Masters of Research

Project 1

The Role of Acetylation in the Non-Homologous End Joining Pathway

Diana Walsh

This project is submitted in partial fulfilment of the requirements for the award of the MRes

Table of Contents

1.0 Abstract

DNA double-strand breaks (DSBs) are known to be the most deleterious of lesions and if left unrepaired, may result in apoptosis or gross chromosomal rearrangements. The non-homologous end joining (NHEJ) pathway is the main repair pathway for DSBs, and defects in the pathway have been shown to cause genomic instability and contribute to tumourigenesis. It has been shown that post-translational modifications play a large role in the control and regulation of the NHEJ pathway, and therefore present attractive targets to develop new therapies. Several components of the NHEJ pathway, including XLF, XRCC4 and DNA Ligase IV have been suggested to undergo acetylation in response to DNA damage, and it is thought that this may play a role in the regulation of the NHEJ pathway. Following investigation, it was found that these components may become acetylated *in vitro*, however *in vivo*, the results were more elusive. The results suggest that XLF and XRCC4 are basally acetylated prior to DNA damage, but may become hyper-acetylated in response to a DSB. If acetylation is confirmed to play a regulatory role in the process of NHEJ, it may provide a target which may be exploited in order to develop novel cancer treatments and therapies.

2.0 Introduction

2.1 DNA Damage and the Repair Process

DNA damage and double strand breaks are known to be a frequent occurrence within eukaryotic cells. Double strand breaks (DSBs) are considered to be the most deleterious of lesions, and it has been estimated that 10 DSBs occur in each cell, every day (1). DSBs can occur through exposure to environmental factors such as ionizing radiation and genotoxic chemicals. They may also occur as a result of certain endogenous processes including normal cell metabolism, DNA replication and gene rearrangements in lymphocytes, known as 'VDJ recombination' (2). In order to maintain genomic integrity and cell viability, cells possess two main mechanisms of repairing damaged DNA; homologous recombination (HR) and the non-homologous end joining (NHEJ) pathway (1,3). These two pathways complement each other, and are sometimes interchangeable; however they function optimally under different circumstances (2). Failure to repair DNA can lead to permanent cell cycle arrest, apoptosis or mitotic cell death as a result of the loss of genomic material (1). In addition, defects in the DNA repair process can lead to gross chromosomal rearrangements, loss of genomic integrity, and can ultimately contribute to tumourigenesis (1,2).

Homologous recombination (HR) maintains genomic integrity by promoting the accurate repair of DSBs through the use of a template strand of DNA (3). For this reason, HR is the primary method of repair during the late S and G2 phases of the cell cycle; when sister chromatids are readily available (4).

The NHEJ pathway is the primary route of repair during Go and G1 phases of the cell cycle (3,4). It is referred to as 'non-homologous' as it carries out repair without the use of a

template strand of DNA; for this reason this pathway is error-prone, and sequence information can often be lost when the strands are re-joined (5).

2.2 The Non-Homologous End Joining Pathway

The NHEJ pathway is not only essential for the repair of DSBs, but also for the re-ligation of DNA ends generated during the process of VDJ recombination. As a result of this, individuals who have defects in the NHEJ pathway also suffer immunodeficiency, as well as hypersensitivity to ionizing radiation (4,5). NHEJ provides an efficient mechanism of DSB repair throughout all stages of the cell cycle, but functions optimally during the Go, G1 and early S phases (5). Following a DSB, both DNA termini are bound by a heterodimeric complex known as Ku70/80 (5). Ku70/80 then translocates inward along the DNA, and recruits and activates DNA-dependent protein kinase, catalytic subunits (DNA PKcs) to both of the DNA ends (2,5). This causes the subsequent activation of the serine/threonine protein kinase function of the complex, thereby promoting the alignment of the two DNA ends. The DNA PKcs interact with each other across the DSB in order to create a molecular bridge across the gap, and join the two DNA ends together (7). This then stimulates the autophosphorylation of the DNA PKcs, resulting in an alteration in the conformation and the dynamics of the DNA PKcs (7). This triggers the recruitment of the Artemis protein along with additional end-processing factors, such as Polymerase X (pol X), which work to produce compatible DNA ends for successful ligation (7). To complete the process, the XRCC4-DNA Ligase IV complex and XLF work together in order to ligate the two ends together and repair the lesion (Figure 1).

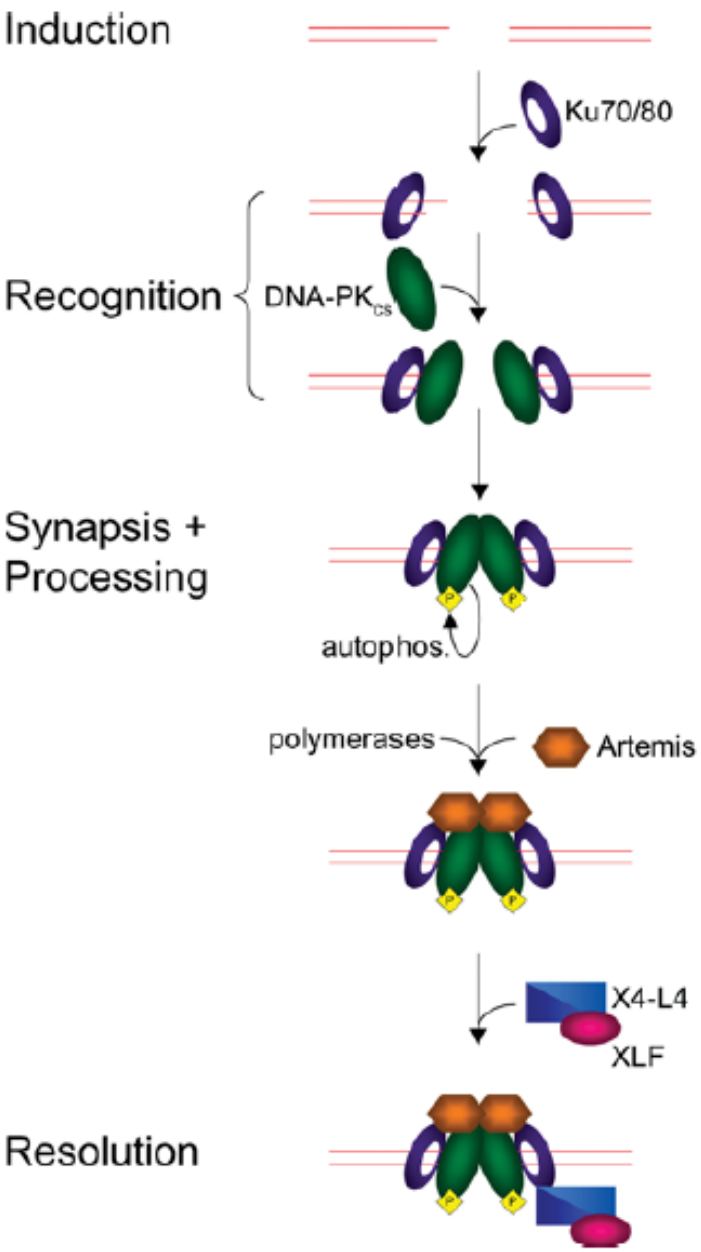


Figure 1. A Diagram to show the Process of NHEJ in Mammals (Modified from AJ Hartlerode et al (2009). The process of NHEJ in mammals occurs following the incidence of a DSB. The Ku 70/80 heterodimer binds to the DNA termini, and translocates inwards, allowing the recruitment of DNA PKcs. Following this, the DNA PKcs are phosphorylated and autophosphorylated, which triggers the recruitment of Artemis, and other polymerases. Finally the XRCC4-DNA Ligase IV complex and XLF co-operate to complete the ligation, and repair the break.

2.3 XRCC4-DNA Ligase IV Complex and NHEJ

DNA Ligase IV and XRCC4 are essential for the successful repair of DSBs through the NHEJ pathway. Following DNA damage, XRCC4 and DNA Ligase IV form a tight complex within the cell, and interact with other proteins involved in the NHEJ pathway in order to ligate the two DNA ends (10). An unusual aspect to the XRCC4-DNA Ligase IV complex is that it is essential to the successful function of NHEJ, however it does not play a role in other DNA repair pathways (10). The fact that the XRCC4- DNA Ligase IV complex is specific to NHEJ suggests that regardless of its intrinsic ability to bind DNA, the XRCC4- DNA Ligase IV complex is not capable of recognizing DSBs; it must be recruited to the sites of damage by additional NHEJ components (10).

XRCC4 is a 36KDa protein known to form homodimers which are capable of interacting with DNA (9,12). Structural studies have shown XRCC4 to be made up of three main domains; a globular N-terminal domain, possibly involved in DNA contacts, a coiled coil arm domain which mediates dimerization and DNA Ligase IV interactions, and finally, a domain extending to the C terminus (figure 2) (9). XRCC4 is said to activate the DNA Ligase IV complex, but importantly, is also essential to ensure DNA ligase IV stability (9). It has been previously noted that XRCC4 is phosphorylated by DNA PKcs in vivo following ionizing radiation (10). It can therefore be suggested that NHEJ could potentially be regulated through various post-translational modifications such as phosphorylation, or acetylation of various components of the pathway including XRCC4.

Human DNA Ligase IV is a 91 kDa protein which is responsible for catalysing the ligation of lesions in the phosphodiester backbone of DNA (10). It is a member of the ATP-dependent DNA ligases, however its roles are specific to NHEJ and V(D)J recombination (9,10). In the absence of DNA Ligase IV, NHEJ does not occur, which suggests that DNA Ligase IV is an

essential component of NHEJ, and that it cannot be substituted for by other members of the ligase family (13). DNA Ligase IV functions to fully ligate DNA ends, even if they are incompatible; this function is stabilised and stimulated by its interactions with XRCC4 and XLF (10). XRCC4 is required for the stability of DNA Ligase IV, and forms a tight association with DNA Ligase IV *in vivo* (9,10).In addition, XRCC4 is not only necessary for the stability of DNA Ligase IV, but has also been shown to stimulate adenylation and the overall activity ofDNA Ligase IV *in vivo* and *in vitro* (9). Inactivation of DNA Ligase IV or XRCC4 causes immunodeficiency and radiosensitivity as a consequence of the resulting defective NHEJ (11,12). Due to the fact that XRCC4 is phosphorylated during NHEJ, it would be interesting to determine whether DNA Ligase IV undergoes any post-translational modifications which have regulatory roles within the NHEJ pathway. If so, these modifications may be targeted in order to potentially develop novel therapies and treatments to treat illnesses such as cancer.

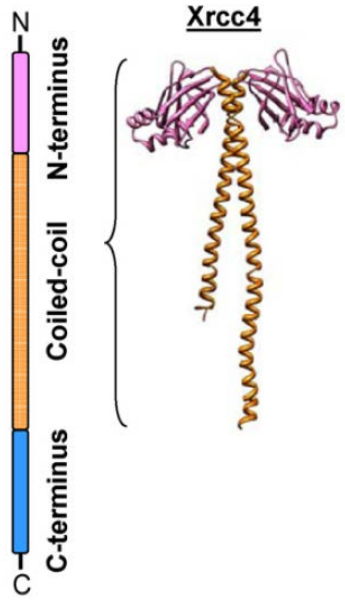
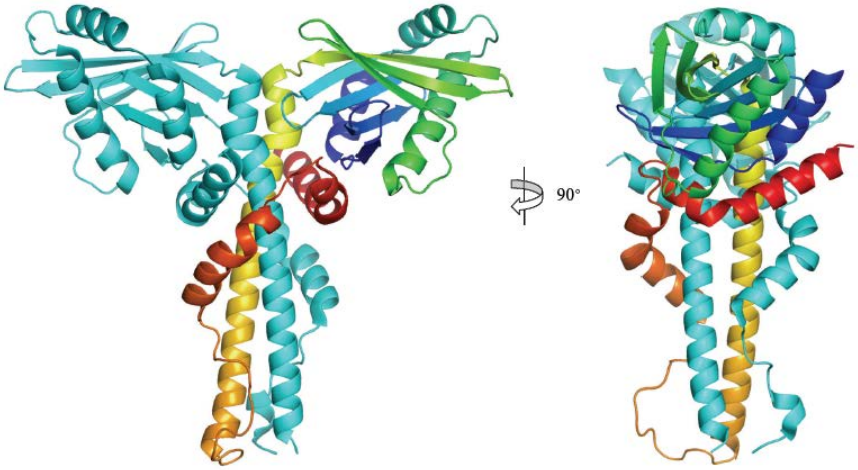


Figure 2: The Atomic Structure of the truncated XRCC4 Dimer, modified from Recuero-Checa et al. (2009). XRCC4 is made up of a globular N terminus, shown in pink, a protruding coiled-coil domain shown in orange, which follows through to the C-terminus (not shown).

2.4 XRCC4-Like Factor (XLF) and NHEJ

XRCC4-Like Factor (XLF) has recently been discovered to interact with XRCC4, the LX complex and has been found to be co-recruited to DSBs along with other NHEJ components (11). XLF has been shown to be an essential factor in the NHEJ pathway, as patients with loss of function mutations in XLF show radiosensitvity and immunodeficiency (10,11).



XLF was shown to share structural similarity to XRCC4; it contains a globular N-terminal domain, and a C-terminal coiled-coil domain (figure 2) (12).

Figure 3.Crystal structure of XLF/Cernunnos.Ribbon diagram of the XLF/Cernunnos dimer.Adapted from Ochi T et al. (2010). One protomer is rainbow colour going from N-terminus (blue) to C-terminus (red).

2.5 Histone Acetyl Transferases and Histone Deacetylases

Histones are the basic packaging unit of DNA, and they readily undergo post-translational modifications in order to regulate the transcription and expression of certain genes (18). These modifications form a certain pattern, known as the 'histone code' which plays a major role in the processes of DNA replication and also in the vital repair pathways which become activated after DNA damage (18, 19). There is increasing evidence to suggest that certain PTMs may contribute significantly to both normal and aberrant cell behaviour and function (18). The PTMs are initiated by histone modifying enzymes known as histone acetyltransferases (HATs) and histone deacetylases (HDACs) (20). HATs act to acetylate certain lysine residues on both histone and non-histone targets through the transfer of an acetyl group from acetyl coenzyme A (19). Acetylation of non-histone targets results in alterations in enzyme activity and protein-protein interactions causing variations in signal transduction (20). Acetylation of histone targets on the other hand, leads to changes in histone-DNA and histone-protein interactions consequentially affecting the accessibility of DNA and the ability of transcription factors to bind the DNA (19). Acetylation is known to

be a dynamic process which plays a major role in the regulation of various processes including DNA replication, and in the DNA damage response pathway (18). The control comes from both the processes of acetylation and deacetylation, and achieving the correct balance between the two modifications (19).

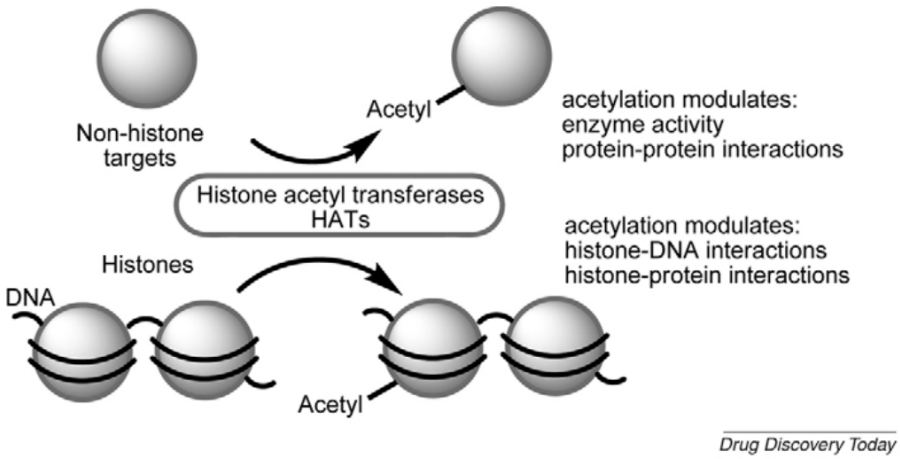


Figure 4. The interaction between HATs and their acetylation targets, from Dekker FJ et al. (2009). Families of HATs are known to acetylate both histone and non-histone targets, resulting in alterations in enzyme activity, protein-protein interactions and the interactions between histones, DNA and proteins. Acetylation in this manner can control and regulate signal transduction, and affect the accessibility of DNA through the acetylation of various targets (16).

HATs are organised into different families based on their structural homology (19). Two of the most extensively studied families are the GNAT family, which includes PCAF and GCN5, and the P300/CBP family which includes P300 and CBP (18, 19). PCAF and GCN5 share 73% homology, while P300 and CBP share 60% homology (18). However PCAF and P300 for example only share minimal sequence and structural homology, suggesting they have different functions and targets (19). Due to the fact that different HATs may have a variety of different targets and functions, the acetylation of XLF, XRCC4 and DNA Ligase IV will be examined in the presence of PCAF, GCN5, P300 and CBP in order to assess which HAT, if any, results in the highest level of acetylation.

2.7 The role of Post Translational Modifications and Acetylation in NHEJ

Post translational modifications (PTMs) are an essential and fundamental means by which protein function and control can be regulated. For any given protein, the chemical properties, activity, localisation as well as its stability may be altered through the action of various PTMs. (19). Of all the PTMs, phosphorylation is the most widely studied and understood. However, methylation, ubiquitination and acetylation, among other PTMs are also known to play a major role in a cellular signalling network which works to regulate various aspects of protein function and activity (19, 20).

Acetylation was originally identified to be important for the modification of histones during transcription (19). For example, it has been shown that following exposure to damaging irradiation, it was observed that several upregulated genes coding for protection to UV, showed increased levels of histone H3 and histone H4 acetylation (21). It is now known that histone acetylation and chromatin modification through acetylation are essential functions during DNA repair; however not a vast amount is known regarding the effects of acetylation events on the function and regulation of DNA repair (21). Histones are acetylated and deacetylated at lysine residue sites in the N-terminal tail region of proteins (22). These modifications are catalysed by Histone Acetyl-transferases (HATs) and Histone deacetyl-transferases (HDACs). Acetylation has also been found to occur in non-histone proteins, and is now known to take place in over 80 transcriptional factors, nuclear regulators and cytoplasmic proteins. In this way, acetylation can be considered to be an essential factor in both the regulation within the nucleus as well as within the cytoplasm (22).

Acetylation at lysine residues occurs within an array of different proteins, and has been discovered to cross-talk with other PTMs in order to create a complex modification system which may enable dynamic regulation of cellular signalling (23). For example, the acetylation

of histones H3 and H4 have been found to crosstalk with the phosphorylation of serines 10 and 28 during signalling events within the cell (23). Due to the fact that phosphorylation is known to occur during DNA damage and NHEJ, and that histone acetylation also plays an essential role during these pathways, it can be questioned as to whether additional acetylation events, possibly linked with phosphorylation may occur during NHEJ.

A previous study performed by Dr. Boris Kysela exposed evidence that several non-histone components of the NHEJ pathway undergo acetylation in response to DNA damage, in particular the XRCC4-DNA Ligase IV complex (Figure 5).

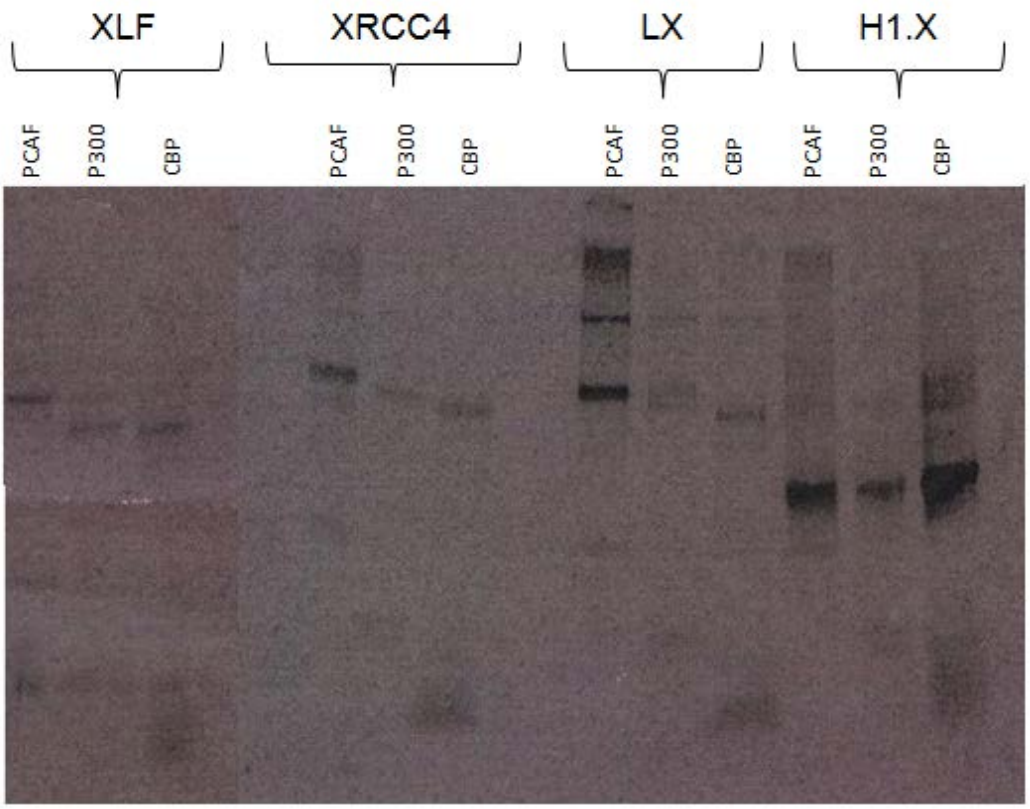


Figure 5: A figure to show the levels of acetylation following DNA damage in several NHEJ proteins by three different histone acetyl transferases (HATs) (Courtesy of Dr. Boris Kysela). XLF, XRCC4, DNA Ligase IV (LX) and histone H1.X have undergone acetylation by histone acetyl transferases, PCAF, P300 and CBP with radioactive C^{14} acetyl coenzyme A as outlined in section 3.3.3. Unfortunately, no size markers were available for this piece of data, however the evidence provided by the exposure suggests that the bands correlate to the proteins, as labelled. Normally, H1.X, the positive control for acetylation would be seen around the 35kDa mark, XLF would be seen at 37kDa, XRCC4 would be seen at 55kDa, and finally DNA Ligase IV would be expected at 96kDa. With these approximations, there are probable bands present for all NHEJ proteins with all HATs, suggesting that they are acetylated. Particularly strong bands can be seen for the DNA Ligase IV complex with PCAF, as well as H1.X with all three HATs.

Therefore, the aim of this project is to determine whether acetylation occurs within the XRCC4-DNA Ligse IV complex, XLF, and H1.X in response to DNA damage, and if so, to map the sites of acetylation in order to identify singular sites of acetylation, as well as clustered sites. If acetylation is found to play an essential role in the NHEJ pathway, new potential targets may be identified and could potentially be exploited in order to create new treatments and therapies to treat cancer. In this way, further information can be gathered regarding the role of acetylation in the NHEJ pathway.

3.0 Methods

3.1 Cells and Cell Culture

3.1.1 MRC5VA Cells

MRC5VA cells were used throughout this research project, and were provided by Dr. Sarah Blair-Reed. MRC5VA cells are embryonic human lung fibroblast cells which have been immortalised by SV40 transformation.

3.1.2 Cell Maintenance

In order to maintain sterility, all procedures involving the cells were carried out under sterile conditions in a class 2 tissue culture hood (HoltenLaminAir). In addition, all solutions and equipment used while handling the cells were obtained sterile, or had been autoclaved prior to use. Cells were split and maintained routinely every 3 days by Dr. Sarah Blair-Reed.

