6 cells per 10 cm tissue culture dish and left to multiply overnight. The following day a fresh transfection buffer including warm optimem, oligofectamine and siRNA to 180, 90, 45 and 22.5 nM final concentration was made in an RNAase free tube and incubated at room temperature for 30 minutes. Cells were washed and resuspended in optimem before adding the transfection buffer (dropwise), mixing by agitation the cells were then left for several hours. Cells were incubated overnight in 2 ml of McCoys medium + 20 % Fetal Calf Serum (FCS) (no

penicillin/streptomycin). Knockdown efficiency was analysed using a Western blot, using antibody PALB2 246A.

Table 2.3 SiRNA sequence used during knockdown of endogenous PALB2

Forward	G G A G A A U A U C U G A A U G A C A d T d T
Reverse	U G U C A U U C A G A U A U U C U C C d T d T

2.8 Immunoprecipitation

Cell lysates were created using 8 x 10^6 cells per IP. The cells were treated with IP lysis buffer and incubated on ice for 30 minutes. After incubation the solution was centrifuged for 5 minutes at 10,000 x g and the supernatant was kept. Using antibody concentrations 2, 4, 6, 8 or 10 µg, FLAG antibody was added to the cell lysate and left to rotate overnight at 4 °C. Finally, 25 µl of protein G beads were added (50 µl of 50 % beads + NETN solution) to each lysate and left to rotate at room temperature for another 4 hours.

To remove additional cell debris from the sample the beads were centrifuged and washed with 1 % NETN solution (20 mM Tris-Cl, 100 mM NaCl, 0.5 mM EDTA, 0.5 % v/v NP40) twice. Before Western blot analysis the control samples were quantified using the BioRad Bradford assay, control and immunoprecipitated cell lysates were then mixed with 4 x loading dye ready for Western blot analysis.

3. Results

3.1 Assessing PALB2 size and level of expression in lymphoblastoid cell lines derived from six family members

PALB2 mutations were located by DNA sequencing before the start of this project. The project began by analysing protein levels from the patient cell lines using Western blotting. Cells from each family member were assayed, in addition to a control 'normal' sample. In order to do this, cells were lysed to enable a full protein profile screen and a Western blot was performed. The nitrocellulose was probed with antibodies PALB2 246A (figure 3.1a) and PALB2 247A (figure 3.2a) for PALB2 analysis and aprataxin was used as a 50 µg loading control (figures 3.1b and 3.2b). Figure 3.1b suggest that the loading of protein in Figure 3.1a was not consistently the same across all lanes, particularly lanes 1 and 2. This is probably a consequence of unequal loading rather than being a fault of the loading control used, aprataxin, as this was regularly and frequently used in the laboratory, was reliable and was unrelated to the samples being analysed. β-actin, however, a highly conserved protein present in all eukaryotic cell lines (Hanukoglu *et al.* 1983), could have been used and in future work perhaps should be used as it is a generally acknowledged loading control.
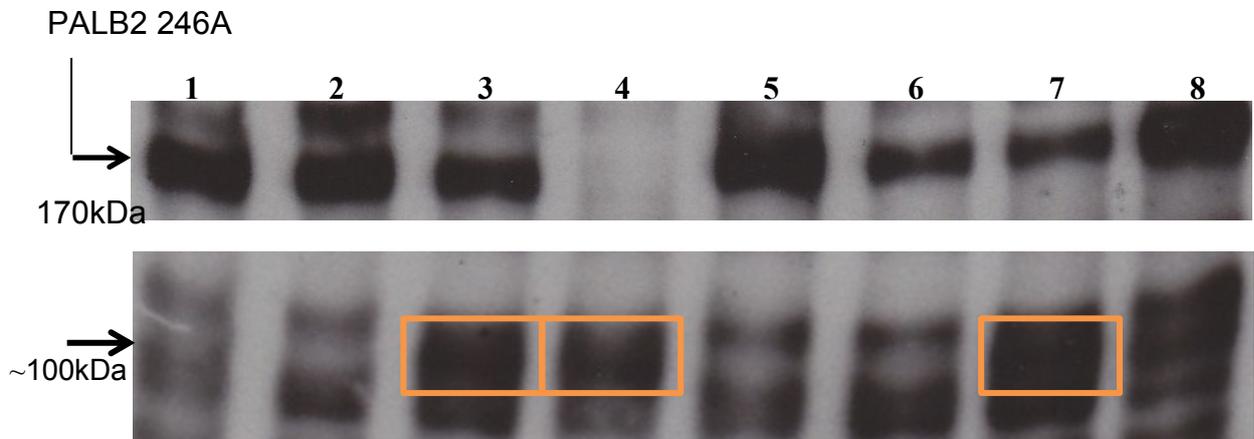
Figure 3.1a. The top panel demonstrates the presence of PALB2 in each sample at ~170kDa using PALB2 246A antibody. The bottom panel shows the presence of a band at ~100 kDa in lanes 3, 4 and 7, thought to represent the presence of the del/ins mutant in these samples. Lanes 1 and 8 are control samples, Lane 2 is the Father and lane 3 is the Mother. Lanes 4, 5, 6 and 7 represent protein expression from the compound heterozygous FA child, the homozygous un-mutated child, the X6 carrier and del/ins carrier, respectively.

Figure 3.1b. Aprataxin was used as a 50 µg loading control per sample. A band at approximately 40 kDa is seen in lanes 1 – 8, above.

PALB2 247A
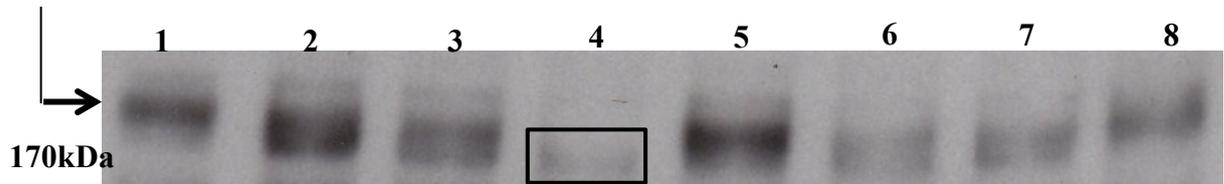
Figure 3.2a. Demonstrates the presence of PALB2 in each sample at ~170kDa using PALB2 247A antibody. Lanes 1 and 8 are control samples, Lane 2 is the father and lane 3 is the mother. Lanes 4, 5, 6 and 7 represent protein expression from the compound heterozygous FA child, the homozygous un-mutated child, the X6 carrier and del/ins carrier, respectively.

Aprataxin
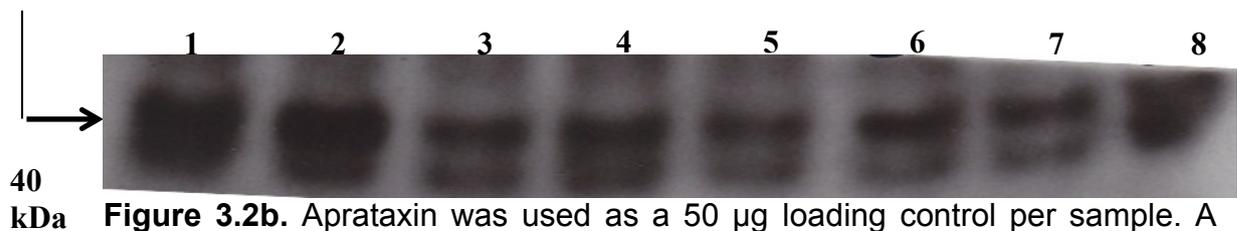
Figure 3.2b. Aprataxin was used as a 50 µg loading control per sample. A band at approximately 40 kDa is seen in lanes 1 – 8, above.

Figure 3.1a shows 2 panels, antibody PALB2 246A has been used to detect the presence of full length PALB2 (top panel) and a truncated (~100kDa) PALB2 mutant (bottom panel). Figure 3.2a illustrates PALB2 expression at ~170 kDa using antibody PALB2 247A. In all figures lanes 1 and 8 are the normal control samples loaded for comparison purposes. It should be noted that the Bradford assay was used to achieve loading of 50 μg of cell lysate per well, however as observed in figure 3.1b and 3.2b loading of these samples varies.

Lanes 1 and 8 are control samples which show unequal loading (figure 3.1b) however, figure 3.1a shows abundant expression of full length PALB2 in each of these lanes therefore, confirming that PALB2 should be abundantly expressed in 'normal' lymphoblastoid cells.

As observed in figure 3.1b, lane 2 was loaded with a higher concentration of cell lysate than that in lanes 3, 5, 6 and 7 which appear to be loaded equally. This has been taken in to account when analysing the rest of this blot. Lane 4 in the top panel of figure 3.1a shows no full length PALB2 expression, yet in the bottom panel of figure 3.1a a band can be observed within the orange box at ~100 kDa when compared with the control in lane 1. Additionally, figure 3.1b shows that loading in lane 4 is acceptable, and therefore the absence of full length PALB2 expression in lane 4, Fig 3.1a (top) is genuine as is the truncated version of approximately 100 kDa in Fig1a (bottom). The top panel of figure 3.1a shows abundant expression of full-length PALB2 in lanes 2, 3 and 5 with slightly reduced expression observed in lanes 6 and 7. The bottom panel of figure 3.1a (~100 kDa) shows a lot of background antibody binding however, it is possible to see additional bands at approximately 100

kDa in lanes 3, 4 and 7 all containing lysates of cells from individuals carrying a mutation that could give rise to this truncated form.

Similar to the loading control found in figure 3.1b, figure 3.2b also shows a loading disparity. In lanes 1 and 2 of figure 3.2b aprataxin is more abundantly expressed whereas lanes 3-8 display more equal loading of cell lysates. There is acceptable level of loading in lane 4 in figure 3.2b suggesting that the absence of full length PALB2 in lane 4 of figure 2a is genuine. Figure 3.2a aims to demonstrate the loss of some full-length PALB2 expression in samples 2 and 6, known from previous sequencing experiments to be the X6 carriers. However since the loss of exon 6 is only a small in-frame, 24 amino acid deletion, detection of this small reduction in band size is difficult. Lanes 2, 3 and 5. 6 and 7 in figure 3.2a show unevenness in the resolution of this X6 band however lane 4 shows a band in the appropriate place (highlighted) which could be this 24 amino acid deletion. When taking in to account the blots in figure 3.1a and 3.2a as well as the unequal loading controls in 3.1b and 3.2b, mutant PALB2 is implicated as being expressed at similar levels to the wild-type. This, in addition to the data collected from sequencing confirms the presence of mutant PALB2 in our patient samples.

3.2 Analysis of transfected and selected U2OS Flp-In/T-REx™ cell lines to confirm the presence of the expected PALB2 construct

The wild type and mutant cDNA copies of PALB2 (whose sequences were verified before this project began) had been cloned into the pcDNA5/FRT/TO-FLAG vector. Five experimental cell lines containing pcDNA5/FRT/TO-FLAG-PALB2 (pcDNA5-PALB2) mutants and wild type vector, one pcDNA5/FRT/TO only cell line (U2OS transfected cells with the construct not containing any PALB2, known as vector only) and one parental (U2OS host cells not transfected with any construct) were analysed after transfection and cellular cloning. The vector only and parental cell lines were used as controls; if amplicons were generated from either control sample it would show non- specific primer annealing to genomic DNA outside of the construct. Two cDNA constructs contained representations of the two mutations found in the patient (X6 and del/ins) two mutations from EUFA1341 (Y551X and revertant X4) (Xia *et al.* 2006b) and one un-mutated wild-type PALB2. After transfection the mutant colonies were left to grow until they were stable enough for experimental use (confirmation was made by visual inspection). As an initial assessment, an agarose gel was run following PCR amplification (figure 3.3). Three primer pairs were specifically created for their location within the sequence. First, the binding site of forward primer 77143F is located at the 5' end of PALB2 cDNA sequence. By pairing this with PALB2 reverse primer at base pair 1975R, amplicons from WT sequence would be expected to be ~1975 bp long and include the vector/PALB2 sequence junction (including the whole of exon 4) (lanes 1 – 8) (see appendix 10). This primer pair is particularly useful for identifying the X4 deletion (lane 4); since this mutant deletes approximately 1.5 kb we would expect to see a band at ~500 bp. As predicted lane 4 produced a band in the desired location preliminarily confirming X4 mutagenesis. X6, del/ins,

Y551X and WT amplicons using the same primers are shown in lanes 5 – 8, their respective bands show a 2000 bp amplicon proving that, each of these did not have the exon 45 deletion. The second primer pair (1513F and 1803R) (lanes 9 – 16) confirmed the loss of exon 4 in X4 only mutant cells. The large deletion removes the binding site of forward primer 1513F (lane 11), when combined with the reverse primer, located at position 1803R, a 337 bp amplicon was generated in all other mutants (lanes 12 – 15), which gave further confirmation of presence of PALB2 in each cell line (further sequence details of primer pair 1513F/ 1803R can be found in appendix 8). The third and final primer pair's amplicon included exon 6. With the exon 6 deletion mutation a smaller product, by 72bp, was expected and indeed was observed, (lane 21). With forward and reverse annealing sites at 2143F and 2764, respectively, the successful mutants from X4, del/ins and Y551X were predicted to produce a normal ~664 bp amplicon, as seen in lanes 20, 22, 23 and 24 (further details on primer sequence are available in appendix 9). The parental cell line (without the construct), vector only (cells transfected with just the vector, no PALB2) and no template control were used as controls for each primer pair. These results show correct amplification of PALB2 sequence in the respective cell lines consistent with the presence of the individual *PALB2* mutations.

Table 3.1. A Table to show the correct use of primer pair to view PALB2 mutant sequences during gel electrophoresis and DNA sequencing.

Mutant	Primer pair required	Figure
Del/ins (our patient)	1513F, 1803R	3.4
Y551X (EUFA 1341)	1513F, 1803R	3.6
X4 (EUFA 1341)	77143F, 1975R	3.8
X6 (our patient)	2143F, 2763R	3.9
Wild type	1513F, 1803R; 77143F, 1975R; 2143F, 2763R	3.5, 3.7 and 3.10
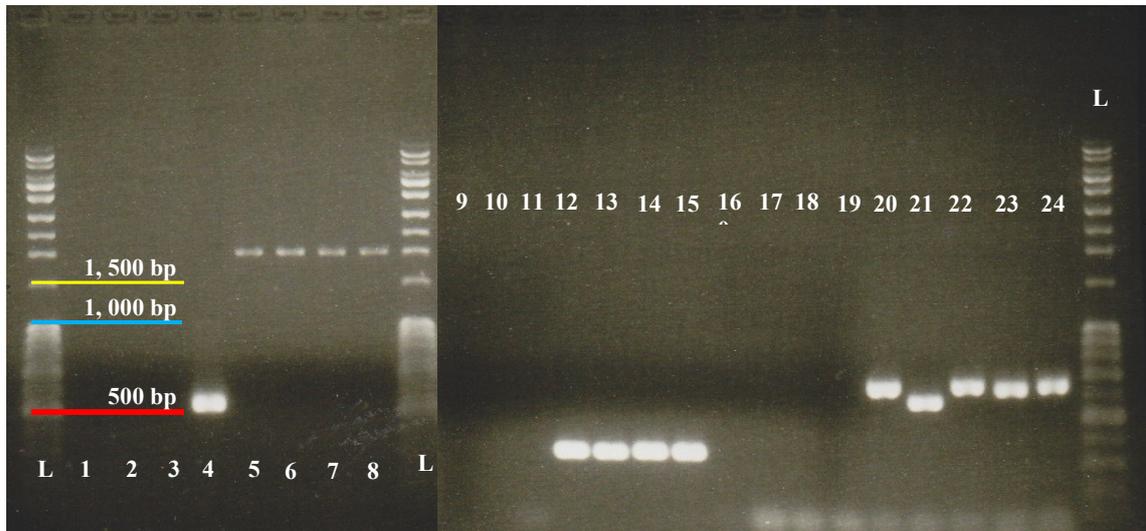
Figure 3.3 Aragose gel electrophoresis after site directed mutagenesis.
Lanes 1-8 represent amplicons produced using primer pair 77143F and 1975R, in order they are: 1= No template control, 2 =parental U2OS cells, 3 = vector only, 4= revertant X4, 5 = patient X6, 6 = patient del/ins, 7 = Y551X, and 8 = WT. Lanes 9-15 illustrate the amplicons replicated using primer pair 1513F and 1803R, in order they are: 9 = parental U2OS cells, 10 = vector only, 11 = X4, 12 = X6, 13 = del/ins, 14 = Y551X, 15 = WT, 16 = no template control. Lanes 17-24 represent PCR products using primer pair 2143F and 2763R, in order they are: 17 = NTC, 18 = parental U2OS cells, 19 = vector only, 20 = X4, 21 = X6, 22 = del/ins, 23 = Y551X, 24 = WT.

