1 Introduction

1.1 The Ubiquitin-Conjugating System

The ubiquitin-conjugating system (UCS), describes a system in which intracellular proteins are selectively targeted by the covalent ligation of a 76-amino acid residue called ubiquitin. Typically, ubiquitinated proteins are targeted for degradation by the 26S proteasome – a major enzyme that catalyses the degradation of intracellular proteins – in this, its main capacity (in which the overall system is usually referred to as the ubiquitin-proteasome system (UPS) (Figure 1)), the UCS plays a crucial role in a number of basic cellular events such as cell signalling, signal transduction, metabolism and the immune and inflammatory response (Pagano (1997), Ben-Neriah (2002)). It is believed that the UCS is involved in the rapid degradation of 30% or more of newly made proteins within a cell (Schubert *et al.* (2000)). However, more recently the ubiquitin signal has also been connected with many other cell processes that function independently of the proteasome, including endocytosis, vesicle fusion, DNA repair, transcriptional silencing and ribosomal function (Weissman (2001)). It has been shown that up to 20% of cellular proteins are conjugated to ubiquitin at any one time within a cell (under standard cellular conditions) (Welchman *et al.* (2005)). The emergence of protein modification by ubiquitin as a critical regulatory process in virtually all aspects of cell biology has been acknowledged by the scientific community: the 2004 Nobel Prize in Physiology or Medicine was awarded for the discovery of Ubiquitin-mediated proteolysis. The fates of the different ubiquitinated protein substrates in all of these intracellular processes are predominantly determined by the number of ubiquitin molecules involved, and the 'linkage' that exists between them.

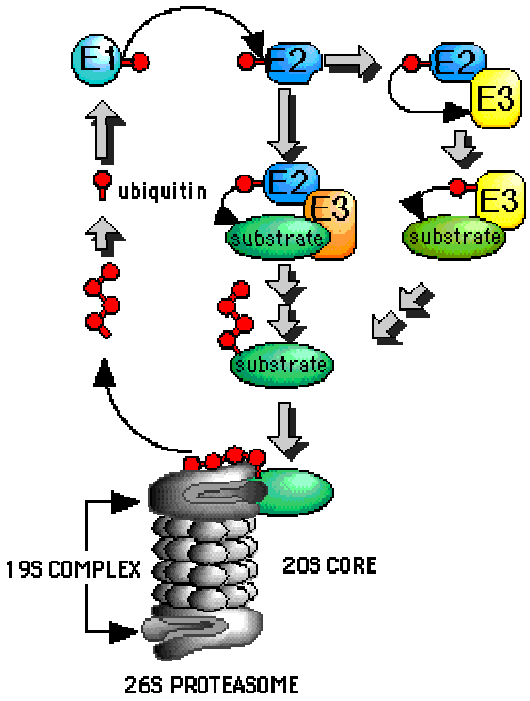


Figure 1 – Schematic diagram of the ubiquitin-proteasome system describing the sequential action of the three main enzymes – E1, E2 and E3 – in recruiting protein substrates to the 26S proteasome. E1 – ubiquitin-activating enzyme, E2 – ubiquitin-conjugating enzyme, E3 – ubiquitin ligase. (Image by W. Hilt, Stuttgart University).

Ubiquitin protein ligation is catalysed by the sequential actions of three main enzymes; a ubiquitin-activating enzyme (Enzyme(E)1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin ligase (E3). The transfer of ubiquitin to a substrate protein is initiated by an E1 activation step. The activation reaction is dependent on ATP (Adenosine TriPhosphate), the hydrolysis of which is catalysed by the E1 activating enzyme, to form a high-energy mixed anhydride intermediate - ubiquitin adenylate - between the C-terminus of ubiquitin and AMP (Adenosine MonoPhosphate). This is then nucleophilically attacked by a conserved cysteine residue of the E1 enzyme, which activates the ubiquitin through the formation of a thiolester linkage (covalent bond formed with a sulfhydryl group (-SH) and the elimination of the elements of water) with the C-terminal glycine residue (G76) of ubiquitin (Schindelin (2005)). Once the ubiquitin is activated, the E1-ubiquitin thiol ester complex undergoes ‘transthiolation’ with the active-site cysteine residue of an E2 ubiquitin-conjugating enzyme forming an E2-

ubiquitin thiol ester complex (Argiles *et al.* (1998)). The ubiquitin can then be transferred from the E2 active site to the ε-amino group (ε for epsilon, describing the fifth atom in a side chain) of the substrate protein's lysine residue by way of isopeptide bond formation (Hershko *et al.* (1983)), but only with the assistance of an E3 ubiquitin-protein ligase. Three major classes of E3 have been identified, termed the HECT (Homologous to E6-associated protein C-Terminus), RING finger (Really Interesting New Gene) and U-box E3s. HECT E3s have their name derived form the archetypical E3 involved in the ubiquitination (leading to the degradation) of the tumour suppressor protein p53 by the human papilloma virus' E6 protein and are characterised by a central cysteine domain within the HECT motif, which acts as an acceptor for ubiquitin. RING finger and U-box E3 motifs, who differ only by the latter's relative structural instability owed to a lack of key residues required for metal chelation, do not appear to have a direct catalytic site for ubiquitin (Ardley & Robinson (2005), Eddins & Pickart (2005)). Thus it follows that if the E3 involved in the particular ubiquitin conjugation cascade belongs to the HECT domain family, substrate protein ubiquitination is achieved by the ubiquitin firstly being transferred to the active site cysteine residue of the E3; whereas if the E3 belongs to the RING finger or U-box domain families, the ubiquitin is transferred directly to the substrate protein's amino group from the E2 enzyme.

It is important to note that the ubiquitin conjugation system described above follows a hierarchical organisation. Only a single E1 is responsible for the initial activation and transfer of activated ubiquitin, whilst the E2 family of enzymes consists of over 20 members, all with differing substrate preferences and subcellular locations. Conjugation to the substrate proteins is in turn assisted by hundreds of E3 enzymes which fine-tune the specificity set forth by the E2s (Pickart (2001)). Thus it can be seen, that the combinatorial diversity of a set of E2s interacting with a large variety of different E3s endows the ubiquitin-proteasome system with an exquisite specificity in the recognition of potential substrates among thousands of cellular proteins.

1.1.1 Polyubiquitination

Most often, substrate conjugation through the cascade of enzymes outlined above results in polyubiquitin chain formation. Specific lysine residues of each ubiquitin molecule in the extending chain serves as a site for further ubiquitination. The ubiquitin molecule itself contains seven lysine residues at amino acid positions 6, 11, 27, 29, 33, 48, and 63 (Figure 2).

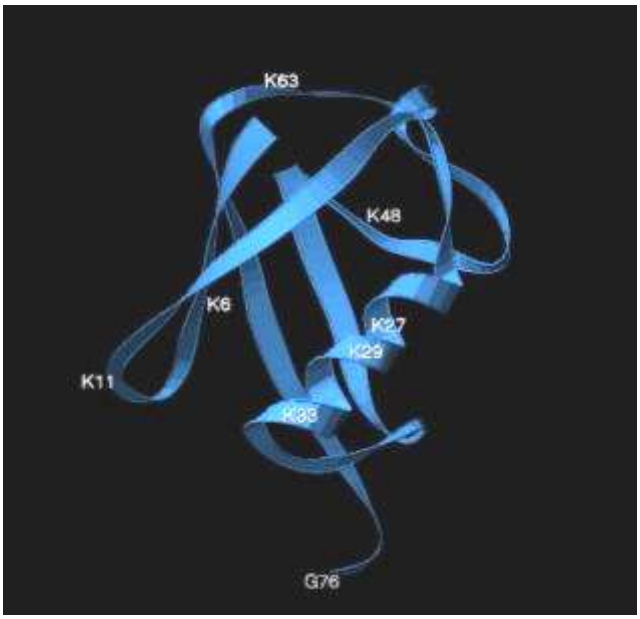


Figure 2 – Ribbon structure of ubiquitin showing internal lysine residues (K) and the C-terminal glycine residue (G76) (www.ks.uiuc.edu/training/tutorials/vmd/tutorial-html/img89.gif).

Polyubiquitin chains linked throughout via lysine(K)48-glycine(G)76 isopeptide bonds represent the predominant *in vivo* targeting signal for 26S proteasomal degradation (a multicatalytic protease complex that degrades substrates into small peptides and amino acids – see below) of substrate proteins in the UPS (Pickart & Fushman (2004)). Work done by Thrower *et al*. elucidated that in lysine 48 (Lys^{48})-linked chains the critical number of ubiquitin moieties required to recruit the 26S proteasome was four (tetraubiquitin) (Thrower *et al*. (2000)). This is due to the spatial arrangement of hydrophobic patches located at Lys^{9}, Ile^{44} and Val^{70} within the tetraubiquitin unit being essential for the efficient recognition by the proteasome (Ulrich (2002)). As well as Lys^{48}-linked chains, there is evidence to support that atypical chain formation through Lys^{29} and Lys^{11} could also be competent proteasomal

targeting signals, as both chains are able to bind to the 26S proteasome *in vitro* (Pickart & Fushman (2004)). In contrast, Lys^{63}-linked chains act as non-proteolytic signals in several intracellular pathways, such as DNA repair, transcriptional regulation, protein trafficking and ribosomal protein synthesis (Weissman (2001)). One example is the regulation of a DNA polymerase processivity factor called PCNA (Proliferating Cell Nuclear Antigen); PCNA is part of a DNA damage tolerance pathway which is crucial in the error free repair of damaged DNA. After initial mono-ubiquitination of PCNA, its activation requires the synthesis of a Lys^{63}-linked polyubiquitin chain – this is brought about by the Ubc13/ Mms/ Rad 5 complex, where UBC13/ Mms are an E2/ UEV (Ubiquitin E2 Variant – an E2 homologue minus the active site cysteine residue) complex and Rad 5 is the RING E3 which targets the E2/ UEV complex to PCNA to allow Lys^{63}-linked chain formation (Eddins & Pickart (2005)). The less frequently used Lys^{27} and Lys^{6} polyubiquitin chains additionally provide functions that are distinct from proteasomal targeting. Lys^{27}-linked polyubiquitin chains have shown not to induce proteasomal degradation directly, but specifically promote the association of BAG-1 (Bcl-2-Asscociated athanoGene-1) with the 26S proteasome (Pickart & Fushman (2004)). In this case, Lys^{27} polyubiquitin chain formation allows BAG-1 to act as a coupling factor between misfolded proteins bound by molecular chaperone Hsp70 and the 26S proteasomal complex (Rechsteiner (2005)). Not much work has been done on Lys^{6}-linked polyubiquitin chains, though the fact their formation is catalysed by the Brca1 (**Br**east-**ca**ncer susceptibility gene 1)/ Bard1 (**B**RCA-**a**ssociated **R**ING **d**omain 1) E3 heterodimer, which seem to associate at sites of DNA double strand break repair after exposure to ionising radiation, indicates a role in DNA repair (Morris & Solomon (2004)). Taken together, the generation of these different polyubiquitin chains (6 have been discussed above) provides an important level of complexity to facilitate all the different roles of the UCS.

1.1.2 Monoubiquitination and Endocytic Trafficking

As outlined above, the cascade of enzymes that make up the UCS usually give rise to polyubiquitin chain formation, the nature of which (lysine linkage) determines whether the proteasome is recruited in the predominant UPS pathway, or a non-proteolytic signal is established in an alternate intracellular pathway. In addition to Lys^{63}, Lys^{27} and Lys^{6} polyubiquitin chains, mono-ubiquitination has also been recently elucidated as an important non-proteolytic signal – most notably in the endocytic trafficking of proteins, specifically mammalian growth factor receptors targeted to the lysosyme.

In mammalian cells, many plasma proteins are ubiquinated in response to ligand binding. These include the RTKs (Receptor Tyrosine Kinases) EGFR (Epidermal Growth Factor Receptor), PDGFR (Platelet Derived Growth Factor Receptor), FGF (Fibroblast Growth Factor) and HGF (Hepatocyte Growth Factor – also referred to as Met) (Marmor &Yarden (2004)). When the ligand binds to them, their tyrosine kinase activity is activated catalysing autophosphorylation which is then recognised by an E3 ligase, such as the RING finger c-Cbl – which is thought to be the main E3 that mediates RTK ubiquitination (Marmor &Yarden (2004)). For those targeted to the lysosome, this results in mono-ubiquitination at multiple lysine residues on the substrate RTK, which promote its internalisation by recruiting multiple sorting adapters which recognise two hydrophobic patches on the mono-ubiquitin molecules around residues Phe^{4} and Ile^{44}, through a wide variety of ubiquitin-binding motifs, e.g. UIM (Ubiquitin-Interacting Motif) and UBA (UBiquitin-Associated) domains (Marmor & Yarden (2004)). Once internalised these activated receptors are sorted into clathrin-coated pits by a multiprotein complex. The most stringent requirement of ubiquitin in the progression through the endocytic process is at the trafficking from early to late endosome/ MVB (Multi-Vesicular Bodies), where Cbl sustained ubiquitination ensures that the RTK is sorted into the MVB and not the recycling endosome, which delivers back to the plasma membrane. The sorting of the receptors into the internal vesicles of the MVB requires the sequential engagement of several multi-protein complexes such as ESCRT (Endosomal Sorting Complex Required for Transport) and its related components, which all have the ability to bind ubiquitin. This then allows the ubiquitinated receptors to be passed down from one complex to another, which with the invagination of the limiting membrane of the MVB, secures them in

internal vesicles. Finally, the MVB organelles gradually accumulate lysosomal acidic proteases and fuse with the lysosome, resulting in degradation of the contents of the internal vesicles.

Thence it not only can be seen that mono-ubiquitination as well as polyubiquitination play large roles in key cellular events, but it is also important to note that ubiquitin is a major player in the two major systems controlling protein degradation within Eukaryotic cells: lysosomal and proteasomal.

1.1.3 A Fourth UCS Enzyme?

Very recent investigations has lent significant credence to work first done in 1999 by Koegl *et al.* which indicated the existence of a forth enzyme in the UCS enzyme cascade (E1, E2 and E3 being the first three – see above) termed E4 by the authors (polyubiquitin chain conjugation factor). It is hypothesised that E4 complexes may regulate the selection of lysine residues used for ubiquitin-ubiquitin linkages during polyubiquitin-chain assembly, and could also determine whether mono- or polyubiquitination is to be implemented. This would suggest that the fates of proteins in the UCS can be regulated even after E3 ubiquitin ligase involvement, which would endow the system with an extra level of complexity to control its wide ranging intracellular roles, at the level of ubiquitin-chain elongation. UFD2 (Ubiquitin Fusion Degradation model 2) defines the first identified family of E4s in humans, which is characterised by a conserved C-terminal U-box (refer to section 1.1). Biochemical and genetic studies have revealed that UFD2 binds to substrates conjugated with one to three ubiquitin molecules, and catalyses the addition of further ubiquitin moieties in the presence of E1, E2 and E3s, yielding polyubiquitinated substrates that are targeted for the 26S proteasome (Hoppe *et al.* (2005)). In addition, work done by Saeki *et al.* (2004) has shown that yeast UFD2 (UFD-2) catalyses a 'linkage switch' from Lys^{29}, used for mono-ubiquitination, by further elongation of the ubiquitin chain through Lys^{48}. A different type of E4 enzyme is represented by p300 which does not contain a U-box motif, it has been shown to polyubiquitinate mono-ubiquitinated species of the tumour suppressor p53 in collaboration with the E3 enzyme MDM2 (Murine Double Minute clone 2 oncoprotein) (Hoppe *et al.* (2005)).