3.1.3 Trypsinisation of Confluent Cells

Upon reaching confluence, the media was withdrawn from the flask with a sterile 10ml pipette. 10ml of phosphate buffered saline (PBS; 2.7mM KCl, 137mM NaCl, 10mM phosphate buffer, pH 7.4) was added to the flask in order to rinse the cells. PBS was then removed carefully with a 10ml pipette, after which approximately 3ml of trypsin-ethylenediaminetetraacetic acid (EDTA; 0.25%:0.02%, 37°C) solution was added to the flask. The flask was then placed in a humidified incubator at 37°C for approximately 3mins or until the cells have become dislodged from the bottom of the flask. The flask was then knocked strongly against the palm of the hand in order to further dislodge the cells. When the majority of cells were dislodged, 7ml of DMEM was added to the flask, and was gently pipetted up and down, with a sterile pipette, and finally removed from the flask and added to a fresh 15ml

centrifugation tube. The tube was then centrifuged for 3 mins at 15,000RPM in order to pellet the cells.

3.1.4 Transfecting Cells with FLAG-Ligase IV

The MRC5VA cells were seeded into 4 flasks, each with approximately $2x10^6$cells. The cells were left to grow in a humidified incubator at 37ºC overnight. The following day, 90µl of serum free Dulbecco's Modified Eagle's Medium (DMEM; containing 1000mg/L glucose) (Sigma-Aldrich) was mixed with 3µl FuGENE HD transfection reagent (Roche), and left in the fume hood for 5 minutes at room temperature. 2µg of FLAG-Ligase IV DNA (see section 3.2) from a successfully transformed bacterial colony was added to the mix, and left for 15 minutes at room temperature under the fume hood. The mixture was then added to the MRC5VA cells, and they were left in an incubator at 37ºC for approximately 48 hours.

3.1.5 Pre-incubation of cells with Sodium Butyrate and Irradiation

Prior to irradiation, 2mM of Sodium butyrate was added to each flask containing the transfected cells. The flasks were then placed in a humidified incubator at 37°C for 6 hours. Cells were then trypsinised**.**

3.1.6 Irradiation of Cells

The cells were then irradiated with 30Gy ionising radiation (Caesium^{137} source). There were 4 different time points for the cells; No irradiation (0), Irradiated and immediately frozen (0IR), Irradiated and frozen after 30 minutes (30), and finally Irradiated and frozen after 1 hour (1hr). These time points were so that the levels of target protein acetylation could be analysed at various time points after DNA damage has occurred. The samples were

centrifuged at 15,000 RPM for 3 minutes after their stated time point, washed in PBS, and centrifuged once more at 15,000 RPM for 3 minutes before storing at -20ºC.

3.2 Preparation of Flag-tagged Ligase IV DNA for Transfection of Cells

3.2.1 Restriction Digest

An insert of Ligase IV DNA in pFASTBAC-HTC vector (Invitrogen) was made available by Dr. Boris Kysela. Restriction digests were carried out in order to cut the Ligase IV gene out of the pFASTBAC vector, so that it may be inserted into pFLAG-CMV vector (Sigma). 5µg of Ligase IV – pFASTBAC DNA was digested in a 1.5ml micro-centrifuge tube with 10 x Multicore Buffer (Promega), 1.0µl EcoRI (Promega, 80u/µl), 1.0µl of KpnI (Promega), and topped up to 25µl with nuclease free water. pFLAG-CMV vector (Sigma) was also digested as described above but with the restriction enzymes EcoRI and KpnI. The two reactions were then incubated at 37ºC for 2 hours.

To ensure digestions were successful, 5x loading dye (50% glycerol, 0.05% bromophenol blue) was added to each 1.5ml micro-centrifuge tube, after which 10µl of each was loaded on to a 1% agarose gel, containing 0.001% ethidium bromide, along with a 1kb ladder (Promega). The gel was run at 150V for 30 minutes. The gel was then visualised under UV light and the appropriate bands were then excised from the gel. The DNA was extracted as outlined below.

3.2.2 Gel Extraction

In order to extract the Ligase IV gene fragment, and the digested pFLAG from the agarose gel, the QIAquick Gel Extraction kit (Qiagen) was used according to the manufacturer's instructions.

3.2.3 Ligation

Ligation reactions were carried out in order to ligate the DNA Ligase IV insert into the pFLAG vector. This was done in a 10µl reaction containing 10x buffer (New England Biolabs), 1µg pFLAG vector, 3µg Ligase IV insert, 1 unit of T4 DNA ligase (New England Biolabs), and topped up with nuclease free water. The reaction was then vortexed briefly, and stored at 4ºC.

3.2.4 Transformation of E.Coli with pFLAG-Lig IV

One microlitre of DNA was incubated on ice with 10µL competent Escherichia coli in a 1.5mL microcentrifuge tube for 30 minutes. They were subsequently subjected to a 1 minute heat shock at 42°C and returned to ice for 2 minutes. To each microcentrifuge tube, 250µl of SOC medium was added and placed in an orbital shaker for 30 minutes at 37°C. Following incubation, 125µL of bacterial solution was spread evenly over an LB-Agar plate containing 100µg/ml ampicillin and incubated at 37°C overnight. Along with pFLAG-DNA Ligase IV DNA, a control pUC19 vector and bacteria only control were set up.

3.2.5 DNA minipreps from bacterial suspensions

In order to extract the DNA from the transformed bacteria, a Miniprep kit from Roche was used according to the manufacturer's instructions.

3.2.6 Identifying Successfully Transformed Bacterial Colonies

In order to determine which bacterial colonies had been transformed successfully, 15 colonies were chosen from the FLAG-DNA Ligase IV plate, and placed into universal tubes containing approximately 3ml Luria-Bertani broth. The universal tubes were then shaken vigorously on a rotary shaker for 16 hours.

Double digests were then carried out on each of the FLAG-Ligase IV samples as outlined in Section 3.21 i, with the use of the NotI and XbaI restriction enzymes. The digests were visualised as outlined in Section 3.21 i, with the successfully transformed bacterial colonies showing two separate bands instead of one.

3.3 Protein Extraction and Acetylation

3.3.1 Protein Extraction

In order to extract the protein from the irradiated cells (section 3.1.6), the cell pellets were resuspended in twice the volume of the cell pellet of phosphoRIPA buffer (50mM Tris/HCl pH 7.5, 1mM EGTA, 1mM EDTA, 50mM sodium fluoride, 5mM sodium pyrophosphate, 1mM sodium orthovanadate, 0.27M sucrose, 1% Triton X-100, 0.1% β-mercaptoethanol, and protease inhibitor cocktailand 2mM Sodium Butyrate). The resuspensions were left on ice for 10 minutes, and were then centrifuged at 17,000RPM for 30 minutes at 4°C. The protein-containing supernatant was then transferred to a fresh 1.5ml microcentrifuge tube. Following this, the concentrations of the proteins were measured, by adding 500µl of water to 500µl of Bradford Reagent (Sigma), and 5µl of protein in a cuvette. The absorbance was recorded at 595nm, using a Genova spectrophotometer against a bovine serum albumin standard. The protein samples were then aliquoted into 15µg samples for use in western gels, and 1mg

samples for use in immunoprecipitation experiments. The samples were then stored at -80ºC for future use.

3.3.2Western Blotting

3.3.2.1 SDS-PAGE

15µg samples of XLF, XRCC4, Ligase IV, and H1-X were denatured by incubating at 100°C for 5 minutes in SDS loading buffer (65mM Tris/HCl pH8, 10% (v/v) glycerol, 2.3% (w/v) SDS, 0.01 bromophenol blue, 1% DTT). The proteins were then separated by electrophoresis on 9% polyacrylamide gels prepared using ProtoGel reagents. This involved preparing a resolving gel segment, made from 30% acrylamide Protogel (National Diagnostics), 0.1% Sodium Dodecyl Sulphate (SDS), 10% AmoniumPersulphate (0.1g/ml), Tetramethylethylenediamine (TEMED) and nuclease free water. The gel was poured and topped up with Isopropanol and left to set. The Isopropanol was washed away and replaced with a stacking gel, which was made up of 3% polyacrylamide, 50mM Tris, 0.1% SDS, 10% APS, and TEMED; 0.7mm combs were inserted immediately after pouring, and the gel was allowed to set. The 15µg protein samples were then loaded onto the gels along with a protein ladder, and run in an electrophoresis tank (BIORAD) containing 10x Running Buffer (10x Tris-Glycine-SDS PAGE Buffer, National Diagnostics) at 150v for 1 hour.

3.3.2.2 Transfer from SDS-PAGE to PDVF Membrane

The resolved proteins from the gel were transferred to a piece of polyvinylidene fluoride (PDVF) membrane (Millipore). Prior to transfer, the PDVF membrane was soaked in 100% methanol, and then immersed in CAPS buffer (10mM CAPS, pH11, 10% Methanol) along with gel and 4 sheets of filter paper in order to perform a wet transfer, using apparatus from

Amersham at 0.23mA for 90 minutes at 4°C. The membranes were then blocked in 5% milk in PBS and 0.1% Tween overnight at 4°C.

3.3.2.3 Western Blotting

Upon completion of transfer to the PDVF membrane, the membrane was inserted into a 50ml centrifuge tube which contained 5ml of 5% Milk in PBS/Tween and the appropriate primary antibody (See Table 1). It was ensured that the side of the membrane which contained bound proteins was facing towards the inside of the tube. The tube was then placed on to a tube roller, and was left at room temperature for 1 hour. Following this, the membrane was then washed 3 x 10 mins in PBS-Tween, and transferred to fresh tube containing 1% milk in PBS-Tween solution with the secondary antibody (see table). The membrane was left rolling at room temperature for 1 hour. The membrane was then washed again in PBS-Tween for 3 x 10 mins, after which 1ml of ECLPlus reagent (Amersham) was applied. The membrane was wrapped in Saran wrap, and secured into a developing cassette. The membrane was then exposed and developed on to X-Ray photographic imaging film (Kodak). The films were developed using an X-Ray film developer (Konica Minolta).

Primary Antibody	Species	Dilution	Secondary Antibody	Dilution
Anti-Panacetyl Lysine	Rabbit	1:1000	Goat anti Rabbit-HRP (DAKO)	1:5000
Anti-H1.X (Abcam)	Rabbit	1:1000	Goat anti Rabbit-HRP (DAKO)	1:5000
Anti-XLF	Rabbit	1:1000	Goat anti Rabbit-HRP (DAKO)	1:5000
Anti-XRCC4	Rabbit	1:1000	Goat anti Rabbit-HRP (DAKO)	1:5000
Anti-HDAC3	Mouse	1:1000	Rabbit anti Mouse-HRP (DAKO)	1:5000
Anti-DNA Ligase IV	Rabbit	1:1000	Goat anti Rabbit-HRP (DAKO)	1:5000
Anti-FLAG (Abcam)	Mouse	1:1000	Rabbit anti Mouse-HRP (DAKO)	1:5000
Anti-GFP (Abcam)	Goat	1:1000	Rabbit anti Goat-HRP (DAKO)	1:5000
Anti-α tublin (Abcam)	Rabbit	1:10,000	Goat anti Rabbit –HRP (DAKO)	1:10,000

3.3.3 *In Vitro* Acetylation of XLF, XRCC4, LX, and H1-X Proteins

Protein samples of XLF, Ligase IV- XRCC4 complex (LX) and H1.X were provided by Dr. Darren Arbon. A master mix of 2 x reaction buffer was made up, containing 100mMHepes, pH8(Sigma), 20% glycerol (Sigma), 2mM Dithiothreitol (DTT) (Sigma), 2mM4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) (Sigma), 10mM Sodium Butyrate (Sigma) and nuclease free water. 16 reactions were set up in total, which all

contained 1µg of protein, 1µg of a particular histone acetyl transferase (HAT), in addition to 1µl of acetyl coenzyme A, 15µl of 2 x reaction buffer, and 12µl of nuclease free water in order to create a 30µl reaction. The combinations of proteins and HATs were as follows;

HAT	**Protein**
PCAF	XLF
PCAF	LX
PCAF	H1.X
P300	XLF
P300	LX
P300	H1.X
CBP	XLF
CBP	LX
CBP	H1.X
GCN5	XLF
GCN5	LX
GCN5	H1.X

The reactions were set up in duplicate; half of the reactions contained C^{14}-Acetyl coenzyme A, while the other half of the reactions contained non-radioactive Acetyl Coenzyme A. The reactions were incubated at 30°C for 1 hour. The samples were then run on a 4-20% gradient Mini-Protean precast gel (BioRad) at 150V for 1 hour. The gels were stained overnight with Page Blue (Fermentas). Bands which corresponded to XLF, XRCC4, Ligase IV and H1.X were excised from the non-radioactive gel, and sent for analysis by mass spectrometry (Advanced Mass Spectrometry Facility, University of Birmingham). The radioactive gel was placed between a piece of filter paper andSaran wrap after which it was vacuum dried at 80°C for 60 minutes on a Gel Master Gel Dryer Vacuum System (Welch).The dried gel was then

placed face down onto a storage phosphor screen (Molecular Dynamics), and left to expose for 1 week. The results were visualised using a STORM Phosphoimager (Molecular Dynamics).

3.3.4*In Vivo* Acetylation

In Vivo acetylations of XLF, XRCC4, Ligase IV, and H1.X were examined in order to determine whether these proteins were acetylated following exposure to ionizing radiation. Cells containing FLAG-tagged XLF and XRCC4 and GFP-tagged H1.X were provided by Dr. Darren Arbon. Cells were transfected with Ligase IV as outlined in section 3.1.4.

3.4 Co-Immunoprecipitation

3.4.1 Immunoprecipitation for GFP-tagged proteins

In order to carry out the immunoprecipitation, 20µl of protein G slurry was washed twice with phosphoripa buffer (excluding protease inhibitor cocktail) (see section 3.3.1) 1mg of protein was added to the washed beads, and left for 1 hour with end-over-end rotation at 4°C. The beads were then centrifuged at 3000RPM for 2 minutes, after which the supernatant was collected and placed into a fresh 1.5ml microcentrifuge tube. At this stage, GFP primary antibody was then added to the protein lysates containing GFP-tagged proteins, and they were then subject to end-over-end rotation at 4°C overnight. The following day, 10µl of beads were washed twice in phosphoripa buffer (excluding protease inhibitor cocktail), after which the lysate-antibody mixture was added, and mixed by end-over-end rotation at 4°C for 3 hours. The beads were then pelleted by centrifuging at 3000 RPM for 2 minutes, and the supernatant was discarded. 250µl of phosphoripa buffer (excluding protease inhibitor cocktail) was used to resuspend the beads, and

the mix was the transferred to a spinX centrifuge tube filter (Costar, 0.45µm pore size). The spin column was then centrifuged at 7000RPM for 1 minute, and the supernatant was discarded; this step was repeated twice more with additional amounts of 250µl phosphoripa buffer for each wash. In order to elute the proteins, 35µl of 1x sample buffer (Biorad) was added to the samples, and heated to 70°C for 5 minutes, after which the samples were centrifuged for 7000RPM for 1 minute. The tubes were then stored at -20°C; the results were visualised by running the samples on SDS-PAGE gels, as outlined in section 3.3.

3.4.2 Immunoprecipitation for FLAG-tagged proteins

Prior to use, anti-FLAG agarose beads (Sigma) were rolled at room temperature and vortexed gently in order to dislodge the beads. For each immunoprecipitation, 30µl of anti-FLAG agarose beads were added to a 1.5ml micro centrifuge tube, and centrifuged at 13,000RPM for 1 minute, after which the supernatant was carefully removed and discarded. 1ml of Tris was added to each tube, and centrifuged once more at 13,000RPM for 1 minute, and the supernatant was discarded. This was repeated 3 times in order to thoroughly wash the beads. 1mg of the FLAG-tagged proteins was added to the labelled micro centrifuge tubes containing the beads, and made up to 1ml with phosphoripa buffer (see section 3.2.1, XI). The tubes were then mixed by end-over-end rotation at 4°C for 2 hours. The contents of the tubes were then transferred to a spinX centrifuge filter (Costar, 0.45µm pore size), and centrifuged at 7000RPM for 1 minute. The flow-through was retained in labelled 1.5ml micro centrifuge tubes, and stored at -20°C. Approximately 1ml of Tris buffered saline (1 M Tris pH 7.5, 5 M NaCl) was added to the spinX filter and was then centrifuged at 7000RPM for 1 minute. The supernatant was discarded, and the process repeated 3 times. The spinX filter was then transferred to a fresh 1.5ml micro centrifuge tube, after which 1 x loading dye was added. The tube was then incubated at 100°C for 5 mins, and then centrifuged at 7000RPM

or 1 minute, and finally stored at -20°C. To visualise the results, the contents of the tubes were then run on western gels (as outlined in section 3.3.2).

4.0 Results

4.1 *In Vitro* Acetylation of NHEJ Components

In order to examine acetylation *in vitro*, protein samples of XLF, XRCC4, DNA Ligase IV and H1.X were individually combined with various HATs, and radioactive C^{14} acetyl co enzyme A. Linker histone H1.X was used as a positive control for acetylation. The HATs tested included PCAF, P300, CBP and GCN5 (see section 2.5). The proteins were also tested for the basal levels of acetylation which may occur in the absence of a HAT. No acetylation was observed for XLF or XRCC4 with any of the HATs tested (data not presented). In addition, no acetylation was observed when DNA Ligase IV (labelled LX complex) was combined with P300, CBP or GCN5 *in vitro*, however low levels of acetylation were observed for DNA Ligase IV when combined with PCAF (Figure 6). These low levels of acetylation could only be observed with 3µg and 4µg of radioactive acetyl co A, suggesting that higher levels of acetyl co A are required to acetylate DNA Ligase IV. Strong bands were observed when H1.X was acetylated using PCAF, which should be expected. Increasing the levels of acetyl co A did not have any significant effects on the levels of acetylation for H1.X, showing that acetylation of H1.X readily occurs at lower concentrations of acetyl CoA. No acetylation was observed when H1.X was combined with acetyl co A alone, confirming that PCAF was responsible for the acetylation observed (Figure 6).

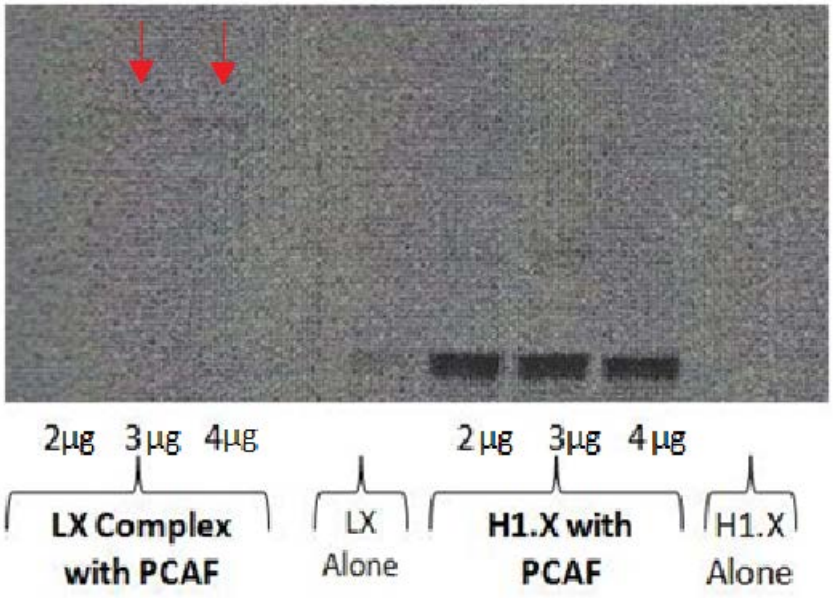


Figure 6.*In Vitro* Acetylation of LX complex and H1.X with PCAF.The LX complex and H1.X were acetylated with varying concentrations of radioactive C^{14}Acetyl Coenzyme A; H1.X was used as a positive control for acetylation. 2µg, 3µg and 4µg refers to the amount of radioactive acetyl coenzyme A used. Unfortunately, no size markers could be obtained for this data. With regard to LX, very slight bands can be seen at 3µg and 4µg of acetyl coA, marked by the two red arrows; this suggests that there is some acetylation occurring. There is a band in the lane for LX alone, however this is suspected to be due to slight spill over from the H1.X-PCAF lane, as the band is not the correct size to be LX. H1.X and the LX complex do not show any bands in the absence of HATs, and so it can be concluded that any acetylation occurring is due to the presence of PCAF. As can be seen, there are three very strong bands which shows H1.X is strongly acetylated in the presence of PCAF at varying concentrations of acetyl co.A.

4.1.1 Mass Spectrometry Results

Due to the fact that no acetylation was observed for XRCC4 and XLF with the use of radioactive acetyl co-A, it was thought that perhaps the reason was because the assay was not sensitive enough, or needed to be optimised. Therefore in order to ensure accurate results, a more sensitive approach was undertaken using mass spectrometry to detect the presence of acetylation. By examining acetylation through more than one technique, a greater degree of confidence can be applied when making conclusions. XLF, XRCC4, LX and H1.X were incubated with non-radioactive acetyl coA and analysed through mass spectrometry in order to determine specific levels of acetylation, and to map the particular sites of acetylation (Tables 1, 2, 3 and 4).

4.1.1a Mass Spectrometry Results for XLF

Lysine	PCAF	P300	CBP	GCN5	XLF alone
63		✓(low)			
160	✓(high)		✓(low)		
197	✓(low)	✓(low)	✓(low)		✓(low)
290		✓(low)			

Table 1: A table to show the sites of acetylation in XLF following incubation with acetyl co A and various HATs. The red ticks indicate that acetylation was present at the various lysine residues and with the HATs listed. XLF alone lists the sites where XLF is acetylated basally, in the absence of HATs. High and low indicate the level of acetylation. As can be seen, the only apparent high acetylation was found at lysine 160 with PCAF..

4.1.1b Mass Spectrometry Results for XRCC4

Lysine	PCAF	P300	CBP	GCN5	XRCC4 Alone
102		✓(low)		✓(low)	
169	✓(low)	✓(high)			
187	✓(high)	✓(med)		✓(high)	✓(high)
188	✓(high)	✓(med)	✓(low)	✓(high)	✓(high)
197	✓(low)	✓(high)	✓(high)		
309	✓(low)	✓(high)	✓(high)		

Table 2: A Table to show the sites of acetylation in XRCC4, following incubation with acetyl co A. As with Table 1, the red ticks indicate sites of acetylation. Lysines 187 and 188 show high levels of acetylation with PCAF, and GCN5, however these sites are also acetylated in XRCC4 alone, and therefore acetylation is most likely not due to the presence of the HATs. XRCC4 shows high levels of acetylation at lysine 197 and 309 in the presence of P300 and CBP.

4.1.1c Mass Spectrometry Results for Ligase IV

Lysine	PCAF	P300	CBP	GCN5	Ligase IVAlone
613	✓(high)		✓(low)	✓(med)	✓(low)

Table 3: A table to show the sites of acetylation in Ligase IV, following incubation with acetyl co A. Ligase IV was found to have only one acetylation site at lysine 613; it is highly acetylated by PCAF at this site, and moderately by GCN5. Low basal levels of acetylation were also found in ligase IV alone at this site.

4.1.1d Mass Spectrometry Results for H1.X

Lysine	PCAF	P300	CBP	H1.X Alone
20			✓(med)	
23		✓(high)	✓(high)	✓(med)
34	✓(med)	✓(high)	✓(high)	✓(low)
35	✓(med)		✓(high)	
47		✓(high)		✓(high)
69		✓(high)		
75		✓(low)		
76	✓(low)	✓(low)		
90		✓(low)	✓(low)	✓(low)
94	✓(high)	✓(med)		✓(low)
115		✓(med)	✓(high)	✓(high)
119		✓(low)	✓(low)	
120		✓(low)	✓(low)	
143	✓(high)	✓(low)	✓(high)	✓(med)
145	✓(high)	✓(low)	✓(high)	✓(low)
146		✓(low)	✓(med)	
147		✓(low)		
159		✓(high)	✓(low)	
160		✓(high)	✓(low)	
195		✓(med)		
196		✓(low)		
198		✓(low)		
199			✓(low)	
202	✓(low)			✓(low)
207		✓(low)	✓(low)	
210		✓(low)		

Table 4: A table to show the sites of acetylation in H1.X, following incubation with acetyl co A. H1.X is acetylated at many various lysine sites both alone, and in the presence of a variety of HATs. It is highly acetylated by CBP at lysine 35, and is highly acetylated at lysines 69, 159 and 160 by P300. All other acetylation sites are either moderate or low levels of acetylation, or have basal levels of acetylation shown by H1X alone.