To further validate the correct transfections, Sanger sequencing was carried out. Table 3.1 presents the primer pairs which were specifically chosen to detect PALB2 mutated regions for each cell line. Figures 3.5, 3.7 and 3.10 showing sequencing in normal cells across the regions of the mutations 559, Y551X and revertant (exon 4 del) and exon 6 del reveal normal sequence in each case. Full electropherograms are available in appendix 1 (for figure 3.10), appendix 2 (for figure 3.5), appendix 3 (for figure 3.7), appendix 4 (for figure 3.8), appendix 5 (for figure 3.4), appendix 6 (for figure 3.9). Figure 3.8 demonstrates the correct sequence of the exon 4 deleted mutation (X4) by showing exon 3 reading directly into exon 5 in the electropherogram position 457. Similarly, figure 3.9 shows the correct sequencing indicating deletion of exon 6; exon 5 transcribes directly into exon 7 close to position 361 in the electropherogram. Correct insertion of the del/ins was confirmed by sequencing in figure 3.4, where two adenines have been replaced with a guanine close to position 169 in the electropherogram. Similarly, for Y551X, figure 3.6 displays the point at which one thymine residue has been mutated to an adenine, truncating the protein at a.a position 551, close to position 133 on the electropherogram. The full electropherograms are available in the appendix along with a BLAST analysis for each of the data sets to confirm that no other mutants were present. Use of sequencing has therefore, showed the presence of the correct sequence in the correct cell lines.
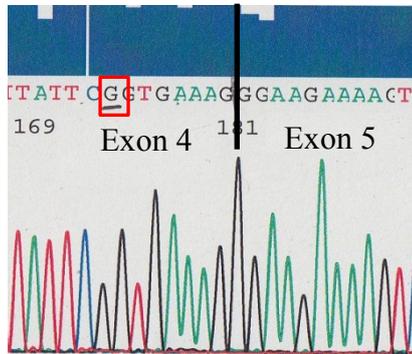
Figure 3.4 Electropherogram of mutant, del/ins. The highlighted guanine replaces the two original adenines 7 bases before exon 5.
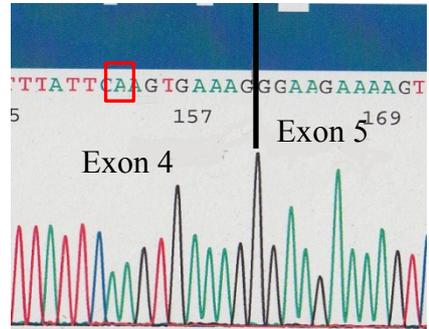
Figure 3.5 Electropherogram of wild type displaying the unmutated adenine residue at position 1676 (a.a. 559. The highlighted adenine is not replaced by two guanine residues, as with the del/ins mutant.
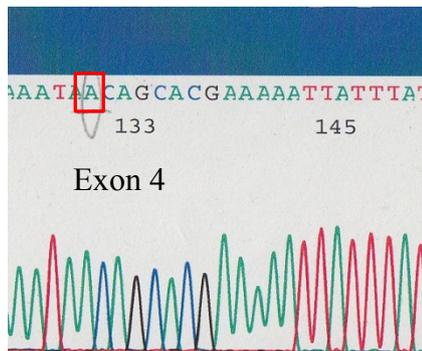
Figure 3.6 Electropherogram of mutant, Y551X. The highlighted adenine replaces the two original thymine, in turn producing the codon 'TAA', a stop codon at residue 551.
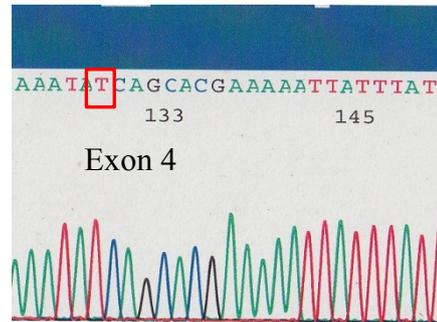
Figure 3.7 Electropherogram of wild type displaying the unmutated form of residue 551 where codon TAT codes for tyrosine. The highlighted thymine is not replaced by an adenine in this mutant.
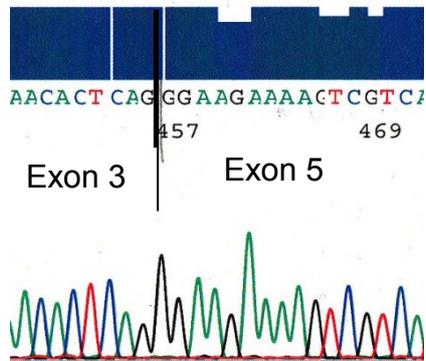
Figure 3.8 Electropherogram of mutant X4, displaying the skipped exon 4. The codon CAG of exon 3 leading, in frame, straight into GGA of exon 5
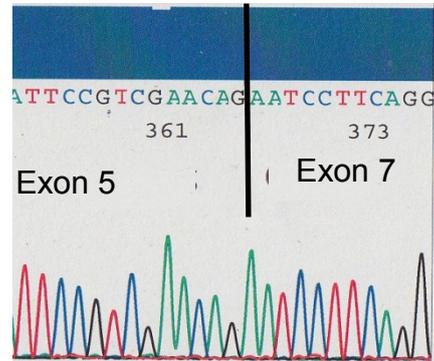
Figure 3.9 Electropherogram of mutant X6, displaying the skipped exon 6. The codon CAG of exon 5 leading, in frame, straight into AAT of exon 7.
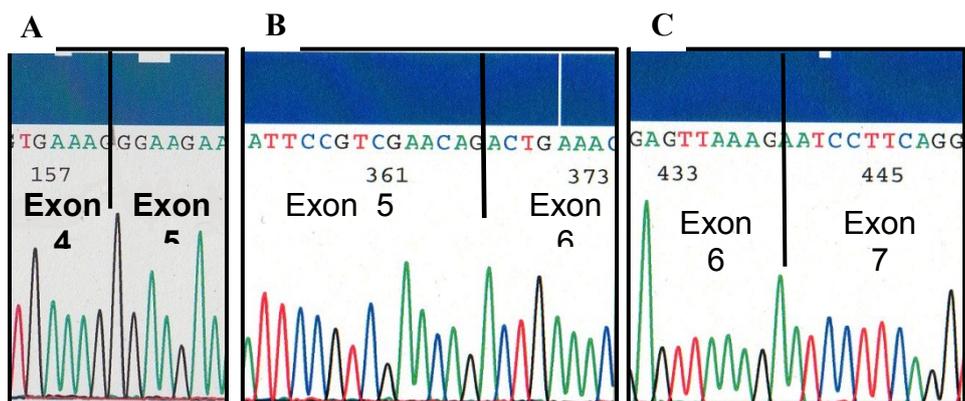
Figure 3.10 Electropherogram of wild type displaying the presence of exon 4 and exon 6. In figure A, exon 4 leads into exon 5. B shows exon 5 leading into exon 6 and exon 6 into exon 7 in figure 3.8C.

3.3 Testing the tetracycline-controlled transcriptional activation system using a time course of doxycycline induction.

PALB2 mutant proteins were expressed in genetically identical cells i.e. isogenic cells, in order that genetic differences in the cells would not be a confounding factor when comparing the properties of the PALB2 mutations. The system used for this purpose was based on the Flp-In U2OS T-REx™ (Invitrogen 2011) human osteosarcoma cell line. These cells contain an Flp Recombination Target (FRT) site stably integrated into the human genome. Co-transfection of these cells with an expression vector containing the gene of interest (PALB2) and an FRT site linked to the hygromycin resistance gene (selectable marker), together with a Flp-recombinase plasmid results in targeted integration of the expression vector at the FRT site in the human genome. Regulation of expression of the integrated gene of interest (PALB2) is controlled by two copies of the tet operator 2 (TetO2) elements of the *E. Coli* Tn10-encoded tetracycline (Tet) resistance operon embedded in the cytomegalovirus (CMV) promoter (Gossen and Bujard 1992, Yao *et al.* 1998).

The efficiency of the induction system was tested by harvesting the cells following induction with doxycycline (a tetracycline derivative) for 24 and 48 hours alongside an un-induced control for each cell line. Using Western blotting with anti-FLAG Ab, PALB2 protein expression levels were analysed against induction time in order to assess whether the transfections were successful, and also to assess the optimal timing of doxycycline induction. Figure 3.11a shows the results of PALB2 expression in WT (lanes 1-3), X6 (lanes 4-6), del/ins (lanes 7-9), Y551X (lanes 10-12), X4 (lanes 13-15) and a control cell line transfected with pcDNA5/FRT/TO vector only (lanes 16-18) without the PALB2 sequence by Western blot analysis after doxycycline induction. The blot has been split into two, the top panel shows location at ~170 kDa

where endogenous, WT and X6 PALB2 are expected; bands at ~100 kDa are shown in the lower panel where Y551X, del/ins and X4 were predicted to be; vector only would not be expected to respond in this blot since there is no PALB2 protein in the construct. In figure 3.11b, senataxin was used to validate results as a 50 µg loading control. Figure 3.11a clearly illustrates a successful induction of PALB2 in the exon 4 deleted cell line as it is highly responsive to the induction system. A similar yield is also shown from the Y551X cell line, and strong bands can be seen ~100 kDa. The del/ins cells (lanes 7-9) show bands at the expected (~100 kDa) location however, the WT and X6 bands also show similar bands at ~100 kDa, it is therefore unclear as to whether or not these bands are a consequence of cross-reactivity from the FLAG anti-body or the induced expression as a direct result of doxycycline induction. The observed band in lane 7 is less abundant than that which can been seen in lanes 8 and 9, therefore it can be assumed that this doxycycline induction was successful in the del/ins mutant cell line, however further testing will be needed to confirm this. Induction in the X6 and WT cell lines is less consistent and further analysis would need to be undertaken to investigate their lack of expression after both 24 and 48 hour doxycycline induction.

This test has shown that two of the mutants, X4 and Y551X, had effective tetracycline induction systems and are able to express large amounts of PALB2 mutant protein. Induction of the del/ins mutant potentially shows some increased expression after doxycycline induction. Finally the WT and X6 cell lines did not respond in this experiment.
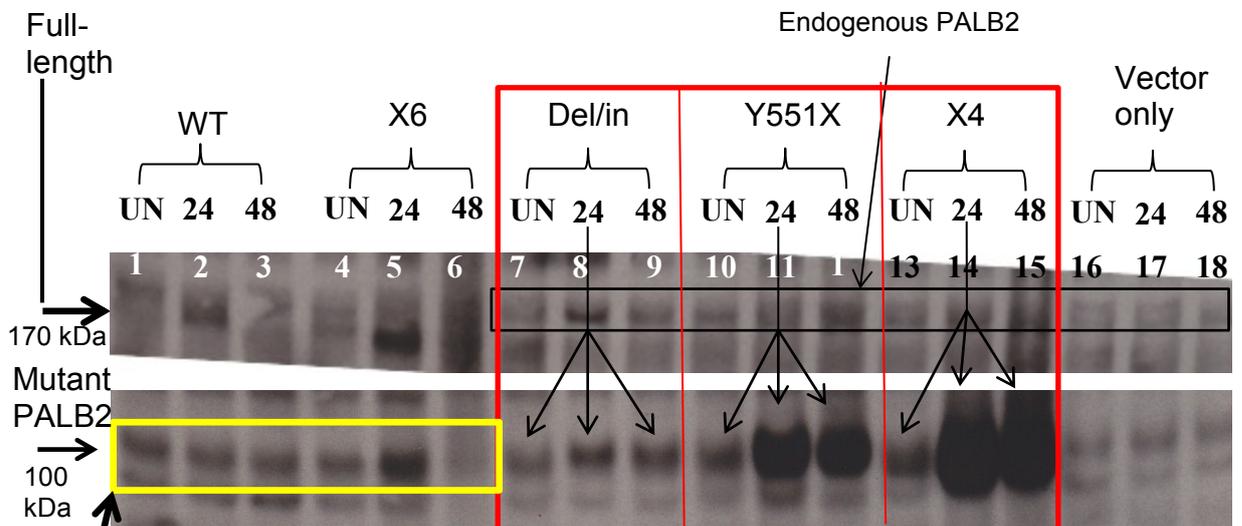
Figure 3.11a Western blot using anti-FLAG after doxycycline induction. All mutants have an un-induced control, a 24 and a 48 hour induced sample. Lanes 1-3, 4-6, 7-9, 10-12, 13-15 and 16-18 represent samples from WT, X6, del/ins, Y551X, X4 and pcDNA vector, respectively.
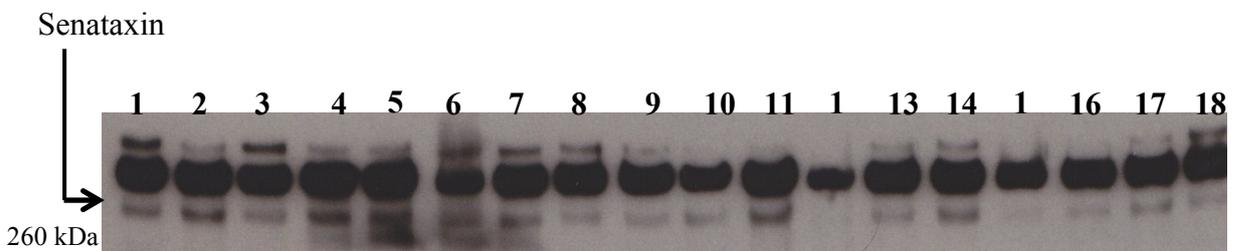
Figure 3.11b Western blot using anti-sentataxin as the 50 µg loading control during doxycycline induction analysis. All mutants have an un-induced control, a 24 and a 48 hour induced sample. Lanes 1-3, 4-6, 7-9, 10-12, 13-15 and 16-18 represent samples from WT, X6, del/ins, Y551X, X4 and pcDNA vector, respectively.

3.4 Choosing mutant cell lines

To enable further experimentation with cell lines displaying similar expression levels, colonies were picked based on varied morphologies, size and location within the petri dish. All colonies were picked in sterile conditions within a tissue culture hood. Following stable growth, 18 of these colonies from each mutant were chosen, induced and analysed on a Western blot using aprataxin as a 50 µg loading control and an un-induced aliquot as a control. Firstly, in figure 3.12 the deleted exon 4 revertant shows variable levels at the expected (~100 kDa) size using antibody PALB2 247A. This antibody was chosen because PALB2 246A epitope binding (residues 200-250) site was lost with the deletion of exon 4 (412-1884). From this, three cell lines were chosen and kept for further experimental use, lanes 4, 6, and 17 show cells kept for this purpose. Fig 3.13 shows X6 responding to doxycycline despite not doing so previously. Blotting with anti-FLAG, this experiment showed a defined band (the lower band of the two present), which is not visible in the un-induced control, in all induced samples, just below the PALB2 band, at ~160 kDa (figure 3.13). The anti-FLAG antibody was used because mutant X6 is difficult to differentiate from endogenous PALB2 on the blot as X6 is only 24 a.a shorter, anti-FLAG however will not bind with endogenous PALB2 therefore X6 can be clearly identified. Using this blot the parental cell lines from lanes 3, 11 and 12 were chosen as 'normal' expressers. Similarly, the blot for the del/ins mutant (figure 3.14) revealed mixed expression levels from all induced samples and, as expected, no expression from the un-induced control (lane C) using antibody PALB2 246A, chosen because the epitope binding site (residues 200-250) was present in this cell line. The three chosen cell lines include 2, 15 and 18 which were grown up and utilised during further experiments. The Y551X mutant, as shown in figure 3.15, illustrates eighteen

successful inductions and one, lesser expressed un-induced control (lane C) after blotting with antibody PALB2 246A, chosen because the epitope binding site (residues 200-250) was present. In keeping with the other mutants three 'normal expressers' were chosen using the blot, here cell lines from lanes 4, 6 and 10 were kept for further experimental use. Finally, in contrast to the other mutants the WT cell lines showed no indication of increased expression (figure 3.16) after blotting with anti-FLAG. Anti-FLAG had to be used as, similar to mutant X6, when using antibodies PALB2 246A or 247A it becomes impossible to differentiate between endogenous and mutant PALB2. Nonetheless, three cell lines were chosen from these cell lines in order to establish a cause for these expression issues, the cell lines came from lanes 6, 11 and 17.

Endogenous
PALB2

170 kDa

X

100 kDa
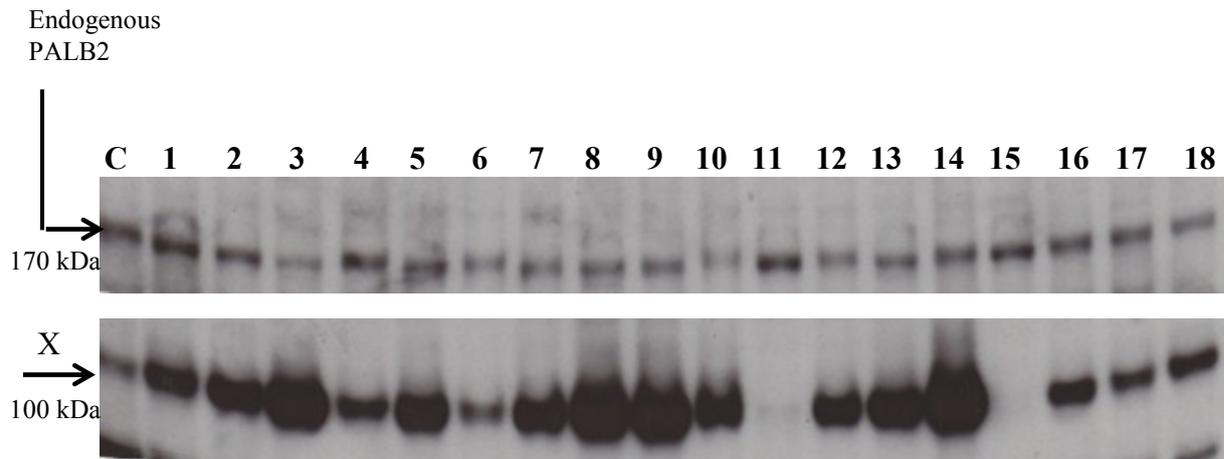
Figure 3.12 Western blot using antibody PALB2 247A (the deletion removes the epitope detected by PALB246A) showing expression of mutant cells X4 after doxycycline induction. Full-length (top) and truncated (bottom) PALB2 bands can be seen, X4 band lies 491 amino acids below the endogenous at ~100 kDa. Lane C illustrates the un-induced control cell line and lanes 1-18 represent colonies 1-18, respectively.