Most of the evidence so far unearthed relevant to E4s tend towards ubiquitinated substrate protein modification to Lys^{48}-linked polyubiquitin chains of four to six ubiquitin molecules – which promotes optimal binding to RAD23 and DSK2 (two well described yeast proteins) - which are the key mediators in the delivery of the substrates to the 26S proteasome. Thence, the extra level of complexity/ control endowed upon the UCS by E4 polyubiquitin chain conjugation factors may only be of particular relevance to the UCS's key role in recruiting the proteasome, i.e. the UPS.

1.1.4 The Proteasome (inv. UPS)

Almost all proteins that are damaged, abnormal or foreign (viral) are degraded by the 26S proteasome – a single, highly conserved 2.5 MDa (MegaDalton) multisubunit enzyme which is ATP-dependent (Glickman (2000)). The 26S proteasome (also known as the haloenzyme) is made up of at least 45 subunits (Figure 3), which can be broken down into two major subcomplexes: the 20S core particle (CP), and the 19S regulatory complex (RC).

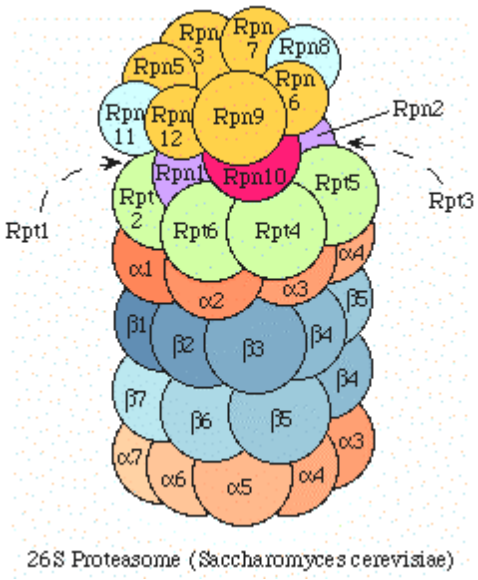


Figure 3 – Schematic representation of the 26S proteasome's subunits in *S.cerevisiae* displaying the relative positions of the six homologous ATPases (Rpts 1-6) and eleven non ATPase subunits of the 19S regulatory complex, as well as the two alpha (α) and beta (β) rings of the 20S core particle. (www.genome.ad.jp).

The CP (20S) is barrel shaped structure made up of four rings of seven subunits each, termed α and β rings. The two inner β-rings (outer being α) realise the proteolytic active sites facing inwards into the sequestered proteolytic inner chamber (which can be compared to the size of serum albumin (Rechsteiner (2005))), while the α-rings seal off the chamber from the external solvent (Groll *et al.* (1997)). One regulatory particle is attached to the surface of either (outer) α-ring of the CP to form the complete 26S proteasome (Figure 3). It should be noted here that what has been deemed the 26S proteasome by conventional literature, could in fact exist as the 30S proteasome *in vivo*; with one 19S RC disassociating from one of the α-rings during the purification process (Dahlmann (2005)).

The 19S RC itself can be broken down into two multisubunit substructures – a lid and a base. The base contains six homologous ATPases (S7, S4, S6, S10b, S6' and S8 - mammalian subunit nomenclature following the chronology of the yeast subunit nomenclature which was discovered first (Rpts 1-6)) which are thought to unfold polyubiquitinated substrates and translocate them into the CP (Braun *et al.* (1999)) The remainder of the base is made up of three non-ATPase subunits (S2, S1 and S5a (Rpns 1,2 and 10)) which can all bind polyubiquitin chains or UBL (UBiquitin-Like) domains. The lid of the RP is made up of a further 8 non ATPase subunits, the majority of which contain PCI domains (so called for their occurrence in Proteasome, Cop9 signalosome, and the eukaryotic Initiation factor 3 subunits) which are thought to mediate subunit-subunit interactions. The most important subunit in the lid seems to be a metalloisopeptidase – S13/ Rpn 11 – that removes ubiquitin chains from the tagged substrate prior to its translocation into the inner chamber for degradation (Rechsteiner (2005)).

The lid and base of the 19S RC orchestrate four of the six steps required in the effective proteolysis of a ubiquitinated protein: (1) polyubiquitinated substrate association with S1, S2 and S10 subunits; (2) substrate unfolding through the base's six ATPases; (3) detachment of polyubiquitin chain from the substrate by S13, and (4) translocation of the polypeptide into the 20S central chamber by threading through a channel in the 20S proteasome's α-ring (a process thought to be controlled by the 19S base's six ATPases as

they progressively unravel substrates – 'pumping' the substrate in a C- or N-terminal direction) (Rechsteiner (2005)).

The fifth and decisive step in ubiquitinated protein proteolysis is peptide bond cleavage, the mechanism of which relies on the positioning of N-terminal threonine residues within the seven subunits of the β-rings of the 20S CP. This residue is positioned just at the open cleft between two layers of β-sheets (subunits within the α-rings realise an extra α-helix in this position), and owed to its nucleophillic 'attacking' properties forms the proteolytically active sites within the 20S CP (Dahlmann (2005)). For reasons that are unclear at present, only three of the seven threonine residues fully function as active sites within each of the β-ring subunits (subunits β1, β2 and β5), each with their own cleavage site specificity: β1 – caspase-like; β2 – trypsin-like, and β5 – chymotrypsin-like (Groll *et al.* (2005)). On average, these three (x 2) threonine active sites allow the 20S CP to produce cleavage fragments with a length of seven to eight amino acids, with a range in length from three to twenty-five (Dahlmann (2005)). The peptide products of the proteasome are short lived, and do not accumulate in the cell under normal (healthy) physiological conditions. These peptides are most probably hydrolysed by an array of downstream proteases and aminopeptidases. The release of the peptide products defines the final step (step 6) in effective proteolysis of ubiquitinated proteins, and includes the release of ubiquitin, polyubiquitin chains and ubiquitin still attached to short, residual chains of amino acids (Hershko and Ciechanover (1998)). The latter two of these products are substrates for certain DeUbiquitinating Enzymes (DUBs) (S13/ Rpn 11 of the proteasome's 19S RC belongs to this family) which are able to cleave the isopeptide bonds that ubiquitin forms with itself in polyubiquitin chains, and lysine residues of its substrate proteins. DUBs in this instance ensure that ubiquitin is recycled back into the cell's free pool of monomeric ubiquitin, a process which is critical in maintaining the UCS's overall efficiency and thus effective functioning.

1.2 Deubiquitinating Enzymes (DUBs)

Based on the mechanism of catalysis, the 561 proteases realised within the human body (Puente & Lopez-Otin (2004)) are divided into five classes – aspartic, metallo, serine, threonine and cysteine proteases. DUBs are generally of the cysteine variety, although there are some that belong to the metallo class. It has been suggested that humans express approximately 79 functional DUBs (Nijman *et al.* (2005)).

Cysteine proteases rely on a catalytic triad of residues for their hydrolase activity – a defining cysteine harbouring the all important thiol group, an adjacent histidine which assists deprotonation of the cysteine, and an aspartate residue which polarises the histidine. The general biochemical nature of cysteine protease catalysis can be summed up thus: (1) the carbonyl (carbon atom double bonded to an oxygen atom) of the scissile (a peptide bond that is hydrolysed by a peptidase) peptide bond between ubiquitin and the target is nucleophilically attacked by the DUBs cysteine; (2) the oxyanion (an oxygen containing, negatively charged polyatomic ion) containing intermediate is stabilised by the main chain of the catalytic cysteine and a glutamate, glutamine or asparagine – the so-called oxyanion hole; (3) the target protein is released, and a covalent ubiquitin/ DUB intermediate is formed; and (4) a water molecule reacts with this intermediate to separate and release the ubiquitin and DUB (Nijman *et al.* (2005)).

Unlike cysteine proteases, metalloproteases generally employ a Zn^{2+} atom stabilised by two histidines and an aspartate to bind and polarise a water molecule which can form a (noncovalent) substrate intermediate. Proton transfer from another water molecule can then release the DUB (Nijman *et al.* (2005)).

DUBs have four distinct functions within the UPS (encompassing the UCS): they are responsible for processing inactive ubiquitin precursors, they are able to remove ubiquitin from proteins inappropriately targeted to the proteasome (a suggested proofreading mechanism (Lam *et al.* (1997)) - this includes small intracellular nucleophiles such as glutathione and polyamines which are of considerable abundance and thus have the potential to rapidly use up free ubiquitin, DUBs efficiently remove ubiquitin from its conjugates prior to proteolysis of the substrate protein – thus allowing the ubiquitin to be recycled, and they

disassemble unanchored ubiquitin oligomers that can compete with ubiquitinated substrates for the ubiquitin-binding sites within the proteasome's 19S RC – which would hinder its protein turnover (Amerik and Hochstrasser (2004)).

DUBs specifically cleave ubiquitin-linked molecules after the terminal carbonyl of the final glycine residue at position 76. The target bond is generally an isopeptide bond, except with ubiquitin C-terminally extended precursors, where a standard peptide bond requires hydrolysis.

The DUBs belonging to the cysteine protease class (see above) can be further sub-categorised into four classes based on their ubiquitin protease domains. The largest and most diverse of these subclasses are the UBiquitin-Specific Proteases (UBPs (Amerik and Hochstrasser (2004)) or USPs (Nijman *et al.* (2005))). The second ubiquitin-specific cysteine protease subclass is made up of the Ubiquitin Carboxyl-Terminal Hydrolases (UCHs), which were among the first DUBs to be described. The remaining two known subclasses have been discovered only recently; a bioinformatics approach led to the identification of the Ovarian TUmour (OTU) subclass (Makarova *et al*. (2000)), and the fourth cysteine protease DUB subclass, for which ataxin-3 is the only demonstrated member to date, is characterised by a domain called the Josephin domain (Amerik and Hochstrasser (2004)) or Machado-Joseph Disease (MJD) protein domain (Nijman *et al.* (2005)). DUBs belonging to the metalloprotease class all have a ubiquitin protease domain called JAMM (**JA**B1/ **M**PN/ **M**ov34 metalloenzyme), and is represented by the S13/ Rpn 11 subunit of the 19S RC.

1.2.1 Ubiquitin-Specific Proteases (UBPs/ USPs)

It has been postulated that the reason the UBP subclass outnumber the other 4 DUB (sub)classes by such a large degree (UCH and MJD by around 13:1, and the OTU and JAMM subclasses by 4:1 (Nijman *et al.* (2005)) is that the UBPs coevolved in an intimate relationship with the ubiquitin E3 ligases, so as the number of E3s increased during evolution, so did the number of UBPs (Semple (2003)).

Crystal structures of ubiquitin and a human UBP – the Herpesvirus-Associated Ubiquitin-Specific Protease (HAUSP) – has indicated an active site in a deep cleft between

two of three major globular domains, harbouring the protease's catalytic triad – a nucleophillic cysteine (Cys 223) and a His box containing a histidine and an aspartate (His 464 and Asp 481). In free HAUSP, the Cys 223 and His 464 are too far apart for a 'productive' interaction, whereas ubiquitin binding causes a major change in the conformation of the catalytic cleft, bringing them within binding distance (Hu *et al.* (2002)).

An important example of a UBP DUB within the UPS is that of isopeptidase T (Ubp 14 in yeast), one of the enzymologically best characterised DUBs (Wilkinson *et al*. (1995)). It acts almost exclusively on unanchored ubiquitin chains, and its yeast homologue is singularly responsible for the bulk of 'free' ubiquitin chain disassembly in vivo. Isopeptidase T disassembles K^{48}-linked polyubiquitin chains (refer to section 1.1.1) starting at the proximal end of the chain (the end that contains a free carboxyl-terminus) in a sequential exo mechanism, and though K^{48}-linked chains are preferentially targeted, isopeptidase T can also cleave ubiquitin polymers in head to tail linkage, such as occurs in the polyubiquitin precursor, albeit less efficiently. It should be noted that isopeptidase T cannot act on polyubiquitinated protein substrates directly; another DUB must first release the chain from the substrate (Wilkinson *et al*. (1995)).

The substrate diversity within this subclass of approximately 58 DUBs (Nijman *et al*. (2005)) can be indicated by another two well known examples of UBP roles within the UCS and UPS. The first example is of ubiquitin retrieval within the endocytic trafficking process (refer to section 1.1.2). Owing to the relative longevity of ubiquitin in vivo, it requires recovering from the involuting membrane proteins prior to complete vesticulation – the yeast DUB involved in this process is Doa4. Doa4's localisation to the endosome can be blocked by the elimination of factors ('E' factor mutants) from the large ESCRT III assembly at the endosome surface. This strongly suggests that ESCRT III helps to recruit Doa4 to the late endosome and direct it toward monoubiquitinated membrane proteins after they have been committed to inclusion in the involuting membrane (Amerik *et al.* (2000)). The human homologue of Doa4 is thought to be UBPY owing to its ability to bind Hbp (Hrs-binding protein), which is, along with Hrs, involved in endocytic trafficking in mammalian cells (Kato *et al.* (2000). The second example is of the UBP found in complex with the proteasome's 19S RC component (refer to section 1.1.4) - USP 14 (Ubp 6 in yeast) (Borodovsky *et al*. (2001)) -

where it works in conjunction with S13/ Rpn 11 to deubiquitinate proteins that are destined for degradation. Specifically, immediately after the substrates have been unfolded by the 19S RC base's six ATPases, and just before they are actively translocated into the 20S inner proteolytic chamber. The importance of this enzyme has also been intriguingly demonstrated by the fact that USP14 mutant mice develop ataxia (inability to coordinate movement) (Wilson *et al*. (2002)).