4.2 *In Vivo* Acetylation of NHEJ components

4.21 *In Vivo* Acetylation of XLF

The *in vitro* acetylation had been tested thoroughly; however *in vitro* experiments occur under a controlled environment, and sometimes do not accurately reflect what occurs *in vivo*. *In vivo* there are many more variables which cannot be artificially created; for example

protein-protein interactions and modifications when in the presence of various cellular content may play an essential role which would be neglected *in vitro*.For these reasons, it was necessary to examine the acetylation of XLF within the cell, *in vivo*. FLAG-tagged XLF was over-expressed in the 5VA-MRC cell line. The cells were exposed to 30Gy of radiation, and collected at various time points. In order to visualise the presence of acetylation, western blots were performed, and blotted with the Panacetyl Lysine (PanacK) antibody; an antibody which is specific for any type of acetylation in the sample. The FLAG-XLF showed faint bands when blotted with PanacK (figure 7a), however when the FLAG-XLF was immunoprecipitated, no reactivity could be observed with the panacK antibody (figure 7b).It was suggested that this may have been due to the fact that there was too much unspecific binding, and therefore, the IP was then repeated with a high salt wash to remove any unwanted and unspecific interactions. The necessary controls were performed to rule out false positives and false negatives; the FLAG-tagged XLF was blotted with XLF antibody to confirm presence of over-expressed XLF (figure 7a), and blotted with α-FLAG to confirm that the XLF was indeed FLAG-tagged (figure 7a). Finally, the western was blotted with α-tubulin, to ensure all samples tested were of equal loading size (figure 7a).

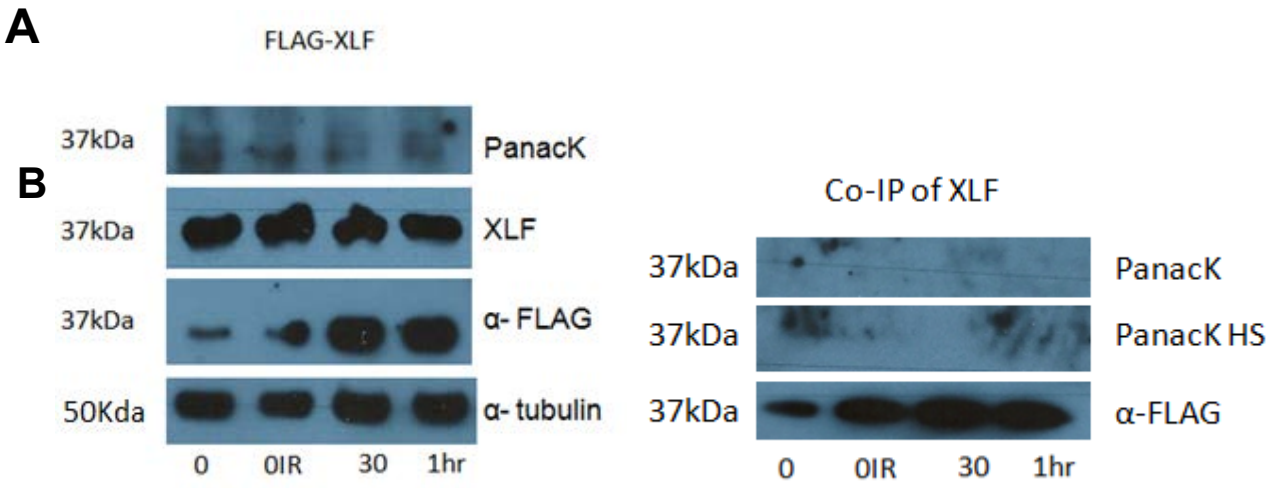


Figure 7.*In Vivo* Acetylation of XLF and Co-Immunoprecipitation of XLF: **A**: When blotted with panacK, FLAG-tagged XLF showed some faint bands, suggesting low levels of acetylation were present. The FLAG-tagged XLF was then blotted with the XLF antibody which confirmed the presence of overexpressed XLF with 4 very clear bands. The FLAG-tagged XLF was shown to be FLAG-tagged by blotting with α-FLAG antibody, confirmed by 4 distinct bands. Finally the 4 bands of equal size in the α-tubulin lane show that the FLAG-tagged XLF was loaded equally through these experiments. **B:**Following co-immunoprecipitation of the FLAG-tagged XLF, western blots with the PanacK antibody did not show any bands. The IP was repeated with a high salt (HS) wash, however upon blotting with the PanacK antibody, no bands were seen. The same blot was blotted with α-FLAG antibody to ensure that the IP had been successful, and as can be seen from the row labelled α-FLAG, bands can be seen for all of the time points of XLF.

4.2.1 *In Vivo* Acetylation of XRCC4

In order to examine the acetylation of XRCC4 *in vivo*, FLAG-tagged XRCC4 was over-expressed in the 5VA-MRC cell line. The cells were exposed to 30Gy of radiation, and collected at various time points, as outlined in section 3.1.5 for the *in vivo* acetylation of XLF. The same westerns were performed, using the PanacK antibody to blot for acetylation. XRCC4 is known to run around the 55kDa mark, so all bands shown are around this size.

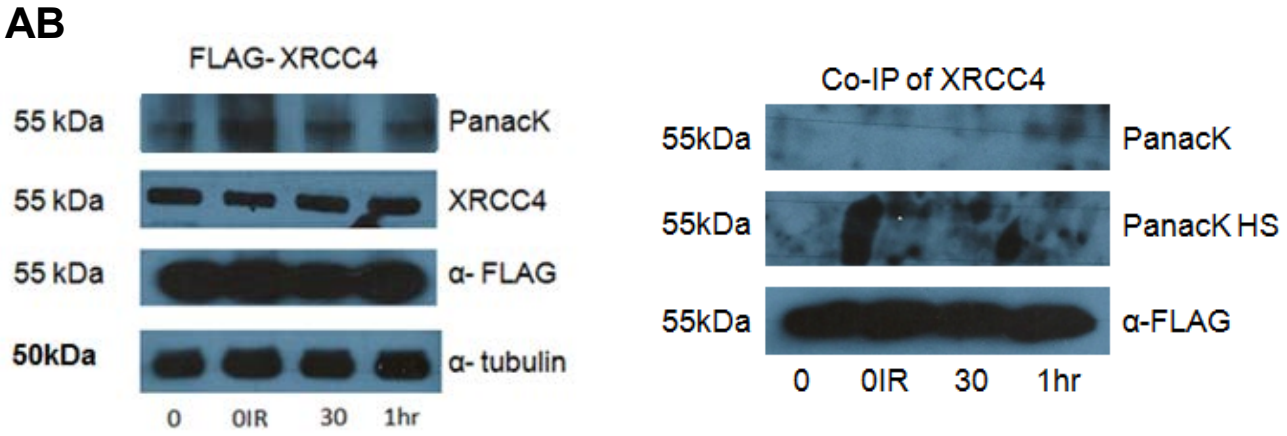


Figure 8: ***In Vivo*** **Acetylation of XRCC4 and Co-immunoprecipitation of XRCC4**When blotted with panacK, strong bands were seen around the 55kDa mark, suggesting that XRCC4 was acetylated. The FLAG-tagged XRCC4 was then blotted with the XRCC4 antibody to show the presence of overexpressed XRCC4, which showed 4 very clear bands, as expected. The FLAG-tagged XRCC4 was also blotted with α-FLAG antibody to confirm that the XRCC4 was FLAG-tagged, and this was clearly shown by the 4 large bands. Finally the FLAG-tagged XRCC4 was blotted with α-tubulin antibody to show equal levels of loading, and this was confirmed by the 4 distinct bands of approximately equal sizes. **(B)**Following Co-IP of the FLAG-tagged XRCC4, western blots with the PanacK antibody did not show any bands for the first three time points, however at the 1hr time point seemed to show a band. The IP was repeated with a high salt (HS) wash; however upon blotting with the PanacK antibody, no bands could be seen. The same blot was blotted with α-FLAG antibody to ensure that the IP had been successful, and as can be seen from the row labelled α-FLAG, the bands are all clearly FLAG-tagged.

4.2.2*In Vivo*Acetylation of Ligase IV

In order to investigate acetylation in Ligase IV, an experiment was carried out in order to express FLAG-tagged Ligase IV in the 5VA-MRC cell line. However, the transfections proved to be unsuccessful, and due to time constraints, the establishment of DNA Ligase IV stably overexpressing cell lines could not be completed.

4.3.3 *In Vivo* Acetylation of H1.X

Acetylation of H1.X was explored by introducing GFP-tagged H1.X into the 5VAMRC cell line, and exposing them to 30Gy of radiation. The cells were then collected at the same time points, as outlined in section 3.1.5. Levels of acetylation were determined before and after co-immunoprecipitation. H1.X is known to run around the 35kDa mark.

A

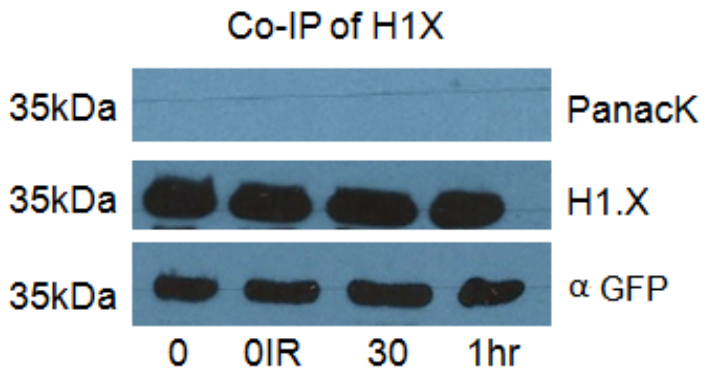


Figure 9: Co-immunoprecipitation of H1.X Following co-IP of the GFP tagged H1.X, western blots with the PanacK antibody did not show any bands. This was a particularly odd result, as H1.X is normally used as a positive control for acetylation. The cells successfully overexpressed H1.X, as 4 clear bands were observed when blotted with H1.X antibody. In addition, it was confirmed that H1.X was GFP-tagged, as 4 distinct bands were seen when blotted with GFP antibody.

5.0 Discussion

5.1 *In Vitro* Acetylation

In order to investigate the levels of acetylation in XLF, XRCC4, DNA Ligase IV and H1.X in response to DNA damage, *in vitro* assays were carried out. It was important to establish the degree of acetylation for these components of the NHEJ pathway primarily *in vitro* due to the fact that the complexities of interactions which occur within a cellular environment involve numerous variants which may affect acetylation. For this reason, if only the *in vivo* effects of acetylation were investigated, it would be difficult to conclude with confidence that the acetylation was occurring as a direct result of the various HATs tested. Therefore, in order to simplify the reaction, and reduce the number of external variables within the experiment, the levels of acetylation were investigated through the use of radioactive acetyl coA and samples of XLF, XRCC4, DNA Ligase IV and H1.X in protein form.

Upon completion of the *in vitro* assays with, it was found that no acetylation could be observed for XLF or XRCC4 alone or in the presence of any HAT tested. It was thought that perhaps the proteins had not been given a sufficient incubation period in the presence of acetyl co-A and the HATs, however it seemed that additional time had no effect on the levels of acetylation observed. These negative results do not necessarily mean that these components are not acetylated *in vitro*. For example, the process of incubating the proteins in the presence of radioactive acetyl co A and the subsequent exposure on to the radioactive plates may not pick up on particularly low levels of acetylation. In addition, the proteins were only tested in the presence of 2µg of acetyl co-A, and this may not have been of a sufficient concentration to acetylate the proteins to a detectable degree. An example of this can be found in the *in vitro* acetylation of the LX complex, which primarily did not show any acetylation, however when tested in the presence of varying amounts of acetyl co-A, LX seemed to show low levels of acetylation only with 3µg and 4µg of acetyl co-A, and not with 2µg as the other proteins were tested with (figure 6). For this reason, it would be beneficial to examine the *in vitro* levels of acetylation of XLF and XRCC4 in the presence of higher concentrations of acetyl co-A.

H1.X was used as a positive control for acetylation, and showed positive results when tested in the same manner as the other proteins. However, H1.X is known to be heavily acetylated in the presence of HATs, and therefore may show acetylation at much lower concentrations of acetyl co A. The LX complex showed faint bands in the presence of 3µg and 4µg of acetyl co-A, as can be observed in figure 6. This suggests that perhaps increasing the concentration of acetyl co-A may lead to higher levels of acetylation.

5.2 Mass Spectrometry Analysis of *in vitro* Acetylation

Due to the fact that previous methods to observe *in vitro* acetylation were not as successful as previously hoped, a more sensitive approach was employed in order to determine whether acetylation could be observed. The proteins were incubated with non-radioactive acetyl co-A, and then analysed through mass spectrometry.

5.2.1 XLF

XLF showed low levels of acetylation in the presence of P300 and CBP, but not with GCN5. XLF did show that at lysine 160, there were high levels of acetylation in the presence of PCAF; suggesting that XLF is in fact acetylated *in vitro*, but perhaps not to the degree that may have been detectable through the previous method. The mass spectrometry analysis also showed a low basal level of acetylation in XLF at lysine residue 197 (table 1), however, it is possible that a certain degree of acetylation may arise as a result of overexpressing the protein. For example, in Baculovirus expression systems, post-translational modifications may occur, and similarly this may happen also through the process of overexpressing proteins in mammalian cells. Therefore, although there is supposedly a low basal level of acetylation, it cannot be said with confidence that this occurs *in vivo*.

5.2.2 XRCC4

XRCC4 showed high levels of acetylation in the presence of both PCAF and GCN5 (table 2). However, XRCC4 was also shown to be highly acetylated alone at these sites, and therefore it is difficult to know whether the acetylation is occurring as a result of the presence of the HAT. XRCC4 was also shown to have high levels of acetylation at lysines 197 and 309 in the

presence of both P300 and CBP. Furthermore, XRCC4 did not seem to be acetylated at these sites when no HAT was present. This suggests that XRCC4 can be acetylated *in vitro* and may therefore become acetylated in response to DNA damage by P300 and CBP *in vivo*.

5.2.3 DNA Ligase IV

When DNA Ligase IV was examined for *in vitro* acetylation through mass spectrometry analysis, it was found that there was only 1 lysine residue which was acetylated. Lysine residue 613 was shown to be highly acetylated by PCAF, moderately by GCN5, and only slightly by CBP. However, low levels of acetylation were also present in DNA ligase IV alone. This may suggest that the acetylation is not as a result of being in the presence of the HATs. However, due to the fact that there are only low levels of acetylation in the absence of a HAT, and high acetylation in the presence of PCAF, it is possible that DNA ligase IV is basally acetylated to a low degree, but may become hyper-acetylated in the presence of PCAF.

5.3*In Vivo* Acetylation

The *in vitro* results received for XLF, XRCC4 and LX were suggestive, but not conclusive. There are many disadvantages to testing *in vitro* including the fact that it can be difficult to extrapolate results to relate to the endogenous interactions and the biology of the whole organism. Therefore it was important to test levels of acetylation in response to DNA damage *in vivo*.

5.3.1 XLF Acetylation *In Vivo*

MRC-5VA cells which were cultured to overexpress FLAG-XLF were exposed to 30Gy of radiation, and were collected at various time points in order to assess levels of acetylation in response to DNA damage. When blotted with PanacK, faint bands correlating to XLF were observed, suggesting that there was some acetylation present. This was confirmed by ensuring that FLAG was present. The band for 0ir seemed to be slightly darker in comparison to the 30min and 1hr time points, suggesting that XLF is acetylated immediately in response to DNA damage, and may become slightly deacetylated with time. However, the cells that had not been irradiated were also shown to be acetylated. It therefore is questionable as to whether the acetylation is as a result of exposure to DNA damage or not.

Immunoprecipitations were carried out to further confirm whether acetylation was present. However, no reactivity could be observed with the panacK antibody (figure 7b). It is possible that this was due to a high level of unspecific binding and background interactions. The IP was repeated with a high salt wash to remove any unspecific binding, but it appeared that this also removed the interactions of interest. Therefore the IP results were inconclusive.

5.3.2 XRCC4 Acetylation *In Vivo*

Cells which overexpressed FLAG-XRCC4 were subject to the same procedures as outlined for XLF in order to determine levels of acetylation following DNA damage. When blotted with the panacK antibody, there were 4 strong bands around the 55kDa mark, suggesting that there was acetylation present at all 3 time points. There also appeared to be acetylation in the sample that had not been exposed to irradiation, suggesting that like XLF, XRCC4 also had a basal amount of acetylation present prior to DNA damage. However, the bands seemed to

darken at the 0ir time point, which implies that XRCC4 undergoes further acetylation in response to DNA damage, as with XLF. This was then examined through immunoprecipitation, where conversely no bands were observed for any time point, except for a very faint band at 1hr. This seems feasible, however the results from the IP were very dirty, and therefore to confirm this with confidence, a cleaner IP would be required. High salt washes were performed in order to attempt to tidy up the results, however as with XLF, the washes appeared to remove all interactions of interest.

5.3.3 H1.X Acetylation *In Vivo*

H1.X is a histone known to undergo high levels of acetylation and deacetylation, and was therefore used as a positive control for acetylation. This was confirmed by the *in vitro* results, however when investigated *in vivo,* the results unusually showed a negative result for acetylation. The IP was repeated several times; however the negative result remained the same. Clear bands were seen at the 17kDa mark, however due to the fact that H1.X usually runs at the 35kDa mark, it was concluded that these bands did not correlate to H1.X.

5.4 Final Conclusions

After careful analysis of all *in vitro* and *in vivo* results, there can be several conclusions made, and evidence has been provded which would suggest that further research into this area would provide interesting results.

Mass spectrometry analysis of XLF *in vitro* showed high levels of acetylation at one lysine residue in the presence of PCAF. In addition, *in vivo* studies showed that following DNA damage, XLF presented a positive result for acetylation, particularly at 0ir; immediately

following damage. These results imply that following DNA damage, XLF may become acetylated at lysine 160 by PCAF. The implications this may have with regard to the NHEJ pathway are elusive, however with additional verification, this provides reason to further investigate whether this step is important for the correct function of the DNA damage response. It must not be ignored however, that acetylation was observed in the absence of DNA damage, both in the mass spectrometry results, and in the *in vivo* result, which implies that perhaps XLF is acetylated to a low degree under normal conditions, but it does not rule out the possibility that it may become further acetylated in the presence of DNA damage. Similarly, XRCC4 showed comparable results when examined *in vitro* and *in vivo* for acetylation in that high levels of acetylation were reported from the mass spectrometry analysis at two lysine residues by CBP and P300. *In Vivo* results showed that XRCC4 was acetylated at all 3 time points following irradiation, but also when it had not been irradiated. Like XLF, XRCC4 showed a slightly darker band at 0ir when compared to 0, suggesting that it has a basal level of acetylation under normal physiological conditions, and may become hyper-acetylated in response to DNA damage. This evidence implies that perhaps the post-translational modifications of XLF and XRCC4 in response to DNA damage may have a similar role in ensuring successful repair through NHEJ. The immunoprecipitation results were not successful in providing any clear conclusions;however after observing all results received from this procedure, it seems that the results received are unreliable, as H1.X showed a negative result for acetylation.

Unfortunately, due to time constraints, it was not possible to examine the acetylation of DNA Ligase IV in response to DNA damage. However the *in vitro* results suggest that DNA Ligase IV can be highly acetylated by PCAF at lysine residue 613. As with XLF and XRCC4, acetylation was found to be present in the protein in the absence of any HATs, and therefore it seems that under normal conditions, it is possible that DNA Ligase IV is similar to XRCC4

and XLF in that it has a basal level of acetylation, and may become hyper-acetylated at lysine residue 613 by PCAF in response to DNA damage. DNA Ligase IV is known to form a heterodimer with XRCC4 and to play an essential role in the NHEJ pathway, and therefore it does not seem unreasonable to hypothesize that the levels of acetylation of DNA Ligase IV may mimic the levels of acetylation seen in XRCC4.

In conclusion, in order to make firm conclusions regarding the acetylation of these components of the NHEJ pathway, it is important that a more specific antibody is used.The panacK antibody was not ideal, and often produced dirty western blots with high levels of unspecific binding. This made it difficult to confirm findings, and therefore with a more specific antibody for acetylation, it may be possible to generate more reliable results. In addition, as the immunoprecipitation reactions were unsuccessful in providing any reliable results, the method used in this project should be reviewed in order to optimise the reactions, or to identify the reason for their failure. There is sufficient evidence provided in this study to suggest that DNA Ligase IV is acetylated in response to DNA damage, and therefore this should be examined in the same manner as XLF and XRCC4; however it is hypothesiezed, that DNA Ligase IV will have similar results to that of XLF and XRCC4, particularly as DNA Ligase IV forms a heterodimer with XRCC4. Finally, the evidence provided implies that these components of the NHEJ pathway are acetylated under normal physiological conditions, but may undergo further acetylation in response to DNA damage and this may have an important role to ensure successful repair of damaged DNA.