Anti-FLAG
background

X
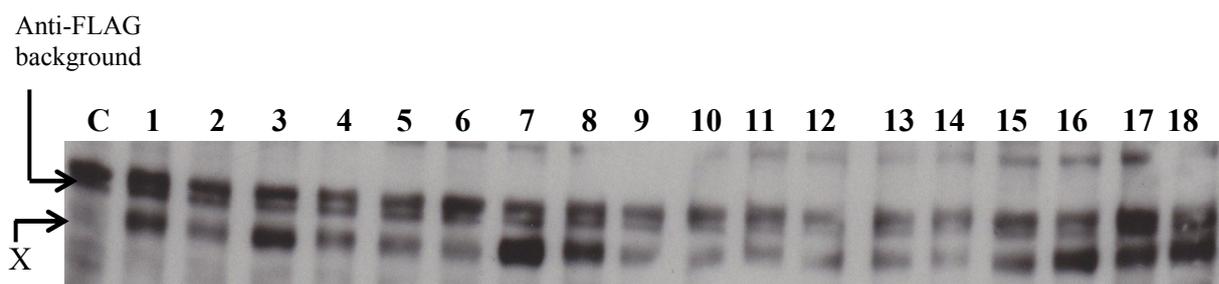
Figure 3.13 Western blot using anti-FLAG showing expression of mutant cells X6 after doxycycline induction. Mutant X6 PALB2 bands can be seen, 24 amino acids below where the wild type would express. Anti-FLAG generates large amounts of background activity, here a band lies at ~170 kDa where PALB2 would be expected. Lane C illustrates the un-induced control cell line and lanes 1-18 represent colonies 1-18, respectively.
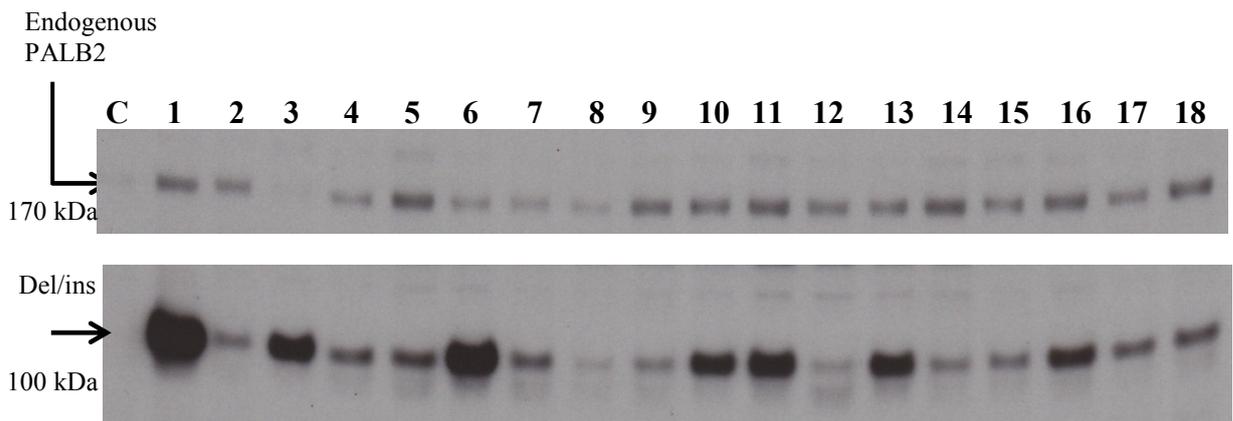
Figure 3.14 Western blot using antibody PALB2 246A showing expression of mutant cells del/ins after doxycycline induction. Full length (top) and truncated (bottom) PALB2 bands can be seen, del/ins band lies 627 amino acids below full length at ~100 kDa (bottom). Lane C illustrates the un-induced control cell line and lanes 1-18 represent colonies 1-18, respectively.
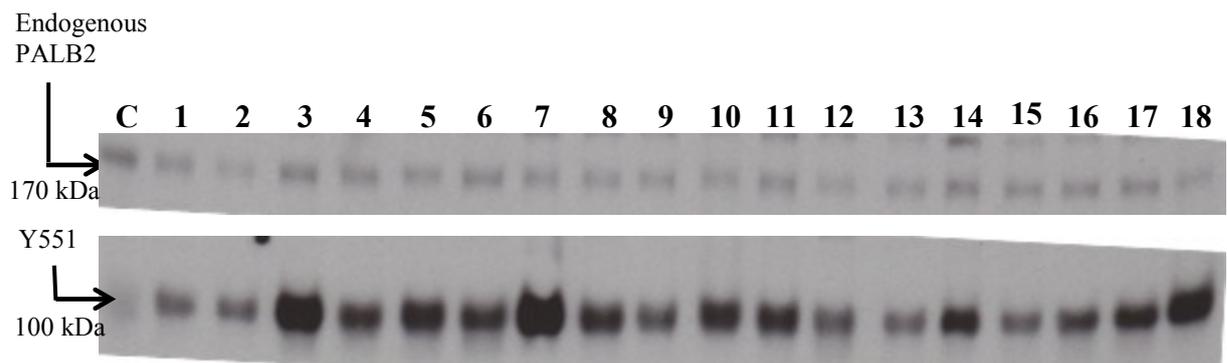
Figure 3.15 Western blot using antibody PALB2 246A showing expression of mutant cells Y551X after doxycycline induction. Full length (top panel) and truncated (lower panel) PALB2 bands can be seen, Y551X band lies 635 amino acids below the full length at ~100 kDa (lower panel. Lane C illustrates the un-induced control cell line and lanes 1-18 represent colonies 1-18, respectively.
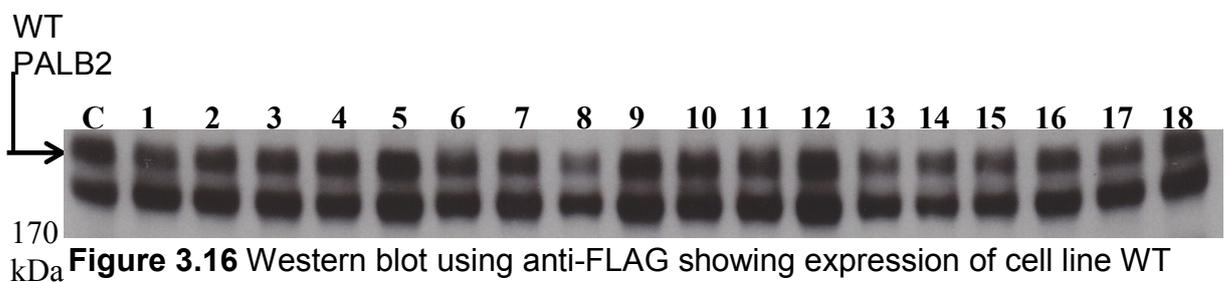
Figure 3.16 Western blot using anti-FLAG showing expression of cell line WT after doxycycline induction. WT PALB2 bands can be seen ~170 kDa. Lane C illustrates the un-induced control cell line and lanes 1-18 represent colonies 1-18, respectively.

3.5 Knockdown of endogenous PALB2 using siRNA to the 3' UTR.

Endogenous PALB2 is present in each mutant cell line therefore it is essential to knockdown the endogenous protein leaving only the mutant variants to ensure that all binding activity seen in further experiments is a direct cause of the mutant PALB2. Initially, the protocol was optimised with respect to duration of treatment and concentration of siRNA. To do this the del/ins cell line was used as it was straightforward to separate endogenous (~170 kDa) and mutant expression (~100 kDa). After siRNA transfection cells were lysed and run on a Western blot to assess how effective treatment had been. The membrane was blotted with PALB2 246A (figure 3.17a) with aprataxin as a 50 µg loading control (figure 3.17b) and developed using ECL. As seen in figure 3.17a endogenous PALB2 was successfully knocked down in all siRNA treated samples (lanes 2-13) as all control samples (lanes 1 and 14-17) were still displaying endogenous PALB2 expression, confirming that this siRNA is specific to our protein of interest, PALB2.

3.6 Anti-FLAG immunoprecipitation using mutant cell line del/ins.

Immunoprecipitation using anti-FLAG to pull down the FLAG containing PALB2 mutant protein and other proteins in complex was achieved as seen in figure 3.18. Protein expression is seen at ~100 kDa in lanes 3 and 4, and faint bands in lanes 5 and 6, all of which represent successful pull down of the del/ins mutant immunoprecipitating with 2, 4, 6 and 8 µg of FLAG antibody (lanes 3-6, respectively). Lane 1, 2 and 7 represent the control samples, lane 1 is a doxycycline induced whole cell lysate, lane 2 shows an un-induced lysate with antibody and protein G sepharose beads, and lane 7 shows the induced lysate with protein G beads and no antibody. The expression of endogenous PALB2 in the no-bead control validates the specificity of the antibody to the mutant since endogenous PALB2 is expressed in this sample

and not in the rest. Lane 2, where the un-induced control is present shows a faint band within the 100 kDa region, this control was used to compare how much protein was available for IP before doxycycline induction. Lane 7 proves the specificity of the antibody to the protein since no antibody was added to the solution no protein would be expected to pull-down. This experiment has proved that immunoprecipitation of mutant proteins is possible using anti-FLAG. Furthermore, it also shows that 4 µg of anti-FLAG is the optimal concentration to immunoprecipitate using this protocol. Moving on from this it would be possible to use this Western and blot for BRCA1 and BRCA2 to assess whether the complex has changed to accommodate this mutant.

Protein interaction analysis using immunoprecipitation with anti-FLAG enabled the identification of the optimal treatment conditions before carrying out additional experiments. Four separate reactions with incrementally higher antibody concentration (2, 4, 6 and 8 µg) were set up and precipitated using protein G sepharose beads. Western blot analysis was used to assess which proteins were pulled down in complex with the FLAG containing PALB2 (Figure 3.18). Three controls are present, one doxycycline induced del/ins mutant whole cell lysate (lane 1) and one un-induced del/ins mutant with protein G sepharose bead precipitation (Lane 2) and a beads only control used without any antibody (lane 7). The no-bead control displays a band at the location of endogenous PALB2 (~170 kDa) whereas the un-induced control with beads only shows a band in the 100 kDa region. The immunoprecipitated cells incubated at an antibody concentration of 2, 4, 6 and 8 µg are shown in lanes 3, 4, 5, and 6, respectively. The upper panel on figure 3.18 displays the protein ~170 kDa, the position of endogenous PALB2. Lanes 2-7 show a lack of antibody binding at this location which proves that the FLAG antibody used to

pull down PALB2 is specific to the mutant protein rather than the endogenous. The lower panel in figure 3.18 represents the length of the protein immunoprecipitated with anti-FLAG at approximately 100 kDa, in keeping with the length of the target del/ins mutant. Despite the higher concentration of antibody used for immuniprecipitation in lanes 5 and 6, it seems there has been a loss in binding efficiency in these samples, it could be assumed that this is consequence of an artefact or loading error. Since this was an optimisation step, it can be confirmed that 4 µg of antibody is sufficient to produce the desired protein binding over 48 hour incubation period.
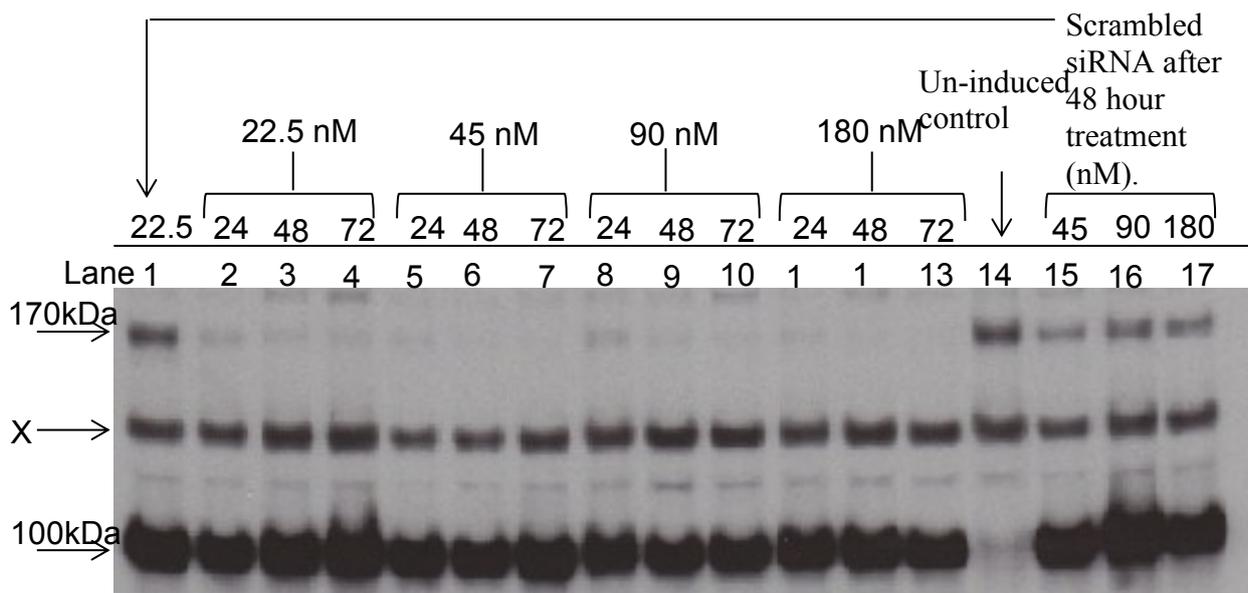
Figure 3.17a Western blot using FLAG antibody after siRNA knockdown of endogenous PALB2 in del/ins mutant cell line. Lanes 2-4 show knockdown efficiency using 22.5nM of siRNA after 24, 48 and 72 hours, respectively. Lanes 5-7 show knockdown efficiency with 45nM siRNA after 24, 48 and 72 hours, respectively. Lanes 8-10 show knockdown efficiency with 90 nM after 24, 28 and 72 hours, respectively. Lanes 11-13 show knockdown after 24, 48 and 72 hours, respectively, when using 180 nM of siRNA. Lanes 1 and 15-17 demonstrate a positive control using scrambled siRNA with doxycycline induced del/ins cells. Lane 14 shows the negative, uninduced control. X marks a cross reacting band.

Aprataxin

40kDa

Figure 3.17b Aprataxin loading control during Western blot after siRNA knockdown.

Endogenous PALB2

170 kDa

Del/ins

100 kDa
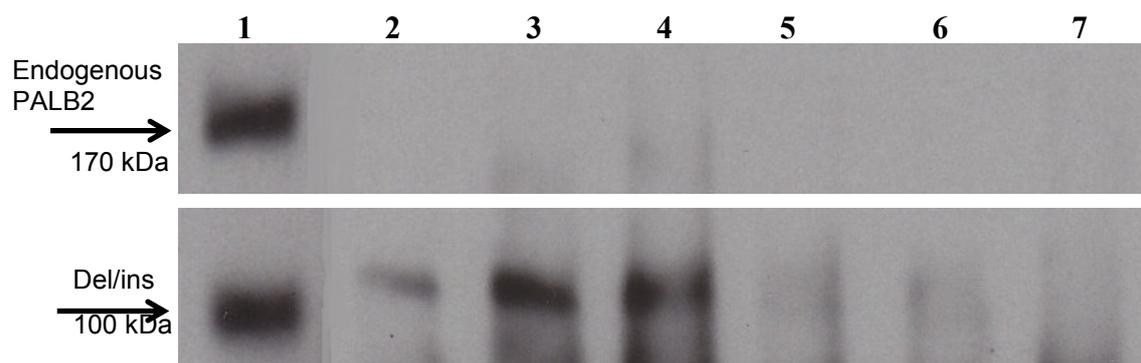
Figure 3.18. Western blot using antibody PALB2 246A after immunoprecipitation with anti-FLAG. Lane 1 displays a whole cell lysate, lane 2 shows a second un-induced control with beads, lanes 3-6 are protein samples using 2, 4, 6 and 8 µg of antibody during immunoprecipitation and lane 7 represents the 'beads only' control.

4 Discussion

Described here is a case of Fanconi anaemia with compound heterozygous alleles for PALB2/FANCN (identified as patient 10). The paternal mutation is c.2586+1G>A that leads to skipping of exon 6 in the mRNA and the in frame deletion of all 24 amino acids of exon 6 (abbreviated here as 'X6'). Transcripts from this allele are translated from exon 5 into exon 7 in-frame (See figure 4.1). The maternal mutant allele deletes two adenosines at nucleotides 1676 and 1677, replacing them with a guanine (1676_1677delAAinsG, abbreviated here as 'del/ins'). This results in a translational frame shift that truncates the PALB2 protein at amino acid 559, which is also changed from glutamine to arginine (p.Gln559Argfs*2).
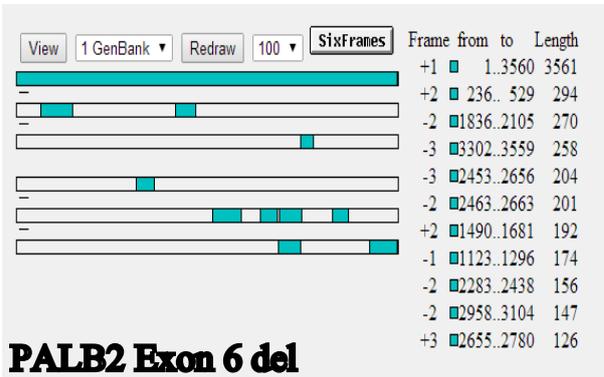
Figure 4.1 A figure to show the open reading frame with and without exon 6 Results acquired from NCBI: ORF Finder
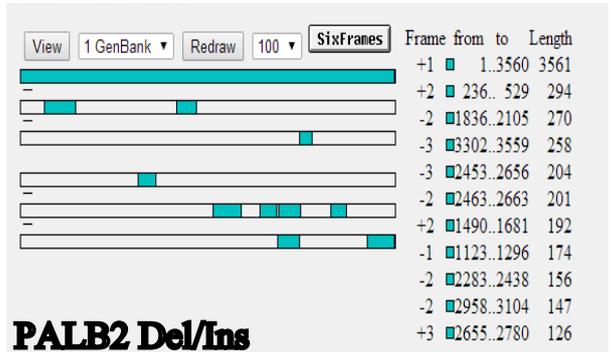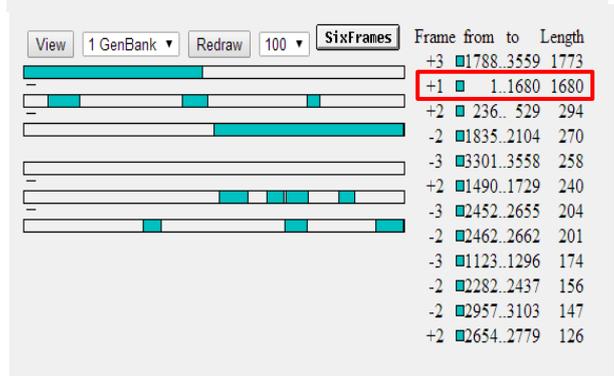
Figure 4.2. A figure to show the frame shift incurred by the del/ins mutant. Results acquired from NCBI: ORF Finder.