1.2.2 Ubiquitin C-Terminal Hydrolases (UCHs)

The human UCH subclass of cysteine protease DUBs consists of at least four mammalian isozymes which all share close homology in their catalytic domains. X-ray crystallographic results of UCH catalytic core structures reveal a catalytic triad that matches very closely to that of a classical cysteine protease (refer to section 1.2), and the three-dimensional folds of these segments are nearly indistinguishable from those of the UBP subclass, with superimposing catalytic residues (Amerik and Hochstrasser (2004)). Also, similarly to the UBP subclass, active site residues within the free form UCH are not in catalytically competent conformations. Both subclasses seem to undergo a conformational change when bound to ubiquitin which brings the cysteine, histidine and aspartate catalytic residues into play by either eliminating steric (effects of atomic arrangements) obstructions in the active site cleft (UCH relevant – see below) or changing their relative positions (UBP relevant – refer to section 1.2.1) (Amerik and Hochstrasser (2004)). The obvious biological advantage of UCHs and UBPs being dependent upon ubiquitin binding for their hydrolase activity, is that they will be proteolytically inactive against non-substrate cellular proteins.

UCHs active site nucleophile cysteine is positioned at the bottom of a narrow groove in the enzyme's surface. When in free form (not ubiquitin-bound), a specific UCH residue aliphatic (noncyclic) side chain occupies part of this groove by arcing over it (Johnston *et al.* (1999)). This so called active site 'crossover loop' becomes ordered when the UCH binds to ubiquitin, which increases the diameter of the loop allowing access to the catalytic triad. However, even at its maximally open state, the diameter is no greater than 15Å (Johnston *et al.* (1999). Biochemical studies have shown that the UCH subclass of DUBs preferentially cleave small adducts or unfolded polypeptides from the C-terminus of ubiquitin. This seems to

be directly owed to the 15Å size restriction of the 21 residue arcing peptide segment (loop) over the active site, which is much smaller than the majority of folded proteins. In addition to the active site obstruction which enhances UCH substrate selection there are extensive, highly specific interactions realised between the UCH subclass and ubiquitin. These include numerous van de Waals interactions, twenty hydrogen bonds, and a salt bridge (electrostatic bond) between Arg(arginine)^{72} of ubiquitin and an aspartate residue (Johnston *et al.* (1999)).

Despite this biochemical data (and being the first described DUBs), their specific functions within the UCS/ UPS remain poorly understood. They are thought to mainly cleave ubiquitin from inappropriately ubiquitinated polypeptides such as intracellular nucleophiles like glutathione and polyamines (Larsen *et al*. (1998)). They have also been shown to be involved in the processing of the ubiquitin precursors, which are translated fused to ribosomal protein precursors or head-to-tail ubiquitin-linked multimers with an additional amino acid on the last ubiquitin (Larsen *et al* (1998)). Though the vast majority of work on the UCH subclass points to a substrates no longer than 20-30 amino acids in length (Amerik and Hochstrasser (2004)), recent work by Misaghi *et al.* (2005) proposes a catalytic model for UCH-L3 which would allow the hydrolysis of larger ubiquitin conjugates.

Four mammalian isozymes have so far been established (Mayer and Wilkinson (1989), Osawa *et al.* (2001)), whose expression is tissue specific (Wilkinson *et al.* (1992)) and developmentally regulated (Schofield *et al.* (1995)). UCH-L1 is predominantly expressed in the cytoplasm of neural and neuroendocrine cells, where its malfunction is thought to bring on the symptoms of Parkinson's disease (refer to section 1.3 and 1.5 for advanced discussion). UCH-L3 is localised in the haematopoietic tissues, whereas UCH-L2 is widely expressed, but at much lower levels than either UCH-L1 or UCH-L3 (Wilkinson *et al.* (1992)). Very little is known of the fourth isozyme – UCH-L4 – though its high sequence similarity with UCH-L3 does give some insight (Osawa *et al.* (2001)).

Additionally, recent work has highlighted a potential specific function of a UCH within the UPS. Wicks *et al.* (2005) has identified a UCH in the 19S RC (also recently referred to as the PA700 activator complex) of the bovine 26S proteasome. This UCH, originally named UCH37, but also known as UCH-L5 (Nijman *et al.* (2005)), has shown the ability to cleave 'branched' ubiquitin chains from the distal end, it does not however display activity toward

linear ubiquitin dimers. Though a ubiquitin editing function has been postulated (Wicks *et al.* (2005)), whereby it deubiquitinates proteins that have been mistakenly ubiquitinated (recognised by having shorter ubiquitin chains) to halt their degradation; compelling evidence is yet to be provided.

1.2.3 The Ovarian Tumour (OTU)- Related Proteases

The otu gene is involved in the development of the *Drosophila melanogaster* ovary, where it is thought to help regulate the translation of certain RNA transcripts (Goodrich *et al.* (2004)). Sequence similarities were originally found between the *Drosophila* otu gene and those encoding viral cysteine proteases (Makarova *et al.* (2000)).

A member of the OTU family of DUBs – human otubain 2 – has had its crystal structure unveiled (Nanao *et al.* (2004)). It realised a five-stranded β-sheet positioned between two helical domains, in which amino residues Cys^{51}, His^{224} and Asn(asparagine)^{226} appear to form the unorthodox catalytic triad. Despite the lack of sequence similarity, the active site of otubain 2 displays almost identical geometries to those of the UBP and UCH subclass, though a critical hydrogen bond between the His^{224} and Asn^{226} is required to stabilise this unorthodox cysteine protease catalytic triad (Nanao *et al.* (2004)). Similar to the UCH and UBP free enzyme conformations, otubain 2 may also be in a self-inhibited state owed to the helix α3 loop spatially restricting the active site (Amerik and Hochstrasser (2004)).

Recent work has confirmed that proteins containing the OTU domain have DUB activity. In one study, a 100-KDa cytoplasmic protein called Cezanne which negatively regulates NF-kB (Nuclear Factor-kappa B), displayed *in vitro* DUB activity upon linear polyubiquitin translation products, isopeptide-linked polyubiquitin chains and ubiquitin-protein conjugates (Evans *et al.* (2003)).

1.2.4 Machando-Joseph Disease Proteins (or Josephin) Domain Protease (MJDs)

A bioinformatics search for other classes of DUB cysteine proteases identified ataxin-3. Instability of a CAG (glutamine coding) nucleotide repeat within the ataxin-3 gene leads to Machando-Joseph Disease – a hereditary neurological condition. Expansion of the tri-nucleotide repeat leads to protein misfolding, which results in aggregation and cellular toxicity (Nijman *et al.* (2005)).

ataxin-3 displays three typical DUB properties: (1) it deubiquitinates ubiquitin-AMC (ubiquitin-7-amido-4-methylcoumarin); (2) it is able to disassemble ubiquitin-lysosyme conjugates; and (3) ubiquitin aldehyde (Ubal) – a potent DUB inhibitor – is readily bound by it (Burnett *et al* (2003)). Very recent work also shows that the characteristic cysteine protease catalytic triad is conserved within ataxin-3 (Nicastro *et al.* (2005)). In evolutionary terms, MJDs are likely to be relatively late additions to the UCS/ UPS, as no yeast homologues have been uncovered.

1.2.5 JAMM Motif Proteases (Metalloprotease Subclass)

The S13/ Rpn 11 subunit (also known as POH1) of the 20S proteasome's 19S RC (refer to section 1.1.4) represents this class of DUBs. Mutation studies have revealed that DUB activity is almost completely attributable to POH1 (Nijman *et al.* (2005)). The sequence of the distinct motif typified by POH1 was named the JAMM domain to distinguish it from a broader group of proteins which contain an MPN motif, of which JAMM is a subtype (Maytal-Kivity *et al.* (2002)).

This metalloprotease motif realises two conserved His residues and an Asp residue that coordinate a zinc ion in the active site. Additionally, there is a conserved Ser (Serine) residue positioned between the two histidines which hydrogen bonds to a Glu (Glutamate) residue, which is thought to function in general JAMM acid-base catalysis (Maytal-Kivity *et al.* (2002)).

Cytidine deaminase is a well characterised metalloenzyme whose overall structure superimposes well onto that of the AF2198 JAMM domain protein, and also utilises a zinc ion

in catalysis, and thus presents clues into the mechanism of isopeptide bond hydrolysis employed by POH1. Cytidine deaminase catalysis proceeds thus: the zinc ion polarises a water molecule which then nucleophilically attacks a carbon atom in the cytidine pyrimidine ring. This results in the formation of a tetrahedral intermediate that rapidly collapses due to instability, releasing the reaction products. It is likely that JAMM proteases including POH1, use a similar zinc dependent mechanism in bond hydrolysis (Snider *et al.* (2002)). Unlike most other proteases described, AF2198 lacks peptide-binding site elements (Amerik and Hochstrasser (2004)), which lends credence to the contention that POH1 is only active when incorporated into the larger heteromeric 19S RC complex. Indeed, neighbouring 19S subunits are known to participate in ubiquitin-protein conjugation binding.

All the main subclasses of DUBs within mammalian cells have been generally discussed with specific examples where appropriate. Ubiquitin carboxyl-terminal hydrolase L-1 (UCH-L1) of the UCH subclass of cysteine proteases DUBs is the main focus of this study, and will now be considered in detail.

1.3 Ubiquitin Carboxyl-Terminal Hydrolase L1 (UCH-L1)

Ubiquitin C-terminal hydrolase L1 (UCH-L1), previously known as protein gene product 9.5 (PGP 9.5), is a cysteine protease of the UCH subclass (refer to section 1.2.2). UCH-L1 is one of the most abundant proteins in the brain (accounting for 1-5% of total soluble protein), with immunohistological experiments demonstrating its exclusive localisation within neurons (Wilkinson *et al.* (1989)). The UCH-L1 gene maps to 4p14 (chromosome four, **p**etit (short) arm, major band fourteen) (Edwards *et al.* (1991)), containing nine exons (coding sequences) and eight introns (intervening non-coding sequences), that span approximately 10kb (**k**ilo**b**ase = 1000 nucleotide bases) in this genomic region (Leroy *et al* (1998a), Day *et al.* (1990)) – to be discussed in greater detail later (refer to section 1.8). Northern blot analysis (primary electrophoresis of ribonucleic acid) revealed a 1.3kb transcript broadly represented in all the brain tested, with higher levels evident in the substantia nigra (Leroy *et al.* (1998a)). UCH-L1 is a 223 amino acid cysteine protease that contains a classical active site catalytic triad composing of cysteine, histidine and aspartate residues which are only activated once ubiquitin is bound, inducing a conformational change within the enzyme that eliminates a steric obstruction over the active site residues (refer to section 1.2.2). However, this 21 residue arcing peptide 'loop' once ordered by the ubiquitin-induced conformational change, still restricts potential UCH-L1 substrates to a 15Å size (up to 20-30 amino acids). This biochemical data (Johnston *et al.* (1999), Amerik and Hochstrasser (2004)) seems to supplement the early work done by Larsen *et al.* (1998), who showed that UCH-L1 rapidly and preferentially cleaves small 'leaving groups' such as amino acids and oligopeptides from the C-terminus of ubiquitin *in vitro*, but not larger 'leaving groups' such as proteins. These findings indicate a physiological role of UCH-L1 in the recycling of ubiquitin; to hydrolyse inappropriately conjugated intracellular nucleophiles such as glutathiones and polyamines, which are abundant by-products of cellular metabolism (Pickart and Rose (1985)). Additionally, Larsen *et al.* (1998) also showed that ubiquitin gene products could also be hydrolysed (very slowly) by UCH-L1, indicating a further possible physiological role in the generation of free monomeric ubiquitin from ubiquitin precursors (proproteins). The size restriction placed upon UCH-L1 substrates by the active site 15Å 'crossover loop', seems to

prohibit deubiquitination of ubiquitin-protein conjugates or disassembling of polyubiquitin chains.

1.3.1 UCH-L1's Isoleucine93Methionine Mutation (Ile93Met) and Parkinson's Disease

Parkinson's disease is the most prevalent neurodegenerative disease in the world affecting 1 in 500 people (Clough *et al.* (2003). It is characterised by four main disabling symptoms – paucity of movement, rigidity, rest tremor and postural instability (the pathophysiology and epidemiology of Parkinson's disease is discussed in greater detail in section 1.5).

Heritability of Parkinson's disease has come to the forefront with the identification of rare gene mutations in familial Parkinson's disease and common genetic risk factors for idiopathic Parkinson's disease. Ten regions of the genome are now linked/ associated with the phenotype (Table 1). Of these ten regions, the genes associated with the UCS – namely *UCH-L1*, *α-synuclein* and *Tau* – were of the first to be uncovered; most prominent of these was *UCH-L1* in a study carried out by Leroy *et al.* in 1998(b).

Chromosome	Locus	Gene	Range of Age (years) at Onset (mean)	Reference
AD				
1p36	PARK9	Unknown	N/A (65)	Hicks *et al*. (2001)
2p13	PARK3	Unknown	36-89 (58)	Gasser *et al*. (1998)
4p14-15	PARK5	*UCH-L1*	49-51 (50)	Leroy *et al*. (1998b)
4p15	PARK4	Unknown	24-48 (30)	Farrer *et al*. (1999)
4q21	PARK1	*α-synuclein*	20-85 (46)	Polymeropoulos *et al*. (1997)
12p11.2-q13.1	PARK8	Unknown	38-68 (53)	Funayama *et al*. (2002)
12q23-24.1	SCA2	*Ataxin-2*	19-61 (39)	Gwin-Hardy *et al*. (2000)
14q32.1	SCA3	*Ataxin-3*	31-57 (42)	Gwin-Hardy *et al*. (2001)
17q21-22	FTDP-17	*Tau*	25-76 (49)	Hutton *et al*. (1998)
19q13	DYT12	Unknown	12-45 (23)	Kramer *et al*. (1999)
AR				
1p35-36	PARK6	Unknown	32-48 (41)	Valente *et al*. (2001)
1p36	PARK7	Unknown	27-40 (33)	Van Duijn *et al*. (2001)
6q25.2-27	PARK2	*Parkin*	6-58 (26)	Kitada *et al*. (1998)

Table 1 – Indicates the main reported mutations or genetic loci associated with familial Parkinson's disease (taken and modified from Skipper & Farrer (2002)).

- *Abbreviations:* AD, autosomal dominant; AR, autosomal recessive; FTDP-17, frontotemporal dementia and Parkinsonism linked on chromosome 17; LRRK2, leucine rich repeat kinase 2; N/A, data not available; PINK, PTEN induced putative kinase 1.

The majority of the work now being done on UCH-L1 was initiated by this investigation. The coding region of the UCH-L1 gene was sequenced in probands from 72 families with Parkinson's disease, and consequently identified a missense mutation in one proband of German pedigree. It was realised in the fourth exon of the UCH-L1 gene, and took the form of an isoleucine to methionine amino acid substitution at position 93 – this corresponds to a C to G change at nucleotide position 277 in exon 4. Further mutation analysis revealed that the affected brother of the proband also carried the Ile93Met mutation. The group's control study comprised of 500 chromosomes with 204 originating from German backgrounds; none carried the Ile93Met change. This mutation was however not seen to be 100% penetrant; the father of the two affected individuals, and the presumed carrier of

Ile93Met, did not develop Parkinson's disease (Figure 4). The reason for this incomplete penetration will be postulated later (refer to section 1.3.2).