References

1. Rothkamm, K, Kruger, I, Thompson, LH, Lobrich, M. (2003) Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. **Molecular and Cellular Biology**, Vol. 23, No.16 p.5706-5715.
2. Polo, SE, Jackson, SP. (2011) Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. **Genes Dev**. 2011 25: 409-433.
3. AJ Hartlerode, R Scully. (2009) Mechanisms of double strand break-repair in mammalian somatic cells. BioChem J, 423: 157-168.
4. Lieber, MR, Gu, J, Lu, H, Shimazaki, N, Tsai, AG. (2010) Nonhomologous DNA End Joining (NHEJ) and Chromosomal Translocations in Humans. **SubcellBiochem**. 2010 ; 50: 279–296.
5. M.M Vilenchik, AG Knudson, (2003). Endongenous DNA double-strand breaks: production, fidelity of repair and induction of cancer. **Proceedings of the National Academy of Sciences of the United States of America.**Vol.100, no 22 pp.12871-12876.
6. Ferguson DO, Sekiguchi JM, Chang S, Frank KM, Gao Y, DePinho RA, Alt FW. The nonhomologous end joining pathway of DNA repair is required for genomic stability and the suppression of translocations. (2000) **ProcNatlAcadSci USA.** Jun 6; 97(12):6630-3.
7. Dobbs TA, Tainer JA, Lees-Miller SP. A structural model for regulation of NHEJ by DNA-PKcsautophosphorylation. **DNA repair (Amst)** (2010) 10;9 (12):1307-14.
8. Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system
9. Recuero-Checa MA, Dore AS, Palomo, EA, Calzada AR, Scheres SHW, Maman JD, Pearl LH, Llorca O. (2009) Electron microscopy of XRCC4 and the DNA ligase IV-XRCC4 DNA repair complex. **DNA Repair**. 8: 1380-1389.
10. Ahnesorg, P, Smith, P, Jackson, SP. (2008) XLF Interacts with the XRCC4- DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. **Cell** 124, 301-313.
11. Barraud A, Fondaneche MC, Sanal O, Plebani A, Stephan JL, Hufnagel M, le Deist F, Fischer A, Durandy A, de Villartay Jp, Revy P. Cernunnos, a novel nonhomologous

end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell.* 2006 Jan 27; 124 (2):287-99

12. Ropars V, Drevet P, Legrand P, Baconnais S, Amram J, Faure G, Marquez JA, Pietrement O, Guerois R, Callebaut I, Le Cam E, Revy P, de Villartay JP, Charbonnier JB. Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end joining. **ProcnatlAcadSci USA** (2011) 108(31):12663-8.
13. L. Chen, K. Trujillo, P. Sung, A.E. Tomkinson, Interactions of the DNA ligase IV-XRCC4 Complex with DNA ends and the DNA-dependent protein kinase, **J. Biol. Chem.** 275 (2000) 26196-26205.
14. Lee, KJ, Jovanovic, M, Udayakumar, D, Bladen, CL, Dynan, WS. (2004) Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system.
15. Ochi T, Sibanda BL, Wu Q, Chirgadze DY, Bolanos-Garcia VM, Blundell TL. (2010) Structural Biology of DNA Repair: Spatial Organisation of the Multicomponent Complexes of Nonhomologous End Joining. **Journal of Nucleic Acids.** 10.4061/2010/621695.
16. Hentges, P, Ahnesorg, P, Pitcher, RS, Bruce, CK, Kysela, B, Green, AJ, Bianchi J, Wilson, TE, Jackson, SP, Doherty AJ. (2006) Evolutionary and Functional Conservation of the DNA Non-homologous End-joining Protein, XLF/Cernunnos **Journal of Biological Chemistry**. Vol 281, No 49, pp 37517-37526.
17. Riballo, E, Woodbine, L, Stiff, T, Walker, SA, Goodarzi, AA, Jeggo, PA. (2009) XLF-Cernunnos promotes DNA Ligase IV-XRCC4 re-adenylation following ligation. **Nucleic Acids Research** Vol. 37, No. 2 10.1093.
18. Vempati RK, Jayani RS, Notani D, Sengupta A, Galande S, Haldar D. p300-mediated Acetylation of Histone H3 Lysine 56 Functions in DNA Damage Response in Mammals. **J BiolChem** (2010) 10;285(37): 28553-28564.
19. Dekker FJ, Haisma HJ. (2009) Histone acetyl transferases as emerging drug targets. **Drug Discovery Today.**Vol 14. No. 19/20.
20. Matthias P, Yoshida M, Khochbin S. (2008) HDAC6 a new cellular stress surveillance factor. **Cell Cycle**;7:7–10. [PubMed: 18196966]

21. u Y, Teng Y, Liu H, Reed SH, Waters R. UV irradiation stimulates histone acetylation and chromatin remodeling at a repressed yeast locus. *ProcNatlAcadSci U S A*. 2005;102:8650–8655.
22. Yang, XJ, Seto, E. (2008). Lysine Acetylation: Codified cross talk with other posttranslational modifications. **Cell.**31(4): 449-461.
23. Averbeck, NB, Durante, M. (2010). Protein Acetylation Within the Cellular Response to Radiation. **Journal of Cellular Physiology.** 226:962-967.

Masters of Research

Project 2

The Identification of Human Disease Genes through the use of Exome Sequencing

Diana Walsh

This project is submitted in partial fulfilment of the requirements for the award of the MRes

Table of Contents

1.0 Abstract

Exome sequencing combined with techniques such as autozygosity mapping provides a useful method to examine variants for autosomal recessive diseases. This approach has been used to investigate and identify the causal variants in two Pakistani families with Ichthyosis and two Israeli-Arab families with Cerebral palsy.

Autosomal Recessive Congenital Ichthyosis (ARCI) is a rare and genetically heterogeneous disorder characterized by hyperkeratosis in addition to dry, scaly skin. Exome sequencing data for two affected individuals from two Pakistani families were analysed. Candidate variants were screened for in several additional members from each of the families to determine if the mutation segregated. It was found that a mutation in *ABCA12* segregated in all affected members across both families. Linkage analysis studies performed suggested that the variant had originated from a common ancestor, which implied that the families may have been distantly related.

Cerebral palsy is characterised by non-progressive abnormalities in posture and motor function as a result of a defect in the development of the nervous system. Exome sequencing data was provided for two individuals from two Israeli-Arab families with Cerebral palsy. A novel variant found in *HPDL* is believed to play a causative role in the Cerebral palsy investigated; however this gene has not been well characterised and so collecting evidence to support this theory will be challenging. The successful identification of disease genes and the improved understanding of their methods of pathogenesis, new potential targets may be discovered through which new treatments, therapies, and methods of diagnosis may be developed.

2.0 Introduction

2.1 Consanguinity and Recessive Disease

The concept of consanguinity refers to a union in which the couple are known to share genes inherited from one or more common ancestors, and consequently, their offspring will contain segments of their genome which are homozygous (1). Many major populations have a preference towards close-kin marriages due to the social and economic advantages associated; this is particularly true in various parts of India, Pakistan and in many Muslim countries (1). An additional factor which exerts a major influence on the preference for a consanguineous relationship is religion; which is demonstrated by the geographical distribution of consanguinity (figure 1). For example, in most Muslim countries from the Middle East and Pakistan, over 50% of the marriages are said to be consanguineous (2). Recent studies have shown that approximately 10.4% of the 6.7 billion global population are related as second cousins or closer (3).

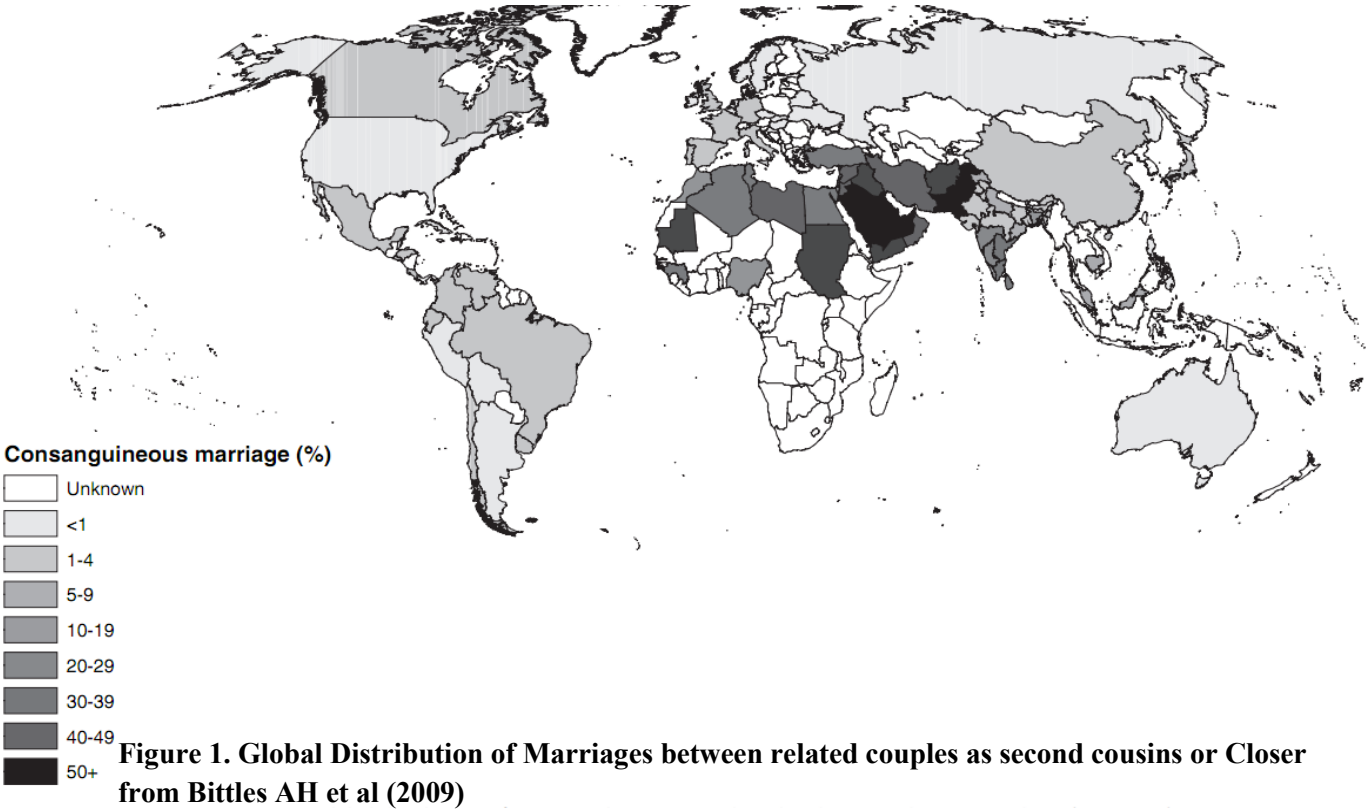


Figure 1. Global Distribution of Marriages between related couples as second cousins or Closer from Bittles AH et al (2009)

There are many adverse genetic implications associated with a consanguineous marriage, particularly when assessing the proportion of identical gene pairs which will be present in the progeny, as they are inherited from a common ancestor (1). The majority of consanguineous marriages are between first cousins, which results in the sharing of approximately 12.5% of genes (4). As a result, the children will be homozygous for 6.25% of gene loci, and thus the coefficient of inbreeding (F) is equal to 0.0625 (4). The clinical consequences of this are that in consanguineous populations, it is expected that there will be an increased incidence of autosomal recessive disease (8). According to *Bittles et al. (2007),* the genetic risk of an autosomal recessive disorder being expressed in the offspring of a consanguineous marriage is inversely proportional to the frequency of the disease allele within the total gene pool (5). In addition, longer histories of consanguinity within a family are expected to exhibit a much higher degree of homozygosity in the offspring, in comparison to a single case of consanguinity. For this reason, the risk of acquiring a recessive disease also increases with the length of the history of consanguinity within a given family (8).

Many rare disease genes have been identified, and their chromosomal locations have been mapped through the investigation of consanguineous families with multiple affected individuals (5). Recent advances in technological abilities to rapidly sequence the whole genome with high-throughput have provided a considerable opportunity to discover and analyse various genotype-phenotype correlations (5,6).

2.2 Autozygosity Mapping

The term ‘autozygosity’ refers to the situation in which two alleles at a particular locus originate from a common ancestor as a result of a consanguineous mating. It is not uncommon for the progeny of a close-kin marriage to acquire an autosomal recessive

disorder as a result of the inheritance of two disease alleles which are identical by descent (IBD) (figure 2). To investigate the genotypes of such individuals, autozygosity mapping has become a useful tool, which makes use of the inbreeding coefficient to search for an autozygous region within the genome of the affected individuals, and so provides an alternative method to identify disease associated single nucleotide polymorphisms (SNPs) (7).

The use of autozygosity mapping to characterise and identify genes in autosomal recessive disorders remains a powerful and effective technique with which information may be generated in order to facilitate the diagnosis of patients and to indicate the potential functional role of the disease gene (7).

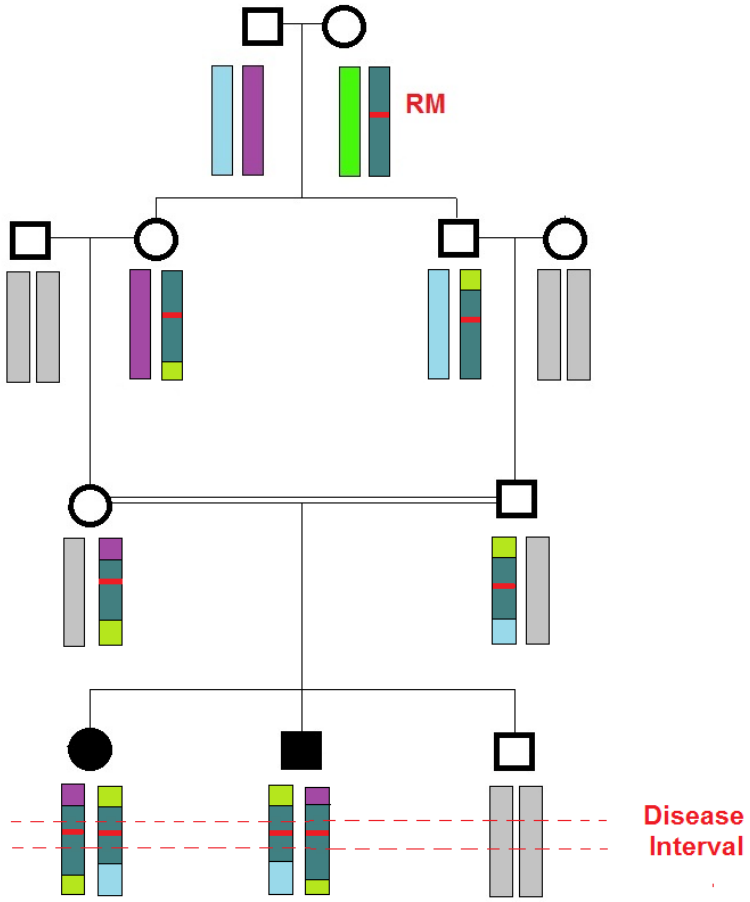


Figure 2. A diagram to show the principal of autozygosity in consanguineous families The recessive mutation (RM) is shown in red. Through the pattern of random inheritance, the mutation is passed on through the generations. The offspring of a consanguineous marriage can inherit two disease alleles which have originated from a common ancestor, thereby causing an autosomal recessive disorder.

2.3 Sequencing and Exome Analysis

The term 'Genomics' refers to the systematic study on a whole-genome scale for the identification of genetic contributions to human conditions (9). Following the completion of the human genome project in 2001, the field of genomics has been advancing exponentially, as the costs are decreasing, while the ability to sequence multiple reads in parallel at high-throughput is continually progressing. For example, in 2001, the first human genome was sequenced at a cost of $2.7 billion, and took 13 years to complete (9). Today, an entire human

genome can be sequenced over the course of a few days, and is expected to cost approximately $10,000 in the near future (9). These advances have instigated a paradigm shift in the approaches with which both rare and common diseases are investigated (10).

Despite the technological advances, and the on-going decline in the cost of sequencing, Whole genome sequencing (WGS) still remains prohibitively expensive to use frequently (10). Although WGS provides the highest sensitivity and accuracy when genotyping an individual, there is a considerable amount of redundancy to this method. The expenses associated with WGS and the vast amount of data provided could be considered to be in excess of what is required to generate the same conclusion. For this reason, employing targeted sequencing approaches as opposed to sequencing the entire genome of an individual may provide higher sequence coverage of the regions of interest, at a fraction of the cost (9).

2.3.1 Exome Sequencing

Whole exome sequencing provides a less expensive alternative to WGS, involving the sequencing of only the protein-coding genes as opposed to the entire genome of an individual. The protein-coding genes make up approximately 1% of the human genome, however despite this, studies suggest that the protein-coding genes may harbour up to 85% of harmful mutations with large effects on disease-associated traits (10). For this reason, although exome sequencing does not screen the whole genome, current research suggests that there is a high probability that a causative mutation may be found within the coding region of the genome.

Exome sequencing is performed by hybridizing genomic DNA to oligonucleotide probes which cover specific regions of the human exome (24). These particular regions are then

sequenced at high throughput, using second generation sequencing technologies. The three major platforms used include one by Agilent, one by Roche/Nimblegen and finally another platform designed by Illumina. The main principle remains the same across all three platforms, in that they make use of biotinilated oligonucleotide probes which are designed to be complementary to the exome targets in question (24).

The success of exome sequencing depends on several factors; firstly it depends on the mutation being present within the region of the genome examined. In addition, should the mutation be found, successful conclusions depend on the ability to identify the gene as harbouring the causative variant, and having sufficient sample sizes and evidence to eliminate the possibility of it being a random variation (11). The identification of a gene which harbours a mutation for a given disease is important in order to understand the genetic basis of the disorder. Through understanding and appreciating the role of a mutation in the acquisition of a particular disease, knowledge may be applied to carry out further research in order to develop potential therapies and treatments.

2.4 Main Aim of Project

The main aim of this project is to use whole exome sequencing data to attempt to find and identify the causative mutation for several autosomal recessive diseases. Through PCR and Sanger sequencing, this project aims to show that the mutation identified segregates with all tested family members, and that the mutation is not present in the general population. Once identified, to further confirm that the mutation is the causative factor for the disease, microsatellite markers will be used to confirm linkage with the mutation across the family which is therefore consistent with the hypothesis.

3.0 Materials and Methods

3.1 Patient Data

Consanguineous families from Pakistan and Saudi Arabia were examined and analysed in this project for Ichthyosis and Cerebral palsy. Permission was received from all of the families studied. For each of the families, extracted DNA was supplied by Dr. Neil Morgan, and made up to a concentration of 20ng/µl for all affected, and the majority of unaffected family members.

3.2 Chemicals and Reagents

10 x TBE Buffer (Tris-borate/EDTA)	National Diagnostics
Agarose	Invitrogen
Ethanol	Fischer Scientific
Ethidium Bromide	Sigma
Genescan-500 LIZ Size Standard	Applied Biosystems
Hi-Di Formamide	Applied Biosystems
PCR Reagents	
Primers	Sigma
1 kb DNA ladder (1µg/µl)	Invitrogen
Biomix Red	Bioline
GC Rich Solution	Roche
Solutions for Sequencing Reactions	
BigDye Terminator	Applied Biosystems
Big Dye Sequencing Buffer	Applied Biosystems

3.3 Polymerase Chain Reaction (PCR)

For each gene, polymerase chain reaction (PCR) was carried out, primers were designed either manually ('by eye'), or by using the program exonprimer, (http://ihg.gsf.de/ihg/ExonPrimer.html).

3.1.3a Standard PCR amplification for Candidate Genes

PCR amplifications were carried out in 25µl reactions containing 20ng of DNA, 1.5mM Biomix Red, and 5pmol of both the forward and reverse primers. To ensure contents were thoroughly mixed, the microtitre plate was spun in a centrifuge at 1000 RPM for 30 seconds. For each PCR carried out, a negative control was included which involved the addition of water instead of DNA to ensure that the samples were not contaminated. Some primers used in the PCR amplifications had a high GC content, and therefore to optimise these reactions, 0.5mM GC Rich Solution was used. The stages of the PCR cycle can be found in table 1.

Temperature	Time		Purpose
94°C	3 mins		Initial Denaturation
94°C	1 min		Melting
60°C	1 min	**30 Cycles**	Annealing
72°C	1 min		Extension
72°C	5 mins		Final Extension

Table 1. The Stages of a Standard PCR Cycle

3.1.3b Analysis of Products from Standard PCR Cycle

The results of the PCR amplification reactions were run on 1% (w/v) Agarose gels containing Ethidium Bromide. 5µl of 1 kb ladder size marker was loaded alongside the samples to allow product sizes to be identified. 5µl of PCR product was then loaded and the DNA was then separated through electrophoresis in 1 X TBE buffer at 150V. The PCR products were then visualised using a UV light transilluminator.

3.1.3c PCR Amplification of Microsatellite Markers

PCR amplifications of microsatellite markers were routinely carried out in 10µl reactions which contained 20ng of genomic DNA, as well as 5mM Biomix Red and 2.0 pmol of the forward and reverse primers. The microtitre plate was then placed into the PCR machine, where the following PCR stages were carried out (table 2).

Temperature	Time		Purpose
95°C	5 mins		Initial Denaturation
95°C	30 secs		Melting
55°C	30 secs	**28 cycles**	Annealing
72°C	30 secs		Extension
72°C	5 mins		Final Extension

Table 2: PCR stages for the amplification of Microsatellite Markers

3.1.3d Analysis of Microsatellite Marker Data

In order to analyse the microsatellite data, the samples were run on the ABI 3730 DNA Analyzer. Prior to this, the markers were diluted 1 in 15 with dH_20 after which 1.5µl of the diluted product was added to 10µl of Hi-Di Formamide and 0.04µl Genescan-500 LIZ size standard. The results and product sizes were analysed using Genemapper v3.0 software (Applied Biosystems).

3.4 Sequencing of PCR Products

3.1.4a PcR product clean-up

An equal volume of microCLEAN was added to 2.4µl of PCR product obtained following section 3.1.3a in a microtitre plate, and centrifuged for 40 minutes at 4000RPM. Following this, the supernatant was removed by inverting the plate on to a paper towel, and then spinning in the centrifuge at 500RPM for 30 seconds.

3.1.4b Sequencing Reactions

To each well, a sequencing buffer was added which included 2µl Big Dye Sequencing buffer (Applied Biosystems), 0.5µl Big Dye reaction mix (Applied Biosystems) and 5.5µl dH_20. 2pmol/µl of the appropriate forward or reverse primer was then added, and was put into the PCR machine with cycling conditions of:

- 96^{o}C 30 seconds
- 50 ^{o}C 15 seconds x 30
- 60 ^{o}C 4 minutes

3.1.4c Sequencing Reactions Clean-up

The sequencing reaction was then cleaned up by adding 2µl of EDTA (0.125M) to each well along with 30µl of 100% Ethanol, and spinning at 2000 RPM for 20 minutes. The supernatant was removed again as before, by blotting on a paper towel, and spinning upside down for 30 seconds at 500 RPM. The pellet was left to dry for 5 minutes before adding 90µl of 70% Ethanol to each well. The plate was then spun at 2000 RPM for 10 minutes, after which the supernatant was removed as outlined earlier. Finally the pellet was resuspended in 10µl of Hi-Di Formamide, and denatured at 95°C for 3 minutes, and snap chilled on ice for 5 minutes. The microtitre plate containing the samples was then loaded on to an ABI Prism 3700 Sequencer by Dr. Dean Gentle. The resulting DNA sequences were analysed with BioEdit Sequence Alignment Editor, or ChromasLite.

3.5 Analysis of Whole Sequencing Data

Whole Exome Sequencing data was provided by Prof. Eamonn Maher in collaboration with a group in Guy's Hospital, London. For each disease studied, the 25,000 variants received from the exome sequencing data were screened and analysed in order to exclude those which were present in the dbSNP131 or the 1000 Genomes project. In addition, synonymous variants were excluded as they do not alter the protein encoded.On average, the data was narrowed down from approximately 25,000 variants to 50 potential candidate variants. Candidate genes for Ichthyosis were identified through the observation that *ABCA12* and *ALOXE3* were already known to be associated with the disease. Whereas the candidate genes for the Cerebral palsy family were prioritised through linkage data provided from earlier research by Dr. Neil Morgan.

3.6 Screening Controls for Candidate Gene

To ensure that sequence variants were not present in the general population, the variants were screened in a panel of control samples of either Asian or Saudi Arabian controls. The controls tested depended on the population from which the variant was found. The control DNA samples were provided by the West Midlands Regional Genetics Laboratory.

4.0 Results

4.1 Ichthyosis

Autosomal Recessive Congenital Ichthyosis (ARCI) is a rare and genetically heterogeneous disorder characterized by hyperkeratosis in addition to dry, scaly skin. ARCI consists of three major subtypes, among which the range of clinical features and severity of disease varies. The subtypes include Harlequin Ichthyosis (HI), which is the most severe and devastating form of Ichthyosis, and is often fatal in the majority of affected neonates (12). Lamellar Ichthyosis (LI) and Congenital Ichthyosiform erythroderma (CIE) make up the further two subtypes, and while they are devastating, they are not as severe as HI. The HI patients present with critical and significant clinical features at birth, including severe ectropion, eclabium and flattened ears, in addition to large plate-like scales over the entire body (13,14). The patients who suffer from CIE on the other hand exhibit fine scales in addition to variable erythroderma, while LI patients show less severe erythroderma, but have thick, dark scales over the entire body (13,14).

4.1.1 Clinical Features of Two Families with Ichthyosis

The affected individuals of family 1 presented with fine white scales over the entire body, in addition to dry skin and hyperkeratosis of the hands, feet, elbows and knees (figure 3). The hair of the patients seemed very dry, brittle and thin. Deep pits were present in the toenails, and they seemed to be dystrophic. The individuals could not sweat, and therefore were heat intolerant.