4.1 Assessing PALB2 protein level in 6 family members

I tested PALB2 expression in patient whole cell lysates using Western blot and antibody PALB2 246A (figure against antibody PALB2 246A figure 3.1b) and PALB2 247A (figure against antibody PALB2 247A figure 3.1b). Figure 3.1 and 3.2 confirmed the presence of PALB2 from two mutant alleles in the patient, from different single mutations in each parental cell line and two other children. A fourth child had normal PALB2 consistent with homozygosity for wild type alleles.

Figure 3.1b and 3.2b show a variation in sample loading which caused difficulty when observing the relevant bands, since this was the first Western blot performed in the laboratory my knowledge was limited, therefore I was unaware that this Western blot should have been repeated. This has taught me that checking the loading controls during a Western blot is imperative during future experiments.

4.2 Generation of Cell Lines Expressing Individual Mutant or WT PALB2 Proteins

In keeping with previous Fanconi anaemia related PALB2 experiments (Xia *et al.* 2006a, Zhang *et al.* 2009b) cell lines were generated by site directed mutagenesis in order to study the separate PALB2 mutations in the patients. The pcDNA5-FLAG construct used here was specifically chosen for its tetracycline induction system. Using an efficient induction system expression of the recombinant protein can be monitored and controlled. Cell lines that contained mutations (Including a revertant sequence) from another patient EUFA1341 (Xia *et al.* 2006b) were also generated for comparison with the germline *PALB2* mutations of my patient (patient 10). The EUFA1341 mutation Y551X and a revertant mutation from the same cell line that deletes exon 4 (X4) were chosen instead of the other 6 previously published cases

for a number of reasons; first, the Y551X allele resulted in a truncated protein whereby the C-terminal WD40 repeats are removed, the consequence of which is similar to that of the del/ins mutation in my patient which suffers a frame shift and truncation at codon 559. Second, the in-frame exon 4 deletion revertant mutation of EUFA1341 retains both the N-terminus and the C-terminus, including the WD40 repeats, and retains BRCA1 and BRCA2 binding capacity and some PALB2 activity. Finally, the EUFA1341 mutants are a good reference for comparison with other studies because they have been used widely in the literature.

4.3 Sanger Sequencing

Proof of successful mutagenesis was shown by DNA sequencing (figure 3.3) showing that amplicons are a result of the transfected pcDNA5-PALB2 construct rather than endogenous PALB2. The importance of sequencing was demonstrated during the first attempt of sequencing where some colonies were found to have random mutated artefacts, in addition to un-mutated regions where mutations were expected. It is unclear whether this is a result of sample mix-up, mutagenesis failure or inaccurate sequencing however further sequencing of alternative colonies proved that mutagenesis had been successful in the picked colonies. Sequencing analysis, as seen in figures 3.4 – 3.10, identified the correct mutations in the correct cell lines and therefore, successful mutagenesis. From this stable U2OS cell lines expressing the correct PALB2 mutants in the correct sample have been confirmed.

4.4 Doxycycline Induction

The tetracycline-controlled transcriptional activation system, as part of the original pcDNA5-PALB2 vector, was used to control expression of each mutant. This induction system was first described by Gossen and Bujard (1992), by manipulating

the Tn*10* tetracycline repressor system found in *E.coli*. The tetracycline repressor (*tet*R) controls expression of the tetracycline resistance genes. In the presence of tetracycline (or a suitable derivative) *tet*R is left unbound to promoters within the operon therefore, resistance genes are allowed to overexpress. A construct was developed whereby any core promoter fused to this system would be stimulated into over expression of the target gene. Protein translation will lie almost dormant until the concentration of tetracycline is rapidly increased in the cell. Once induced this system can increase expression to over five orders of magnitude.

Figure 3.11 displays varied responses to the tetracycline induction system. For the over expressers, in particular X4, the difference in expression between the un-induced (lane 13) and induced (lanes 14 and 15) clearly shows that not only does the construct work, but that it is extremely efficient in this cell line. A similar pattern was found in the Y551X (lanes 10-12) mutant which again, showed effective expression in this cell line between induced (lanes 11 and 12) and uninduced (lane 10). The construct seemed to be less active in the X6 (lanes 4-6) and del/ins (lanes 7-9) cells, although bolder bands are seen at ~170 kDa and ~100 kDa, respectively, in some cell lines on the Western blots. Results from X6 (lanes 4-6) and del/ins (lanes 7-9) indicate a partially inducible construct as the increase in expression is not as notable as that from X4 (lanes 13-15) and Y551X (lanes 10-12) cell lines. Induction of WT (lanes 1-3) indicated little to no change in expression compared to that over the un-induced control. This is somewhat of a concern as the wild type cell line was to be used as a control against the other mutants. Further testing was required to be able to provide resolution of this problem. There are a number of reasons why all cell lines displayed varied results after doxycycline induction; firstly, the stability of the host cell

lines is a major factor since an unstable lineage can cause ineffective transcription and inevitably, cell death. However, the U2OS host, as used in all mutant cell lines, is one recommended by the T-REx™ user manual (Invitrogen 2011) making this an unlikely cause. Additionally, the expression levels of the protein also rely on the number of cells available, the fewer the cells, the less available DNA to transcribe from. Inadequate volume of cells is an unlikely cause for lack of expression as a 50 µg loading control was used (figure 3.11b), showing equal expression across all lanes. Furthermore, cells were visually assessed for confluence in the tissue culture dish before doxycycline induction (Berens and Hillen 2003). Figure 3.4 outlines how effective the doxycycline induction was using cell lines containing a truncated PALB2 mutant construct (Y551X and X4), whereas the longer proteins were less susceptible to induction (WT and X6). Taken together it would seem that there is another reason for the lack of expression in certain cell lines, proposing the question whether over expression of full-length PALB2 is toxic to the cell.

4.5 Choosing mutant cell lines
It is important for further experiments that all cell lines are expressing PALB2 at similar levels. Figures 3.12 – 3.16 display the various expression levels shown between different cultures in all five experimental cell lines. In keeping with the Western blot shown in figure 3.11, expression varied between cultures within the same cell line after 48 hour doxycycline induction. X4 (figure 3.12, lanes 1-18) and Y551X (figure 3.15, lanes 1-18) constructs displayed high levels of protein expression after doxycycline induction compared to the un-induced control (lane C), these results are similar to the expression levels shown in figure 3.11. X6 induction (figure 3.13) indicated expression at the expected band (~160kDa) in all induced

cultures (lanes 1-18) with no expression in the un-induced (lane C), confirming the success of the construct in this cell line. In contrast to the previous induction test (figure 3.11) where del/ins displayed weak induction efficiency, figure 3.14 proved certain strains of del/ins were inducible using this construct. Again, there was no notable change in WT expression after doxycycline induction in all 18 colonies, further supporting the theory that over expression of WT PALB2 could be toxic to the cell. Other explanations include the use of antibody when using PALB2 246A or PALB2 247A during Western blotting. With the exception of wild type all other mutants are expressed at different sizes to endogenous PALB2 (~170kDa), therefore making it impossible to differentiate between the expression of the WT and endogenous PALB2 when using one of these antibodies. Additionally, the M2 FLAG antibody used to blot during these experiments generated a lot of background 'noise' on the blots, it is therefore difficult to differentiate between what is a direct effect of our induction system and a result of this 'noise'.

4.6 Small Interfering RNA

By knocking down endogenous PALB2, it was possible to ensure any further interactions between PALB2 and other proteins within the cell were a direct effect of the recombinant, mutant protein. Firstly, the optimal treatment conditions needed to be established, including concentration of siRNA and duration of treatment. Using the del/ins cell line siRNA concentration levels were staggered at 22.5 nM, 45 nM, 90 nM and 180 nM for 24, 48 and 72 hours. Five controls included, four scrambled RNA samples transfected with 22.5 nm, 45 nm, 90 nm and 180 nm for 48 hours each and one un-induced control. To test this protocol the del/ins mutant was chosen as it is distinctively smaller than endogenous PALB2 and, therefore, there would be no

dispute whether or not any interaction was a reflection of mutant or endogenous PALB2 activity. Successful knockdown using this siRNA was achieved, as shown in figure 3.17, lanes 2 - 13. Using antibody PALB2 246A, high levels of del/ins expression was detected at ~100kDa. Additionally, absence of knockdown is displayed in all control samples (lanes 1, 14-17). These results suggest a successful knockdown of endogenous PALB2 with simultaneous induction of mutant del/ins. Using these results optimal conditions for further experiments were established and it was concluded that transfection of siRNA in a 45 nM concentration for 48 hours produced the required level of PALB2 knockdown. Figure 3.18 shows equal loading of 50 µg using aprataxin (~40 kDa).

4.7 Immunoprecipitation

Immunoprecipitation using anti-FLAG to pull down the FLAG containing PALB2 mutant protein and other proteins in complex was achieved as seen in figure 3.19. Protein expression is seen at ~100 kDa in lanes 3 and 4, and faint bands in lanes 5 and 6, all of which represent successful pull down of the del/ins mutant immunoprecipitating with 2, 4, 6 and 8 µg of FLAG antibody (lanes 3-6, respectively). Lane 1, 2 and 7 represent the control samples, lane 1 is a doxycycline induced whole cell lysate, lane 2 shows an un-induced lysate with antibody and protein G sepharose beads, and lane 7 shows the induced lysate with protein G beads and no antibody. The expression of endogenous PALB2 in the no-bead control validates the specificity of the antibody to the mutant since endogenous PALB2 is expressed in this sample and not in the rest. Lane 2, where the un-induced control is present shows a faint band within the 100 kDa region, this control was used to compare how much protein was available for IP before doxycycline induction. Lane 7 proves the specificity of the

antibody to the protein since no antibody was added to the solution no protein would be expected to pull-down. This experiment has proved that immunoprecipitation of mutant proteins is possible using anti-FLAG. Furthermore, it also shows that 4 µg of anti-FLAG is the optimal concentration to immunoprecipitate using this protocol. Moving on from this it would be possible to use this Western and blot for BRCA1 and BRCA2 to assess whether the complex has changed to accommodate this mutant.

4.8 Troubleshooting

Time was a difficult factor from the outset of this project, 17 weeks to create and grow five stable mutant cell linages with similar expression levels and perform numerous experimental methods was ambitious. With this in mind preliminary and optimisation experiments were used to avoid spending time constructing the same experiment numerous times. For instance, after site directed mutagenesis the cells were allowed to grow up on a petri dish, after days of growth twelve colonies were picked in to wells of a 96 well plate. PCR and DNA sequencing was performed on one of the twelve colonies picked, when one of the sequence analyses returned unsuccessful another colony was promptly chosen from the 96 well plate, PCR amplified and sequenced, this time round analysis proved a successful mutant was present. Having these additional colonies ready created a more efficient process whereby the mutant cells were available without going through the lengthy mutagenesis and transfection processes again. Additionally, during DNA sequencing each mutant was sequenced with both the forward and the reverse primer for the relevant pair to provide two separate results for the same mutant to ensure the correct mutants were in place. This proved to be useful during analysis when identifying the unsuccessful mutant by checking both electropherograms for confirmation.

Other preliminary experiments undertaken included the doxycycline induction, a test which identified the optimal time needed to induce the cells. Establishing optimal conditions of doxycycline treatment early on meant cells were not induced for longer than necessary, therefore saving time. More importantly, optimisation of induction conditions meant the construct was induced long enough to be able to effectively increase PALB2 expression. As shown in figure 3.11, cells treated for 48 hours did not display much increase in expression compared to those treated for 72 hours, this result indicates a plateau effect therefore underlying the importance of this experiment in enabling optimal use of time and efficiency of the construct. Similarly, the siRNA transfection was performed with two variables, concentration of siRNA and duration of treatment. The same principle applies here, by optimising these two variables both time and resources would be saved upon completion of further siRNA treatments. Optimisation of IP experiments by staggering antibody concentration during immunoprecipitation meant resources could be saved by not overusing the antibody during each IP. Finally, further troubleshooting was addressed by using different antibodies depending on the mutant strain used, for example the deleted exon 4 in mutant X4 removes a 491 amino acid sequence (residues 412-1884) which also deleted the epitope to PALB2 246A antibody, located at residues 200-250. It was not until one Western blot failed to show any X4 activity after doxycycline induction that this was brought to light, however PALB2 247A provided an epitope available for X4 binding. As previously addressed other antibody issues arose during the course of this project, mostly from the unspecific binding of FLAG. Figure 4.3 shows a blot run after siRNA transfection using the same lysates as those shown in figure 3.17 using PALB2 246A. However in Fig 4.3 instead of indicating knockdown

of endogenous PALB2 this blot suggests that there is still protein expression. The conclusion is that the 170kDa band in Fig 4.3 is a consequence of non-specific binding by the FLAG Ab. In order to avoid this issue arising again another FLAG antibody was ordered for future experiments.

4.9 Applications to the Field

4.9.1 Is PALB2 toxic to the cell?

Although it is unfinished, this work has brought to light many questions regarding the functionality of PALB2 and its mutants. Firstly, the over expression of full-length PALB2 has been indicated as potentially toxic to the cell, after evaluating the available literature it would seem that overexpression of PALB2 has not previously been reported as being toxic to the cell; furthermore, Reid *et al*. (2007), Xia *et al.* (2006b), Zhang *et al.* (2009a) and Zhang *et al*. (2009b) all published results where recombinant PALB2 was over expressed in U2OS host cells, in each case their experiments were successful and produced valid data. In addition, PALB2 was also
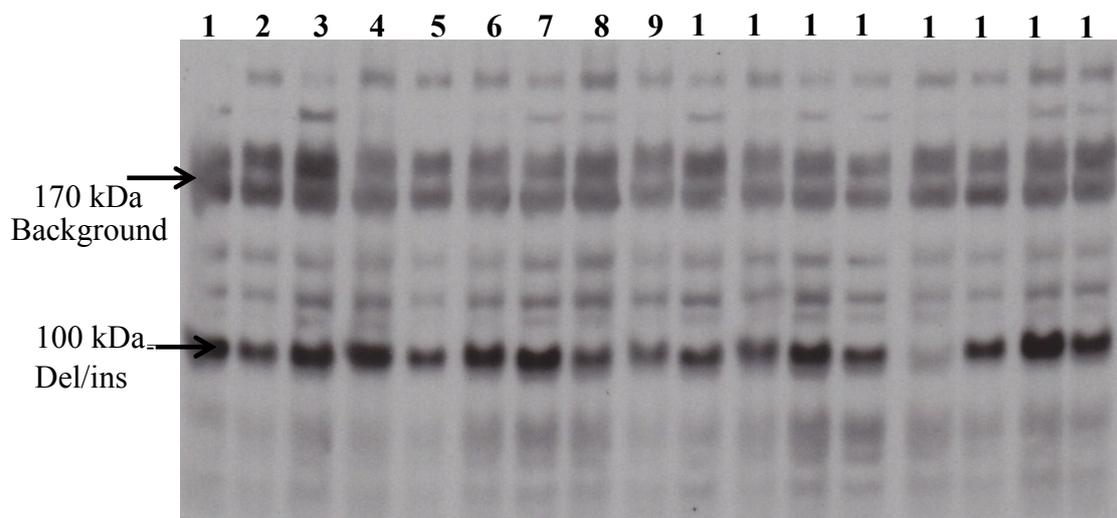
Figure 4.3. Western blot after knockdown of endogenous PALB2 and detection of induced PALB2 using anti-FLAG. The bands at ~170 kDa, where the endogenous PALB2 would be expected, are not present using PALB2 246A (Fig 3.17) but appear to be present here using FLAG Ab. The signal here at ~170kD in Fig 4.3 is a consequence of unspecific binding of FLAG Ab.

expressed in HeLa, 293/293T cells, non-neoplatic, epithelial cells, fibroblasts, LCLs, SF9 insect cells, and lymphocytes without any mention of any toxic responses to the cell (Oliver *et al.* 2009).

4.9.2 Does del/ins function in the FA pathway?

Serra *et al* (2012) recently published a case from a FA patient homozygous for mutant del/ins (1676_1677DelAA/InsG), the same as the mutation in my patient the investigated here. Their patient exhibited the most severe form of FA phenotype with embryonal tumours, Wilms, neuroblastoma and a medulloblastoma as well as abnormal thumbs, horseshoe kidney, sacral dysplasia and uterine bicorn. Importantly, Serra *et al.* (2012) noted that they did not detect any notable PALB2/FANCN activity in patient cells as measured by MMC induced chromosome breakage analysis (100% of the metaphase cells showed >10 breaks after MMC exposure). However, their report lacked much information on the functionality and activity of this mutant. Located at the C-terminus of PALB2, seven WD40 repeats form a seven bladed locking mechanism whereby the N-terminus of the first blade binds to the C-terminal tiers of the seventh blade, thereby stabilising PALB2. Both patients 8 and 9 (see table 1.1) exhibit heterozygous Y1183X alleles, a mutant PALB2 protein which truncates 3 codons upstream of the termination codon. Y1183X causes the loss of WD40 repeat domain and in turn, the locking mechanism, since the loss of only the final 3 codons is enough to destabilise PALB2 in these patients, it would seem logical to assume that del/ins produces similar, destabilising effects. Despite this, research into the effect of embryonic development in PALB2 knockout mice suggests that although these truncated proteins are no longer stable, they are still generating some form of activity from the N-terminal, coiled coil region. Both

studies by Rantakari *et al.* (2010) and Bouwman *et al*. (2011) clearly suggest that embryonic lethality within early gestation was a direct result of PALB2 knockout. All 10 patients listed in table 1.1 present either homozygous or compound heterozygous mutants of PALB2 but yet survived passed birth, taken together this data suggests that hypomorphic PALB2 activity is present in the identified mutants.