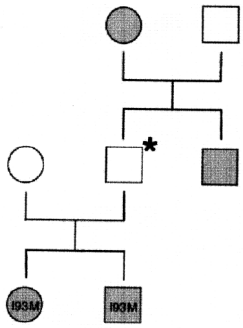


Figure 4 – The family in which the Ile93Met mutation was discovered (taken from Leroy *et al.* (1998b)), demonstrating the Ile93Met's penetrance (squares = males, shaded = Parkinson's disease). Note the unaffected father, presumed to be the carrier of Ile93 Met. Genotypes were only available from the most recent generation, with the results shown.

In the same investigation, the catalytic activity of the mutant protein was then compared to that of the wild-type (Wt) on standard substrates (ubiquitin ethyl ester and ubiquitin-7-amido-4-methycoumarin (Ub-AMC)). Results obtained indicated a roughly 50% reduction in the Ile93Met protein's V_{MAX} (maximum velocity of enzymic reaction), with no difference in K_m (Michaelis' constant – a substrate affinity measurement). This indicated a 50% reduction in the catalytic activity of the Ile93Met protein.

Though in the early stages of the understanding of the biochemistry of the UPS and UCH-L1's role within it, the potential significance of the results obtained in this investigation were appreciated; an abundant protein in the brain, with a significant role in the 'ubiquitin-dependent proteolytic pathway' (as the authors referred to it) had realised a coding region mutation, resulting in reduced enzymatic activity, indicating a causative role in the pathogenesis of autosomal dominant Parkinson's disease - which it was strongly associated with. Leroy *et al.* postulated that the reduced catalytic activity could affect the cleavage and turnover of the unknown substrate(s), leading to aggregation of the substrate(s) over time, causing severe aberration in the processing and degradation of proteins, which would then bring about the neuronal degeneration apparent in Parkinson's disease.

Once Leroy *et al.* had reported these findings, several groups around the world (Harhangi *et al.* (1999), Maraganore *et al.* (1999), Wintermeyer *et al.* (2000), Zhang *et al.* (2000)) initiated work to uncover further Parkinson's disease patients carrying the UCH-L1 Ile93Met amino acid substitution. None of these groups, or any group since, has identified the mutation in any Parkinson's disease cases (or any other individual). This indicates that though the mutation is a highly penetrant one (~80% - from what can be assumed – refer to Figure 4) that seriously affects UCH-L1's hydrolase activity, it is an extremely rare cause of autosomal dominant Parkinson's disease.

Furthermore, more recent work done by Nishikawa *et al.* (2003) has verified the early enzyme kinetic work done by Leroy *et al.* (1998b). Using the same substrate that Leroy *et al.* had employed in part of their work – Ub-AMC (ubiquitin-7-amino-4-methylcoumarin) – this group observed the hydrolase activity of the Ile93Met mutant was 45.6% of that of the Wt UCH-L1. This confirms UCH-L1's considerable loss of hydrolase activity when harbouring the Ile93Met amino acid substitution. Structural model analysis of the UCH-L1 enzyme suggests a highly plausible reason for the mutant's substantial loss in hydrolase activity (Figure 5); it clearly shows that residue 93 is proximal to the active site nucleophile – Cys90 – and is thus in a good position to affect the active site's catalytic activity.

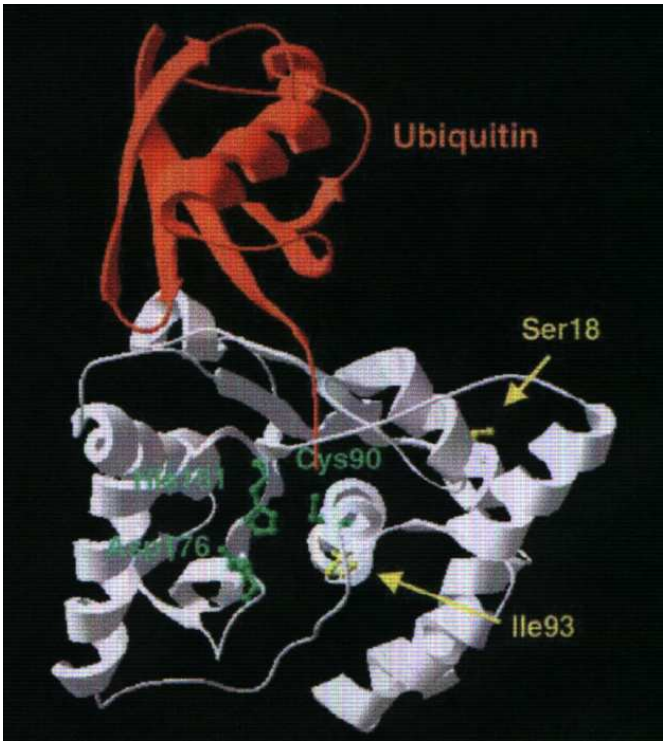


Figure 5 – A structural model for the UCH-L1-ubiquitin complex. The sequence of UCH-L1 is shown, mapped onto the determined structure of the highly homologous (52% identity) UCH-L3, complexed to the inhibitor ubiquitin aldehyde (SwissPdbViewer V.3.7b2) (Johnson *et al.* (1997), Johnston *et al.* (1999)). Residue 93 is proximal to the active site nucleophile (Cys90), while Ser18 (refer to section 1.3.2) is distal from the active site and from the ubiquitin binding site (Johnston *et al.* (1997), Johnston *et al.* (1999)).

It is postulated that (as discussed above) UCH-L1's critical function to catalyse the hydrolysis of C-terminal esters and amides of ubiquitin (Larsen *et al.* (1998)), which allows ubiquitinated intracellular nucleophiles and ubiquitin proproteins to be processed, maintains cellular levels of 'free ubiquitin' (i.e. ligatable to further substrate proteins at the C-terminus) (Larsen *et al.* (1998)). This decreased hydrolase activity of the Ile93Met UCH-L1 could therefore result in reduced levels of free ubiquitin, that may well have adversely affected the normal degradation of certain proteins in the brains of the two German sibs, particularly in the substantia nigra region, where *UCH-L1* is transcribed at higher levels (Leroy *et al.* (1998b)).

Parkin-associated endothelin receptor-like receptor (Pael-R) and O-glycosylated α-synuclein are two such proteins whose degradation could be affected; as they have both been

identified as substrates for parkin (Imai *et al.* (2001)) – which is a UPS E3 enzyme (refer to section 1.1). Without normal levels of free ubiquitin, parkin (discussed below – refer to section 1.4.2) would no longer be able to remove misfolded proteins such as Pael-R from the endoplasmic reticulum (ER), and thus protect neurons from ER-mediated stress induced cell death (Imai *et al.* (2001)); a possible route that could generate the stresses realised upon the brain of a Parkinson's disease sufferer, specifically in the dopaminergic neurons of the substantia nigra where *UCH-L1* is transcribed at higher levels (Leroy *et al.* (1992)). The degradation of O-glycosylated α-synuclein, as a parkin substrate, would also be affected by a reduction in the availability of monomeric ubiquitin – leading to its build up and deposition. Intriguingly α-synuclein (discussed below – refer to section 1.4.1 and 1.5.3) has been found to be a major component of Lewy Bodies (the major pathological hallmark of Parkinson's disease - refer to section 1.5.3) in Parkinson's disease patients (Spillantini *et al*. (1990)), as has UCH-L1 itself (Doran *et al.* (1983)).

1.3.2 UCH-L1's Serine18Tyrosine Polymorphism (Ser18Tyr) and Parkinson's Disease

As mentioned earlier, the fact that Leroy *et al.* (1998) reported a missense mutation (Ile93Met) in exon 4 of the UCH-L1 gene in a case of familial Parkinson's disease, spurred a series of similar investigations which failed to uncover further individuals carrying this mutation. However, a more common sequence variation in exon 3 of the UCH-L1 gene was uncovered (Maraganore *et al*. (1999), Gasser *et al*. (1999), Wintermeyer *et al*. (2000), Zhang *et al*. (2000)).

Maraganore *et al.* (1999) were the first group to uncover this sequence polymorphism in a study which was initiated as a linkage analysis on a large well-documented family with Parkinson's disease. Their early data from this scan strongly suggested a locus for autosomal dominant familial Parkinson's disease on chromosome 4p – a region containing many candidate genes (Farrer *et al.* (1999)). However, knowing that the gene for UCH-L1 was located on the short arm of chromosome 4, and that Leroy *et al.* (1998b) had very recently reported a Parkinson's disease-associated mutation within this gene, they undertook a cDNA sequencing strategy of *UCH-L1* in four individuals, and identified a coding region and

expressed polymorphism - Ser18Tyr. They subsequently found that the novel polymorphism occurred on ~20% of Caucasian chromosomes, and determined it was outside their 'obligate candidate region'. Further investigation into this polymorphism was halted in the study. However, importantly the group had uncovered a novel polymorphism in exon 3 of the UCH-L1 gene – a C-to-A transition at nucleotide 54 – which results in a serine to tyrosine substitution at amino acid position 18 (Ser18Tyr).

Where Farrer *et al.* (1999) had left off, several investigations (Maraganore *et al*. (1999), Gasser *et al*. (1999), Wintermeyer *et al*. (2000), Zhang *et al*. (2000)) went on to confirm the Ser18Tyr polymorphism's prevalence in various populations, but also investigated its association with sporadic and familial Parkinson's disease. A reduced risk was observed for carriers of the Tyr allele in an American population in which half the controls and Parkinson's disease cases reported having at least one parent of German origin (PD (Parkinson's disease) cases – *n* = 139, Controls – *n* = 113. OR (odds ratio) = 0.53) (Maraganore *et al*. (1999)). This protective effect was subsequently replicated in two Japanese populations (PD cases – *n* = 160, Controls – *n* = 160. OR = 0.73) (Zhang *et al*. (2000), Satoh *et al*. (2001)) and one German (PD cases – *n* = 229, Controls – *n* = 200. OR = 0.57) (Wintermeyer *et al*. (2000)). Though three other studies could not reproduce this effect in Australian (PD cases – *n* = 142, Controls – *n* = 142. OR = ns (not significant)) (Mellick *et al*. (2000)), Italian (Savettieri *et al*. (2001)) and French (PD cases – *n* = 114, Controls – *n* = 93. OR = ns) (Levecque *et al*. (2001)) Caucasian populations. The lack of association consistency of the Ser18Tyr polymorphism's Tyr allele protection against Parkinson's disease can be explained if the results from the studies above are related to sporadic and familial Parkinson's disease, and ethnic origin is taken into account.

The protective effect of the Tyr allele was only found among German and Japanese sporadic cases (Wintermeyer *et al*. (2000), Zhang *et al*. (2000), Satoh *et al*. (2001)), and also in an American group where nearly half the studied patients has at least one parent of German origin (Maraganore *et al*. (1999)) – indicating a possible German founder effect (a change in the American population's Ser18Tyr frequency directly owed to the establishment of a German subpopulation). Moreover, the fact that all the studies which observed a protective influence only involved sporadic cases suggests that this polymorphism has no

impact on familial Parkinson's disease. This has led to the hypothesis that the full effect of another major UPS gene in the potential pathology of familial cases, may obscure the more moderate impact of the Ser18Tyr protective effect (Levecque *et al.* (2001)). Additionally, as one may expect, Parkinson's disease protection does seem to be dependent on the Tyr allele dosage; that is, homozygotes are at significantly lower risk (relative risk 0.31) than are heterozygotes (relative risk between 0.55 and 0.81) (Levecque *et al.* (2001), Maraganore *et al.* (1999), Satoh and Kuroda (2001)). Furthermore, two other studies have shown that the Tyr allele is inversely associated with early onset sporadic Parkinson's disease (PD cases – n = 209. p = 0.03) (Elbaz *et al.* (2002)), (PD cases – n = 160. p = 0.043) (Wang *et al.* (2002)), i.e. as well as protecting against sporadic Parkinson's disease outright, the UCH-L1 Ser18Tyr polymorphism is also indicated in the more discreet effect of delaying the age of onset.

The predicted location of residue 18 on UCH-L1's surface (refer to Figure 5) indicates that it is not situated near either the active nor ubiquitin binding site (Johnson *et al.* (1997), (1999)). This discounts the simple explanation for Ser18Tyr's protective influence; that it opposes the Ile93Met reduction in hydrolase activity. Additionally, residue 18 is only one of a few amino acids which are not conserved between and human and other mammals (horse, mouse, and rat have Ala at position 18), which indicates non-involvement in UCH-L1 normal hydrolysis activity, and suggests the existence of a distant pathological UCH-L1 activity. This pathological activity seems to have recently been uncovered (Liu *et al.* (2002)).

Liu *et al.* (2002) discovered a novel *in vitro* ubiquityl ligase activity of the UCH-L1 dimer, that does offer a simple mechanistic explanation for the observation that the Ser18Tyr polymorphism reduces susceptibility to sporadic Parkinson's disease. This ligase activity was an unexpected outcome from cell culture experiments which indicated that UCH-L1 was responsible for the ubiquitination of α-synuclein-ubiquitin conjugates. Other predominantly hydrolytic enzymes have been seen to have a ligation mechanism that catalyses amide bond formation, though usually only under extreme conditions *in vitro*. This, as the study indicates, may be the first example of catalysis (by a hydrolase) of amide bond formation under physiologically reasonable conditions – which strongly supports its occurrence *in vivo*. UCH-L1's dimerisation dependent ubiquityl ligase activity is mechanistically reasonable, and importantly is still dependent on its hydrolase activity; once a substrate ubiquitin C-terminal

amide or ester (for instance) has been hydrolysed, a ubiquitin/ UCH-L1 intermediate is formed (refer to section 1.2) which then, instead of reacting with water (step 4 in cysteine protease catalysis), react directly with a nucleophillic lysine of another ubiquitin to produce a ubiquitin-ubiquitin amide bond (Liu *et al.* (2002)). In this process, ubiquitin does not have to be activated using ATP (which is ordinarily required in the UCS enzyme cascade – refer to section 1.1), as the ubiquitin will already be bound post C-terminal amide/ ester hydrolysis.