Figure 3. Clinical Features for Several Patients from Family 1 with Ichthyosis: (A) Proband from family 1 shows dry, cracked skin and hyperkeratosis on the bottom of the feet. (B) Another member of family 1 exhibits fine, white scales in addition to erythroderma over entire body. (C) Severe hyperlinear palms on an additional member of family

The affected individuals in family 2 similarly showed fine white scales over the entire body in addition to rough, dry skin (figure 4). The patients all had generalized pruritis and exhibited erythematous and exczematous lesions. All patients had hypohidrosis, and as a result, suffered from severe heat intolerance. Despite their conditions, all patients had normal mental health.

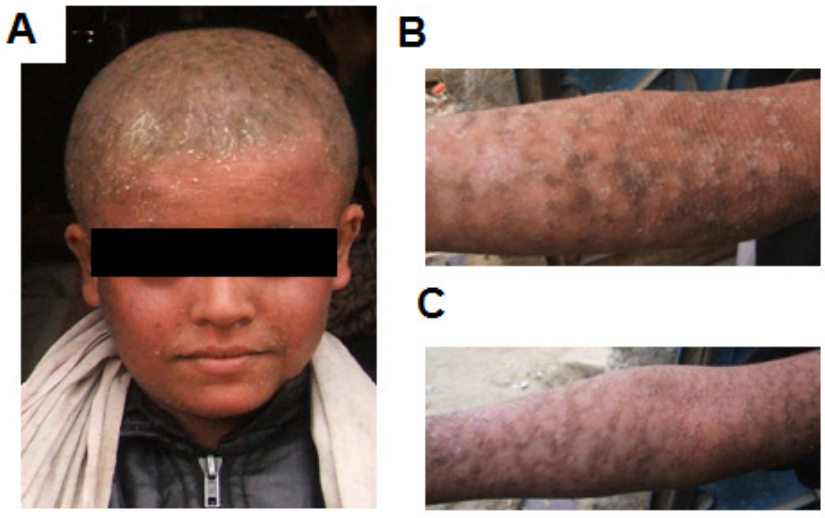


Figure 4. Clinical Features for Proband from Family 2

(A) The patient exhibits yellowish adherent scales and dry skin on the scalp. In addition, the patient displays fine white scales over the entire body and erythroderma. Hyperkeratosis can be observed on the feet, knees and elbows (B,C). The patient and the other affected individuals in the family all have hypohidrosis and have severe heat intolerance. The affected patients in family 2 also have difficulty breathing and allergic rhinitis.

4.1.2 Methods

4.1.2a Patient Information

Two consanguineous Asian families from the remote area of Sindh, a province within Pakistan, were ascertained and provided samples to be studied by King's College London. The anonymous pedigrees of the families can be found in section 4.1.2b. For both families, DNA was available for most of the affected individuals, at least one of their parents, and some unaffected individuals. The range of clinical symptoms varied among the affected individuals, however all presented with dry, scaly skin and hyperkeratosis to the elbows, knees, feet and hands.

4.1.2b Pedigree Information for Families

Family 1:

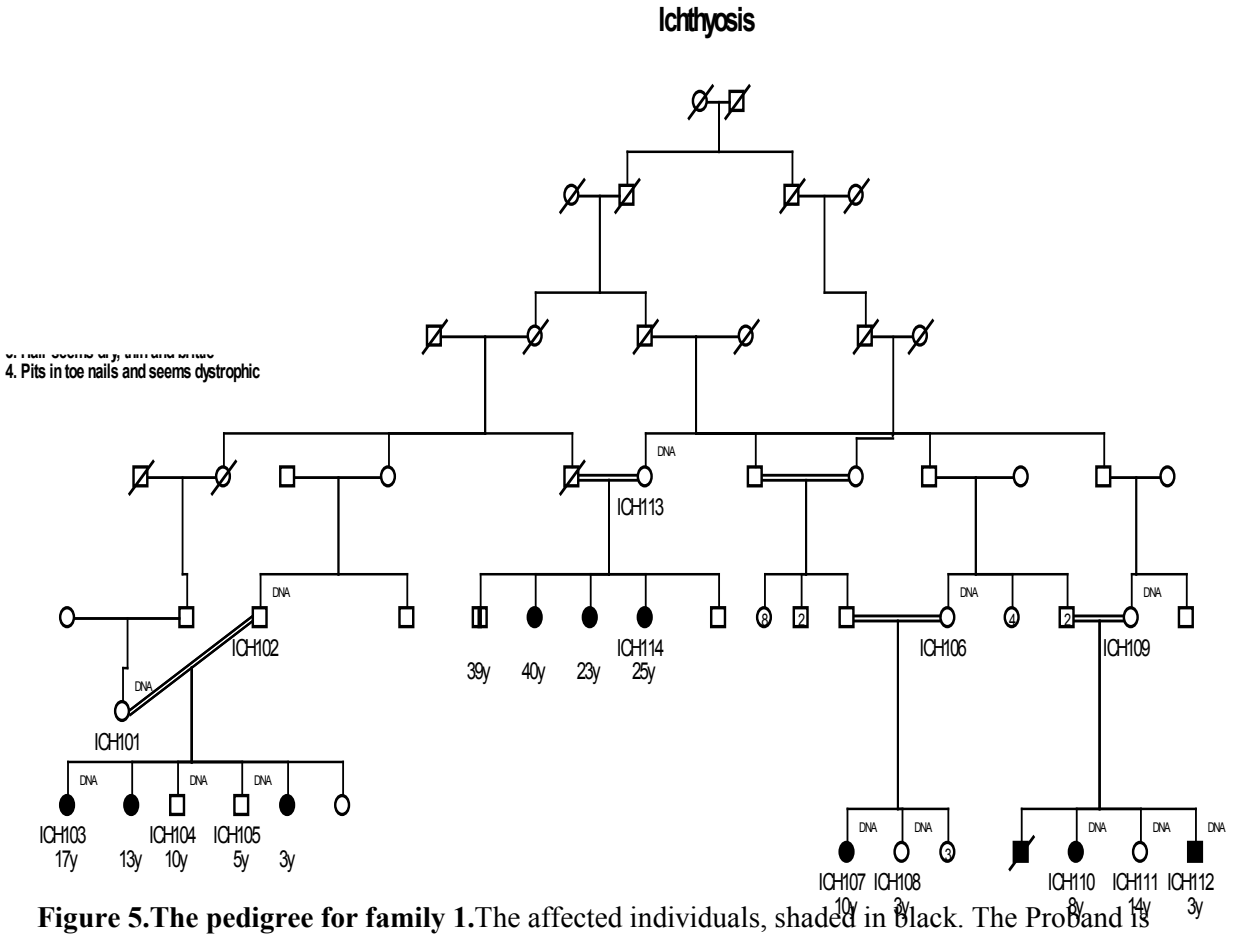
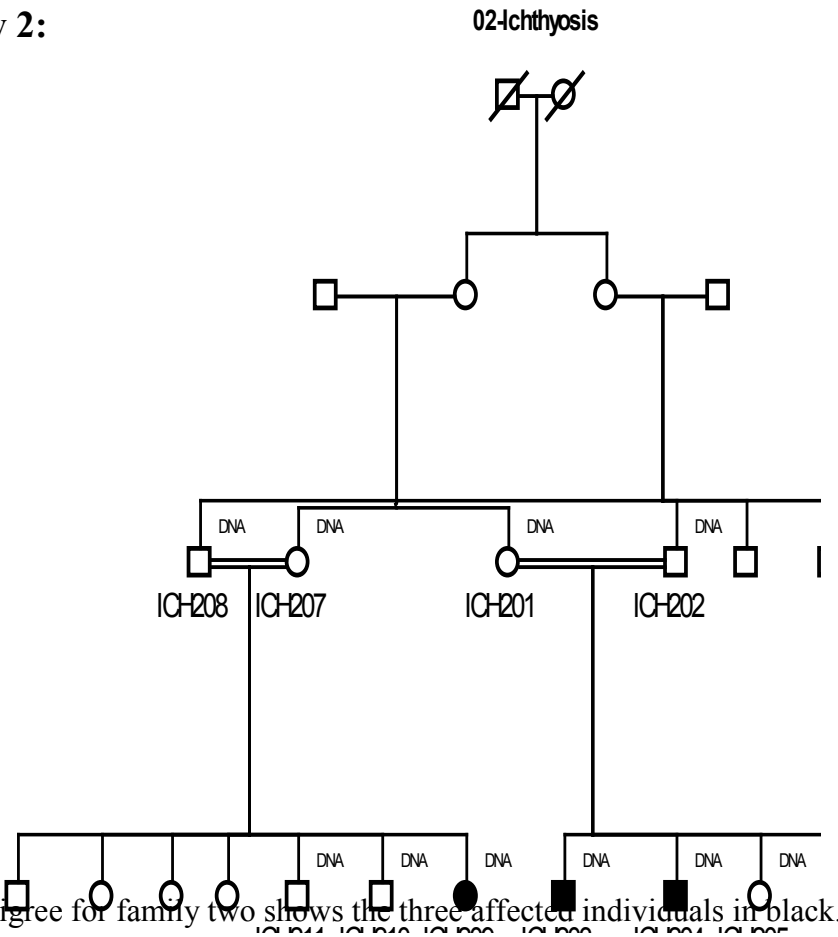


Figure 5.The pedigree for family 1. The affected individuals, shaded in black. The Proband is marked ICH107.

Family 2:



The pedigree for family two shows the three affected individuals in black. The Proband is shown at ICH204.

4.1.2d Mutation Analysis of Candidate Genes

Primers were designed for *ABCA12* using the UCSC genome browser to view the full sequence, and identify the location of the mutation. Primers were designed to flank either side of the coding exon harbouring the mutation. The program ExonPrimer(http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html) was used to facilitate the design of the primers, which were ordered from Sigma Aldrich.

These primers were used to amplify regions of interest of *ABCA12* from DNA of the two families (see methods section 3.1.3)

4.1.2e Markers Analysed in the Families

Microsatellite markers flanking*ABCA12* were used to establish linkage within the two families. The following microsatellite markers were chosen: D2S2944, D2S2382, D2S164 and D2S434. PCR reactions were carried out and analysed, as outlined in methods section 3.1.3.

4.1.3 Results

4.1.3a Analysis of Exome Sequencing Data

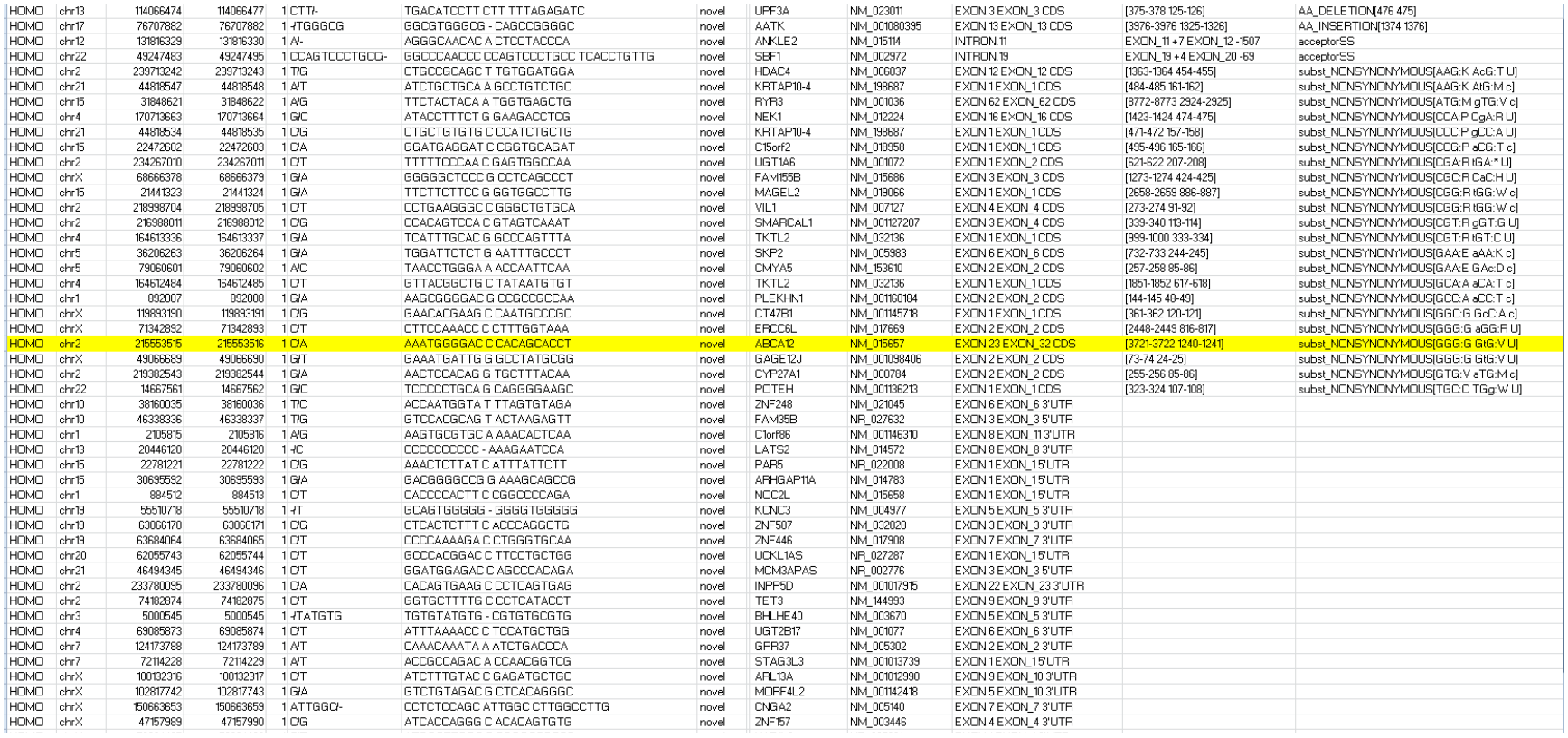
Whole exome sequencing data was provided for the probands, ICH107 and ICH204. Through process of elimination, the novel and non-synonymous changes were identified. This lowered the number of candidate variants from the original 29,403 to 51 variants in family 1, and was reduced from 29,526 down to 51 variants in family 2.

There are 6 genes currently known to be associated with Ichthyosis; *ALOXE3*, *ALOX12B*, *TGM1*, *CYP4F22*, *NIPAL4* and *ABCA12*. Upon analysing the exome sequencing data for both families, it was noticed that two genes had already been described in current literature, and had been shown to play a causative role in the disorder. Both family 1 and family 2, although thought to be unrelated, carried the same mutation in 1 particular gene; *ABCA12*. For this reason, *ABCA12* was identified as the primary candidate gene.

A. Exome Sequencing Data for Family 1

S0146	HOMO	chr1	223339937	223339938	1	A/-	TCTTGAAAAA A TGCTATTTAA	novel	DNAH14	NM_001373	EXON.21 EXON_23 CDS	[3396-3397 1132-1133]	FRAMESHIFT[4515 1147]	1132-4516
S0146	HOMO	chr13	99420855	99420855	1	-/GCG	GGCGGCGGCG - CCGGGGGCGG	novel	ZIC5	NM_033132	EXON.1 EXON_1 CDS	[1075-1075 358-359]	AA_INSERTION[663 664]	358-359
S0146	HOMO	chr16	1763392	1763393	1	G/-	GACCCAGCGG G TGAGCGCAGG	novel	EME2	NM_001010865	EXON.1 EXON_1 CDS	[163-164 54-55]	FRAMESHIFT[444 72]	54-445
S0146	HOMO	chr16	87327911	87327911	1	-/CTG	GCTGCTGCTG - ATGCTCCTGC	novel	FAM38A	NM_001142864	EXON.17 EXON_17 CDS	[2232-2232 744-744]	AA_INSERTION[2521 2522]	744-744
S0146	HOMO	chr19	7026698	7026701	1	AAG/-	TCTCAGCGAG AAG GACACACAGA	novel	ZNF557	NM_001044387	EXON.4 EXON_4 CDS	[64-67 21-23]	AA_DELETION[430 429]	21-23
S0146	HOMO	chr1	223395054	223395054	1	-/TA	TATATATATA - GAGAGAGATA	novel	DNAH14	NM_001373	INTRON.26	EXON_28 +21435 EXON_29 -1	donorSS	
S0146	HOMO	chr15	26785471	26785472	1	T/G	GATATTTTGG T TTTTTTTTTA	novel	WHAMML2	NR_026589	INTRON.2	EXON_2 +1 EXON_3 -3463	acceptorSS	
S0146	HOMO	chr17	7691063	7691064	1	G/T	CGCCAGGTGA G CCCCTGCCTG	novel	KDM6B	NM_001080424	INTRON.9	EXON_9 +4 EXON_10 -83	acceptorSS	
S0146	HOMO	chr2	208753705	208753706	1	A/G	ATCCCAGGTC A CCCTTACCTT	novel	C2orf80	NM_001099334	INTRON.6	EXON_6 +7 EXON_7 -8661	acceptorSS	
S0146	HOMO	chr6	36574244	36574245	1	T/G	GGGTAGCCTG T AATAAAAAAG	novel	STK38	NM_007271	INTRON.10	EXON_10 +1382 EXON_11 -3	donorSS	
S0146	HOMO	chr6	37527553	37527554	1	C/T	TTTTTTCTCC C TGTAGAGAAA	novel	FTSJD2	NM_015050	INTRON.5	EXON_5 +1456 EXON_6 -5	donorSS	
S0146	HOMO	chr6	37553342	37553343	1	C/T	CCTCTACCTG C AGGGTGAAGG	novel	FTSJD2	NM_015050	INTRON.20	EXON_20 +1394 EXON_21 -2	donorSS	
S0146	HOMO	chr12	11135293	11135294	1	T/C	TAGCTGAATC T AATAGCTTTG	novel	TAS2R43	NM_176884	EXON.1 EXON_1 CDS	[801-802 267-268]	subst_NONSYNONYMOUS[AGA:R gGA:G U]	267-268
S0146	HOMO	chr12	11135302	11135303	1	T/G	CTAATAGCTT T GCAGAACATG	novel	TAS2R43	NM_176884	EXON.1 EXON_1 CDS	[792-793 264-265]	subst_NONSYNONYMOUS[AAA:K cAA:Q c]	264-265
S0146	HOMO	chr12	11135622	11135623	1	A/T	GTTTCCTTCA A ATTCTTTTGT	novel	TAS2R43	NM_176884	EXON.1 EXON_1 CDS	[472-473 157-158]	subst_NONSYNONYMOUS[TTT:F TaT:Y c]	157-158
S0146	HOMO	chr15	97489840	97489841	1	A/C	ATTATTTTGC A ACAGAAGAGT	novel	SYNM	NM_145728	EXON.4 EXON_4 CDS	[3751-3752 1250-1251]	subst_NONSYNONYMOUS[CAA:Q CcA:P c]	1250-1251
S0146	HOMO	chr16	2171025	2171026	1	T/C	GGTCGGGTTT T GGTGGGCGTC	novel	CASKIN1	NM_020764	EXON.18 EXON_18 CDS	[2343-2344 781-782]	subst_NONSYNONYMOUS[AAA:K gAA:E c]	781-782
S0146	HOMO	chr17	7958632	7958633	1	G/A	GGGAAGCCGG G CAGGTACCGA	novel	ALOXE3	NM_001165960	EXON.6 EXON_7 CDS	[969-970 323-324]	subst_NONSYNONYMOUS[CCC:P tCC:S U]	323-324
S0146	HOMO	chr19	3710946	3710947	1	C/T	CGGCCAGCTT C GGCAGACAGG	novel	APBA3	NM_004886	EXON.2 EXON_2 CDS	[315-316 105-106]	subst_NONSYNONYMOUS[GAA:E aAA:K c]	105-106
S0146	HOMO	chr19	6405416	6405417	1	T/C	AGGGAGCGGA T GCCTCCCTCA	novel	SLC25A23	NM_024103	EXON.6 EXON_6 CDS	[711-712 237-238]	subst_NONSYNONYMOUS[ATC:I gTC:V c]	237-238
S0146	HOMO	chr2	215553515	215553516	1	C/A	AAATGGGGAC C CACAGCACCT	novel	ABCA12	NM_015657	EXON.23 EXON_32 CDS	[3721-3722 1240-1241]	subst_NONSYNONYMOUS[GGG:G GtG:V U]	1240-1241
S0146	HOMO	chr2	217267648	217267649	1	T/G	GAGGGCTTTC T CGTCGCAGGG	novel	IGFBP5	NM_000599	EXON.1 EXON_1 CDS	[94-95 31-32]	subst_NONSYNONYMOUS[GAG:E GcG:A c]	31-32
S0146	HOMO	chr2	231971163	231971164	1	G/A	GCGGCAGTCC G CGGGTGGGGG	novel	B3GNT7	NM_145236	EXON.2 EXON_2 CDS	[489-490 163-164]	subst_NONSYNONYMOUS[GCG:A aCG:T c]	163-164
S0146	HOMO	chr3	132226544	132226545	1	T/A	CATCTGGATC T TCTCTCTAGC	novel	ASTE1	NM_014065	EXON.3 EXON_3 CDS	[295-296 98-99]	subst_NONSYNONYMOUS[AAG:K AtG:M c]	98-99
S0146	HOMO	chr3	53812526	53812527	1	C/T	AACTATTTGC C GGGAAGACCC	novel	CACNA1D	NM_000720	EXON.45 EXON_46 CDS	[5532-5533 1844-1845]	subst_NONSYNONYMOUS[CGG:R tGG:W c]	1844-1845
S0146	HOMO	chr6	29563231	29563232	1	A/C	GGGATAAAAA A CACGACTCCA	novel	MAS1L	NM_052967	EXON.1 EXON_1 CDS	[426-427 142-143]	subst_NONSYNONYMOUS[TTT:F gTT:V c]	142-143
S0146	HOMO	chr9	83737131	83737132	1	G/A	AAAGAAGTCC G CATCAAGCTT	novel	FLJ43859	NM_001145197	EXON.4 EXON_4 CDS	[2235-2236 745-746]	subst_NONSYNONYMOUS[GCA:A aCA:T c]	745-746
S0146	HOMO	chr12	117177822	117177823	1	T/G	CAATTTTTTT T GGGGGGTAAA	novel	TAOK3	NM_016281	EXON.3 EXON_3 5'UTR			
S0146	HOMO	chr13	110165640	110165641	1	A/-	AAAAAAAAAA A TTGACCGCTA	novel	ING1	NM_005537	EXON.1 EXON_4 5'UTR			
S0146	HOMO	chr1	37999363	37999363	1	-/AA	AAAAAAAAAA - CAATGACTTC	novel	EPHA10	NM_173641	EXON.3 EXON_3 3'UTR			
S0146	HOMO	chr13	90804050	90804051	1	T/-	TTTTTTTTTT T GGCGTTTGCA	novel	MIR17HG	NR_027350	EXON.2 EXON_2 5'UTR			
S0146	HOMO	chr16	68567791	68567792	1	T/C	TTTGTTCCAC T CACAGATTTC	novel	PDXDC2	NR_003610	EXON.26 EXON_26 3'UTR			
S0146	HOMO	chr16	87754286	87754287	1	C/-	GAAGACCCCC C GCTTCCTCAC	novel	C16orf81	NR_024347	EXON.3 EXON_3 5'UTR			
S0146	HOMO	chr16	87754378	87754378	1	-/C	CGAAGACCCC - GCTTCCTCAC	novel	C16orf81	NR_024347	EXON.3 EXON_3 5'UTR			
S0146	HOMO	chr17	8602497	8602498	1	C/T	CTTCCAGACT C CGTTAGGACC	novel	SPDYE4	NM_001128076	EXON.1 EXON_1 5'UTR			
S0146	HOMO	chr19	5631514	5631515	1	C/G	CGGATCCACG C GCAAGGACAC	novel	C19orf70	NM_205767	EXON.1 EXON_1 5'UTR			
S0146	HOMO	chr19	7871909	7871912	1	GGG/-	GAGATGGGGG GGG CGGGGGCAGC	novel	LRRC8E	NM_025061	EXON.3 EXON_3 3'UTR			
S0146	HOMO	chr19	8461050	8461051	1	C/T	GGCTGGCTGT C CTGGCCCCAC	novel	PRAM1	NM_032152	EXON.10 EXON_10 3'UTR			
S0146	HOMO	chr2	218848384	218848385	1	C/T	TCATAGGGCC C AGCCCTCTAG	novel	TMBIM1	NM_022152	EXON.12 EXON_12 3'UTR			
S0146	HOMO	chr2	226224543	226224544	1	G/A	GATACAACTT G CCAAATGCTT	novel	KIAA1486	NM_020864	EXON.6 EXON_6 3'UTR			
S0146	HOMO	chr22	42589669	42589675	1	GCGCCC/-	GCCCGCGCCC GCGCCC CGCACACGCT	novel	SULT4A1	NM_014351	EXON.1 EXON_1 5'UTR			
S0146	HOMO	chr2	55499856	55499857	1	G/-	GAGGAGGGGG G CACTCTCCCT	novel	CCDC88A	NM_001135597	EXON.1 EXON_1 5'UTR			
S0146	HOMO	chr2	58239915	58239916	1	C/T	AGCATTCTTA C ACACTACACA	novel	FANCL	NM_001114636	EXON.14 EXON_14 3'UTR			
S0146	HOMO	chr2	63703600	63703601	1	C/G	CACCACCGCA C GGCGCCCAAG	novel	LOC388955	NR_003131	EXON.1 EXON_1 3'UTR			
S0146	HOMO	chr3	131305753	131305754	1	C/T	TAGCGGGCAG C GTCCCATGTG	novel	LOC729375	NR_024252	EXON.3 EXON_3 3'UTR			
S0146	HOMO	chr5	136342033	136342033	1	-/CA	ACAACAACAA - AAAAAAAACA	novel	SPOCK1	NM_004598	EXON.11 EXON_11 3'UTR			
S0146	HOMO	chr6	29906560	29906560	1	-/ATTTGTTCATGCCT	GAGTGGCAAG - TCCCTTTGTG	novel	HLA-G	NM_002127	EXON.8 EXON_8 3'UTR			
S0146	HOMO	chr6	31794617	31794618	1	G/T	GCAAAGTGGG G AGCCCTTCTA	novel	LY6G6C	NM_025261	EXON.3 EXON_3 3'UTR			
S0146	HOMO	chr6	35215886	35215887	1	C/T	AGCCGCAAAG C CTCAATCTAA	novel	TCP11	NM_018679	EXON.3 EXON_3 5'UTR			
S0146	HOMO	chr6	37558886	37558887	1	C/G	CCTGCTTGAT C CTGATCAAGG	novel	C6orf129	NM_138493	EXON.4 EXON_4 3'UTR			
S0146	HOMO	chr7	1546993	1546994	1	C/T	CCAGCCGGTG C GGGAGCCCCT	novel	MAFK	NM_002360	EXON.3 EXON_3 3'UTR			
S0146	HOMO	chr7	4965235	4965236	1	G/A	CAGAGCCTGG G GGGCGCGGCG	novel	MMD2	NM_001100600	EXON.1 EXON_1 5'UTR			
S0146	HOMO	chr8	95592411	95592411	1	-/T	TTAAAATGTT - ACCCCCATAT	novel	KIAA1429	NM_183009	EXON.13 EXON_13 3'UTR			