4.10 Future Work
Firstly, to further test the efficiency of the induction construct in WT cell lines Western blot would be used using a different branded FLAG antibody if, as before, there is no change in the expression of un-induced compared with the induced samples. If this antibody also proves to be unspecific, another option would be to siRNA treat WT to knockdown endogenous PALB2 since we have already proved that the siRNA transfections work, Western blot analysis with antibody PALB2 246A or 247A would generate bands at ~170 kDa where WT expression should lie, since endogenous PALB2 would be knocked down the resulting bands would be an effect of doxycycline induction, representing the induced expression of WT PALB2.

By inducing cells with siRNA and doxycycline, knockdown of endogenous PALB2 and expression of the mutant PALB2, followed by subsequent reciprocal co-immunoprecipitation and Western blot experiments, it will be possible to study the implications of X6, X4, Y551X and del/ins expression on PALB2 binding affinity with all known binding partners of PALB2. To check the integrity of our experimental procedures our results would be compared to that published by Xia *et al*. (2006b) Sequential experiments would involve understanding whether the PALB2 mutants retain their ability to localise to the nucleus after IR/MMC treatment. Immunofluorescence would visualise the movements of any tagged protein both

before and after IR/MMC treatment, using the two results to compare PALB2 concentration in the nucleus with results from wild-type. Subcellular fractionation could further aid in assessing the localisation of mutant PALB2 within the nucleus, for example, does the complex locate to the foci of DNA double stranded breaks like WT PALB2? (Xia *et al*. 2006a, Sy *et al.* 2009, Zhang *et al.* 2009b). Flowcytometry can separate cells by size and position in the cell cycle, if DNA DSBs remain unrepaired when the cell reaches the G2 check point it will remain in G2 until the DSB is repaired and only then will it progress onto mitosis. An ineffective DSB repair pathway caused by inactive mutant PALB2 would result in a high volume of cells at G2 arrest after IR/MMC treatment, as shown by the flowcytometer. Analysis of PALB2 expression and other interacting partners can also be achieved using a flowcytometer. Additionally, cell apoptosis and proliferation using fluorescent antibodies bound to cell surface proteins can be monitored with using flowcytometry. After assessing the effect of mutant PALB2 on the cells, testing for protein binding sites, checking localisation, expression levels, and effect on homologous recombination, WT will be transfected into the same cell lines as the mutant to see if PALB2 functions and binding partners are restored, thereby confirming the loss of function from mutant alleles.

5. Conclusion

PALB2/FANCN is one of the 15 currently known FA associated proteins, functionally involved in the DNA double stranded break repair pathway via homologous recombination (Xia *et al*. 2006a, Sy *et al*. 2009a, Zhang *et al*. 2009a, Zhang *et al* 2009b). Of the 15 currently known FA complementational groups, FA-N is one of the more recently described. Of the mutants used here, X4 and Y551X have been previously described in several publications and their consequences upon cellular activity have been well documented (Xia *et al*.2006a, Sy *et al*. 2009a, Zhang *et al*. 2009a, Zhang *et al*. 2009b). Cells from the *PALB2* Del/ins and X6 from a patient was investigated here. Mutant del/ins has previously been reported (Serra *et al*. 2012) in a homozygous PALB2 patient, however the impact this mutation had on PALB2 function was not studied fully by these authors. Interestingly, my patient was diagnosed with lymphoma, a disease not recognised as being associated with FA. My project set out to create five cell line each carrying a different PALB2 mutation mutant in order to evaluate their effect on the DSB repair pathway at the cellular level, paying particular attention to the patient 10 copied PALB2 mutants (X6 and del/ins). The size and expression of PALB2 mutants from patient 10, both parents and two siblings, and the homozygous wild-type alleles in one sibling were confirmed using Western blot. Site directed mutagenesis followed by transfection of the construct, that included a tetracycline-controlled transcriptional activation system, in frame FLAG sequence and one of four mutant or the wild-type PALB2 alleles, was confirmed using PCR amplification as well as DNA sequencing. Activity of the tetracycline regulatory system was established and cell lines were chosen based on their PALB2 mutant expression. Endogenous PALB2 knockdown was achieved in del/ins cell line using siRNA and confirmed with Western blotting against antibody

PALB2 246A. Immunoprecipitation of the del/ins mutant protein using anti-FLAG in knocked down del/ins cell lines was confirmed using Western blot with antibody PALB2 246A. All experimental results are preliminary and much work is still to be done to be able to draw a definitive conclusion on the impact each mutant has upon PALB2 activity and, in turn, how this affects the DSB pathway. Moreover, this information can be used to further understand the link between PALB2 activity and lymphoma development in the two affected individuals described here.

Bibliography

American Cancer Society. (2011). **Global Cancer Facts and Figures 2nd Edition** [online]. Atlanta: American Cancer Society. Available from: http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-027766.pdf [Accessed: 15-06-2014]

Bastard, C. Tilley, H. Lenormand, B. Bigorgne, C. Boulet, D. Kunlin, A. Monconduit, M. Piguet, H. (1992). Translocations involving Band 3q27 and Ig Gene Regulations in Non-Hodgkin Lymphoma. **Blood**, 79: (10): 2527-2531.

Berens, C. and Hillen, W. (1998). Gene regulation by tetracyclines. Constrains of resistance in bacteria shape TetR for application in eukaryotes. **European Journal of Biochemistry**, 270: (15): 3109-3121.

Bleuyard, J. Buisson, R. Masson, J. Esashi, F. (2012). ChAM, a novel motif that mediates PALB2 intrinsic chromatin binding and facilitates DNA repair. **EMBO Reports**, 13: (2): 135-141.

Buisson, R. Dion-Cote, A. Coulombe, Y. Launay, H. Cai, H. Stasiak, A. Stasiak, A. Xia, B. Masson, J. (2010). Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating HR. **Nature Structural and Molecular Biology**, 17: (10): 1247-1254.

Buisson, R. Niraj, J. Pauty, J. Maity, R. Zhoa, W. Coulombe, Y. Sung, P. Masson, J. (2014). Breast Cancer Proteins PALB2 and BRCA2 Stimulate Polymerase η in Recombination-Associated DNA Synthesis at Blocked Replication Fork. **Cell Reports**, 6: (3): 553 – 564.

Bouwman, P. Drost, R. Klijn, C. Pieterse, M. van der Gulden, H. Song, J. Szuhai, K. Jonkers, J. (2011). Loss of p53 partially rescues embryonic development of *Palb2*

knockout mice but does not foster haploinsufficiency of *Palb2* in tumour suppression. **Journal of Pathology**, 224: (1): 10-21.

Campo, E. Swerdlow, S. Harris, N. Pileri, S. Stein, H. Jaffe, E. (2011). The 2008 WHO Classification of Lymphoid Neoplasms and Beyond: Evolving Concepts and Practical Application. **Blood**, 117: (19): 5019 – 5032.

Cancer Research UK. (2013). **Non-Hodgkin Lymphoma Cancer Statistics Report** [Online]. London: Cancer Research UK. Available from: http://publications.cancerresearchuk.org/downloads/Product/CS_REPORT_NHL.pdf [Accessed: 15-06-2014].

Chen, G. Tisayakorn, S. Jorgenson, C. D'Ambrosio, L. Goudreault, M. Gingras, A. (2008). Mechanisms of Signal Transduction: PP4R4/KIAA1622 Forms a Novel Stable Cytosolic Complex with Phosphoprotein Phosphatase 4. **The Journal of Biological Chemistry** [Online]. 283. Available at: http://www.jbc.org/content/283/43/29273.full [Accessed: 04/04/14].

Clark, S. Rodriguez, A. Snyder, R. Hankins, G. Boehning, D. (2012). Structure – Function of the Tumour Suppressor BRCA1. **Computational and Structural Biotechnology Journal** [Online]**. 1: (1).** Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3380633/. [Accessed: 23/06/2014]

Constantinou, A. 2010. Rescue of replication failure by Fanconi Anaemia proteins. Chromosoma. 121 (1), 21 – 36.

Cybulski, K. and Howlett, N. (2011). FANCP/SLX4: a Swiss army knife of DNA interstrand crosslink repair. **Cell Cycle**, 10: (11): 1757-1763.

de Winter, J. van der Weel, L. de Groot, J. Stone, S. Waisfisz, Q. Arwert, F. Scheper, R. Kruyt, F. Hoatlin, M. Joenie, H. (2000). The Fanconi Anaemia Protein FANCF

Forms a Nuclear Complex with FANCA, FANCC, and FANCG. **Human Molecular Genetics**, 9: (18): 2665-2674.

Dosanjh, M. Collins, D. Fan, W. Lennon, G. Albala, J. Shen, Z. Schild, D. (1998). Isolation and characterisation of RAD51C, a new human member for the RAD51 family of related genes. **Nucleic Acids Research**, 26: (5): 1179-1184.

Ekstrom-Smedby, K. (2006). Epidemiology and etiology of non-Hodgkin lymphoma – a review. **Acta Oncologica**, 45: (3): 258-271.

Fulop, V. Jones, D. (1999). B-propellers: Structural Rigidity and Functional Diversity. **Current opinion in Structural Biology,** 9: (6): 715-721.

Galkin, V. Wu, Y. Zhang, X. Qian, X. Yu, X. Hever, W. Luo, Y. Egelman, E. (2006). The Rad51/RadA N-terminal domain activates nucleoprotein filament ATPase activity. **Structure**, 14: (6): 983-992.

Garcia-Higuera, I. Kuang, Y. D'Andrea, A. (1999). Fanconi Anaemia Proteins FANCA, FANCC and FANCG/XRCC9 Interact in a Functional Nuclear Complex. **Molecular and Cellular Biology**, 19: (7): 4866-4873.

Garcia-Higuera, I. Taniquchi, T. Ganesan, S. Meyn, M. Timmers, C. Hejna, J. Grompe, M. D'Andrea, A. (2001). Interaction of the Fanconi Anaemia Proteins and BRCA1 in a Common Pathway. **Molecular Cell**, 7: (2): 249-262.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. **Proceedings of the National Academy of Science USA,** 89: (12): 5547 – 555.

Gingras, A. Caballero, M. Zarske, M. Sanchez, A. Hazbun, T. Fields, S. Sonenberg, N. Hafen, E. Raught, B. Aebersold, R. (2005). A Novel, Evolutionarily Conserved

Protein Phosphatase Complex Involved in Cisplatin Sesitivity. **Molecular and Cellular Proteomics,** 4: (11): 1725-1740.

Hanukogle, I. Tanese, N. Fuchs, E. (1983). Complimentary DNA sequence of a human cytoplasmic actin: interspecies Divergence of 3' non-coding regions. **Journal of Molecular Biology,** 163: (4): 673-678.

Hayakawa, T. Zhang, F. Hayakawa, N. Ohtani, Y. Shinmyozu, K. Nakayania, J. Andreassen P. (2010). MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. **Journal of Cell Science**, 123: (7): 1124-1130.

Hill, D. Wang, S. Cerhan, J. Davis, S. Cozen, W. Severson, R. Hartage, P. Wacholder, S. Yeager, M. Chanock, S. Rothman, N. (2006). Risk of non-Hodgkin lymphoma (NHL) in relation to germline variation in DNA repair and related genes. **Blood**, 108: (9): 161-167.

Hodgkin, T. (1832). On Some Morbid Appearances of the Absorbent Glands. **Medico-Chirurgical Transactions,** 17: 68-114.

Howlett, N. (2007). Fanconi Anaemia: Fanconi anaemia, breast and embryonal cancer risk revisited. **European Journal of Human Genetics**, 15: (7): 715-717.

Invitrogen. 2012. **Flp-In™ T-REx™ Core Kit** [Online]. Life Technologies. Available from: http://tools.lifetechnologies.com/content/sfs/manuals/flpintrex_man.pdf [Accessed 07/05/2014].

Ishiai, M. Kitao, H. Smogorzewska, A. Tomid, J. Kinomura, A. Uchida, E. Saberi, A. Kinoshita-Kikuta, E. Koike, T. Tashiro, S. Elledge, S. Takata, M. (2008). FANCI phosphorylation functions as a molecular switch to turn on the Fanconi Anaemia Pathway. **Nature Structural and Molecular Biology**, 15: (11): 1138-1146.

Joenje, H. Lo ten Foe, J. Oostra, A. van Berkel, C. Rooimans, M. Schroeder-Kurth, T. Wegner, R. Gille, J. Buchwald, M. Arwert, F. (1995). Classification of Fanconi anaemia patients by complementation analysis: evidence for fifth genetic subtype. **Blood,** 86: (6): 2156-2160.

Kerckaert, J. Deweindt, C. Tilley, H. Quief, S. Lecocq, G. Bastard, C. (1993). LAZ3, a novel Zinc-Finger Encoding Gene, is Disrupted by Recurring Chromosome 3q27 Translocations in Human Lymphomas. **Nature Genetics**, 5: (1): 66-70.

Kim, J. Kee, Y. Gurtan, A. D'Andrea A. (2008). Cell Cycle Dependent Chromatin Loading of Fanconi Anaemia Core Complex by FANCM/FAAP250. **Blood**, 111: (10): 5215 – 5222.

Kupfer, G. Naf, D. Suliman, A. Pulsipher, M. D'Andrea, A. (1997). The Fanconi Anaemia Proteins FAA and FAC Interact to Form a Nuclear Complex. **Nature Genetics,** 17: (4): 487-490.

Kutler, D. Auerbach, A. (2004). Fanconi anaemia in Ashkenazi Jews. **Familial Cancer**, 3: (3-4): 241–248.

Kutler, D. Auerbach, A. Satagopan, J. Giampietro, P. Batish, S. Huvos, A. Goberdhan, A. Shah, J. Singh, B. High Incidence of Head and Neck Squamous Celll Carcinoma in Patients with Fanconi anaemia. (2003) **Archives of otolaryngology – Head and Neck** Surgery, 129: (1): 106-112.

Levitus, M. Rooimans, M. Steltenpool, J. Cool, N. Oostra, A. Mathew, C. Hoatlin, M. Waisfisz, Q. Artwert, F. de Winter, P. Joenje, H. (2004). Heterogeneity in Fanconi anaemia: evidence for 2 new genetic subtypes. **Blood**, 103: (7): 2498-2503.

Invitrogen. (2011). T-REx™ System. A Tetracycline-Regulated Expression System for Mammalian Cells [Online]. Life Technologies. Available at:

http://tools.lifetechnologies.com/content/sfs/manuals/trexsystem_man.pdf [Accessed: 21/04/2014].

Ling, C. Ishiai, M. Ali, A. Medhurst, A. Neveling, K. Kalb, R. Yan, Z. Xue, Y. Oostra, A. Auerbach, A. Hoatin, M. Schindler, D. Joenie, H. de Winter, J. Takata, M. Meetei, A. Wang, W. (2007). FAAP100 is Essential for Activation of the Fanconi Anaemia Associated DNA Damage Response Pathway. **EMBO Journal**, 26: (8): 2104-2014.

Liu, J. Zheng, Q. Deng, Y. Cheng, C. Kallenbach, N. Lu, M. (2006). A Seven-Helix Coiled Coil. **Proceedings of the National Academy USA**, 103: (42): 15457-15462.

Lobitz, S and Velleuer, E. (2006). Guido Fanconi (1892-1979): a jack of all trades. **Nature**, 6: (11): 892-898.

Lupas, A. (1996). Coiled coils: New Structures and New Functions. **Trends in Biochemical Sciences**, 21: (10): 375-382.

Mann, W. R. Venkatraj, V. S. Allen, R. G. Liu, Q. Olsen, D. A. Adler-Brecher, B. Mao, J.-I. Weiffenbach, B. Sherman, S. L. Auerbach, A. D. (1991). Fanconi anaemia: evidence for linkage heterogeneity on chromosome 20q. Genomics, (9), 329-337.

Meetei, A. Sechi, S. Wallisch, M. Yang, D. Young, M. Joenie, H. Hoatlin, M. Wang, W. (2003a). A Multi-Protein Nuclear Complex Connects Fanconi Anaemia and Bloom Syndrome. **Molecular and Cellular Biology**, 23: (10): 3417-26.

Meetei, A. de Winter, J. Medhurst, A. Wallisch, M. Waisfisz, Q. van de Vrugt, H. Oostra, A. Yan, Z., Ling, C. Bishop, C. Hoatlin, M. Joanie, H. Wang, W. (2003b). A Novel Ubiquitin Ligase is Deficient in Fanconi Anaemia. **Nature Genetics**, 35: (2): 165-170.

Meetei, A. R. Medhurst, A. L. Ling, C. Xue, Y. Singh, T. R. Bier, P. Steltenpool, J. Stone, S. Dokal, I. Mathew, C. G. Hoatlin, M. Joenje, H. de Winter, J. P. Wang, W.

(2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anaemia complementation group M. Nature Genetics, 37: (9): 958-963.

Miki, T. Kawamata, N., Hirosawa, S. Aoki, N. (1994). Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5, encodes a Kruppel-like zinc-finger protein. **Blood**, 83: (1): 26-32.

Ogawa, T. Yu, X. Shinohara, A. Egelman, E. (1993). Similarity of the Yeast RAD51 filament to the bacterial RecA filament. **Science**, 253: (5103): 1896-1899.

Oliver, A. Swift, S. Lord, C. Ashworth, A. Pearl, L. (2009). Structural Basis for Recruitment of BRCA2 by PALB2. **EMBO Reports**, 10: (9): 990-996.

Osorio, A. Boglio, M. Fernandez, V. Barroso, A. de la Hoya, M. Caldes, T. Lasa, A. Ramon y Cajal, T. Santamarina, M. Vega, A. Quiles, F. Lazaro, C. Diez, O. Fernandez, D. Gonzalez-Sarmiento, R. Duran, M. Pigueras, J. Marin, M. Pujol, R. Surralles, J. Benitez, J. (2013). Evaluation of rare variants in the new Fanconi anaemia gene ERCC4 (FANCQ) as familial breast/ovarian cancer susceptibility alleles. Human Mutation, 34: (12): 1615-1608.