Liu *et al.* (2002) go on to report the Ser18Tyr (serine being the major allele) mutant as bringing about a five-fold reduction in UCH-L1's ligase activity at high concentrations, this linked to the fact that the variant displayed comparable ligase activity to that of the Wt enzyme at lower levels, does indeed indicate a protective route for the Ser18Tyr mutant. It is postulated that the Ser18Tyr variant's observed five-fold decrease in ligase activity (at high concentrations), would not elevate the concentration of cytoplasmic α-synuclein as much as the Wt protein; whose higher ligase activity would inhibit its 'normal' degradation by the potential production of undegradable, Lys^{63} linked polyubiquitin chains, which act as a non-proteolytic signal inhibiting the proteasome's activity upon α-synuclein. In order to avoid Parkinson's disease, it has become accepted that α-synuclein levels are required to be kept under the 'critical concentration' for oligomerisation (Rochet and Lansbury (2000)), owed to its accumulation resulting in synaptic damage and neurotoxicity via amyloid-like fibril formation and mitochondrial dysfunction, leading to apoptotic-like changes (Giasson *et al.* (2000), Hashimoto *et al.* (1999), Irizarry *et al.* (1998), Tompkins *et al.* (1997), van Duinen *et al.* (1999)). The Ser18Tyr variant's comparable ligase activity to that of the Wt at low levels could also be important, as the basal ligase activity of UCH-L1 would not be disturbed, which may have an important as yet unknown physiological function. The novel ligase activity uncovered by Liu *et al.* (2002) is an attractive hypothesis; for it would suggest that UCH-L1 is not produced solely for its relatively weak hydrolase activity (200-fold less than UCH-L3). It was also suggested by the authors that UCH-L1's dimerisation-dependent ligase activity could have evolved to be regulated posttranslationally and could be regulated by a number of cytoplasmic events such as synaptic vesicle binding, which would be expected to promote dimerisation.

It is also important to note that Liu *et al.* (2002) also investigated the possibility of an indirect interaction between the Ser18Tyr polymorphism and the Ile93Met mutation (refer to section 1.3.1), to help explain the incomplete penetrance of the latter. An explanation was realised in the fact that a 1:1 Wt/ Ile93Met mixture (a model of the affected heterozygotes) showed significantly much higher ligase activity to a 1:1 Ser18Tyr/ Ile93Met mixture (a potential model of the unaffected father). Thus, if ligase activity confers susceptibility, then the Ser18Tyr polymorphism could protect against Ile93Met.

In a very recent study (Naito et al. (2006), the large five-fold decrease in UCH-L1's dimerisation dependent ligase activity compared to that of the wildtype protein, is seen to have a morphological basis in enzyme dimer configuration. Using a small-angle neutron scattering (SANS) approach, this group found that the UCH-L1 dimer was a rotating ellipsoidal, contrasting to the Ser18Tyr dimer, which exhibited a much more globular appearance. This pronounced change in shape lends credence to the fact that the Ser18Tyr enzyme can bring about such a reduction in ligase activity.

1.3.2.1 UCH-L1's Ser18Tyr Polymorphism and Alzheimer's Disease

Alzheimer's disease is the most prevalent neurodegenerative disease in the world, making up over half of all dementia cases in the elderly (Small *et al*. (1997)). The clinical progression of Alzheimer's disease is made up of three main features – progressive memory impairment, progressive cortical dysfunction and neuropsychiatric disturbances (the pathophysiology and epidemiology of Alzheimer's disease is discussed in greater detail in section 1.6).

Recent work done by Xue and Jia (2006) tested the hypothesis that the Tyr allele of the Ser18Tyr polymorphism may also confer protection Alzheimer's disease, specifically in the Chinese Han population. After examining the *UCH-L1* Ser18Tyr polymorphism genotypes in 116 sporadic Alzheimer's disease patients and 123 healthy subjects, the data demonstrated, after stratification by gender, that female Alzheimer's disease patients had significantly less frequencies of the Tyr allele (the Ser18Tyr polymorphism) than the female controls (p = 0.003). From these results, the study does seem to indicate that the Tyr allele confers protection towards sporadic Alzheimer's disease in female subjects within the

Chinese population, however, similar studies are now required in other populations to ascertain the level of importance this polymorphism actually has in the pathogenesis of Alzheimer's disease globally.

1.3.2.2 UCH-L1's Ser18Tyr Polymorphism and Huntington's Disease

In addition to Parkinson's disease and Alzheimer's disease, the UCH-L1's Ser18Tyr polymorphism has also been implicated in Huntington's disease (refer to section 1.6.3) (Naże *et al.* (2002)). The study indicated that the Ser18Tyr variant could have a moderate effect in delaying the age-at-onset of Huntington's Disease , though due to the vast majority of these cases being governed by the expansion size of a triplet CAG repeat mutation in the 'HD' gene, the group was left with only a small proportion (seven) of the initial 138 Huntington's disease-diagnosed patients available for Ser18Tyr association work. Therefore, much larger studies are required if a Ser18Tyr Huntington's disease age-at-onset association of any real significance is to be unveiled.

1.3.3 UCH-L1's Methionine124Leucine Mutation (Met124Leu) and Parkinson's Disease

In addition to the highly penetrant familial Parkinson's disease Ile93Met mutation, and the protective Ser18Tyr polymorphism, one other UCH-L1 mutation has been identified – an A371C *UCH-L1* single nucleotide polymorphism (SNP) in exon 5 - leading to a Met124Leu amino acid substitution (Farrer *et al.* (2000)). This mutation was discovered in two individuals with 'probable' Parkinson's disease, and although this amino acid is conserved across species (*Equus caballus* (wild horse) and *Rattus norvegicus* (brown rat) species have a leucine at this position), the Met 124Leu variant does not seem to completely segregate with familial Parkinson's disease in this German family. Farrer *et al.* thus concluded that this UCH-L1 variant is unlikely to be pathogenic, but without association data or functional analysis of the Met124Leu substitution, the pathogenic potential must remain a possibility.

1.3.4 UCH-L1 and Dementia With Lewy Bodies

In addition to the association of UCH-L1's Ser18Tyr polymorphism with Alzheimer's disease (through the potentially protective influence of the Tyr allele), the influence of UCH-L1 on the pathogenesis of another dementia – dementia with Lewy bodies – has also been investigated. Dementia with Lewy bodies (refer to section 1.6.2) is the second commonest form of dementia and is characterised by the widespread distribution of Lewy bodies (refer to section 1.5.3) and α-synuclein-containing neuritis in the cerebral cortex (Spillantini *et al.* (1998)). With this pathological feature of dementia with Lewy bodies in mind, a study was initiated by Barrachina *et al.* (2005) to test for *UCH-L1* mRNA and protein levels in post mortem frontal cortex of dementia with Lewy bodies cases, compared with age matched controls. Their findings demonstrated down-regulation of *UCH-L1* frontal cortex mRNA and protein, which could indicate that reduced UCH-L1 expression contributes to the abnormal protein aggregation seen in dementia with Lewy bodies.

1.3.5 Gracile Axonal Dystrophic Mouse - UCH-L1 Intragenic Deletion

Further to the UCH-L1 single point mutations discussed above, an in-frame deletion of *UCH-L1* has been reported in a mammalian model of neurodegeneration (Saigoh *et al.* (1999)). The deletion includes exons 7 and 8, and has been realised within the UCH-L1 gene of the gracile axonal dystrophic (*gad*) mouse, which is an autosomal recessive mutant that exhibits sensory ataxia at an early stage, followed by motor ataxia at a later stage. This was pathologically characterised by 'dying-back' type axonal degeneration, and accumulation of ubiquinated deposits along the sensory and motor nervous systems. The *gad* allele encodes a truncated UCH-L1 lacking a segment of 42 amino acids containing the His161 residue – the general base involved in UCH-L1's hydrolysing catalysis. This strongly supports the hypothesis that an abnormal expression or function of UCH-L1, particularly effecting the active site residues – as with the Ile93Met mutation – can accelerate the formation of ubiquitinated protein inclusions (comparable to Lewy bodies) that disrupt cellular homeostasis in neural cells, leading to neurodegeneration. Though this study is generally very applicable to UCH-L1 and the pathological deterioration seen in neurodegenerative diseases, i.e. Parkinson's disease, Alzheimer's disease and dementia with Lewy bodies; the mice do not

specifically exhibit the pathological and clinical features consistent with any of these diseases, however, with such a large portion of the UCH-L1 enzyme having been truncated, it is difficult to draw any specific cross species comparisons (of neurological features) with any real confidence.

The *gad* mouse was also used in another UCH-L1 relevant study (Osaka *et al*. (2003)) in which the authors deemed it analogous to a *UCH-L1* null mutant using polyclonal antibody immunoblotting techniques of brain lysates (no compensatory UCH-L3 isozyme up regulation was identified). Concordant accumulation of UCH-L1 substrates was not detected, though a decrease in mono-ubiquitin was - specifically in the neurons. Furthermore, overexpression of UCH-L1 in the transgenic mice, as well as cultured mouse embryonic fibroblast (MEF) cells, showed a proportional increase in mono-ubiquitin levels. The authors eliminated ubiquitin transcriptional activation in the cells transfected with adeno-*UCH-L1* (UCH-L1 expressed by the adenovirus vectors), and catalysed release of mono-ubiquitin from potential UCH-L1 substrates was also ruled out (though by no means definitely). Using an immunoreactivity approach within neuronal dopamine producing cells, and his-tagged wildtype and two UCH-L1 mutants – one lacking hydrolase activity (UCH-L1^{C90S}) and one with a disrupted ubiquitin binding domain (UCH-L1^{D30K}) - UCH-L1 affinity for ubiquitin was asserted as the critical factor in maintaining these ubiquitin levels. Moreover, further work in this study showed that ubiquitin half life of control MEF cells could be increased to a level comparable to those expressing transfected adeno-*UCH-L1* if a lysosome inhibitor was added (2,3-*trans*-epoxysuccinyl-L-leucylamide-3 methyl butane ethyl ester (EST)). This led to the authors' conclusion that UCH-L1's ability to bind ubiquitin is potentially critical in sequestering it from lysosomal degradation, which in turn maintains its intracellular levels.

1.3.6 Ubiquitin Carboxyl-Terminal Hydrolase-L1 - Discussion

Neuron specific UCH-L1 was one of the first DUBs to be described, though its specific functioning still has to yet to be clearly illuminated. Leroy *et al.* uncovered the Ile93Met mutation within a Parkinson's disease sufferer in 1998, and together with the associated loss in hydrolase function of this mutant (around fifty percent), firmly established UCH-L1 as a critical DUB of the brain. Almost disappointingly (to those in the UCH-L1 field), Larsen *et al.* (1998) confirmed the *in vitro* substrate specificity of UCH-L1, in which only small C-terminal adducts (amides or esters) of ubiquitin could be rapidly cleaved. The ubiquitin proprotein, though generally described in the literature as a UCH-L1 substrate, was shown to be cleaved only very slowly, perhaps indicating an evolutionary redundant role. Though deubiquitinating the abundant by-products of cellular metabolism did endow UCH-L1 with a very important physiological role in maintaining cellular levels of 'free' ubiquitin, the fact that its hydrolase activity was over 200-fold less than other UCH subclass members, meant that its perceived hydrolase function (or disturbance there of) was probably not enough on its own to warrant the pathological decay seen in the German proband.

The breakthrough required came in 2002, when Liu *et al.* reported the novel *in vitro* ubiquityl-ligase activity of the UCH-L1 dimer. Suddenly, UCH-L1 had become a multifunctional enzyme, with an E3-like ligase activity which was mechanistically still dependent on its hydrolase activity. (This is generally understated in DUB/ UCH(-L1) literature, but the only biochemically reasonable proposition for its dimerisation dependent ligase activity to date, depends on its ability to form a ubiquitin/ UCH-L1 intermediate (sometimes referred to as an acyl enzyme due to the loss of a hydroxyl group), after hydrolysis of a C-terminal ubiquitin bond). The Ser18Tyr variant's observed five-fold reduction in this ligase activity was presented as the physiological rationale for the Tyr allele's observed protective effect (with respect to Parkinson's disease) in Japanese and German populations (possibly indicating a protective environmental cofactor), and could also be the main biochemical factor contributing to the allele's observed protective effect seen in female Alzheimer's disease patients in China, along with its proposed delayed age-at-onset effect seen in Huntington's disease patients.

UCH-L1's ability to ubiquitinate such substances as α-synuclein, opened up a vast array of possibilities for its actual physiological role. Only in the past two to three years has the true versatility of the UCS (briefly summarised in section 1.1) been uncovered, with much yet still to be learned (Mayer & Layfield (2005)). UCH-L1 is thought to promote the formation of K^{63}-linked polyubiquitin chains (Liu *et al.* (2002)), which would point towards a role in the UCS away from the UPS and 26S proteasome recruitment.

The work done by Osaka *et al.* (2003) with the *gad* mouse has clouded UCH-L1's overall picture, as it demonstrated the apparent ability of UCH-L1 to raise and maintain ubiquitin levels independent of its hydrolase activity, and thus potentially (from what can be inferred to date) its ligase activity. However, no study of this nature has been reported since, and it should be noted that this investigation was far from exhaustive in ruling out *gad* mouse UCH-L1 substrate hydrolysis. Nevertheless, it certainly threw open to conjecture the true nature of UCH-L1's physiological role. With this in mind, it is worth indicating that a very recent study by Manago *et al.* (2005) has uncovered the first evidence of a relationship between UCH-L1 and a neurotransmitter receptor - specifically P2X - which are widely distributed in the brain, and involved with various biological activities including neurosecretion. The results from this study showed that UCH-L1 overexpression in PC12 cells (Rat adrenal pheochromocytoma (adrenal gland tumour) cell line) potentiated ATP-induced currents due to the activation of these receptors, though again, somewhat perplexingly, through the use of a C90S mutant (refer to section 1.3.4) these effects were deemed to be independent of UCH-L1's hydrolase activity. However, work *in vitro* on one neuronal precursor cell line is not sufficient to pass judgement on the biochemical mode of UCH-L1's effect, though the effect realised on the PX2 receptors does allow speculation that UCH-L1 may play an important role in synaptic activity.

This finding could eventually be seen as a turning point, for so far no specific neurological role has been put forward for UCH-L1. Its loss of hydrolase function, which is most probably linked mechanistically to its ligase activity, has certainly been shown to be pathogenic (Ile93Met), but the current state of knowledge on UCH-L1 goes no further than it as a general 'hydrolyser' of ubiquitinated glutathiones and polyamines to 'prime' its yet unknown E3-like ubiquityl-ligase activity.