Exome Sequencing Data for Family 2

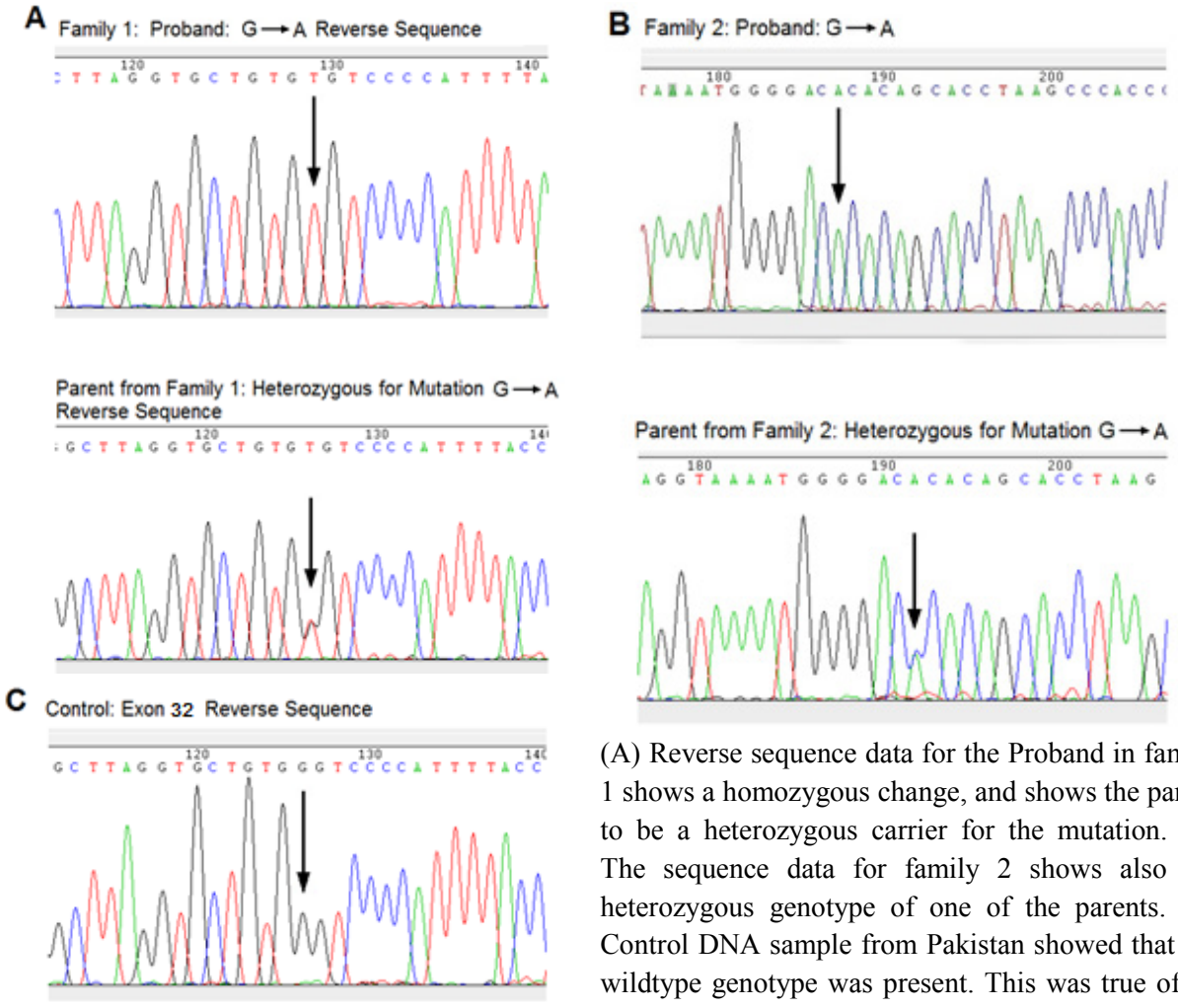


HOMO	chr13	114066474	114066477	1 CTTI-	TGACATCCTT CTT TTTAGAGATC	novel	UPF3A	NM_023011	EXON.3 EXON_3 CDS	[375-378 125-126]	AA_DELETION[476 475]
HOMO	chr17	76707882	76707882	1 -/TGGGCG	GGCGTGGGCG - CAGCCGGGGC	novel	AATK	NM_001080395	EXON.13 EXON_13 CDS	[3976-3976 1325-1326]	AA_INSERTION[1374 1376]
HOMO	chr12	131816329	131816330	1 A/-	AGGGCAACAC A CTCCTACCCA	novel	ANKLE2	NM_015114	INTRON.11	EXON_11 +7 EXON_12 -1507	acceptorSS
HOMO	chr22	49247483	49247495	1 CCAGTCCCTGCC/-	GGCCCAACCC CCAGTCCCTGCC TCACCTGTTG	novel	SBF1	NM_002972	INTRON.19	EXON_19 +4 EXON_20 -69	acceptorSS
HOMO	chr2	239713242	239713243	1 T/G	CTGCCGCAGC T TGTGGATGGA	novel	HDAC4	NM_006037	EXON.12 EXON_12 CDS	[1363-1364 454-455]	subst_NONSYNONYMOUS[AAG:K AcG:T U]
HOMO	chr21	44818547	44818548	1 A/T	ATCTGCTGCA A GCCTGTCTGC	novel	KRTAP10-4	NM_198687	EXON.1 EXON_1 CDS	[484-485 161-162]	subst_NONSYNONYMOUS[AAG:K AtG:M c]
HOMO	chr15	31848621	31848622	1 A/G	TTCTACTACA A TGGTGAGCTG	novel	RYR3	NM_001036	EXON.62 EXON_62 CDS	[8772-8773 2924-2925]	subst_NONSYNONYMOUS[ATG:M gTG:V c]
HOMO	chr4	170713663	170713664	1 G/C	ATACCTTTCT G GAAGACCTCG	novel	NEK1	NM_012224	EXON.16 EXON_16 CDS	[1423-1424 474-475]	subst_NONSYNONYMOUS[CCA:P CgA:R U]
HOMO	chr21	44818534	44818535	1 C/G	CTGCTGTGTG C CCATCTGCTG	novel	KRTAP10-4	NM_198687	EXON.1 EXON_1 CDS	[471-472 157-158]	subst_NONSYNONYMOUS[CCC:P gCC:A U]
HOMO	chr15	22472602	22472603	1 C/A	GGATGAGGAT C CGGTGCAGAT	novel	C15orf2	NM_018958	EXON.1 EXON_1 CDS	[495-496 165-166]	subst_NONSYNONYMOUS[CCG:P aCG:T c]
HOMO	chr2	234267010	234267011	1 C/T	TTTTTCCCAA C GAGTGGCCAA	novel	UGT1A6	NM_001072	EXON.1 EXON_2 CDS	[621-622 207-208]	subst_NONSYNONYMOUS[CGA:R tGA:* U]
HOMO	chrX	68666378	68666379	1 G/A	GGGGGCTCCC G CCTCAGCCCT	novel	FAM155B	NM_015686	EXON.3 EXON_3 CDS	[1273-1274 424-425]	subst_NONSYNONYMOUS[CGC:R CaC:H U]
HOMO	chr15	21441323	21441324	1 G/A	TTCTTCTTCC G GGTGGCCTTG	novel	MAGEL2	NM_019066	EXON.1 EXON_1 CDS	[2658-2659 886-887]	subst_NONSYNONYMOUS[CGG:R tGG:W c]
HOMO	chr2	218998704	218998705	1 C/T	CCTGAAGGGC C GGGCTGTGCA	novel	VIL1	NM_007127	EXON.4 EXON_4 CDS	[273-274 91-92]	subst_NONSYNONYMOUS[CGG:R tGG:W c]
HOMO	chr2	216988011	216988012	1 C/G	CCACAGTCCA C GTAGTCAAAT	novel	SMARCAL1	NM_001127207	EXON.3 EXON_4 CDS	[339-340 113-114]	subst_NONSYNONYMOUS[CGT:R gGT:G U]
HOMO	chr4	164613336	164613337	1 G/A	TCATTTGCAC G GCCCAGTTTA	novel	TKTL2	NM_032136	EXON.1 EXON_1 CDS	[999-1000 333-334]	subst_NONSYNONYMOUS[CGT:R tGT:C U]
HOMO	chr5	36206263	36206264	1 G/A	TGGATTCTCT G AATTTGCCCT	novel	SKP2	NM_005983	EXON.6 EXON_6 CDS	[732-733 244-245]	subst_NONSYNONYMOUS[GAA:E aAA:K c]
HOMO	chr5	79060601	79060602	1 A/C	TAACCTGGGA A ACCAATTCAA	novel	CMYA5	NM_153610	EXON.2 EXON_2 CDS	[257-258 85-86]	subst_NONSYNONYMOUS[GAA:E GAc:D c]
HOMO	chr4	164612484	164612485	1 C/T	GTTACGGCTG C TATAATGTGT	novel	TKTL2	NM_032136	EXON.1 EXON_1 CDS	[1851-1852 617-618]	subst_NONSYNONYMOUS[GCA:A aCA:T c]
HOMO	chr1	892007	892008	1 G/A	AAGCGGGGAC G CCGCCGCCAA	novel	PLEKHN1	NM_001160184	EXON.2 EXON_2 CDS	[144-145 48-49]	subst_NONSYNONYMOUS[GCC:A aCC:T c]
HOMO	chrX	119893190	119893191	1 C/G	GAACACGAAG C CAATGCCCGC	novel	CT47B1	NM_001145718	EXON.1 EXON_1 CDS	[361-362 120-121]	subst_NONSYNONYMOUS[GGC:G GcC:A c]
HOMO	chrX	71342892	71342893	1 C/T	CTTCCAAACC C CTTTGGTAAA	novel	ERCC6L	NM_017669	EXON.2 EXON_2 CDS	[2448-2449 816-817]	subst_NONSYNONYMOUS[GGG:G aGG:R U]
HOMO	chr2	215553515	215553516	1 C/A	AAATGGGGAC C CACAGCACCT	novel	ABCA12	NM_015657	EXON.23 EXON_32 CDS	[3721-3722 1240-1241]	subst_NONSYNONYMOUS[GGG:G GtG:V U]
HOMO	chrX	49066689	49066690	1 G/T	GAAATGATTG G GCCTATGCGG	novel	GAGE12J	NM_001098406	EXON.2 EXON_2 CDS	[73-74 24-25]	subst_NONSYNONYMOUS[GGG:G GtG:V U]
HOMO	chr2	219382543	219382544	1 G/A	AACTCCACAG G TGCTTTACAA	novel	CYP27A1	NM_000784	EXON.2 EXON_2 CDS	[255-256 85-86]	subst_NONSYNONYMOUS[GTG:V aTG:M c]
HOMO	chr22	14667561	14667562	1 G/C	TCCCCCTGCA G CAGGGGAAGC	novel	POTEH	NM_001136213	EXON.1 EXON_1 CDS	[323-324 107-108]	subst_NONSYNONYMOUS[TGC:C TGg:W U]
HOMO	chr10	38160035	38160036	1 T/C	ACCAATGGTA T TTAGTGTAGA	novel	ZNF248	NM_021045	EXON.6 EXON_6 3'UTR		
HOMO	chr10	46338336	46338337	1 T/G	GTCCACGCAG T ACTAAGAGTT	novel	FAM35B	NR_027632	EXON.3 EXON_3 5'UTR		
HOMO	chr1	2105815	2105816	1 A/G	AAGTGCGTGC A AAACACTCAA	novel	C1orf86	NM_001146310	EXON.8 EXON_11 3'UTR		
HOMO	chr13	20446120	20446120	1 -/C	CCCCCCCCCC - AAAGAATCCA	novel	LATS2	NM_014572	EXON.8 EXON_8 3'UTR		
HOMO	chr15	22781221	22781222	1 C/G	AAACTCTTAT C ATTTATTCTT	novel	PAR5	NR_022008	EXON.1 EXON_1 5'UTR		
HOMO	chr15	30695592	30695593	1 G/A	GACGGGGCCG G AAAGCAGCCG	novel	ARHGAP11A	NM_014783	EXON.1 EXON_1 5'UTR		
HOMO	chr1	884512	884513	1 C/T	CACCCCACTT C CGGCCCCAGA	novel	NOC2L	NM_015658	EXON.1 EXON_1 5'UTR		
HOMO	chr19	55510718	55510718	1 -/T	GCAGTGGGGG - GGGGTGGGGG	novel	KCNC3	NM_004977	EXON.5 EXON_5 3'UTR		
HOMO	chr19	63066170	63066171	1 C/G	CTCACTCTTT C ACCCAGGCTG	novel	ZNF587	NM_032828	EXON.3 EXON_3 3'UTR		
HOMO	chr19	63684064	63684065	1 C/T	CCCCAAAAGA C CTGGGTGCAA	novel	ZNF446	NM_017908	EXON.7 EXON_7 3'UTR		
HOMO	chr20	62055743	62055744	1 C/T	GCCCACGGAC C TTCCTGCTGG	novel	UCKL1AS	NR_027287	EXON.1 EXON_1 5'UTR		
HOMO	chr21	46494345	46494346	1 C/T	GGATGGAGAC C AGCCCACAGA	novel	MCM3APAS	NR_002776	EXON.3 EXON_3 5'UTR		
HOMO	chr2	233780095	233780096	1 C/A	CACAGTGAAG C CCTCAGTGAG	novel	INPP5D	NM_001017915	EXON.22 EXON_23 3'UTR		
HOMO	chr2	74182874	74182875	1 C/T	GGTGCTTTTG C CCTCATACCT	novel	TET3	NM_144993	EXON.9 EXON_9 3'UTR		
HOMO	chr3	5000545	5000545	1 -/TATGTG	TGTGTATGTG - CGTGTGCGTG	novel	BHLHE40	NM_003670	EXON.5 EXON_5 3'UTR		
HOMO	chr4	69085873	69085874	1 C/T	ATTTAAAACC C TCCATGCTGG	novel	UGT2B17	NM_001077	EXON.6 EXON_6 3'UTR		
HOMO	chr7	124173788	124173789	1 A/T	CAAACAAATA A ATCTGACCCA	novel	GPR37	NM_005302	EXON.2 EXON_2 3'UTR		
HOMO	chr7	72114228	72114229	1 A/T	ACCGCCAGAC A CCAACGGTCG	novel	STAG3L3	NM_001013739	EXON.1 EXON_1 5'UTR		
HOMO	chrX	100132316	100132317	1 C/T	ATCTTTGTAC C GAGATGCTGC	novel	ARL13A	NM_001012990	EXON.9 EXON_10 3'UTR		
HOMO	chrX	102817742	102817743	1 G/A	GTCTGTAGAC G CTCACAGGGC	novel	MORF4L2	NM_001142418	EXON.5 EXON_10 3'UTR		
HOMO	chrX	150663653	150663659	1 ATTGGC/-	CCTCTCCAGC ATTGGC CTTGGCCTTG	novel	CNGA2	NM_005140	EXON.7 EXON_7 3'UTR		
HOMO	chrX	47157989	47157990	1 C/G	ATCACCAGGG C ACACAGTGTG	novel	ZNF157	NM_003446	EXON.4 EXON_4 3'UTR		

Figure 6.Exome Sequencing Data for Family 1 and Family 2.(A) In family 1, it was identified that *ALOXE3* and *ABCA12* were present in the data, and had already been reported to be associated with Ichthyosis; therefore these genes were highlighted as potential candidates. (B) The *ABCA12* gene was also present in family 2, and it was then noticed that this particular mutation was identical to that of family 1, and therefore this was identified as a strong candidate gene.

4.1.3b Analysis of Sequence Data

Following the sequencing of *ABCA12* with the DNA from the individuals, the electropherograms were analysed to determine whether the mutation segregated in the families. It was found that the affected individuals all showed a homozygous change for the mutation, while all parents consistently showed that they were heterozygous carriers for the mutation. Furthermore, all controls screened for the mutation showed that they all had the wild-type genotype (figure 7).



(A) Reverse sequence data for the Proband in family 1 shows a homozygous change, and shows the parent to be a heterozygous carrier for the mutation. (B) The sequence data for family 2 shows also the heterozygous genotype of one of the parents. (C) Control DNA sample from Pakistan showed that the wildtype genotype was present. This was true of all controls examined.

Figure 7. Sequence Analysis of Ichthyosis Patients.

4.1.3c Autozygosity Mapping

In order to investigate whether the two families sharing the identical mutation in the *ABCA12* gene also shared a common haplotype, autozygosity mapping using microsatellite markers was carried out. The two families with Ichthyosis from Pakistan were analysed with microsatellite markers flanking *ABCA12*. The markers included D2S2944, D2S2382, D2S164, and D2S434.

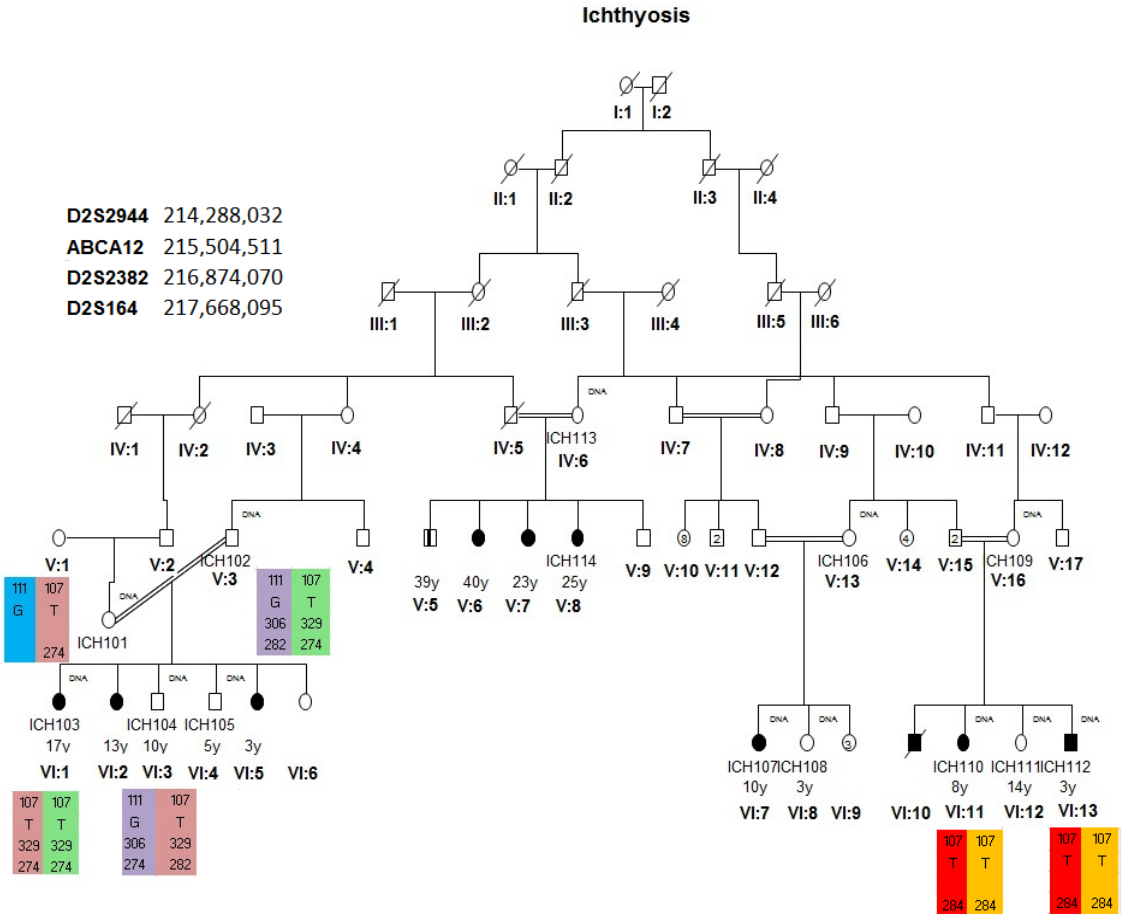


Figure 8.Autozygosity mapping in Family 1 to show linkage of the mutation within the family. The microsatellite markers used are shown in order in bold writing; the alleles for each individual tested are shown in the same order as the markers. The affected individuals show homozygosity for the mutation, and the parents clearly show that they are heterozygous for the mutation. It should be noted that individuals ICH110 and ICH112 show a different haplotype to the other affected individuals in the family. This is most likely due to a recombination event between the *ABCA12* and DSS164.