Park, J. Singh, T. Nassar, N. Zhang, F. Feund, M. Hanenburg, H. Meetei, A. Andreassen, P. (2013). Breast cancer associated mis-sense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. **Oncogene** [online]. Available from: http://www.nature.com/onc/journal/vaop/ncurrent/pdf/onc2013421a.pdf. [Accessed: 22/06/14].

Qiagen. 2006. DNeasy Blood and Tissue Handbook [Online]. Qiagen. Available: http://mvz.berkeley.edu/egl/inserts/DNeasy_Blood_&_Tissue_Handbook.pdf [Accessed 14/05/2014].

Rantakari, P. Nikkila, J. Jokela, H. Ola, R. Pylkas, K. Lagerbahm, H. Sainio, K. Poutanen, M. Winqvist, R. (2010). Inactivation of *Palb2* gene leads to mesoderm differentiation defect and early embryonic lethality in mice. **Human Molecular Genetics,** 19: (15): 3021-3029.

Rass, U. Ahel, I. West, C. (2007). Actions of Aprataxin in Multiple DNA Repair Pathways. **Journal of Biological Chemistry**, 282: (13): 9469-9474.

Reid, S. Schindler, D. Hanenberg, H. Barker, K. Hanks, S. Kalb, R. Neveling, K. Kelly, P. Seal, S. Freund, M. Wurm, M. Batish, S. D. Lach, F. P. Yetgin, S. Neitzel, H. Ariffin, H. Tischkowitz, M. Mathew, C. G. Auerbach, A. D. Rahman, N. (2006). Biallelic mutations in PALB2 cause Fanconi anaemia subtype FA-N and predispose to childhood cancer**.** Nature Genetics, 39: (2): 162-164.

Schroeder, T. Anschútz, F. Knopp, A. (1964). Spontaneous Chromosome abberations in familial panmyelopathy. **Humangenetik**, 1: (2): 194-196.

Serra, A. Eirich, K. Winkler, A. Mrasek, K. Gohring, G. Barbi, G. Cario, H. Schlegelberger, B. Pokora, B. Liehr, T. Leriche, C. Henne-Bruns, D. Barth, T. Schindler, D. (2012). Shared Copy Number Variation in Simultaneous Nephroblastoma and Neuroblastoma due to Fanconi Anaemia. **Molecular Syndromology**, 3: (3): 120-130.

Shen, M. Zheng, T. Lan, Q. Zhang, Y. Zahm, S. Wang, S. Holford, T. Leaderer, B. Yeager, B. Welch, R. Kang, D. Boyle, P. Zhang, B. Zou, K. Zhu, Y. Chanock, S. Rothman, N. (2006). Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma among women in Connecticut. **Human Genetics**, 119: (6): 659-668.

Singh, T. Ali, A. Paramasivam, M. Pradhan, A. Wahengbam, K. Seidman, M. Meetei, A. (2013). ATR-dependent phosphorylation of FANCM at serine 1045 is essential for FANCM functions. **Cancer Research**, 73: (14): 4300-4310.

Smedby, K. Lindgren, C. Hjalgrim, H. Humphreys, K. Schóllkopf, C. Chang, E. Roos, G. Ryder, L. Flak, K. Palmgren, J. Kere, J. Melbye, M. Glimelius, B. Adami, H. (2006). Variations in DNA repair genes ERCC2, XRCC1 and XRCC3 and risk to follicular lymphoma. **Cancer Epidemiology, Biomarkers and Prevention**, 15: (2): 258-265.

Strathdee, C. Duncan, A. Buchwald, M. (1992). Evidence of at least four Fanconi anaemia genes including FACC on chromosome 9. **Nature Genetics**, 1: (3): 196-198.

Sy, S. Huen, M. Chen, J. (2009a). PALB2 is an Integral Componant of the BRCA Complex Required for Homologous Recombination Repair. **Proceedings of the National Academy of Sciences of the United States of America**, 106: (17): 7155-7160.

Sy, S. Huen,M . Chen, J. (2009b). MRG15 is a novel PALB2-interacting factor involved in homologous recombination. **The Journal of Biological Chemistry.** 284: (32): 21127-21131.

Tasto, J. Carnahan, R. McDonald, W. Gould, K. (2001). Vectors and gene targeting modules for tandem affinity purification in *Schizosaccharomyces pombe*. **Yeast,** 18: (7): 657-662.

Tischkowitz, M. and Xia, B. (2010). PALB2/FANCN – recombining cancer and Fanconi anaemia. **Cancer Research**, 70: (19): 7353 – 7359.

Vaz, F. Hanenberg, H. Schuster, B. Barker, K. Wiek, C. Erven, V. Neveling, K. Endt D. Kesterton, I. Autore, F. Fraternali, F. Freund, M. Hartmann, L. Grimwade, D. Roberts, R. Schaal, H. Mohammed, S. Rahman, N. Schlindler, D. Mathew, C.(2010). Mutation of the *RAD51C* gene in a Fanconi anaemia–like disorder. **Nature Genetics**, 42: (5): 406-409.

Xia, B. Sheng, Q. Nakanishi, K. Ohashi, A. Wu, J. Christ, N. Liu, X. Jasin, M. Couch, F. Livingston, D. (2006a). Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. **Molecular Cell**, 22: (6): 719-729.

Xia, B. Dorsman, J. C. Ameziane, N. de Vries, Y. Rooimans, M. A. Sheng, Q. Pals, G. Errami, A. Gluckman, E. Llera, J. Wang, W. Livingston, D. M. Joenje, H. de Winter, J. P. (2006b). **Fanconi anaemia is associated with a defect in the BRCA2 partner PALB2.** Nature Genetics, 39: (2): 159-161.

Xiao, H. Zhang, K. Xia, B. (2008). Defects of FA/BRCA pathway in lymphoma cell lines. **International Journal of Haematology**, 88: (5): 543-550.

Yao, F. Svensjo, T. Winkler, T. Lu, M. Eriksson, C. Eriksson, E. (1998). Tetracycline Repressor, tetR, rather than the tetR-Mammalian Cell Transcription Factor Fusion Derivatives, Regulates Inducible Gene Expression in Mammalian Cells. **Human Gene Therapy,** 9: (13): 1939 – 1950.

Zhang, F. Ma, J. Wu, J. Ye, L. Cai, H. Xia, B. Yu, X. (2009a). PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. **Current Biology**, 19: (6): 524-529.

Zhang, F. Fan, Q. Ren, K. Andreassen, P. (2009b). PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. **Molecular Cancer Research** [online]. Available from: http://mcr.aacrjournals.org/content/7/7/1110.full.pdf+html [Accessed: 12/04/2014]

Appendix

Appendix 1 – WT electropherogram using primer 2143F
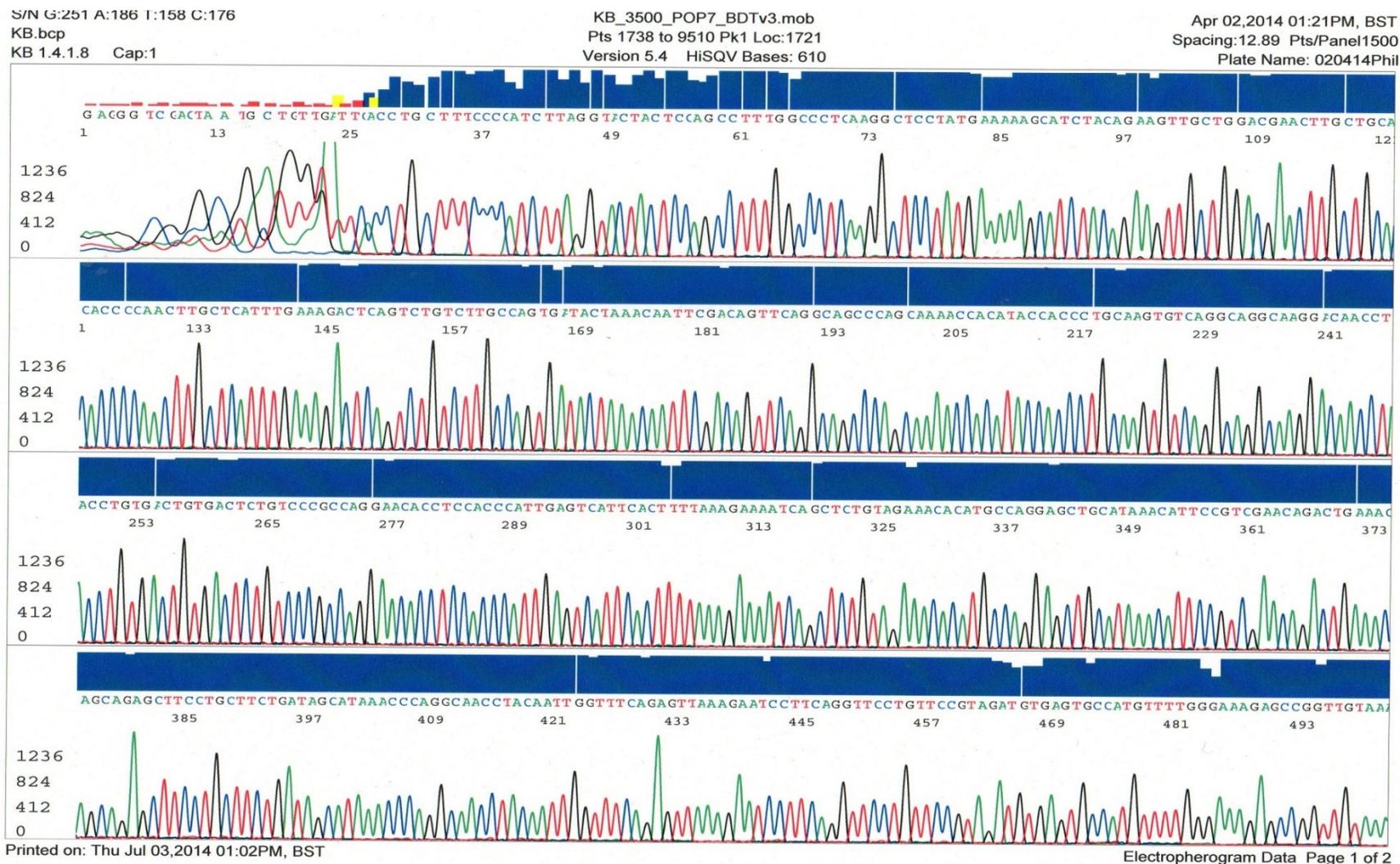
WT 2143F BLAST sequence

Range 1: 2372 to 2986

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand	Frame
1136 bits(615)	0.0()	615/615(100%)	0/615(0%)	Plus/Plus	

Features:

```
Query  1     TTCACCTGCTTTCCCCATCTTAGGTACTACTCCAGCCTTTGGCCCTCAAGGCTCCTATGA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2372  TTCACCTGCTTTCCCCATCTTAGGTACTACTCCAGCCTTTGGCCCTCAAGGCTCCTATGA

Query  61    AAAAGCATCTACAGAAGTTGCTGGACGAACTTGCTGCACACCCCAACTTGCTCATTTGAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2432  AAAAGCATCTACAGAAGTTGCTGGACGAACTTGCTGCACACCCCAACTTGCTCATTTGAA

Query  121   AGACTCAGTCTGTCTTGCCAGTGATACTAAACAATTCGACAGTTCAGGCAGCCCAGCAAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2492  AGACTCAGTCTGTCTTGCCAGTGATACTAAACAATTCGACAGTTCAGGCAGCCCAGCAAA

Query  181   ACCACATACCACCCTGCAAGTGTCAGGCAGGCAAGGACAACCTACCTGTGACTGTGACTC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2552  ACCACATACCACCCTGCAAGTGTCAGGCAGGCAAGGACAACCTACCTGTGACTGTGACTC

Query  241   TGTCCCGCCAGGAACACCTCCACCCATTGAGTCATTCACTTTTAAAGAAAATCAGCTCTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2612  TGTCCCGCCAGGAACACCTCCACCCATTGAGTCATTCACTTTTAAAGAAAATCAGCTCTG

Query  301   TAGAAACACATGCCAGGAGCTGCATAAACATTCCGTCGAACAGACTGAAACAGCAGAGCT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2672  TAGAAACACATGCCAGGAGCTGCATAAACATTCCGTCGAACAGACTGAAACAGCAGAGCT

Query  361   TCCTGCTTCTGATAGCATAAACCCAGGCAACCTACAATTGGTTTCAGAGTTAAAGAATCC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2732  TCCTGCTTCTGATAGCATAAACCCAGGCAACCTACAATTGGTTTCAGAGTTAAAGAATCC

Query  421   TTCAGGTTCCTGTTCCGTAGATGTGAGTGCCATGTTTTGGGAAAGAGCCGGTTGTAAAGA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2792  TTCAGGTTCCTGTTCCGTAGATGTGAGTGCCATGTTTTGGGAAAGAGCCGGTTGTAAAGA

Query  481   GCCATGTATCATAACTGCTTGCGAAGATGTAGTTTCTCTTTGGAAAGCTCTGGATGCTTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2852  GCCATGTATCATAACTGCTTGCGAAGATGTAGTTTCTCTTTGGAAAGCTCTGGATGCTTG

Query  541   GCAGTGGGAAAAACTTTATACCTGGCACTTCGCAGAGGTTCCAGTATTACAGATAGTTCC
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2912  GCAGTGGGAAAAACTTTATACCTGGCACTTCGCAGAGGTTCCAGTATTACAGATAGTTCC
```

Appendix 2 – WT electropherogram using primer 1513F

WT 1513F BLAST sequence

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand	Frame

518 bits(280) 8e-146() 290/294(99%) 4/294(1%) Plus/Plus

```
Features:      Exon 4
Query  15   CACAGG--AAAG-AAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTTTTGCC
71
            ||||||   ||||  |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1736 CACAGGAAAAAGAAAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTTTTGCC
1795


Query  72   AACTTCTAGCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAATATCA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1796 AACTTCTAGCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAATATCA
1855
                                                     Exon 5
Query  132  GCACGAAAAATTATTTATTCAAGTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAGGATTC
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1856 GCACGAAAAATTATTTATTCAAGTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAGGATTC
1915


Query  192  CCTTTCTTGGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCTCCATT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1916 CCTTTCTTGGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCTCCATT
1975


Query  252  TCATAGGGATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTTTCTCAGTATCA   305
            |||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Sbjct  1976 TCATAGGGATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTT-CTCAGTATCA   2028
```

164

Appendix 3 – WT electropherogram using primer 77143F

WT 77143F BLAST sequence

Range 1: 201 to 878

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand	Frame
1223 bits(662)	0.0()	675/680(99%)	5/680(0%)	Plus/Plus	

Features:

```
Query  228  ATGGACGAGCCTCCCGGGAAGCCCCTCAGCTGTGAGGAGAAGGAAAAGTTAAAGGAGAAA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  201  ATGGACGAGCCTCCCGGGAAGCCCCTCAGCTGTGAGGAGAAGGAAAAGTTAAAGGAGAAA

Query  288  TTAGCATTCTTGAAAAGGGAATACAGCAAGACACTAGCCCGCCTTCAGCGTGCCCAAAGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  261  TTAGCATTCTTGAAAAGGGAATACAGCAAGACACTAGCCCGCCTTCAGCGTGCCCAAAGA

Query  348  GCTGAAAAGATTAAGCATTCTATTAAGAAAACAGTAGAAGAACAAGATTGTTTGTCTCAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  321  GCTGAAAAGATTAAGCATTCTATTAAGAAAACAGTAGAAGAACAAGATTGTTTGTCTCAG

Query  408  CAGGATCTCTCACCGCAGCTAAAACACTCAGAACCTAAAAATAAAATATGTGTTTATGAC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  381  CAGGATCTCTCACCGCAGCTAAAACACTCAGAACCTAAAAATAAAATATGTGTTTATGAC

Query  468  AAGTTACACATCAAAACCCATCTTGATGAAGAAACTGGAGAAAAGACATCTATCACACTT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  441  AAGTTACACATCAAAACCCATCTTGATGAAGAAACTGGAGAAAAGACATCTATCACACTT

Query  528  GATGTTGGGCCTGAGTCCTTTAACCCTGGAGATGGCCCAGGAGGATTACCTATACAAAGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  501  GATGTTGGGCCTGAGTCCTTTAACCCTGGAGATGGCCCAGGAGGATTACCTATACAAAGA

Query  588  ACAGATGACACCCAAGAACATTTTCCCCACAGGGTCAGTGACCCTAGTGGTGAGCAAAAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  561  ACAGATGACACCCAAGAACATTTTCCCCACAGGGTCAGTGACCCTAGTGGTGAGCAAAAG

Query  648  CAGAAGCTGCCAAGCAGAAGAAAGAAGCAGCAGAAGAGGACATTTATTTCACAGGAGAGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  621  CAGAAGCTGCCAAGCAGAAGAAAGAAGCAGCAGAAGAGGACATTTATTTCACAGGAGAGA

Query  708  GACTGTGTCTTTGGCACTGATTCACTCAGATTGTCTGGGAAAAGACTAAAGGAACAGGAA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  681  GACTGTGTCTTTGGCACTGATTCACTCAGATTGTCTGGGAAAAGACTAAAGGAACAGGAA

Query  768  GAAATCAGTAGCAAAAATCCTGCTAGATCACCAGTAACTGAAATAAGAACTCACCTTTTA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  741  GAAATCAGTAGCAAAAATCCTGCTAGATCACCAGTAACTGAAATAAGAACTCACCTTTTA
```