In 2002, when Liu *et al.* first uncovered UCH-L1's ligase activity, they also found that UCH-L1 coimmunoprecipitated with α-synuclein in synaptic vesicle fractions. At that time α-synuclein's specific biological role was completely unknown, apart from that it was a major substrate of Parkin – a well documented E3 ligase. However, very recently, an illuminating study done by Chandra *et al.* (2005) revealed that α-synuclein acts as a dualistic molecular chaperone with cysteine-string protein • (CSP•) at the synapse, assisting in the folding and refolding of SNARE proteins which are essential in neurotransmitter release, vesicle recycling and synaptic integrity. When this is borne in mind with the fact that overexpression of *UCH-L1* causes an accumulation of α-synuclein (Liu *et al.* (2002)), and can also activate P2X receptors (Manago *et al.* (2005)) whose subunits have been shown to be located proximal to proteins of the SNARE complex (Barden *et al.* (1999)), an intriguing new role for UCH-L1 at the synapse, in my opinion, begins to unveil itself. Perhaps its ability to ubiquitinate α-synuclein could provide an extra level of control to α-synuclein's role at the synapse, or perhaps it straightforwardly inhibits proteasomal degradation of α-synuclein (with K^{63}-linked polyubiquitin chains) which allows its beneficial presynaptic generation to continue.

This potential role of UCH-L1 could explain the results obtained by Osaka *et al.* (2003) in the *gad* mouse – where overexpression of *UCH-L1* led to increased levels of mono-ubiquitin. UCH-L1 could regulate α-synuclein at the synapse through K^{63}-linked ubiquitination: after each facilitating interaction in the folding/ refolding cycle of the SNARE proteins, post neurotransmitter release, the α-synuclein could be deubiquitinated by a yet unidentified DUB that associates within the SNARE complex, ready to be incorporated again into another 'neurotransmitting cycle'. This would probably cause a proportional amount of ubiquitin to be present at the synapse, dependent upon how much UCH-L1 dimer there is present, governing the amount of α-synuclein being ubiquitinated, and thus deubiquitinated after each cycle. In this hypothetical model, it could also be speculated that UCH-L1 concentration at the synapse (and thence its dimerisation dependent ligase activity) could be a rate-limiting step, in view of α-synuclein potential activation, in neurotransmitter release at the synapse.

Further to this, though the increased ligase activity (five-fold) of the UCH-L1 wildtype dimer may lead to increased susceptibility of sporadic Parkinson's disease and Alzheimer's disease, and possibly an earlier age-at-onset of Huntington's disease, in later life (with

reference to the Ser18Tyr variant's *in vitro* activity and its protective effect in certain populations), what would be very interesting to elucidate, is whether individuals carrying the Ser18Tyr allele realise a reduced level of neurotransmitter activity throughout life, caused by reduced UCH-L1 α-synuclein ubiquitination/ activation. This could then lead to a reduced build up of protein aggregates and their attributable cytoplasmic stresses (refer to section 1.5.3) in later life (depending on certain apparent environmental influences), when protein turnover and degradation pathways become less efficient, thus explaining the potential protective effect.

This potential synaptic role for UCH-L1 also provides a model to help explain Ile93Met pathogenesis in familial Parkinson's disease. As the current model determines (Liu *et al.* (2002)), UCH-L1's hydrolase activity is necessary to form the UCH-L1/ ubiquitin intermediate which is required for ubiquityl-ligase activity. Because hydrolysis of ubiquitin C-terminal amides and esters by UCH-L1 is so rapid (Larsen *et al.* (1998)), a reduction in this activity of around fifty percent, which the Ile93Met mutant exhibits, should not effect synaptic α-synuclein turnover ordinarily – the fact that the German proband harboured this mutation free of illness till 51 years of age strongly supports this (Leroy *et al* (1998)). However, it is very possible that later in life this hydrolysis step could become rate limiting, if it realised such a decrease in enzymatic kinetics. Once a threshold is reached, where α-synuclein cannot be (K^{63}) ubiquitinated by UCH-L1 as quickly as it is deubiquitinated by the 'SNARE complex DUB', then α-synuclein would start to accumulate, reaching the 'critical concentration' described by Rochet and Lansbury (2000) (refer to section 1.3.2), and overload the UPS pathways. This would in turn lead to aggregation of α-synuclein and related proteins (e.g. parkin and synphilin-1 – discussed below), especially within neurons of the substantia nigra where *UCH-L1* is transcribed at higher levels (Leroy *et al.* (1992)), causing synaptic damage and neurotoxicity to this region (in the first instance) via amyloid-like fibril formation and mitochondrial dysfunction, leading to the symptoms observed in Parkinson's disease. Incomplete penetrance (eighty percent from what can be ascertained – see Figure 4) of this mutation in the family of the German proband, as also speculated by Liu *et al.* (2002), could be owed to the protective Ser18Tyr allele. The Ser18Tyr:Ile93Met UCH-L1 dimer model of the unaffected heterozygote showed a significantly reduced ligase activity *in vitro* compared to that of the Wt:Ile93Met dimer model (Liu *et al.* (2002)). This translated *in vivo*, could

potentially have maintained α-synuclein levels at tolerable levels within the UPS owed to the reduction in general α-synuclein recruitment to the synaptic SNARE complex throughout life, brought about by UCH-L1's lower ligase activity in this instance.

Whatever the specific function and mechanisms of UCH-L1 at the synapse/ in the brain, it is fairly clear that UCH-L1's hydrolase/ ligase function, and the variation there of, is intimately linked to Parkinson's disease pathogenesis. A recent study (Barrachina *et al.* (2005)) not only indirectly supports this statement further, but it also correlates with a novel hypothesis put forward describing Lewy bodies (the pathological hallmark of Parkinson's disease) as a protective attempt by the neuron to sequester α-synuclein from the cytoplasm – to maintain levels below the critical concentration (refer to section 1.5.3), which results in a further refined hypothesis. The study clearly showed, through post mortem analysis of Parkinson's disease sufferers' brains, that UCH-L1 protein levels were reduced in the substantia nigra (where UCH-L1 is potentially expressed at higher levels compared to the rest of the brain (Leroy *et al.* (1992)) of only those Parkinson's disease cases exhibiting Lewy body pathology. This does suggest that that Lewy body formation is an active process in the neuron to regain control of cytoplasmic α-synuclein concentration, for it is coupled with apparent UCH-L1 down regulation - an effective route, bearing in mind UCH-L1 as a probable α-synuclein ubiquitinator, to curb any further cytoplasmic α-synuclein neuronal build up. Moreover, together with a down regulation of other deubiquitinating enzymes in response to Lewy body build up – especially those potentially realised in the SNARE complex involved in α-synuclein deubiquitination at the synapse (discussed above) – indicates a physiological response that can viably explain how an individual can survive with Lewy body pathology without ever presenting a disease phenotype.

The very fact that *UCH-L1* is ordinarily transcribed at higher levels in the substantia nigra, seems to be one of the primary reasons why any variation in its functioning affects the neural pathways involved with voluntary movement – as seen in Parkinson's disease – in the first instance. However, variation in UCH-L1 function also seems to be associated with dementias whose pathological features are concentrated elsewhere in the brain, i.e. Alzheimer's disease and dementia with Lewy bodies. In these cases, UCH-L1's variation in

function may only have an effect on these areas of the brain when in combination with yet unspecified genetic and/ or environmental factors.

1.3.7 Summary of *UCH-L1* Neurodegenerative Disease Association

Below is a summary in tabular form of the association that *UCH-L1* has been shown to have with various neurodegenerative diseases: -

Mutation/ Pathogenesis	Neurodegenerative Disease	Effect	Reference
Ile93Met (C277G)	Familial PD	Highly pathogenically penetrant (80%)	Leroy *et al.* (1998b)
Ser18Tyr (C54A)	Idiopathic PD	Protective in American (German founder), Japanese & German populations.	Maraganore *et al.* (1999), Zhang *et al.* (2000), Satoh *et al.* (2001), Wintermeyer *et al.* (2000).
Ser18Tyr (C54A)	Idiopathic AD	Protective in Chinese Han population.	Xue & Jia (2006)
Ser18Tyr (C54A)	HD	Delayed age of onset.	Naze *et al.* (2002)
Met124Leu (A371C)	Familial PD	Potential pathogenic role.	Farrer *et al.* (2001)
↓ Regulation of *UCH-L1* mRNA	DLB	Pathogenic role.	Barrachina *et al.* (2005)

Table 2 – Summarises the association that *UCH-L1* has with various neurodegenerative diseases.

- *Disease Abbreviations*: AD, Alzheimer's disease; DLB, Dementia with Lewy Bodies; HD, Huntington's disease; PD, Parkinson's disease.

1.4 Other UCS Associated Genes Implicated in Parkinson's Disease

In addition to *UCH-L1*, three other genes associated with the UCS (incorporating the UPS) have been strongly implicated in the pathology of Parkinson's disease – these are *α-synuclein* (already discussed above in reference to UCH-L1 – refer to section 1.3.6), *parkin* and *synphilin-1*. Abnormal expression or function of each of their respective proteins seems to cause a failure in the UCS/ UPS, leading to the realisation of the neuronal stresses associated with Parkinson's disease. All three of these gene products have not only been shown to be intimately associated with one another in neuronal cells, but they have all interestingly been localised at the synapse. The Parkinson's disease associated mutations identified in these three genes and their specific association with UCS malfunctions are discussed in turn below:-

1.4.1 *α-Synuclein*

Two point mutations (Alanine(Ala)53Threonine(Thr) and Ala30Proline(Pro)) have so far been identified within the α-synuclein gene in families with dominantly inherited Parkinson's disease (PD cases – *n* = 193, Controls – n = 200. OR = x12.8) (Kruger *et al.* (1998), Polymeropoulos *et al.* (1997)). These cases only account for a small percentage of familial cases of Parkinson's disease, however their discovery led to the finding that Lewy bodies and Lewy neurites (the major pathological hallmark of Parkinson's disease - refer to section 1.5.3) in sporadic Parkinson's disease patients stain very strongly for α-synuclein. This suggests that α-synuclein is a major component of Lewy body filaments, not only in patients carrying mutations in the α-synuclein gene, but also in patients with sporadic Parkinson's disease (Mezey *et al.* (1998), Spillantini *et al.* (1998)).

α-Synuclein being the major component of Parkinson's disease Lewy bodies was initially explained by a relatively simple hypothesis. It postulated that the amino acid changes in the α-synuclein protein associated with Parkinson's disease may favour the β-pleated sheet conformation, which in turn may lead to an increased tendency to form aggregates – possible precursors to Lewy bodies (Goedert *et al.* (1998)). However, more recent hypothesises are relatively more complex, in that they indicate a possible inhibitory interaction with the 26S

proteasome. Most *in vitro* studies indicate that α-synuclein is degraded by the 26S proteasome (Bennett *et al.* (1999)). In one study it was shown that the proteasome degrades both wild type and Ala53Thr (familial) mutant α-synuclein, but the mutant α-synuclein was degraded more slowly, indicating a disturbed protein degradation which could give rise to Lewy body formation (Stefanis *et al.* (2001)). In another study (genetic), it was shown that a complex dinucleotide repeat in the promoter region of the α-synuclein gene was associated with sporadic Parkinson's disease (Kruger *et al.* (1999)). It has been shown in luciferase reporter assays that certain allele variations of this dinucleotide repeat enhances expression of the gene (Touchman *et al.* (2001)). If this is borne in mind with other studies which have shown a decreased 26S proteasome activity induced by *α-synuclein* over-expression *in vitro* (Lee *et al.* (2001)), possibly due to an inhibitory interaction of α-synuclein with a subunit of the 19S regulatory complex (19S RC) as seen in rats (Stefanis *et al.* (2001)) – accumulation, aggregation and disrupted degradation of α-synuclein leading to Lewy body formation is certainly a possibility.

This fits in well with UCH-L1's newly hypothesised control (through its dimerisation dependent ubiquityl-ligase activity) of α-synuclein's recently uncovered function at the synapse (refer to section 1.3.6), which could readily be overwhelmed by *α-synuclein* overexpression – leading to the detrimental effects on the 26S proteasome complex described above. Furthermore, this also lends credence to the hypothesised pathological model for *UCH-L1*'s highly penetrant Ile93Met familial Parkinson's disease mutation (refer to section 1.3.6), as it involves the build up of α-synuclein within the UPS (in later life), which would concordantly bring about the proteasomal pathological routes discussed.

1.4.2 *Parkin*

It has been shown that several mutations in the parkin gene cause autosomal recessive juvenile Parkinson's disease (*n* = 3) (Kitada *et al.* (1998)). Furthermore, a significant proportion of patients with sporadic early-onset Parkinson's disease have also been shown to carry a homozygous mutation in the parkin gene (early onset families – *n* = 73, normal PD families – *n* = 100. *p* = 0.001) (Lucking *et al.* (2000)). Importantly, patients with mutations in the parkin gene do not exhibit the main pathological hallmark of Parkinson's disease – Lewy bodies. This however can be explained by parkin's function in the UPS as an E3 ligase (refer to section 1.1), i.e. if parkin is malfunctioning as a 'classical' E3 ligase (i.e. giving rise to K^{48}-linked polyubiquitin chain formation targeted by the 26S proteasome – refer to section 1.1.1), then there will be a concordant accumulation in parkin's proteolytic target substrates (O-glycosylated α-synuclein (a small modified sub-population there of) (Shimura *et al.* (2001)), CDCrel-1 - a protein associated with synaptic vesicles (Peng *et al.* (2002)), and a transmembrane protein called parkin-associated endothelin receptor-like receptor (Pael-R) (Imai *et al.* (2001))) due to the failure of them being ubiquitinated, which would also prevent their aggregation, and thus thwart Lewy body deposition (Shimura *et al.* (2001), Imai *et al.* (2001)). However, it is hypothesised that the accumulation of these proteins in Parkinson's patients exhibiting mutations in the parkin gene, brings about the severe degeneration of dopaminergic neurons and reactase gliosis in the zona compacta of the substantia nigra seen in this form of the disease (Gasser (2001)). Further to this, studies have shown that the accumulation of parkin's substrates (especially Pael-R) has resulted in unfolded protein stress and induced cell death in neural cell lines (Imai *et al.* (2001)).

These findings suggest that it could be the early step of conformational change and protofibril formation, and not specifically the production of Lewy bodies, which mediate toxicity to the dopaminergic substantia nigra neurons in Parkinson's disease sufferers (Conway *et al.* (2001)). Moreover, it could be postulated that Lewy bodies might actually provide protection through the sequestration of ubiquinated proteins, and toxicity might only occur in the absence of further sequestration capacity (refer to section 1.3.6).

It should be noted here, that although parkin malfunction discussed above may lead to an accumulation of its substrates, which would ordinarily be targeted for degradation, which leads to the neuronal stresses realised in Parkinson's disease sufferers' neurons. It is parkin's role as a functioning classical E3-ligase, localised at the synapse (Fon *et al.*, unpublished data), which may contribute to the pathogenesis hypothesised in *UCH-L1*'s Ile93Met mutation, as any (O-glycosylated) α-synuclein not K^{63}-ligated to ubiquitin by UCH-L1, will be 'K^{48}-tagged' by parkin - which will lead to aggregation and the UPS overload already described (refer to section 1.3).