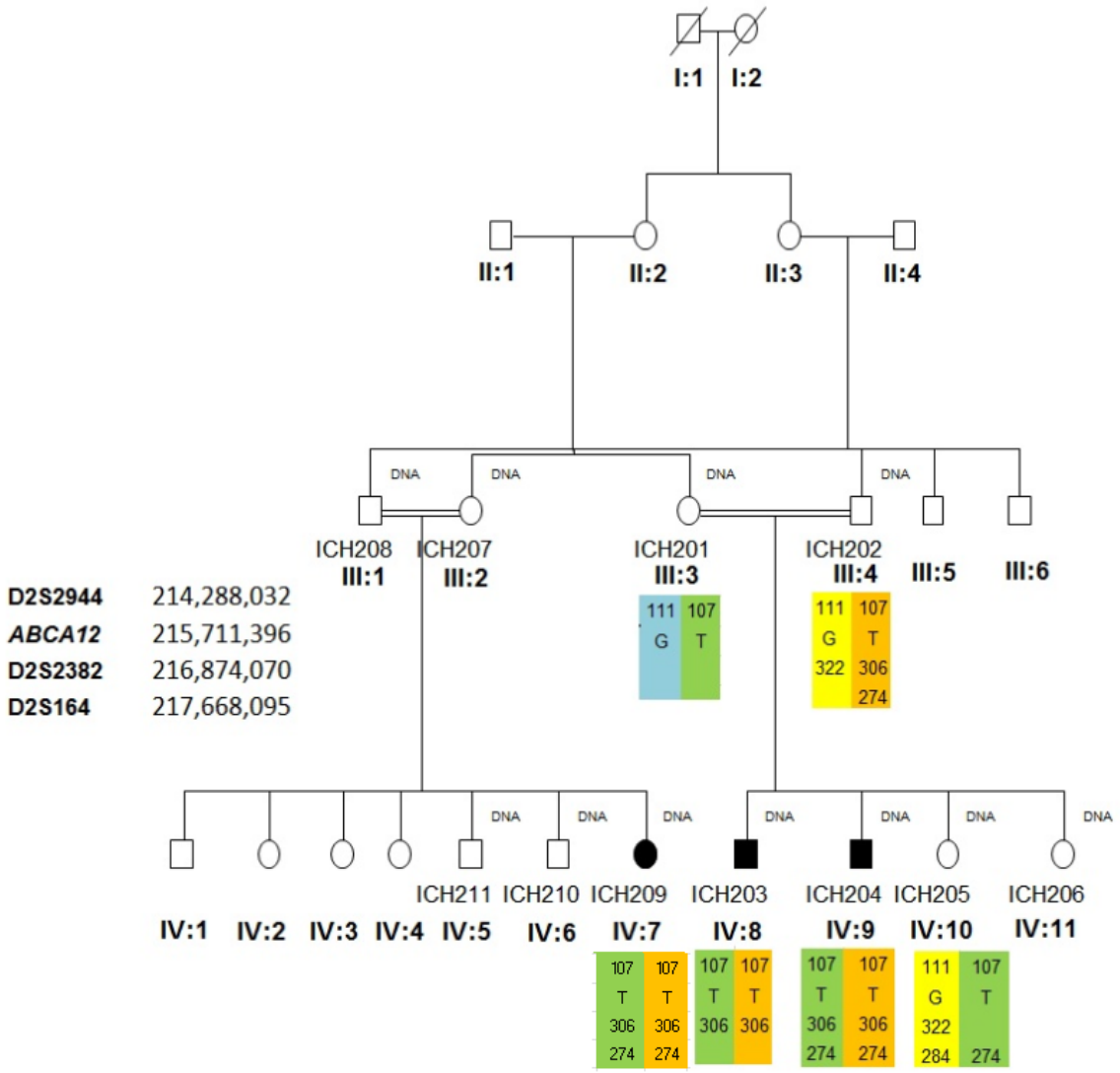


Figure 9:Autozygosity Mapping of Family 2. The microsatellite markers used are shown in bold writing; the alleles for each individual tested are shown in the same order as the markers. Upon observation of this map, and the map of family 1, it seems that the mutation segregates, and is clearly linked between affected individuals.

Marker	Physical Location (Mb)	Family 1 ICH103	Family 1 ICH110	Family 2 ICH203	Family 2 ICH204	Family 2 ICH209
D2S2944	21.4	107	107	107	107	107
ABCA12	21.5					
D2S2382	21.6	329	--	306	306	306
D2S164	21.7	274	284	--	274	274

Table 3.Haplotype Analysis in Pakistani Patients Homozygous for the 1559 G>V mutation

4.1.4 Discussion of Ichthyosis and *ABCA12*

To date, several genes have been shown to be associated with Ichthyosis, including *ALOXE3*, *ALOX12B*, *TGM1*, *CYP4F22*, *NIPAL4* and *ABCA12* (2,4). Mutations in the *ABCA12* gene have been identified in various cases of all three subtypes of Autosomal Recessive Congenital Ichthyosis. The resulting defect in the ABCA12 protein is thought to play a causative role in the development of the disease in these particular cases.

4.1.4a*ABCA12* and the ABCA12 Protein

ATP-binding cassette sub-family A member 12 is a protein that is encoded by the *ABCA12* gene. *ABCA12* is part of the ATP-binding cassette family, which are involved in the regulation of the transport of certain molecules across cell membranes. *ABCA12* has been shown to be expressed in various cells and tissues, including skin cells, placenta, lung, stomach and liver (17). Interestingly, normal function of *ABCA12* has been reported to be essential for the development of the epidermis, and therefore mutations in this gene may result in skin disorders, such as Ichthyosis (17). The function of *ABCA12*remains relatively elusive; however it is thought that its main role is involved with the transportation of lipids in cells. By taking all this information in to consideration, the ways in which mutations in *ABCA12* may be involved in the pathogenesis of Ichthyosis may be approached.

4.1.4b Predicted Effect of Variant on Protein

ABCA12 is known to be associated with Ichthyosis, and many different mutations within this gene have been documented to be the causative factor in numerous cases of Ichthyosis (figure 10). By analysing where on the gene the mutation occurs, the resulting effect on the ABCA12 protein may be predicted (figure 11), and in this way, the mechanism and genetic basis of the disease may begin to be elucidated.

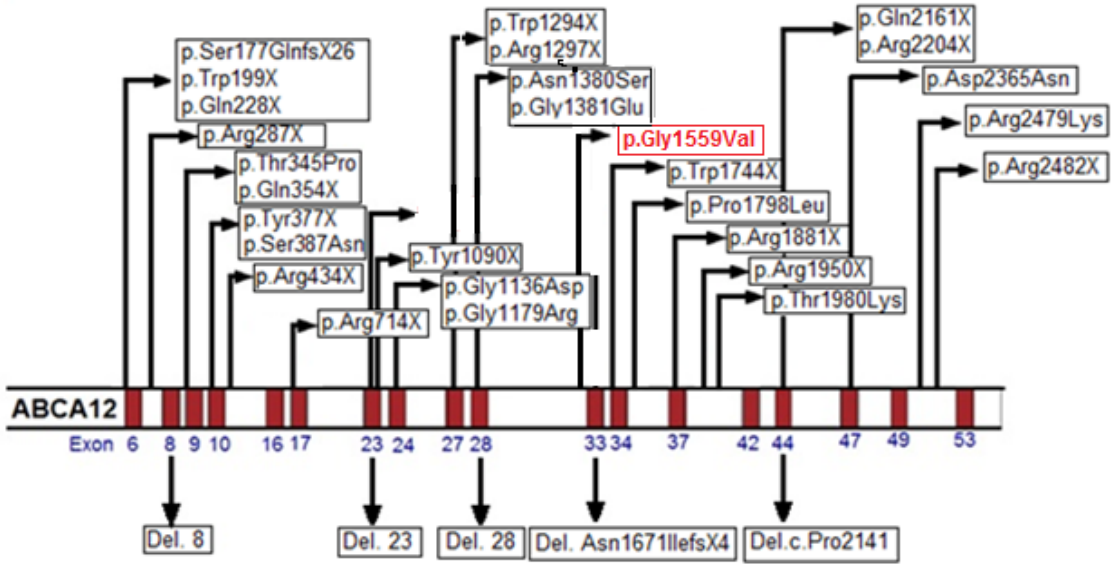


Figure 10. ABCA12 Gene showing the Locations of Several Mutations already Identified, and location of Novel Missense Change (Modified from AC Thomas et al (2006) The exons in the gene are shown in red. Missense changes can be found above the gene, while deletions can be found underneath the gene. The novel missense change identified in the two families is marked in red at exon 32.

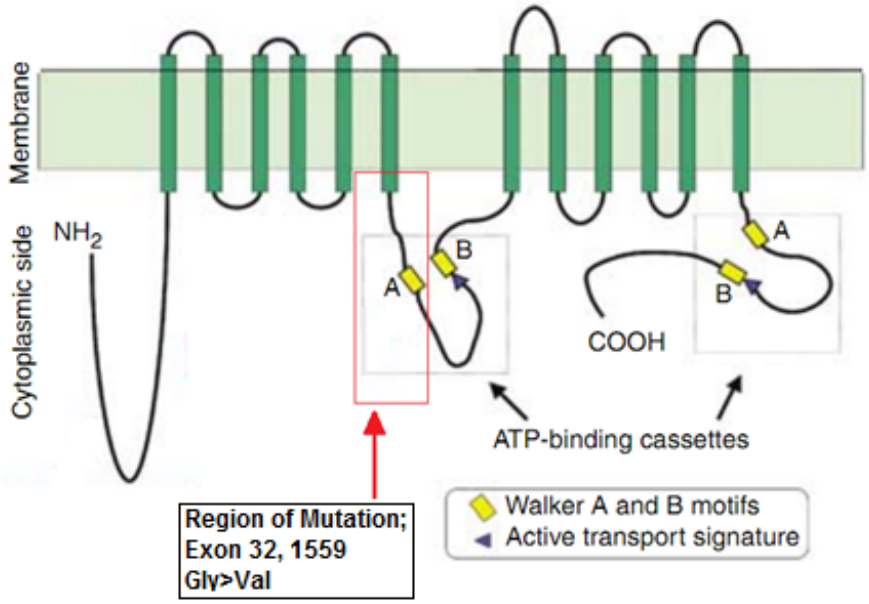


Figure 11. Schematic representation of the ABCA12 protein, and predicted location of Mutation; adapted from Akiyama et al (2006).

In order to assess the predicted effect the mutation will have on the function of the protein, the program called Polymorphism Phenotyping V.2 was used (PolyPhen, Harvard University (16). It takes into account how conserved the gene is, and the amino acid change in order to predict the likely effect on the function of the protein.*ABCA12* is known to be highly conserved among all mammals, zebra fish and chicken, suggesting that it is assential (Figure 12). It is therefore predicted that a mutation in exon 32, at 1559 Gly>Val will be 'probably damaging' (figure 13).

H.sapiens	1535	HHLDEAEVLSDRIAFLEQGGLRCCGSPFYLKEAFGDGYHLTLTKKKSPNL	1584
P.troglodytes	1535	HHLDEAEVLSDRIAFLEQGGLRCCGSPFYLKEAFGDGYHLTLTKKKSPNL	1584
C.lupus	1534	HHLDEAEVLSDRIAFLEQGGLRCCGSPFYLKEAFGDGYHLTLTKKKSPNL	1583
B.taurus	1473	HHLDEAEVLSDRIAFLEQGGLRCCGSPFYLKEAFGDGYHLTLTKKKSPNL	1522
M.musculus	1532	HHLDEAEVLSDRIAFLEQGGLRCCGSPFYLKEAFGDGYHLTLTKKKSPNL	1581
G.gallus	2763	HHLDEAEVLSDRIAFLEHGGLKCCGSPFYLKETFGDGYHLTLTKKKSVFL	2812
D.rerio	2586	HHLDEAEVLSDRIAFLERGGLKCCGSPFYLKDKLAKGYNLTLTKKVETA-	2634

Figure 12. Multiple sequence alignment of *ABCA12.* This sequence alignment of *ABCA12* between species shows that the amino acid present at 1559 is highly conserved among all mammals, zebrafish and chicken.

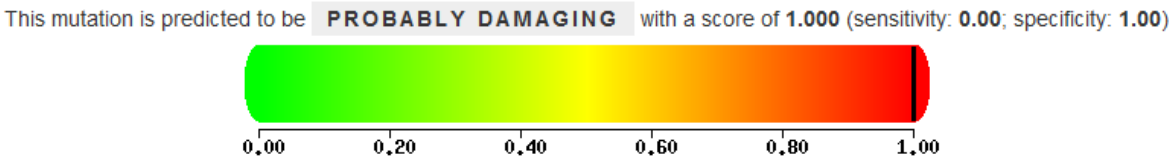


Figure 13. Predicted Effect of 1559 Gly>Val mutation on ABCA12 protein by PolyPhen V.2 Due to the fact that this gene is conserved among all species, and that the amino acid change is relatively significant, the predicted outcome is that this mutation will most likely have a damaging outcome on the ABCA12 protein.

By analysing the results outlined in this study, it is clear that the mutation identified through whole exome sequencing for families 1 and 2 with Ichthyosis, is indeed highly likely to be the causative mutation for the disease. There is some evidence that implies that the mutation in *ABCA12* in these two families has originated from a common ancestor, and been

transmitted to the affected individuals as a result of consanguineous matings. Though both families shared a common allele at D2S2944, no common allele was apparent at the other two microsatellite loci that mapped to the other side of *ABCA12*. Indeed in Family 1 there is evidence of a recombination between *ABCA12*and D2S164 (different alleles in ICH103 and ICH110 at D2S164). Thus it cannot be excluded that the two families, although previously thought to be unrelated and were observed to share a common conserved haplotype at D2S2944, might be distantly related in some way.

4.1.5 Future Work

ABCA12 has been identified to be the causative mutation in two families from Pakistan. It is thought that these families could be distantly related after careful analysis of linkage data collected from the individuals. The exact function of *ABCA12* and the ways in which it contributes to Ichthyosis are still elusive, and therefore it may be beneficial to carry out functional assays on the gene. Furthermore, as this is a novel mutation identified in *ABCA12* it would be interesting to screen more suffers of Ichthyosis, particularly from consanguineous backgrounds in order to determine if this mutation exists in other families.

4.2 Cerebral palsy

Cerebral palsy (CP) is a heterogeneous group of neurological disorders, characterised by non-progressive abnormalities in posture and motor function as a result of a defect in the development of the nervous system (21). It is estimated that 1 in approximately 250 to 1000 neonates develops CP, therefore making it one of the most frequently occurring congenital disabilities (21). Non-progressive forms of symmetrical, spastic CP have been identified,

which show a Mendelian autosomal recessive pattern of inheritance (21). CP can occur in an autosomal dominant, X-linked or autosomal recessive form. There are several different subtypes of CP, which are grouped based on their phenotypes (21). The groups are made up of Spastic CP, Dyskinetic CP which includes the choreoathetotic and dystonic subtypes, and Ataxic CP which includes congenital ataxia and ataxia diplegia subtypes (21,22). Of the three major groups, Spastic CP is the most prevalent, affecting approximately 70% of CP patients (21).

Individuals with Spastic CP often display muscular hypertonicity, and stiffness of the limbs and extremities (21).

A previous study described the case of two consanguineous families from the Mirpur region of Pakistan with non-progressive, autosomal recessive, symmetrical Spastic Cerebral palsy. The study described the mapping of a recessive spastic CP locus to a 5cM chromosomal region located at 2q24-31.1. The narrow region suggested that a mutation in GAD1 could potentially underlie the CP within these families, however it was later discovered that this was not the causative variant (21). This study therefore aimed to usewhole exome sequencing data to attempt to identify the causative variant responsible for their disease. Successful identification of the causal variant may provide insight into the possible mechanisms of pathogenesis, in addition to furthering knowledge and understanding of this disorder.

4.2.1 Methods

DNA was provided by two related families from the Mirpur region of Pakistan. The pedigrees of the families can be found in figure 14.

4.3.1a Patient Information and Pedigrees

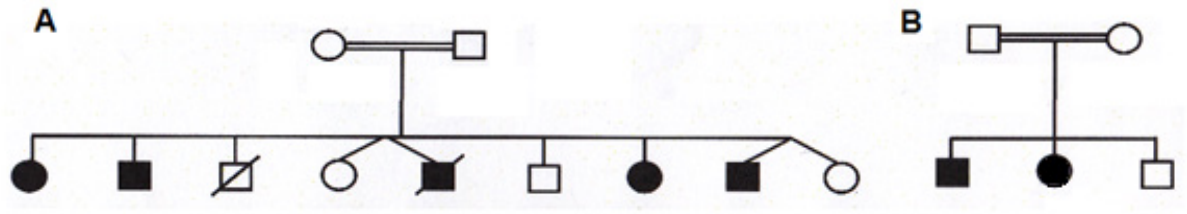


Figure 14. Pedigrees of the two families from Pakistan with CP(A) Family 1 from Pakistan with 5 affected individuals. (B) Family 2, related to family 1 with 2 affected individuals.

4.3.1b Analysis of Exome Sequencing Data

HOMO	chr1	45,565,933	45,565,934	1 T/C	TCCCCCACAC T TTTGCGCTGG	novel	HPDL	NM_032756	EXON.1 EXON_1 CDS	[526-527 175-176]	subst_NONSYNONYMOUS[	175-176
HOMO	chr1	57,124,291	57,124,292	1 G/T	CATTAATGGA G CTTCCAGATC	novel	C8A	NM_000562	EXON.7 EXON_7 CDS	[959-960 319-320]	subst_NONSYNONYMOUS[	319-320
HOMO	chr1	93,030,544	93,030,545	1 G/A	ACCATGATAA G CAGAGAAGAC	novel	EVI5	NM_005665	EXON.1 EXON_1 5'UTR			
HOMO	chr1	110,094,464	110,094,465	1 G/-	AATTCTCGGG G CTCTAATGGG	novel	EPS8L3	NM_024526	EXON.19 EXON_21 3'UTR			
HOMO	chr10	18,127,587	18,127,588	1 C/G	TACCTCTCCC C CTGCCACCCT	novel	FAM23A	NM_001098844	EXON.8 EXON_8 3'UTR			
HOMO	chr10	48,010,785	48,010,786	1 T/C	CAAGAGGACC T TGGCCATGTC	novel	RBP3	NM_002900	EXON.1 EXON_1 CDS	[97-98 32-33]	subst_NONSYNONYMOUS[	32-33
HOMO	chr10	102,662,943	102,662,944	1 C/T	ATCTGAGACC C GGTAGTACCG	novel	FAM178A	NR_024245	EXON.1 EXON_1 5'UTR			
HOMO	chr10	106,855,210	106,855,211	1 G/A	CCTTTTGCCC G CTCCATTGAC	novel	SORCS3	NM_014978	EXON.7 EXON_7 CDS	[1159-1160 386-387]	subst_NONSYNONYMOUS[	386-387
HOMO	chr10	120,344,091	120,344,091	1 -/GGCGCT	CGCTGGCGCT - CCTGGGAGCC	novel	PRLHR	NM_004248	EXON.2 EXON_2 CDS	[655-655 218-219]	AA_INSERTION[370 372]	218-219
HOMO	chr11	5,036,705	5,036,706	1 C/T	CACATGTGAA C CACATGTGCT	novel	OR52E2	NM_001005164	EXON.1 EXON_1 CDS	[727-728 242-243]	subst_NONSYNONYMOUS[	242-243
HOMO	chr11	5,418,981	5,418,982	1 C/-	GTATGCCTGA C TCCATGAAGG	novel	OR51I1	NM_001005288	EXON.1 EXON_1 CDS	[338-339 112-113]	FRAMESHIFT[314 119]	112-315
HOMO	chr11	6,697,300	6,697,301	1 C/T	CAATCGTGAT C TGCTCCAGGT	novel	GVIN1	NR_003945	EXON.1 EXON_1 3'UTR			
HOMO	chr12	114,894,353	114,894,354	1 C/T	AGATTCCACA C ATCCGGCACA	novel	MED13L	NM_015335	EXON.26 EXON_26 CDS	[5801-5802 1933-1934]	subst_NONSYNONYMOUS[	1933-1934
HOMO	chr15	63,157,085	63,157,086	1 C/G	CCTGCCGCAG C CGGCCGCCGG	novel	KBTBD13	NM_001101362	EXON.1 EXON_1 CDS	[879-880 293-294]	subst_NONSYNONYMOUS[	293-294
HOMO	chr18	19,997,048	19,997,048	1 -/T	TTTTTTTTTT - AAAAGATAAG	novel	OSBPL1A	NM_018030	EXON.14 EXON_29 3'UTR			
HOMO	chr2	21,104,235	21,104,236	1 G/A	AAATCCAAAG G CAGTGAGGGT	novel	APOB	NM_000384	EXON.14 EXON_14 CDS	[2035-2036 678-679]	subst_NONSYNONYMOUS[	678-679
HOMO	chr2	28,469,331	28,469,337	1 AGAGAG/-	AGAGAGAGAG AGAGAG GGAGAG(	novel	FOSL2	NM_005253	EXON.1 EXON_1 5'UTR			
HOMO	chr2	79,218,480	79,218,483	1 ACA/-	TCACTGAACA ACA CAAGACAACC	novel	REG1P	NR_002714	EXON.1 EXON_1 3'UTR			
HOMO	chr2	79,218,668	79,218,669	1 A/T	AGCCATGCTG A GCAGCAGTGA	novel	REG1P	NR_002714	EXON.1 EXON_1 3'UTR			
HOMO	chr2	94,845,183	94,845,184	1 T/C	TTTTCTTGCT T TCTTTTTCCT	novel	ANKRD20B	NR_003366	EXON.13 EXON_13 3'UTR			
HOMO	chr2	113,974,544	113,974,545	1 C/T	ACATCGCTGG C TGCCCCTTTG	novel	FOXD4L1	NM_012184	EXON.1 EXON_1 3'UTR			
HOMO	chr2	171,383,381	171,383,382	1 C/G	TCCGCAACCT C CTCGAACGCG	novel	GAD1	NM_000817	EXON.2 EXON_2 CDS	[34-35 11-12]	subst_NONSYNONYMOUS[	11-Dec
HOMO	chr2	179,042,686	179,042,687	1 G/A	TCAATGCTCC G TGGTCCTTTG	novel	FKBP7	NM_001135212	EXON.3 EXON_4 CDS	[441-442 147-148]	subst_NONSYNONYMOUS[	147-148
HOMO	chr2	189,651,524	189,651,525	1 G/A	CTCTCTCCTG G CATTCCCCTC	novel	COL5A2	NM_000393	EXON.16 EXON_16 CDS	[1020-1021 340-341]	subst_NONSYNONYMOUS[	340-341
HOMO	chr2	190,039,484	190,039,485	1 C/T	GAACAGCCCA C CTTGGTTACA	novel	WDR75	NM_032168	EXON.13 EXON_13 CDS	[1378-1379 459-460]	subst_NONSYNONYMOUS[	459-460
HOMO	chr22	29,215,414	29,215,415	1 C/T	TTAGTAGAGA C AGGGTTTCAC	novel	SEC14L4	NM_001161368	EXON.12 EXON_12 3'UTR			
HOMO	chr22	29,305,833	29,305,834	1 G/A	AGGTGTGGGG G CAGCTGCACC	novel	PES1	NM_014303	EXON.12 EXON_12 CDS	[1257-1258 419-420]	subst_NONSYNONYMOUS[	419-420
HOMO	chr22	31,211,148	31,211,149	1 C/T	GAGGGCAGCT C CGCTACTCTC	novel	FBXO7	NM_001033024	EXON.4 EXON_6 CDS	[502-503 167-168]	subst_NONSYNONYMOUS[	167-168
HOMO	chr3	5,000,545	5,000,545	1 -/TATGTG	TGTGTATGTG - CGTGTGCGTG	novel	BHLHE40	NM_003670	EXON.5 EXON_5 3'UTR			
HOMO	chr3	197,196,982	197,196,983	1 C/T	AAATGCAGCT C GCAAAGCCGG	novel	SDHAP1	NR_003264	EXON.3 EXON_3 3'UTR			
HOMO	chr3	197,933,768	197,933,769	1 G/A	ATTCCTGCAG G ACTTTATGTG	novel	PIGX	NM_001166304	EXON.3 EXON_3 CDS	[262-263 87-88]	subst_NONSYNONYMOUS[	87-88
HOMO	chr5	36,085,088	36,085,089	1 T/C	CCGAATGAAG T GGAAAGAATG	novel	UGT3A2	NM_001168316	EXON.3 EXON_4 CDS	[399-400 133-134]	subst_NONSYNONYMOUS[	133-134
HOMO	chr6	29,906,560	29,906,560	1 -/ATTTGTTC	GAGTGGCAAG - TCCCTTTGTG	novel	HLA-G	NM_002127	EXON.8 EXON_8 3'UTR			
HOMO	chr6	32,654,658	32,654,659	1 G/A	CATCAATGCT G GGACTTCAGG	novel	HLA-DRB1	NM_002124	EXON.6 EXON_6 3'UTR			
HOMO	chr7	71,976,064	71,976,065	1 A/-	AAAAAAAAAA A GACTGTAGGA	novel	SPDYE7P	NR_003666	EXON.1 EXON_1 3'UTR			
HOMO	chr7	72,056,917	72,056,918	1 A/G	GAGGACAGGC A TCTTCCTTTC	novel	NSUN5P2	NR_033323	EXON.11 EXON_11 3'UTR			
HOMO	chr7	74,953,165	74,953,166	1 C/G	TCGCGAGGTC C CCTCCTGTCC	novel	POM121C	NM_001099415	EXON.1 EXON_1 5'UTR			
HOMO	chr7	100,424,332	100,424,333	1 C/A	AAAACTCTCA C CTGCCCGCTC	novel	MUC12	NM_001164462	EXON.2 EXON_2 CDS	[3768-3769 1256-1257]	subst_NONSYNONYMOUS[	1256-1257
HOMO	chr7	100,424,333	100,424,334	1 C/A	AAACTCTCAC C TGCCCGCTCC	novel	MUC12	NM_001164462	EXON.2 EXON_2 CDS	[3769-3770 1256-1257]	subst_NONSYNONYMOUS[	1256-1257
HOMO	chr9	34,554,640	34,554,641	1 C/T	CCAGGAGTCA C GGTGGAAGCA	novel	CNTFR	NM_001842	EXON.3 EXON_4 CDS	[274-275 91-92]	subst_NONSYNONYMOUS[	91-92
HOMO	chr9	34,600,735	34,600,736	1 C/T	AGTCACAACG C TTCTCAGAAA	novel	C9orf23	NM_148178	EXON.2 EXON_2 3'UTR			
HOMO	chr9	35,032,395	35,032,396	1 G/A	ATTCTTATTC G TGGTTTGGCA	novel	C9orf131	NM_203299	EXON.1 EXON_2 CDS	[144-145 48-49]	subst_NONSYNONYMOUS[	48-49
HOMO	chr9	35,896,618	35,896,618	1 -/ACCATCCC	CACCATCCCC - GCCACGCTCG	novel	HRCT1	NM_001039792	EXON.1 EXON_1 CDS	[334-334 111-112]	AA_INSERTION[115 118]	111-112

Figure 16. Exome Sequencing Data for Proband of related families from Pakistan with CP. Previous studies had shown the variant to be located on chromosome 1, in the region of 20,000,000 to 58,000,000. Therefore *HPDL* and *C8A* were identified and selected as the primary candidates for the causative mutation.