Appendix 4 – X4 electropherogram using primer 77143F
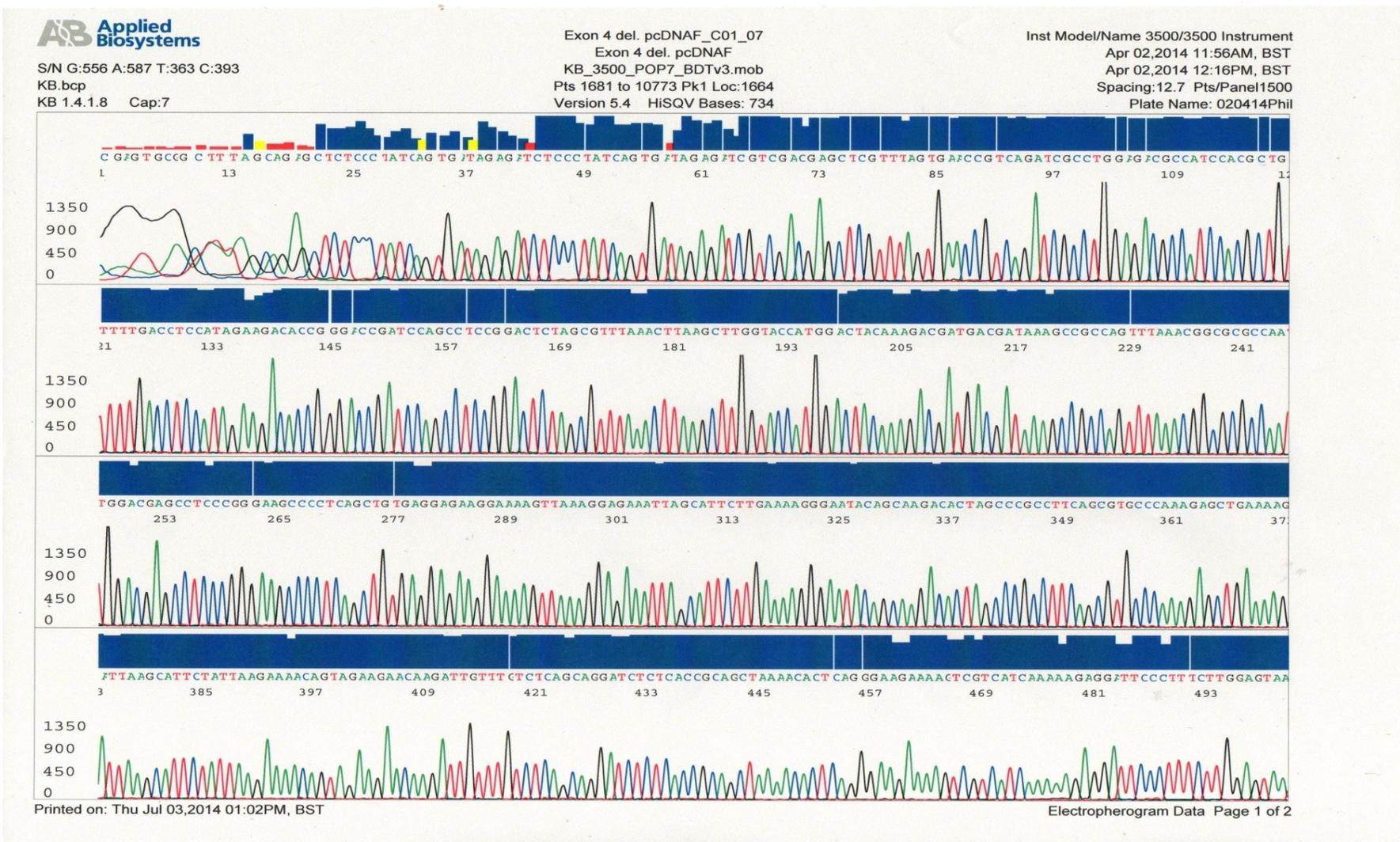
X4 77143F BLAST sequence

201 to 411

Score	**Expect**	**Identities**	**Gaps**	**Strand**	**Frame**

390 bits(211) 5e-107() 211/211(100%) 0/211(0%) Plus/Plus

Features:

```
Query  245  ATGGACGAGCCTCCCGGGAAGCCCCTCAGCTGTGAGGAGAAGGAAAAGTTAAAGGAGAAA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  201  ATGGACGAGCCTCCCGGGAAGCCCCTCAGCTGTGAGGAGAAGGAAAAGTTAAAGGAGAAA

Query  305  TTAGCATTCTTGAAAAGGGAATACAGCAAGACACTAGCCCGCCTTCAGCGTGCCCAAAGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  261  TTAGCATTCTTGAAAAGGGAATACAGCAAGACACTAGCCCGCCTTCAGCGTGCCCAAAGA

Query  365  GCTGAAAAGATTAAGCATTCTATTAAGAAAACAGTAGAAGAACAAGATTGTTTGTCTCAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  321  GCTGAAAAGATTAAGCATTCTATTAAGAAAACAGTAGAAGAACAAGATTGTTTGTCTCAG
```

Exon 3 | Exon 5

```
Query  425  CAGGATCTCTCACCGCAGCTAAAACACTCAG  455
            |||||||||||||||||||||||||||||||
Sbjct  381  CAGGATCTCTCACCGCAGCTAAAACACTCAG  411
```

Score	**Expect**	**Identities**	**Gaps**	**Strand**	**Frame**

568 bits(307) 2e-160() 307/307(100%) 0/307(0%) Plus/Plus

Features: Exon 3 | Exon 5

```
Query  454  AGGGAAGAAAAGTCGTCATCAAAAAGAGGATTCCCTTTCTTGGAGTAATAGTGCTTATTT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1883 AGGGAAGAAAAGTCGTCATCAAAAAGAGGATTCCCTTTCTTGGAGTAATAGTGCTTATTT

Query  514  ATCCTTGGATGATGATGCTTTCACGGCTCCATTTCATAGGGATGGAATGCTGAGTTTAAA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1943 ATCCTTGGATGATGATGCTTTCACGGCTCCATTTCATAGGGATGGAATGCTGAGTTTAAA

Query  574  GCAACTACTGTCTTTTCTCAGTATCACAGACTTTCAGTTACCTGATGAAGACTTTGGACC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2003 GCAACTACTGTCTTTTCTCAGTATCACAGACTTTCAGTTACCTGATGAAGACTTTGGACC

Query  634  TCTTAAGCTTGAAAAAGTGAAGTCCTGCTCAGAAAAACCAGTGGAGCCCTTTGAGTCAAA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2063 TCTTAAGCTTGAAAAAGTGAAGTCCTGCTCAGAAAAACCAGTGGAGCCCTTTGAGTCAAA
```

168

Appendix 5 – del/ins electropherogram using primer 1513F

<u>Del/ins 1513F BLAST Sequence</u>

Alignment statistics for match #1

Score   Expect   Identities   Gaps   Strand Frame

508 bits(275) 5e-143() 282/285(99%) 2/285(0%) Plus/Plus

```
Features:        Exon 4
Query  42    AAAG-AAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTTTTGCCAACTTCTA
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1744  AAAGAAAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTTTTGCCAACTTCTA


Query  101   GCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAATATCAGCACGAAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1804  GCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAATATCAGCACGAAA

                                           Exon 5
Query  161   AATTATTTATTC**G**-GTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAGGATTCCCTTTCTT
             |||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1864  AATTATTTATTC**AA**GTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAGGATTCCCTTTCTT


Query  220   GGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCTCCATTTCATAGGG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1924  GGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCTCCATTTCATAGGG


Query  280   ATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTTCTCAGTATCA    324
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1984  ATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTTCTCAGTATCA    2028
```

170

Appendix 6 – X6 electropherogram using primer 2143F
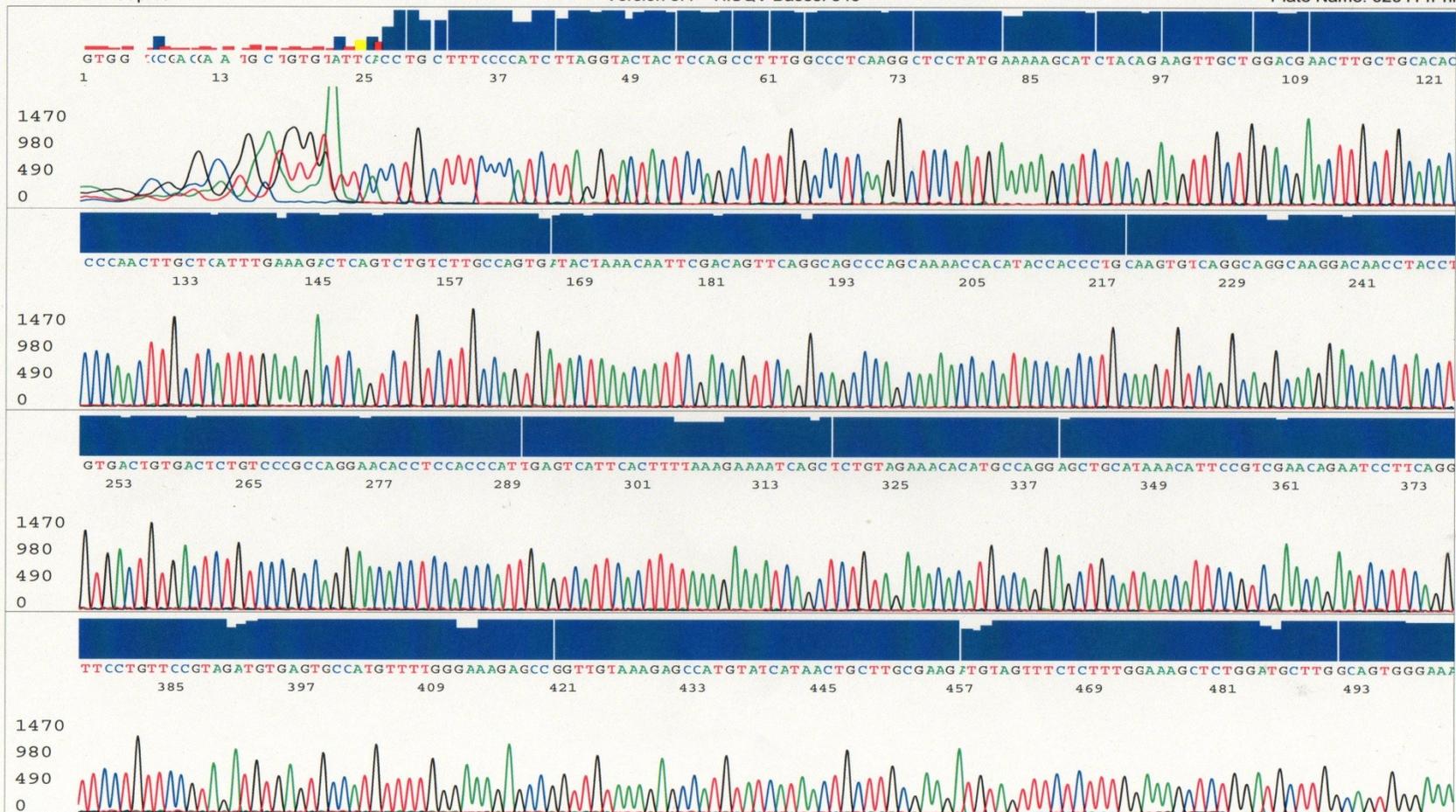
<u>X6 2143F BLAST sequence</u>

Alignment statistics for match #1

Score   Expect   Identities   Gaps   Strand Frame

638 bits(345) 0.0()   347/348(99%) 0/348(0%) Plus/Plus

Features:

```
Query   19    TGTATTCACCTGCTTTCCCCATCTTAGGTACTACTCCAGCCTTTGGCCCTCAAGGCTCCT
              |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2368  TGTGTTCACCTGCTTTCCCCATCTTAGGTACTACTCCAGCCTTTGGCCCTCAAGGCTCCT

Query   79    ATGAAAAAGCATCTACAGAAGTTGCTGGACGAACTTGCTGCACACCCCAACTTGCTCATT
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2428  ATGAAAAAGCATCTACAGAAGTTGCTGGACGAACTTGCTGCACACCCCAACTTGCTCATT

Query   139   TGAAAGACTCAGTCTGTCTTGCCAGTGATACTAAACAATTCGACAGTTCAGGCAGCCCAG
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2488  TGAAAGACTCAGTCTGTCTTGCCAGTGATACTAAACAATTCGACAGTTCAGGCAGCCCAG

Query   199   CAAAACCACATACCACCCTGCAAGTGTCAGGCAGGCAAGGACAACCTACCTGTGACTGTG
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2548  CAAAACCACATACCACCCTGCAAGTGTCAGGCAGGCAAGGACAACCTACCTGTGACTGTG

Query   259   ACTCTGTCCCGCCAGGAACACCTCCACCCATTGAGTCATTCACTTTTAAAGAAAATCAGC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2608  ACTCTGTCCCGCCAGGAACACCTCCACCCATTGAGTCATTCACTTTTAAAGAAAATCAGC

                                                         Exon 5 | Exon 7
Query   319   TCTGTAGAAACACATGCCAGGAGCTGCATAAACATTCCGTCGAACAGA   366
              |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2668  TCTGTAGAAACACATGCCAGGAGCTGCATAAACATTCCGTCGAACAGA   2715
```

Alignment statistics for match #2

Score   Expect   Identities   Gaps   Strand Frame

374 bits(202) 4e-102() 202/202(100%) 0/202(0%) Plus/Plus

Features: Exon 5 | Exon 7

```
Query   364   AGAATCCTTCAGGTTCCTGTTCCGTAGATGTGAGTGCCATGTTTTGGGAAAGAGCCGGTT
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2785  AGAATCCTTCAGGTTCCTGTTCCGTAGATGTGAGTGCCATGTTTTGGGAAAGAGCCGGTT

Query   424   GTAAAGAGCCATGTATCATAACTGCTTGCGAAGATGTAGTTTCTCTTTGGAAAGCTCTGG
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2845  GTAAAGAGCCATGTATCATAACTGCTTGCGAAGATGTAGTTTCTCTTTGGAAAGCTCTGG

Query   484   ATGCTTGGCAGTGGGAAAAACTTTATACCTGGCACTTCGCAGAGGTTCCAGTATTACAGA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2905  ATGCTTGGCAGTGGGAAAAACTTTATACCTGGCACTTCGCAGAGGTTCCAGTATTACAGA

Query   544   TAGTTCCAGTGCCTGATGTGTA   565
              ||||||||||||||||||||||
Sbjct   2965  TAGTTCCAGTGCCTGATGTGTA   2986
```

Appendix 7 – Y551X electropherogram using primer 1513F

Y551X 1513F BLAST Sequence

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand	Frame
520 bits(281)	2e-146()	293/298(98%)	4/298(1%)	Plus/Plus	

```
Features:         Exon 4
Query  12    AGAT-CACAGG--AAAG-AAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTT
             |||| ||||||    ||||  |||||||||||||||||||||||||||||||||||||||||
Sbjct  1731  AGATACACAGGAAAAAGAAAATCAGCCTGCACCCCAGCATCAGATCATTGTGAACCACTT
       1790


Query  68    TTGCCAACTTCTAGCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1791  TTGCCAACTTCTAGCCTGTCGATTGTTAACAGGTCCAAGGAAGAAGTCACCTCACACAAA
                                                              Exon 5
Query  128   TAACAGCACGAAAAATTATTTATTCAAGTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAG
             ||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1851  TATCAGCACGAAAAATTATTTATTCAAGTGAAAGGGAAGAAAAGTCGTCATCAAAAAGAG
       1910


Query  188   GATTCCCTTTCTTGGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCT
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1911  GATTCCCTTTCTTGGAGTAATAGTGCTTATTTATCCTTGGATGATGATGCTTTCACGGCT
       1970


Query  248   CCATTTCATAGGGATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTTCTCAGTATCA
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1971  CCATTTCATAGGGATGGAATGCTGAGTTTAAAGCAACTACTGTCTTTTCTCAGTATCA
       2028
```

174

Appendix 8 - Primer pair 1513F/1803R within the PALB2 sequence.

Sequence also shows in bold where each mutation lies within the amplicon.