Recently it has been determined (Lim *et al.* (2005)), that Parkin not only functions as a classical E3-ligase in the UPS, but it also functions in a non-classical, proteasomal-independent manner in the wider UCS that involves K^{63}-linked polyubiquitin chains. This reveals parkin as a dual-functioning ubiquitin ligase; intriguingly its only 'K^{63} substrate' known to date is synphillin-1 - which is discussed below.

1.4.3 *Synphilin-1*

Synphilin-1 has been identified as both an α-synuclein-interacting protein (Engelender *et al.* (1999)) and a substrate of the non-classical, proteasomal-independent ubiquitin ligase activity of parkin (Lim *et al*. (2005)). This not only supports *synphillin-1* as a good candidate gene in the pathology of Parkinson's disease, but the fact that it has also been localised at the synapse (Ribeiro *et al.* (2002)), suggests a possible role in neurotransmitter release in conjunction with α-synuclein (and potentially UCH-L1) which parkin may well regulate through mono/ polyubiquitination.

With reference to Parkinson's disease pathology, the gene encoding synphilin-1 is located on the long arm of chromosome five, which was indicated as a candidate locus for Parkinson's disease in a study which used a whole genome mapping approach in large cohorts of familial Parkinson's disease patients (markers – *n* = 344, families – *n* = 174; family members – *n* = 870, PD – *n* = 378, unaffected – *n* = 379) (Scott *et al.* (2001)). It has been shown, in cultured cells, that the co-expression of synphilin-1 and α-synuclein results in the formation of Lewy body-like intracytoplasmic aggregations (Engelender *et al.* (1999),

Kawamata *et al.* (2001)), which is further backed up by the fact that synphilin-1 is a component of Lewy bodies in the brains of sporadic Parkinson's disease patients (Wakabayashi *et al*. (2000)).

One group Holzmann *et al*., unpublished data) has identified an Arginine(Arg)642Cys mutation in two independent sporadic Parkinson's disease patients (whilst not in healthy controls). Whether or not this sequence variation disrupts synphilin-1's interaction with α-synuclein at the synapse, or interferes with parkin's potential 'K^{63} regulation' of synphillin-1, is the subject of ongoing research.

The synaptic localisation of UCH-L1, α-synuclein, parkin and synphilin-1, their association with the UCS and individual implication in Parkinson's disease, underlies the importance of determining the true nature and function of their individual interactions with one another. Current state of knowledge in the field alludes to a role for UCH-L1 in controlling α-synuclein's facilitation in neurotransmitter release (hypothesised in section 1.3.6), and establishes a dual-function for parkin in 'ubiquitin-tagging' a modified form of α-synuclein for 26S proteasome degradation, as well as a contrasting regulatory ubiquitin modification role with synphillin-1, presumably in its intriguing, but yet unknown association with α-synuclein. These series of interactions within neuronal cells, as well as the individual enzymes themselves, will need to be studied much further to fully elucidate the true nature of the UCS/ UPS in synaptic neurotransmission pertaining to the pathology of Parkinson's disease.

1.5 Parkinson's Disease

Neurodegenerative diseases are becoming increasingly prevalent with the aging of the general population; a significant demographic change in the human population was realised in the twentieth century, which shifted life expectancy to the upper age ranges. Arguably, the main determinant of 'quality of life' for this aging population is the normal process of neuronal aging in the central nervous system, especially diseases which can accelerate this neuronal loss. Neurodegenerative diseases (as they are collectively known) are amongst the most important contributors to human disability and disease in the world today, with Alzheimer's disease being the most prevalent (refer to section 1.6.1), followed by Parkinson's disease. There is also a large number of rarer neurodegenerative diseases, such as dementia with Lewy bodies (refer to section 1.6.2) and Huntington's disease (refer to section 1.6.3), and though they each only effect a small number of patients, taken together they certainly have a detrimental impact on the aging population.

1.5.1 Physiological Characteristics of Parkinson's Disease

Parkinson's disease is so called owed to the fact that James Parkinson first clinically described the neurological disorder in 1817 in an essay titled "Shaking Palsy" – which literally means 'loss of voluntary movement with tremors'. Parkinson's disease is characterised by four main neurological features (which are also known as the cardinal symptoms or signs): paucity of movement, rigidity, tremor and postural instability. Paucity of movement is one of the most disabling features of the disease (Lyons *et al.* (1997)) and consists of three main entities: slowness of movement (bradykinesia), reduced movement (hypokinesia) and an inability to initiate movement (akinesia). Rigidity is defined as the increase in resistance to passive movements around a joint, but this is mainly considered a sign rather than a symptom of Parkinson's disease, owed to the patient's feeling of 'stiffness' often being attributable to bradykinesia (Jankovic (1992)). Around 75 percent of patients with Parkinson's disease realise rest tremor (Gelb *et al.* (1999)), and this is the symptom that usually causes the patient to seek medical counsel. The rest tremor associated with Parkinson's disease is relatively slow (around four to six hertz (Hz)), and is most prominent in the hands, with it also effecting

the feet and jaws (Sawle (1999)). Postural instability is usually the last of the core symptoms to be realised, though it is deemed the most disabling (Lyons *et al.* (1997)) owing to the danger of the patient falling. As well as these four main symptoms of Parkinson's disease, there are less common associated features which include autonomic dysfunction, cognitive disturbance and dysphagia (difficulty in swallowing) (Dunnet and Bjorklund (1999)).

1.5.2 Parkinson's Disease Epidemiology

As indicated above, Parkinson's disease is the second most common form of neurodegenerative disease. Estimates of the number of new Parkinson's disease cases in a year, or its incidence (generally thought of as the best measure of disease frequency as it is not modified by factors effecting survival (Korell & Tanner (2005)), are in the range of 4-20/ 100'000 per year (Twelves *et al.* (2003), (Rosati *et al.* (1980), (Rajput *et al.* (1984)). This variability is most probably down to a difference in the age distribution of the populations, and varied Parkinson's disease diagnostic criteria between the groups, rather than a true difference in disease frequency. Incidentally, there are around 15'000 new cases in the UK each year (Krugger *et al.* (1998)). The prevalence (total number of cases at one time) of Parkinson's disease in the UK is about one in two hundred people (Krugger *et al.* (1998)), though a widely accepted figure world wide is 200/ 100'000 population (Clough *et al.* (2003)).

1.5.3 Pathology of Parkinson's Disease

The main pathological feature of Parkinson's disease is the degeneration of dopaminergic (dopamine containing) neurons in the substantia nigra (refer to Figure 6), which biochemically leads to the loss of dopamine, as well as its metabolites (including homovanillic acid (HVA) and 3, 4-dihydroxyphenylacetate (DOPAC)), its biosynthetic enzyme (tyrosine hydroxylase (TH)) and the dopamine transporter (DAT) (Jenner and Olanow (1998)). The dopaminergic neurons of the substantia nigra is one of two closely related midbrain groups of dopamine producing cells that make up the brain's dopaminergic diffuse modulatory systems, the other being the ventral tegmental area (Figure 6). The dopamine cells of the substantia nigra projects axons to the striatum, specifically the caudate nucleus and the putamen - which are both part of the basal ganglia in the basal forebrain - crucial in the normal control of voluntary and involuntary movement.

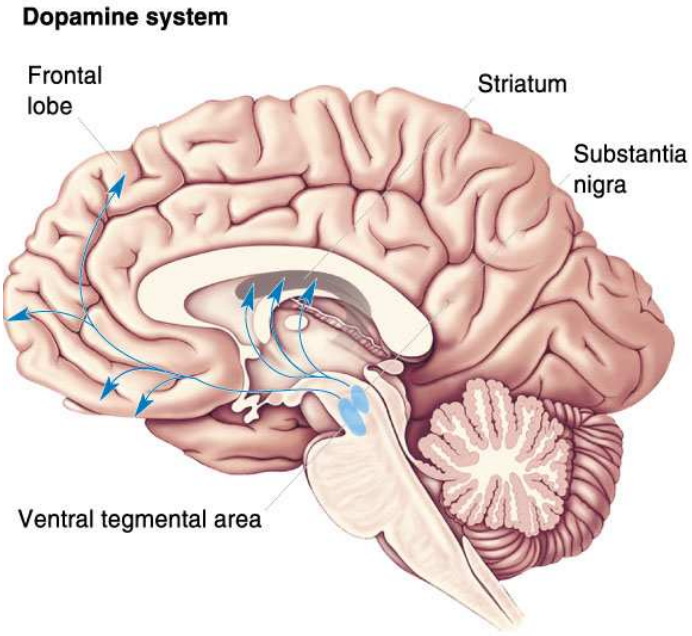


Figure 6 – The dopaminergic diffuse modulatory systems arising from the substantia nigra and the ventral tegmental area in the midbrain (Bear *et al.* (1996)).

Dopamine influences the activity of two main motor pathways within the basal ganglia. The function of these pathways is to control the activity of the globus pallidus interna, through an excitatory (indirect pathway) and an inhibitory (direct pathway) effect. The globus pallidus interna has inhibitory connections to the thalamus, which in turn has excitatory connections to the motor cortex which facilitate movement. Dopamine increases the direct pathway's inhibitory activity, whilst decreasing the indirect pathway's excitatory activity, to give rise to a net reduction in the inhibitory effect of the globus pallidus interna upon the thalamus, allowing the facilitation of movement (Lang & Lozano (1998a), (Lang & Lozano (1998b)).

The progressive loss of dopamine in Parkinson's disease results in an imbalance of normal activity within the basal ganglia's two main pathways: there is a reduction in the direct pathway's inhibitory activity, and an increase in the indirect pathway's excitatory activity. An increase in globus pallidus interna activity is the overall net effect, which consequently inhibits the thalamo-cortical outflow – inhibiting movement (Lang & Lozano (1998a), (Lang & Lozano (1998b)). Though this archetypic schema is currently generally accepted in the field as the basic role of dopamine in facilitating movement, some inconsistencies exist, which have prompted continual modifications and resultant modifications (Blandini *et al*. (2000), Rodriquez *et al*. (2000)).

As well as dopaminergic loss within the substantia nigra, the presence of Lewy bodies (Figure 7) in areas of neuronal degeneration is the pathological hallmark of Parkinson's disease.

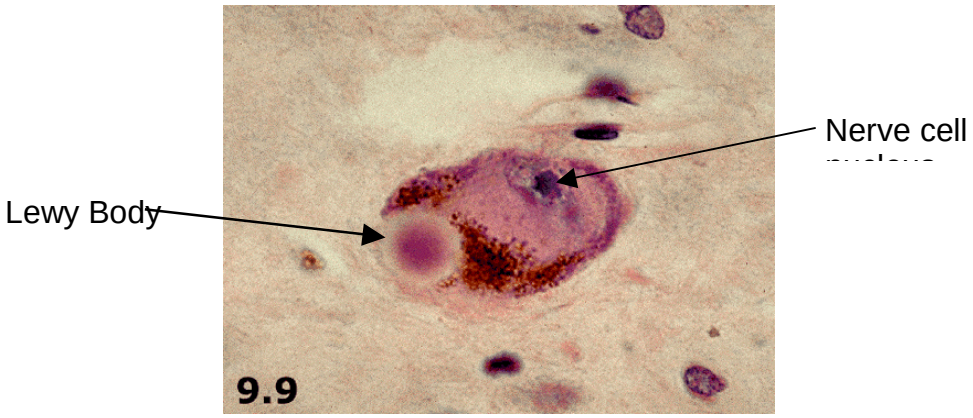


Figure 7 – A substantia nigra neuron with a Lewy body inclusion (www.akronchildrens.org/neuro (2004))

Lewy bodies were first described by Friedrich Lewy in 1912 in the brain stem of patients with 'paralysis agitans' (Parkinson's disease) and are typically spherical inclusions that have a diameter of approximately 4-30μm, with a hyaline eosinophillic core and a pale peripheral halo (Figure 7). They are typically found in the substantia nigra, though they can be widespread, being found in other areas of the brain such as the hypothalamus, sympathetic ganglia, as well as the mesolimbic and mesocortical dopaminergic neurons, where Parkinson's disease neuronal loss also occurs (Gibb and Lees (1988)).

As already described/ discussed above (refer to sections 1.3.1, 1.3.6 and 1.4.1), α-synuclein is a major constituent of Lewy bodies (Spillantini *et al.* (1998)). Furthermore, other UCS α-synuclein-interacting proteins such as UCH-L1 and synphillin-1 have also been shown to be present within Lewy body filaments (Wakabayashi *et al.* (2000), Doran *et al.* (1983)). Towards the end of the last century, Lewy bodies were generally described in the literature as the intracellular pathological structures which were directly responsible for the lethal stresses realised within the dopaminergic neurons of the substantia nigra. However, as discussed above (refer to section 1.4.2), the fact that Parkinson's disease sufferers carrying mutations in the parkin gene did not exhibit Lewy bodies, and that only an accumulation (as opposed to aggregation of parkin's substrates as seen in Lewy bodies) is required to induce neuronal cell

death (Imai *et al.* (2001)), led to an alternative hypothesis. This hypothesis proposed that Lewy bodies arise in a protective capacity, sequestering ubiquitinated α-synuclein in an attempt to reduce its cytoplasmic concentration before it reaches the 'critical levels' described by Rochet and Lansbury (2000), which overload the UPS pathways and give rise to the synaptic damage and neurotoxicity which causes Parkinson's disease associated neuronal decay. Thence, the protective capacity of Lewy bodies has been overwhelmed in Parkinson's disease cases that exhibit them. Moreover, this protective hypothesis explains why occasionally normal healthy individuals harbour Lewy bodies in the brain which are unearthed during autopsy. Furthermore, it is important to note that parkin's ligase activity is probably intimately involved in Lewy body formation, owed to the fact that a fully functioning parkin is required for their formation.

Although Lewy bodies are the pathological hallmark of Parkinson's disease, it is worth mentioning that they are not restricted to these disorders. They are also the hallmark of dementia with Lewy bodies, and have also been described as a secondary pathology in a number of other disorders, including progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, motor neuron disease, Hallervorden-Spatz disease, sporadic and familial Alzheimer's disease, subacute sclerosing panencephalitis, and Down's syndrome (Jellinger (2001)). The significance of these associations remains to be formally determined, though similarly to Parkinson's disease, Lewy body formation could arise in all these neurological disorders as a physiological protective strategy in areas of the brain that are sensitive (such as the substantia nigra in Parkinson's disease) to the particular stresses generated by the relevant intraneuronal impairment.