4.2.1c Mutation Analysis of Candidate Genes

Primers were designed as outlined in section 4.1.2 d for *C8A* and *HPDL* ensuring that the primers were designed to flank either side of the coding exon harbouring the mutation. These primers were used to amplify regions of interest of *C8A* and *HPDL* from DNA of the family (see methods section 3.1.3) Data was then analysed to determine which variant, if any, segregated within the family.

4.2.2 Cerebral palsy Results

4.2.2a Analysis of Exome Sequencing Data

Whole exome sequencing data was provided for an individual from family 1. Through process of elimination, variants which were present in the dbSNP, or 1000 genomes project were excluded. The novel and non-synonymous changes were identified, and through this, the number of candidate variants was reduced from the original 29,431 to 59 variants. Information collected from previous studies on this family was provided by Dr. Neil Morgan; the studies suggested that the causal variant was within the region of chromosome 1 and in the range of 20,000 to 50,000.

4.2.2a Analysis of Sequence Data

The electropherograms received for the *C8A* data were analysed, however upon close inspection, it was found that there was a homozygous change for the mutation present in an unaffected individual. This implied that the mutation did not have a causative role in the development of the disorder. This gene was therefore ruled out as a candidate variant.

The electropherograms were then analysed for *HPDL*, and it was found that this variant segregated within both families (Figure 16).

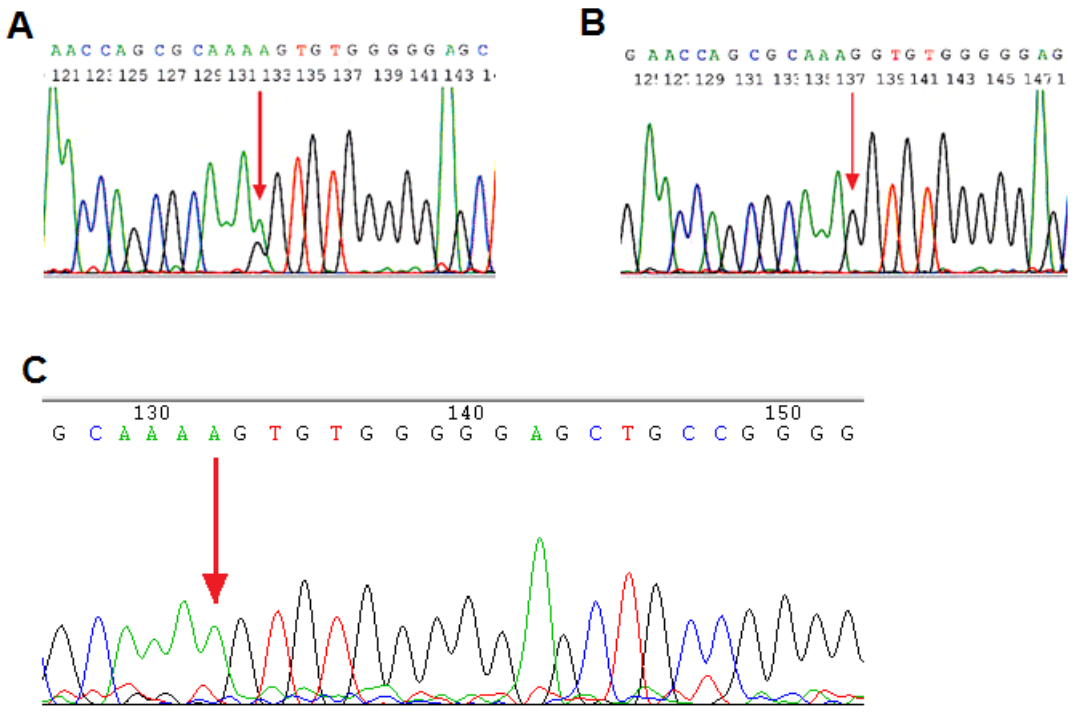


Figure 15.Electropherogram Data for *HPDL* (A) Electropherogram data for parent of Proband; the data clearly shows the parent to be a heterozygous carrier for the mutation (Leu176Pro A>G). (B) Affected individual from family 1, shows a homozygous genotype for the mutation. (C) Control DNA sample screened for the mutation, shows wild type genotype. The electropherogram data confirms that *HPDL* segregates in the two families, and that it is not present in the general population.

4.2.3 Cerebral palsy Discussion

4.2.3a *HPDL* and its associated Proteins

The*HPDL* variant segregated in both families, suggesting that it may be the causative mutation. *HPDL*, also known as 4-hydroxyphenylpyruvate dioxygenase-like, is made up of 1 exon, however the understanding of its functions remain elusive. It is part of the HPD gene family which are known to be involved in the formation of an enzyme known as 4-hydroxyphenylpyruvate dioxygenase. This enzyme is found in abundance in the liver and

kidneys, and it is thought to be involved in the breakdown of tyrosine (23). The precise function of HPDL however remains unknown, and therefore for future work, it would be important to carry out protein assays in order to fully elucidate the function of HPDL in order to determine whether *HPDL* is in-fact the causative mutation, and if so, to understand the role it may play in the occurrence of CP.

4.2.3b Predicted Effect on Protein

The effect that a mutation at Leu176Pro in *HPDL* may have on the resulting protein can be predicted using the program PolyPhen, as was done with *ABCA12.* The amino acid at 176 in *HPDL* is moderately conserved among species (figure 16), and this will be taken into account when predicting the consequences of a mutation at this location on the protein encoded.

H.sapiens	139	RGPFLPGFRPVSSA-PGPGW----VSRVDHLTLACTPGSSPTLLRWFHDC	183
P.troglodytes	139	RGPFLPGFRPVSSA-PGPGW----VSRVDHLTLACTPGSSPTLLRWFHDC	183
C.lupus	139	RGPFLPGFQSVPSA-AGPGW----VSHVDHLTLACTPGSSPTLMRWFRHC	183
B.taurus	139	GGPFLPGFSPVPST-PGPGW----FSHVDHLTLACNPSSSPKLLHWFHDC	183
M.musculus	139	RGSFLPGFRPLPCT-PGPGW----VSHVDHLTLACTSGSSPMLMRWFHDC	183
R.norvecicus	139	RGSFLPGFRPLPCT-PGPGW----VSHVDHLTLACTSGSSPTLMRWFHNC	183
G.gallus	144	RGPFLPGFQPIRGAPPGDGAE---VSHFDHITYVCPRGGTQAALEWYRRC	190
D.rerio	151	QGPFLPGFREVSSGCQGADRSRCPITHFDHITYACPRSSTVHITNWYRRN	200

Figure 16. Multiple Sequence Alignment for *HPDL* This multiple alignment shows that *HPDL* is moderately conserved among species. It does not appear to be conserved in B.taurus, or D.rerio.

The predicted effects of the variant on the resulting proteinfrom PolyPhen suggested that a Leu176Pro variation in *HPDL*would be‘possibly damaging’. This suggests a possibility that the variant would be damaging, but it cannot be used as firm evidence.

4.2.3c Conclusions about *HPDL* and CP

The variant in *HPDL* was found to segregate within the two families from Pakistan with CP. It is difficult to say with confidence however that this is without doubt the causative variant. Although this gene has not been documented before, and it was not present in the controls screened, or in the 1000 genomes database, this is still not sufficient evidence. For example, previous studies on this family suggested that the causative variant lied within a gene known as GAD1; however after investing much time and effort into this gene, it was found that GAD1, despite being a novel variant, and not present in the 1000 genomes database, was not responsible for the disease (21). This highlights some of the difficulties encountered with not only identifying a potential candidate variant, but to prove with 100% confidence that this variant is in fact the causative factor.

4.2.4 Future Work

In order to fully determine with confidence whether this variant is the causative factor for the CP in these two families, further studies investigating the gene should be carried out. For example, the function of the protein encoded by HPDL should be investigated in order to determine if it has any links to the symptoms of the disease, or the pathogenesis of the disease. Furthermore, assays should be carried out in order to determine if the variant has any effect on the stability of the protein *in vivo;* information gathered from procedures such as these may give further insight into the ways in which this variant affects the individual, and the consequences associated with it. In addition, more individuals with similar phenotypes should be examined in order to determine if any of them share the same variant. Successful identification of the gene, and greater knowledge about the protein encoded may give opportunities to improve diagnosis of patients, or to develop genetic counselling in order to give patients an idea of the risks associated with having an affected child.

5.0 Discussion

Recent developments in next generation sequencing technologies have revolutionised the way in which both rare and common diseases are currently investigated. The advances in abilities to cost-effectively, sequence numerous samples with high-throughput, in parallel, have facilitated researchers to identify disease-causing variants, and map their patterns of inheritance (10).

Whole exome sequencing has enabled researchers to focus on critical parts of the genome, thereby permitting deeper sequencing of targeted regions in a greater number of samples. This reduces redundancy of information which may arise as a result of whole genome sequencing, as protein-coding genes, although only constituting approximately 1% of the genome, are known to contain close to 85% of causal variants with large consequences on disease-related phenotypes (10). For this reason, the protein-coding region is usually the first location to be investigated when searching for causal variants associated with disease; therefore whole exome sequencing is a much more appropriate technique to employ than whole genome sequencing.

5.1Whole Exome Sequencing to Identify Disease Genes

Advances in next generation sequencing have revolutionised the ways in which monogenic disorders are approached. For example, single nucleotide polymorphism arrays are able to rapidly ascertain areas of homozygosity within the genome, and this information can be used to facilitate and reduce time required for linkage analysis which will highlight regions of interest (10). Use of linkage analysis data in consanguineous families can help to identify variants with a common ancestor, which are likely to be the causative mutation for a

particular disease. For example, the mutation found in *ABCA12* which was unexpectedly identical between the two Pakistani families with Ichthyosis, gave strong evidence to suggest that the two families were distantly related in some way.

5.2 Limitations of Exome Sequencing

These approaches while possessing many advantageous qualities with regard to throughput are not without their limitations. These processes can often be rather expensive, and take time, and the success rate in identifying the causal variant is not always high. The main problem associated with these methods, is that a large sample size is required to provide reliable data; and due to the rarity of these particular diseases, this is often very challenging. For this reason, consanguineous families often provide a solid foundation for investigating recessive disease, due to the larger sample sizes with common, shared variants. By combining exome sequencing technologies and linkage data with consanguineous families, the task of distinguishing the links between causative variants and disease may be more achievable.

The majority of Mendelian disorders are known to occur as a result of exonic or splice-site variants which result in an alteration in the amino acid sequence of a particular gene (10). In such cases, if sample sizes are of sufficient size, the identification of the causal variant is usually achievable; however limitations of exome sequencing become apparent when the causative mutation is located within an intron or promoter region of a gene. In these cases, the variant would not appear in the exome sequencing data, and would be overlooked as a possibility.

Another frustration with exome sequencing can occur when the causal variant is identified successfully, however it has already been documented in literature surrounding the particular

disease. The international collaboration of the 1000 genomes project is advancing, and the creation of large, public databases of known variants should improve this problem, and save valuable time, effort and money.

Finally, as in the case of *HPDL* identified for the families with Cerebral palsy in this study, on occasion, a variant which is believed to play a causative role in the disease may be discovered; however due to the fact that it has not been documented before, the challenges involved in proving that this is the correct gene are extensive. In order to provide strong evidence, more samples with similar phenotypes are required to determine if any share the mutation. In addition, in order to give light to the mechanisms of pathogenicity of the mutation, functional studies of the protein must be carried out. This can prove to be difficult, particularly when the knowledge surrounding the gene is elusive. Computer predictions, such as PolyPhen,may be used to predict the effect the variant may have on the protein encoded, however these results are not as reliable as an *in vivo* experiment may provide.

5.3 Establishing the Genotype-Phenotype Relationship

One further difficulty in the challenge to identify and understand the role of a causal variant in disease is to establish the relationship between the genotype and the resulting phenotype. Locus heterogeneity for example, is a concept which confuses matters; as variants in unrelated gene loci may contribute to cause a single disease. Furthermore, in some cases, a mutation in one gene may contribute to multiple phenotypes, while certain phenotypes can also be a result of a combination of multiple genes. For this reason, when a causal variant is identified, it can often be an extremely challenging task to gather sufficient evidence to conclude with confidence that the variant is the leading cause of the disease; particularly when the variant is novel, as in the case of *HPDL*.

5.4 Future Work and Prospects

The fact that the processes involved in identifying and confirming a variant as the cause of a disease requires so much research gives a lot of prospects for future work. In the case of *ABCA12*, protein functional assays can be carried out in order to determine the effects on the protein caused by the mutation. *In Vivo* studies can be carried out, to establish whether this mutation may cause Ichthyosis in animal models in order to provide further evidence that it is a causative factor. Furthermore, protein functional assays may provide insight into the pathogenicity of the mutation and elucidate the role played by the variant in the occurrence of Ichthyosis in the two families discussed. By gathering additional samples with similar phenotypes, it can also be investigated as to whether further individuals hold this mutation.

With regard to *HPDL*, a lot of work remains in order to fully determine whether this mutation is the causative factor. Due to the fact that this is a novel mutation, and has not been described in CP before, combined with the fact that there is not much information surrounding the gene, functional studies on the protein encoded will be more difficult. Research must first be carried out in order to establish gene function prior to investigating the effects the mutation may have on protein function. In addition, more samples may be required with similar phenotypes in order to determine if further individuals possess the same mutation. Finding more cases in which *HPDL* is suspected to cause CP will strengthen the evidence that *HPDL* is the causative factor in these cases.

In conclusion, it is clear that a lot of work and effort is involved in order to identify causal variants through whole exome sequencing; however if successful, the prospects and wealth of knowledge that stand to be gained are large. There have been many successes gained from investigating disease through exome sequencing. Successful identification of a causal variant

for such diseases may provide the opportunities to develop new therapies, or improve diagnostic methods. In addition, useful results may facilitate screening of families, which may give informative information regarding the likelihood of having affected offspring. With continued research into this field, more genes will be identified and examined, and this may contribute greatly to the field of genomics and the ways in which we treat genetic diseases.

References

1. Bittles AH, (2002) Endogamy, consanguinity and community genetics. **Journal of Genetics**. Vol. 81, No.3, Dec. 2002.

2. Bittles AH, (2001) Consanguinity and its relevance to clinical genetics. **Clinical Genetics** 2001: 60: 89-98.

3. Bittles AH, (2010) Consanguinity, human evolution, and complex diseases. **PNAS.**doi:10.1073/pnas.0906079106

4. Alfonso-Sanchez MA, Pena JA. (2005). Effects of Consanguinity on Pre-reproductive Mortality: Does Deomgraphic Transition Matter?.**American Journal of Human Biology.** 17:773-786.

5. Bittles AH, (2007) Consanguineous Marriages and Childhood Health. **Developmental Medicine and Child Neurology.** 45: 571–576

6. Hamamy H, Antonarakis SE, Luca L, Sforza C, Temtamy S, Romeo G, Ten Kate LP, Bennett RL, Shaw A, Megarbane A, Duikn CV, Bathija H, Fokstuen S, Engel E, Zlotogora J, Dermitzakis E, Bottani A, Dahoun S, Morris MA, Arsenault S, Aglan M, Ajaz M, Alkalamchi A, Alnageb D, Alwasiyah MK, Anwer N, Awwad R, Bonnefin M, Corry P, Gwanmesia L, Karbani GA, Mostafavi M, Pippucci T, Boscardin ER, Reversade B, Sharif SM, Teeuw ME, Bittles AH. (2011) Consanguineous marriages, pearls and perils: Geneva International Consanguinity Workshop Report. DOI: 10.1097

7. Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT. (2006) Interactive Visual Analysis of SNP Data for Rapid Autozygosity Mapping in Consanguineous Families. **Human Mutation** 27 (10) 1041-1046.

8. Woods GC, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Sharif SM, Karbani G, Ahmed M, Bond J, Clayton D, Inglehearn CF. (2006) Quantification of Homozygosity in Consanguineous Individuals with Autosomal Recessive Disease. **American Journal of Human Genetics.** Vol.78:889-896.

9. Zhang J, Chiodini R, Badr A, Zhang G. (2011) The impact of next-generation sequencing on genomics. **J Genet Genomics.** 2011. March 20; 38(3): 95-109.

10. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can Exome Sequencing Do for You? **J Med Genet** (2011). doi:10.1136/jmedgenet-2011-100223

11. The 1000 Genomes Project Conortium (2010) A Map of Human Genome Variation From Population-Scale Sequencing. **Nature** doiL10.1038.nature 09534.

12. L. Rodríguez-Pazos, M. Ginarte, L. Fachal, J. Toribio, A. Carracedo, A. Vega. (2011) Analysis of TGM1, ALOX12B, ALOXE3, NIPAL4 and CYP4F22 in Autosomal Recessive Congenital Ichthyosis from Galicia (NW Spain): evidence of founder effects.**British Journal of Dermatology.** DOI: 10.1111/j.1365-2133.2011.10454.

13. LAkiyama, M. (2010) ABCA12 Mutations and Autosomal Recessive Congenital Ichthyosis: A Review of Genotype/Phenotype Correlations and of Pathogenetic Concepts. **Human Genome Variation Society.** D01 10.1002/humu.21326.

14. Akiyama, M. (2006) Pathomechanisms of Harlequin Ichthyosis and ABCA Transporters in Human Diseases.**Archives of Dermatology**. 142:914-918.

15. Akiyama M, Sakai K, Sugiyama-Nakagiri Y, Yamanaka Y, McMillan JR, Sawamura D, Niizeki H, Miyagawa S, Shimizu H. (2006) Compound Heterozygous Mutations Including a De Novo Missense Mutation in ABCA12 Led to a Case of Harlequin Ichthyosis with Moderate Clinical Severity. **Journal of Investigative Dermatology**. 126.1518-1523.

16. Dzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. **Nat Methods** 7(4):248-249 (2010).

17. Akiyama M, Nakagiri-YS, Sakai K, McMillan JR, Goto M, Arita K, Abe YT, Tabata N, Matsuoka K, Sasaki R, Sawamura D, Shimizu H. Mutations in Lipid Transporter ABCA12 in harlequin Ichthyosis and functional recovery by corrective gene transfer. **Journal of Clinical Investigation** (2005); 115(7): 1777-1784.

18. Kohan ZG, Antonelli J, Cohen HA, Adar H, Chemke J. Further delineation of the acrocallosal syndrome. **European Journal of Pediatrics**. (1991). 150:797-799.

19. Rehman F, Forshew T, Kurian MA, Cattell G, Farndon P, Brueton L, Maher E. A Novel Locus in a Family with an Acrocallosal-like Phenotype. **Medical and Molecular Genetics, University of Birmingham**

20. A Putoux1, S Thomas, K L M Coene, E E Davis, Y Alanay, G Ogur, E Uz, DBuzas, CGomes, SPatrier, C L Bennett, N Elkhartoufi, M H Saint Frison, L Rigonnot, N Joyé, SPruvost, G E Utine, KBoduroglu, PNitschke, LFertitta, ClThauvin-Robinet, A Munnich, V Cormier-Daire, RHennekam, E Colin, N AAkarsu, C Bole-Feysot, N Cagnard, A Schmitt, NGoudin, SLyonnet, F Encha-Razavi1, J PSiffroi, M Winey, NKatsanis, MGonzales, MVekemans, P L Beales& TAttié-Bitach. KIF7 mutations cause fetalhydrolethalus and acrocallosalsyndromes.**Nature Genetics** (2010) doi:10.1038/ng.826.

21. Lynex CN, Carr IM, Leek JP, Achuthan R, Mitchell S, Maher ER, Woods CG, Bonthon DT, Markham AF. Homozygosity for a missense mutation in the 67lDa isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: parallels with Stiff-Person Syndrome and other movement disorders.**BMC Neurology**. (2004) doi:10.1186/1471-2377-4-20.

22. Coutinho P, Barros J, Zemmouri R, Guimaraes J, Alves C, Chorao R, Lourenco E, Ribeiro P, Loureiro J, Santos JV, Abdelmadijd H, Paternotte C, Hazan J, Silva MC, Prud'homme JF, Grid D. Clinical Heterogeneity of Autosomal Recessive Spastic Paraplegias. **Archives of Neurology**. (1999); 56:943-949.

23. Genetics Home Reference. HPD Genes (2008) accessed 08/08/11. http://ghr.nlm.nih.gov/gene/HPD

24. Clark MJ, Chen Rui, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. Performance comparison of Exome DNA sequencing technologies. 2011 **Nature;** doi: 10.1038/nbt.1975