The each exon is separated with an underline.

atggacgagcctcccgggaagcccctcagctgtgaggagaaggaaaagttaaaggagaaattagcattcttgaaa
agggaatacagcaagacactagcccgccttcagcgtgcccaaagagctgaaaagattaagcattctattaagaaa
acagtagaagaacaagattgtttgtctcagcaggatctctcaccgcagctaaaacactcagaacctaaaaataaa
atatgtgtttatgacaagttacacatcaaaacccatcttgatgaagaaactggagaaaagacatctatcacactt
gatgttgggcctgagtcctttaaccctggagatggcccaggaggattacctatacaaagaacagatgacacccaa
gaacattttccccacagggtcagtgaccctagtggtgagcaaaagcagaagctgccaagcagaagaaagaagcag
cagaagaggacatttatttcacaggagagagactgtgtctttggcactgattcactcagattgtctgggaaaaga
ctaaaggaacaggaagaaatcagtagcaaaaatcctgctagatcaccagtaactgaaataagaactcacctttta
agtcttaaatctgaacttccagattctccagaaccagttacagaaattaatgaagacagtgtattaattccacca
actgcccaaccagaaaaaggtgttgatacattcctaagaagacctaatttcaccagggcgactacagttccttta
cagactctatcagatagcggtagtagtcagcaccttgaacacattcctcctaaaggtagcagtgaacttactact
cacgacctaaaaaacattagatttacttcacctgtaagtttggaggcacaaggcaaaaaatgactgtctctaca
gataacctccttgtaaataaagctataagtaaaagtggccaactgcccacaagttctaatttagaggcaaatatt
tcatgttctctaaatgaactcacctacaataacttaccagcaaatgaaaaccaaaacttaaaagaacaaatcaa
acagagaaatctttaaaatctcccagtgacactcttgatggcaggaatgaaaatcttcaggaaagtgagattcta
agtcaacctaagagtcttagcctggaagcaacctctcctctttctgcagaaaaacattcttgcacagtgcctgaa
ggccttctgtttcctgcagaatattatgttagaacaacacgaagcatgtccaattgccagaggaaagtagccgtg
gaggctgtcattcagagtcatttggatgtcaagaaaaaagggtttaaaaataaaaataaggatgcaagtaaaaat
ttaaacctttccaatgaggaaactgaccaaagtgaaattaggatgtctggcacatgcacaggacaaccaagttca
agaacctctcagaaacttctctcattaactaaagtcagctctcccgctgggcccactgaagataatga<mark style="background:yellow">cttgtct</mark>
<mark style="background:yellow">aggaaggcagttg</mark>cccaagcacctggtagaagatacacaggaaaaagaaaatcagcctgcaccccagcatcagat
cattgtgaaccactttttgccaacttctagcctgtcgattgttaacaggtccaaggaagaagtcacctcacacaaa
tatcagcacgaaaaattatttattc[**aa/g**]gtgaaagggaagaaaagtcgtcatcaaaaagaggattccctttc
ttggagtaatagtgcttatttatccttggatgatgatgctttcacggctccatttcatag<mark style="background:lime">ggatggaatgctgag</mark>
<mark style="background:lime">tttaaa</mark>gcaactactgtctttctcagtatcacagactttcagttacctgatgaagactttggacctcttaagct
tgaaaaagtgaagtcctgctcagaaaaaccagtggagccctttgagtcaaaaatgtttggagagagacatcttaa
agagggaagctgtattttttccagaggaactgagtcctaaacgcatggatacagaaatggaggacttagaagagga
ccttattgttctaccaggaaaatcacatcccaaaaggccaaactcgcaaagccagcatacaaagacgggcctttc
ttcatccatattactttatactcctttaaatacggttgcgcctgatgataatgacaggcctaccacagacatgtg
ttcacctgcttccccatcttaggtactactccagcctttggccctcaaggctcctatgaaaaagcatctacaga
agttgctggacgaacttgctgcacaccccaacttgctcatttgaaagactcagtctgtcttgccagtgatactaa
acaattcgacagttcaggcagcccagcaaaaccacataccaccctgcaagtgtcaggcaggcaaggacaacctac
ctgtgactgtgactctgtcccgccaggaacacctccacccattgagtcattcactttaaagaaaatcagctctg
tagaaacacatgccaggagctgcataaacattccgtcgaacag**actgaaacagcagagcttcctgcttctgatag**
cataaacccaggcaacctacaattggtttcagagttaaagaatccttcaggttcctgttccgtagatgtgagtgc
catgttttgggaaagagccggttgtaaagagccatgtatcataactgcttgcgaagatgtagtttctctttggaa
agctctggatgcttggcagtgggaaaaactttatacctggcacttcgcagaggttccagtattacagatagttcc
agtgcctgatgtgtataatctcgtgtgtgtagctttgggaaatttggaaatcagagagatcag<u>ggcattgttttg</u>
ttcctctgatgatgaaagtgaaagcaagtactactgaagtctggaaatataaaagctgtgcttggcctgacaaa
gaggaggctagttagtagcagtgggacccctttctgatcaacaagtagaagtcatgacgtttgcagaagatggagg
aggcaaagaaaccaattttgatgcccctgaggagactatactaacttttgctgaggtccaagggatgcaaga
agctctgcttggtactactattatgaacaacattgttatttggaatttaaaaactggtcaactcctgaaaaagat
gcacattgatgattcttaccaagcttcagtctgtcacaaagcctattctgaaatggggcttctctttattgtcct
gagtcatccctgtgccaaagagagtgagtcgttgcgaagccctgtgtttcagctcattgtgattaaccctaagac
gactctcagcgtgggtgtgatgctgtactgtcttcctccagggcaggctggcag<u>gttcctggaaggtgacgtgaa</u>

agatcactgtgcagcagcaatcttgacttctggaacaattgccatttgggacttacttctcggtcagtgtactgc
cctcctcccacctgtctctgaccaacattggtctttgtgaaatggtcgggtacagactctcatttgctggctgg
acaaaaagatggaaatatatttgtataccactattcataagttagggtaaagtgaaaacacaattttctggatat
attgggcctcttagtattttttggagttttaaatataaaggagaatatctgaatgacacttaaaatgattgcttg
tttatgtccagacagacttatttttttattctaatgatggtagcaccactgatcttggatgtacatttatgtatac
tttgagaaaaagggttttaggttgattttgtaatttcccacatttgtacatgtgctttttaaaggtgtacataaa
gcttcaaatggcaataaatatttattttatacattcaaaaaaaaaa

Forward Primer: C T T G T C T A G G A A G G C A G T T G

Reverse Primer: T T T A A A C T C A G C A T T C C A T C C

Total sequence length: 313 bp

Appendix 9 - Primer pair 2413/2763R within the PALB2 sequence.

The each exon is separated with an underline.

```
gcaggccgaatggtggatttaattggccggagtttagggcgcgcttggcccgcgtgggtcagctgatcgcgcact
gagggtgcgatcccgggctccccattccttcctggggcgcctccccggcccagggccaactgggtcccggtgtcg
gcaggcctggggtcggcgacggctgctcttttcgttctgtcgcctgcccgatggacgagcctcccgggaagcccc
tcagctgtgaggagaaggaaaagttaaaggagaaattagcattcttgaaaagggaatacagcaagacactagccc
gccttcagcgtgcccaaagagctgaaaagattaagcattctattaagaaaacagtagaagaacaagattgtttgt
ctcagcaggatctctcaccgcagctaaaacactcagaacctaaaaataaaatatgtgtttatgacaagttacaca
tcaaaacccatcttgatgaagaaactggagaaaagacatctatcacacttgatgttgggcctgagtcctttaacc
ctggagatggcccaggaggattacctatacaaagaacagatgacacccaagaacattttccccacagggtcagtg
accctagtggtgagcaaaagcagaagctgccaagcagaagaaagaagcagcagaagaggacatttatttcacagg
agagagactgtgtctttggcactgattcactcagattgtctgggaaaagactaaaggaacaggaagaaatcagta
gcaaaaatcctgctagatcaccagtaactgaaataagaactcacctttaagtcttaaatctgaacttccagatt
ctccagaaccagttacagaaattaatgaagacagtgtattaattccaccaactgcccaaccagaaaaaggtgttg
atacattcctaagaagacctaatttcaccagggcgactacagttccttacagactctatcagatagcggtagta
gtcagcaccttgaacacattcctcctaaaggtagcagtgaacttactactcacgacctaaaaaacattagattta
cttcacctgtaagtttggaggcacaaggcaaaaaaatgactgtctctacagataacctccttgtaaataaagcta
taagtaaaagtggccaactgcccacaagttctaatttagaggcaaatatttcatgttctctaaatgaactcacct
acaataacttaccagcaaatgaaaaccaaaacttaaaagaacaaaatcaaacagagaaatctttaaaatctccca
gtgacactcttgatggcaggaatgaaaatcttcaggaaagtgagattctaagtcaacctaagagtcttagcctgg
aagcaacctctcctctttctgcagaaaaacattcttgcacagtgcctgaaggccttctgtttcctgcagaatatt
atgttagaacaacacgaagcatgtccaattgccagaggaaagtagccgtggaggctgtcattcagagtcatttgg
atgtcaagaaaaaagggtttaaaaataaaaataaggatgcaagtaaaaatttaaacctttccaatgaggaaactg
accaaagtgaaattaggatgtctggcacatgcacaggacaaccaagttcaagaacctctcagaaacttctctcat
taactaaagtcagctctcccgctgggcccactgaagataatgacttgtctaggaaggcagttgcccaagcacctg
gtagaagatacacaggaaaaagaaaatcagcctgcaccccagcatcagatcattgtgaaccacttttgccaactt
ctagcctgtcgattgttaacaggtccaaggaagaagtcacctcacacaaatatcagcacgaaaaattatttattc
aagtgaaagggaagaaaagtcgtcatcaaaaagaggattccctttcttggagtaatagtgcttatttatccttgg
atgatgatgctttcacggctccatttcatagggatggaatgctgagtttaaagcaactactgtctttttctcagta
tcacagactttcagttacctgatgaagactttggacctcttaagcttgaaaaagtgaagtcctgctcagaaaaac
cagtggagccctttgagtcaaaaatgtttggagagagacatcttaaagagggaagctgtattttttccagaggaac
tgagtcctaaacgcatggatacagaaatggaggacttagaagaggaccttattgttctaccaggaaaatcacatc
ccaaaaggccaaactcgcaaagccagcatacaaagacgggcctttcttcatccatattactttatactc[HL]ctttaa
atacggttgcgcctgatgataatgacaggcctaccacagacatgtgttcacctgctttccccatcttaggtacta
ctccagcctttggccctcaaggctcctatgaaaaagcatctacagaagttgctggacgaacttgctgcacacccc
aacttgctcatttgaaagactcagtctgtcttgccagtgatactaaacaattcgacagttcaggcagcccagcaa
aaccacataccaccctgcaagtgtcaggcaggcaaggacaacctacctgtgactgtgactctgtcccgccaggaa
cacctccacccattgagtcattcactttaaagaaaatcagctctgtagaaacacatgccaggagctgcataaac
attccgtcgaacag[actgaaacagcagagcttcctgcttctgatagcataaacccaggcaacctacaattggtt
tcagagttaaag]aatccttcaggttcctgttccgtagatgtgagtgccatgttttgggaaagagccggttgtaa
agagccatgtatcataactgcttgcgaagatgtagtttctctttggaaagctctggatgcttggcagtgggaaaa
actttatacctggcacttcgcagaggttccagtattaca[HLG]gatagttccagtgcctgatgt[/HLG]gtataatctcgtgtg
tgtagctttgggaaatttggaaatcagagagatcagggcattgttttgttcctctgatgatgaaagtgaaaagca
agtactactgaagtctggaaatataaaagctgtgcttggcctgacaaagaggaggctagttagtagcagtgggac
cctttctgatcaacaagtagaagtcatgacgtttgcagaagatggaggaggcaaagaaaaccaattttttgatgcc
ccctgaggagactatactaacttttgctgaggtccaagggatgcaagaagctctgcttggtactactattatgaa
caacattgttatttggaatttaaaaactggtcaactcctgaaaaagatgcacattgatgattcttaccaagcttc
agtctgtcacaaagcctattctgaaatggggcttctctttattgtcctgagtcatccctgtgccaaagagagtga
```

gtcgttgcgaagccctgtgtgtttcagctcattgtgattaaccctaagacgactctcagcgtgggtgtgatgctgta
ctgtcttcctccagggcaggctggcaggttcctggaaggtgacgtgaaagatcactgtgcagcagcaatcttgac
ttctggaacaattgccatttgggacttacttctcggtcagtgtactgccctcctcccacctgtctctgaccaaca
ttggtctttttgtgaaatggtcgggtacagactctcatttgctggctggacaaaaagatggaaatatatttgtata
ccactattcataagttagggtaaagtgaaaacacaattttctggatatattgggcctcttagtattttttggagt
tttaaatataaaggagaatatctgaatgacacttaaaatgattgcttgtttatgtccagacagacttattttttta
ttctaatgatggtagcaccactgatcttggatgtacatttatgtatactttgagaaaaagggtttttaggttgatt
tttgtaatttcccacatttgtacatgtgcttttaaaggtgtacataaagcttcaaatggcaataaatatttattt
ttatacattcaaaaaaaaaaa

Total sequence length: 664bp

Appendix 10 - Primer pair 77143F/1975R within the PALB2 sequence

The each exon is separated with an underline. The region identified by capital letters is a sequence added by the 77143F primer containing restriction enzymes EcoRI and AscI.

gcaggccgaatggtggatttaattggccggagtttagggcgcgcttggcccgcgtgggtcagctgatcgcgcact
gagggtgcgatcccgggctccccattccttcctggggcgcctccccggcccagggccaactgggtcccggtgtcg
gcaggcctggggtcggcgacggctgctcttttcgttctgtcgcctgcccgGTTGTTGAATTCGGCGCGCCA
atggacgagcctcccgggaagccctcagctgtgaggagaaggaaaagttaaaggagaaattagcattcttgaaa
agggaatacagcaagacactagcccgccttcagcgtgcccaaagagctgaaaagattaagcattctattaagaaa
acagtagaagaacaagattgtttgtctcagcaggatctctcaccgcagctaaaacactcagaacctaaaaataaa
atatgtgtttatgacaagttacacatcaaaacccatcttgatgaagaaactggagaaaagacatctatcacactt
gatgttgggcctgagtcctttaaccctggagatggcccaggaggattacctatacaaagaacagatgacacccaa
gaacattttccccacagggtcagtgaccctagtggtgagcaaaagcagaagctgccaagcagaagaaagaagcag
cagaagaggacatttatttcacaggagagagactgtgtctttggcactgattcactcagattgtctgggaaaaga
ctaaaggaacaggaagaaatcagtagcaaaaatcctgctagatcaccagtaactgaaataagaactcacctttta
agtcttaaatctgaacttccagattctccagaaccagttacagaaattaatgaagacagtgtattaattccacca
actgcccaaccagaaaaaggtgttgatacattcctaagaagacctaatttcaccagggcgactacagttccttta
cagactctatcagatagcggtagtagtcagcaccttgaacacattcctcctaaaggtagcagtgaacttactact
cacgacctaaaaaacattagatttacttcacctgtaagtttggaggcacaaggcaaaaaatgactgtctctaca
gataacctccttgtaaataaagctataagtaaaagtggccaactgcccacaagttctaatttagaggcaaatatt
tcatgttctctaaatgaactcacctacaataacttaccagcaaatgaaaaccaaaacttaaaagaacaaaatcaa
acagagaaatcttttaaaatctcccagtgacactcttgatggcaggaatgaaaatcttcaggaaagtgagattcta
agtcaacctaagagtcttagcctggaagcaacctctcctctttctgcagaaaaacattcttgcacagtgcctgaa
ggccttctgtttcctgcagaatattatgttagaacaacacgaagcatgtccaattgccagaggaaagtagccgtg
gaggctgtcattcagagtcatttggatgtcaagaaaaaagggtttaaaaataaaaataaggatgcaagtaaaaat
ttaaacctttccaatgaggaaactgaccaaagtgaaattaggatgtctggcacatgcacaggacaaccaagttca
agaacctctcagaaacttctctcattaactaaagtcagctctcccgctgggcccactgaagataatgacttgtct
aggaaggcagttgcccaagcacctggtagaagatacacaggaaaaagaaatcagcctgcaccccagcatcagat
cattgtgaaccacttttgccaacttctagcctgtcgattgttaacaggtccaaggaagaagtcacctcacacaaa
tatcagcacgaaaaattatttattcaagtgaaagggaagaaaagtcgtcatcaaaaagaggattcccttttcttgg
agtaatagtgcttatttatccttggatgatgatgctttcacggctccatttcatagggatggaatgctgagtttta
aagcaactactgtcttttctcagtatcacagactttcagttacctgatgaagactttggacctcttaagcttgaa
aaagtgaagtcctgctcagaaaaaccagtggagccctttgagtcaaaaatgtttggagagagacatcttaaagag
ggaagctgtattttttccagaggaactgagtcctaaacgcatggatacagaaatggaggacttagaagaggacctt
attgttctaccaggaaaatcacatcccaaaaggccaaactcgcaaagccagcatacaaagacgggcctttcttca
tccatattactttatactccttaaatacggttgcgcctgatgataatgacaggcctaccacagacatgtgttca
cctgctttccccatcttaggtactactccagcctttggccctcaaggctcctatgaaaaagcatctacagaagtt
gctggacgaacttgctgcacaccccaacttgctcatttgaaagactcagtctgtcttgccagtgatactaaacaa
ttcgacagttcaggcagcccagcaaaaccacataccaccctgcaagtgtcaggcaggcaaggacaacctacctgt
gactgtgactctgtcccgccaggaacacctccacccattgagtcattcacttttaaagaaaatcagctctgtaga
aacacatgccaggagctgcataaacattccgtcgaacagactgaaacagcagagcttcctgcttctgatagcata
aacccaggcaacctacaattggtttcagagttaaagaatccttcaggttcctgttccgtagatgtgagtgccatg
ttttgggaaagagccggttgtaaagagccatgtatcataactgcttgcgaagatgtagtttctctttggaaagct
ctggatgcttggcagtgggaaaaactttatacctggcacttcgcagaggttccagtattacagatagttccagtg
cctgatgtgtataatctcgtgtgtgtagctttgggaaatttggaaatcagagagatcagggcattgttttgttcc
tctgatgatgaaagtgaaaagcaagtactactgaagtctggaaatataaaagctgtgcttggcctgacaaagagg
aggctagttagtagcagtgggacctttctgatcaacaagtagaagtcatgacgtttgcagaagatggaggaggc
aaagaaaaccaattttttgatgcccctgaggagactatactaactttgctgaggtccaagggatgcaagaagct
ctgcttggtactactattatgaacaacattgttatttggaatttaaaaactggtcaactcctgaaaaagatgcac
attgatgattcttaccaagcttcagtctgtcacaaagcctattctgaaatggggcttctctttattgtcctgagt

catccctgtgccaaagagagtgagtcgttgcgaagccctgtgtttcagctcattgtgattaaccctaagacgact
ctcagcgtgggtgtgatgctgtactgtcttcctccagggcaggctggcaggttcctggaaggtgacgtgaaagat
cactgtgcagcagcaatcttgacttctggaacaattgccatttgggacttacttctcggtcagtgtactgccctc
ctcccacctgtctctgaccaacattggtcttttgtgaaatggtcgggtacagactctcatttgctggctggacaa
aaagatggaaatatatttgtataccactattcataagttagggtaaagtgaaaacacaattttctggatatattg
ggcctcttagtattttttggagtttttaaatataaaggagaatatctgaatgacacttaaaatgattgcttgttta
tgtccagacagacttatttttttattctaatgatggtagcaccactgatcttggatgtacatttatgtatactttg
agaaaaagggtttttaggttgatttttgtaatttcccacatttgtacatgtgcttttaaaggtgtacataaagctt
caaatggcaataaatatttattttttatacattcaaaaaaaaaaa

<mark>Forward Primer</mark>:

G T T G T T G A A T T C G G C G C G C C A A T G G A C G A G C
C T C C C G G G A A G C

<mark>Reverse Primer</mark>:G A C T C A G G A T T T G C G T A C C T A T G T

Fragment length: 2019 bp