1.5.4 Environmental Factors in Parkinson's Disease

Presently there is clearly very good evidence that several different highly penetrant gene mutations can initiate abnormalities in the function or structure of three main proteins, which are all involved with the UCS at the synapse, which are capable of inducing stresses on the dopaminergic cells in the substantia nigra, leading to Parkinson's disease. It must however be borne in mind that from what is currently known, the vast majority of these cases give rise to autosomal dominant or recessive Parkinson's disease, i.e. familial Parkinson's disease, which only make up a small proportion of those people diagnosed – there is a subset of about 5 to 15% of families with more than one affected family member (Gasser (2001)). The highly penetrant gene mutations which give rise to familial Parkinson's disease have allowed elucidation of the UPS and its saturation as central to the aetiology and pathology of sporadic, as well as familial Parkinson's disease, whose general pathology seem identical, i.e. stresses upon the dopaminergic neurons of the substantia nigra.

Though studies have shown that the UCS genes identified in monogenically inherited forms of Parkinson's disease can also harbour polymorphisms associated (positively or negatively) with sporadic forms of the disease (i.e. *UCH-L1*'s Ser18Tyr polymorphism and *α-synuclein*'s promoter dinucleotide repeat), no consistent findings in every population studied have emerged, and to date no specific gene polymorphism has been unequivocally associated with sporadic Parkinson's disease (Gasser (2001), Tan *et al.* (2000)) (synphilin-1's Arg642Cys mutation still awaits confirmation). Nevertheless, the substantial role for genetic factors in sporadic Parkinson's disease has been clearly demonstrated in monozygotic twin studies (Piccini *et al.* (1997), Burn *et al.* (1992)). Not including genes involved with the UCS, polymorphisms within eleven other genetic loci have been associated with sporadic Parkinson's disease (Table 1), including genes encoding proteins involved in dopamine metabolism, xenobiotic metabolising hepatic enzymes (that are thought to protect from environmental toxins – i.e. debrisoquine-4-hydroxylase – discussed later) and more recently, neuronal survival genes (Gasser (2001), Momose *et al.* (2002)).

Other Genes Associated with Sporadic Parkinson's Disease	Locus
Dopaminergic	
Dopaminergic transporter	5p15.3
Dopamine D2 receptor	11q22-23
Catechol-*O*-methyltransferase	22q12.1
Monoamine oxidase B	Xp11.3
Involved in Detoxification of Metabolites	
Debrisoquine-4-hydroxylase	22q13
N-acetyltransferase	8p23.1-21.3
Heme oxygenase 1	22q12
Quinone oxidorductase 2	6p25
Lipoproteins	
Apolipoprotein E	19q13.2
Neuronal Survival	
Nurr 1	2q22-q23
Brain-derived neurotrophic factor	11p13

Table 3 – Indicates genes (UCS unrelated), and their respective genetic loci, that are associated with sporadic Parkinson's disease (taken and modified from Gasser (2001)).

The fact that none of these numerous genes, nor those discussed from the UCS, have been unequivocally associated with sporadic Parkinson's disease, and that many of these genes are involved in the detoxification of metabolites from environmental toxins, seems to suggest that sporadic Parkinson's disease cases might well be the result of an interaction between one or more 'susceptibility genes' (genetically predisposing these individuals) with other non-genetic causes, i.e. environmental influences.

Several epidemiological studies have been undertaken to determine which environmental factors may be important in the onset of sporadic Parkinson's disease. Rural living has been identified by several surveys to increase the relative risk of the development of Parkinson's disease, although this observation has not been seen consistently (Barbeau *et al.* (1987), Gorrel *et al.* (1996), Fall *et al.* (1999), Semchuk (1992)). The agricultural industry is certainly associated with rural living, and work in this environment has also been suggested as a risk factor for Parkinson's disease (Seidler *et al.* (1996)). These observations have raised the issue as to whether pesticide use may contribute to the increased risk of Parkinson's disease for those in rural areas. An organochloride pesticide – Dieldrin – was found to be a risk factor in one case-control study in Germany, and in another investigation, was found to

be present in 6 out of 20 Parkinson's disease brains and in no controls (Semchuk *et al.* (1991)). Another widely available and commonly used pesticide – Rotenone – which is often used to control fish stocks in reservoirs, has been shown to induce selective dopaminergic cell death, intraneuronal ubiquitin rich protein inclusions (Lewy body-like structures) and a motor deficit in rats (Betarbet *et al.* (2000)).

The work carried out on both of these pesticides is of great significance to the idea that environmental agents may contribute to Parkinson's disease. Furthermore, perhaps the most important observation that has supported the possibility of this environmental factor hypothesis, is the fact that inadvertent exposure to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) can induce parkinsonism (generic term referring to slowness and mobility problems that look like Parkinson's disease, but not referring to the disease itself) in humans within 7 to 14 days (Langston *et al.* (1983)). This study clearly established that an 'environmental' agent could, through specific uptake and conversion mechanisms, target the neurons of the substantia nigra and cause degeneration and a Parkinson's disease-like picture. MPTP is now routinely used as a toxin to model Parkinson's disease in rodents and primates. Importantly, only a small fraction of these individuals exposed to MPTP went onto develop Parkinson's disease-like symptoms, which could indicate that only certain people are genetically predisposed to MPTP toxicity.

The fact that this 'environmental' agent which is a proven cause of parkinsonism in humans does not always bring about its pathological effect, suggests that certain people could be genetically predisposed to MPTP toxicity, i.e. certain people could harbour mutations/ polymorphisms within specific genes involved in MPP+ dopaminergic uptake (active toxin which MPTP is converted to by monoamine oxidase B – refer to Table 1), which could protect or cause susceptibility.

This obviously lends credence to the hypothesis that sporadic Parkinson's disease cases are the result of the interaction of one or more 'susceptibility genes' with environmental influences. What also lends credence to this hypothesis is the fact that one of the most widely studied candidate genes for Parkinson's disease (before those involved in the UCS were concentrated on) was the gene for debrisoquine-4-hydroxylase (CYP2D6) – a member of the cytochrome P-450 family of mixed function oxygenases – which is involved in the

detoxification of a number of xenobiotic compounds. Several DNA-sequence variations are associated with a decreased (or in some cases enhanced) metabolic activity, and were seen to be more common in patients with Parkinson's disease (Armstrong *et al.* (1992), Smith *et al.* (1992)).

Work uncovering the relative importance of certain genes in the susceptibility of Parkinson's disease, will no doubt give us an even clearer understanding of the causes and pathogenesis involved. This will in turn provide us with more specific targets for future treatments that will aid Parkinson's disease modification and neuroprotection.

1.5.5 Parkinson's Disease Therapy

The mainstay of current Parkinson's disease therapy is standard oral administration of L-dopa (a substance that crosses the blood-brain barrier and is transformed into the neurotransmitter dopamine) which can significantly improve motor functioning during the first years of treatment. However, higher doses are required as the neurodegeneration progresses, which leads to deleterious side effects such as dyskinesia (involuntary movements). Overall, L-dopa (and other related drugs) can indeed prevent the apparition of symptoms such as tremor, rigidity, etc, but it does not treat the progress or underlying cause of the disease. However, electrical deep brain stimulation of the subthalamic nucleus reduces dyskinesia and the amount of drugs needed by 50% (Sarkis & Mallet (2005)).

An ideal treatment would be that of gene therapy, which could be administered once and work for a long period of time; both restorative and protective gene therapy strategies have been investigated. The restorative approach consists of a sustained and focused delivery of L-dopa (or dopamine) into the striatum, by transfer of genes encoding the L-dopa (and/ or dopamine) synthetic enzymes. Protective strategies are however the preferred solution, as they prevent the death of the dopaminergic neurons prior to disease onset, but also protect functioning neurons post diagnosis.

Protective strategies include genes that protect against neuronal cell death, including those coding for neurotrophic factors such as GDNF. The genetic transfer of GDNF has been studied in many animal models of Parkinson's disease, and the results have been very

encouraging (Sarkis & Mallet (2005)). GDNF can not only prevent neuronal death of dopaminergic neurons when expressed directly in the neurons of the substantia nigra, but it can also prevent axonal degeneration through 'sprouting', when expressed in the projection area of the degenerating neurons, compensating for any deficit. Recently, in a study carried out by Kordower *et al.* (2005), young adult nonlesioned rhesus monkeys which were treated one week prior with MPTP, had a lentiviral vector expressing GDNF injected into their respective striatum and substantia nigra. This treatment was seen to reverse functional deficits and completely prevent against nigrostriatal degeneration. An added benefit of GDNF use, is that it can be secreted and subsequently captured by the dopaminergic neuron terminals, i.e. it does not need to be expressed inside the neurons themselves.

Both restorative and protective approaches described are viable therapeutic strategies for Parkinson's disease patients.

1.6 Basic Overview of other Neurodegenerative Diseases (Dementias) Associated with UCH-L1

1.6.1 Alzheimer's Disease

As already discussed above, Alzheimer's disease is the most common form of dementia.

The clinical progression of Alzheimer's disease can be divided into three categories:-

1.) Progressive memory impairment,

2.) Progressive cortical dysfunction – separable into three main features:-

i.) Aphasia – loss of the ability to use and understand language (speaking, writing and listening).

ii.) Apraxia – inability to perform purposeful movements in spite of being able to demonstrate normal muscle function.

iii.) Visuospatial dysfunction – geographical and environmental disorientation.

3.) Neuropsychiatric disturbances – separable into four main features:-

i.) Mood disturbance.

II.) Delusions and hallucinations.

iii.) Personality change.

iv.) Disorders of behaviour

The typical clinical scenario of Alzheimer's disease is the gradual, insidious onset of amnesia with difficulties experienced in learning and recall. In the initial stages of the disease the memory impairment is for newly acquired information with recall of remote events being relatively unaffected. This early progression of memory and cognitive compromise is generally not realised by the individual (anosagnosia).

As the disease evolves, cognitive deficits progressively worsen over time, advancing from a relatively benign 'word finding difficulty' (anomia – precursor to aphasia) to an inability to carry out the more demanding tasks of daily living, such as driving and finance management, as abstract reasoning, executive functions and visuospatial skills all become

disrupted, before cognitive impairment finally reaches a point where the patient becomes dependent for feeding and hygiene.

In respect to neuropsychiatric symptoms, personality change is usually the first clinical change seen in Alzheimer's disease, and occurs in three quarters of sufferers. As the disease progresses delusions, hallucinations and other psychotic behaviours may develop, which are often exacerbated by verbal and physical aggressive tendencies.

It is worth noting that motor abnormalities are typically absent in Alzheimer's disease patients until the last few years of the disease; patients with Alzheimer's disease usually survive seven to ten years after the onset of initial symptoms.

As indicated above, Alzheimer's disease is predominantly an age-related phenomenon and accounts for more than half of all dementia cases in the elderly (Small *et al.* (1997)). In 1998 it was estimated that there will be 360,000 new cases of Alzheimer's disease each year, a figure that is expected to rise to 1.14 million incidences by 2048 (Brookmeyer *et al.* (1998)). Reflecting the greater longevity of women, there is a greater prevalence of Alzheimer's disease in women than in men (1.2 - 1.5: 1) (Gao *et al.* (1998)). As expected for a disease whose most important risk factor is increasing age, both the incidence and prevalence of Alzheimer's disease increases dramatically with age – exponentially increasing after the age of 65 (Kukull & Ganguli (2000)).

The major pathological features found in Alzheimer's disease are generalised cerebral atrophy with widespread neuritic amyloid plaques and neurofibrillary tangles – both of which are believed to be central to the pathogenesis of Alzheimer's disease and its associated degenerative cascade.

The pathogenesis of Alzheimer's disease was once thought to be (and still is by a few researchers) entirely due to the abnormal processing or disposition of amyloid (Selkoe (2000)). The amyloid precursor protein (βAPP) is a transmembrane protein composed of an extracellular amino terminal and an intracellular carboxyl terminus. It is initially cleaved by β-secretase at amino acid 671, which then allows γ-secretase to further cleave the protein forming either Aβ40 or Aβ42 depending on whether, respectively, residue 713 or 711 is targeted. This differential proteolytic cleavage of βAPP is normal, however, the Aβ40 species

is more often detected in plasma and cerebral spinal fluid, with Aβ42 ordinarily making up only a small component. In the Alzheimer's disease state it is the Aβ42 species which becomes more prominent, which causes *in vitro* fibrils to be formed more rapidly (those composed of Aβ40 form relatively slower). Owed to this property of the Aβ42 species, it is thought that it may become deposited earlier in the genesis of an insoluble senile plaque, which is then believed to initiate the cascade of events that result in cell destruction through inflammatory responses (Selkoe (2000)).

As well as widespread neuritic amyloid plaques, the other primary pathological feature of Alzheimer's disease is neurofibrillary tangle formation, which is thought to involve the abnormal processing of the axonal phosphoprotein tau. Alternative splicing of the mRNA for tau gives rise to six isoforms within the central nervous system, with either three tandem repeat sequences (3R-tau) or four tandem repeat sequences (4R-tau) in a normal 1: 1 ratio in the human adult brain (van Slegtenhorst *et al.* (2000)). These tandem repeat sequences are essential in tau's main function as a microtubule stabiliser – as they bind the microtubules (Lee *et al*. (2001)). Alternative splicing of tau's mRNA, together with tau's degree of phosphorylation, are thought to be key regulators of microtubule binding, and this has led to the hypothesis that neurodegeneration and pathology ensue when either or both of these mechanisms are disrupted, through mutations or other mechanisms, resulting in abnormal aggregation of tau. In support of this, paired helical fragments which make up the neurofibrillary tangles found in Alzheimer's disease have been shown to be composed of hyperphosphorylated tau (Lee *et al.* (2001)). Additionally, the normal 1: 1 ratio of 3R- and 4R-tau is disrupted in several neurodegenerative disorders, including corticobasal degeneration, progressive supranuclear palsy and argyrophilic grain disease, where 4R-tau is predominant (Dickson (1999)).

There are only two types of pharmacological treatment of Alzheimer's disease currently in use; one is a disease modifying approach using antioxidants, and the second is symptomatic treatment with the use of cholinesterase inhibitors (AChEIs).

Acetylcholine is involved in many aspects of cognition, including memory and attention. Patients with Alzheimer's disease have a cholinergic deficit owed to a reduction in choline acetyltransferase, and a loss of cholinergic neurons in the basal forebrain through

damage caused by the deposition of plaques and tangles. In careful, double-blind, placebo-controlled studies, AChEI administration has been shown to significantly improve memory in Alzheimer's disease patients. However, all AChEIs have similar side effect profiles including increased bowel frequency, nausea and vomiting. Nonetheless, AChEIs do appear to enhance cognitive function in subjects with Alzheimer's disease, and are recommended for the treatment of patients with mild to moderate Alzheimer's disease (Tang-Wai *et al*. (2005)).

The disease modifying approach to Alzheimer's disease treatment is mainly focused on alpha-tocopherol (vitamin E). In a two-year double-blind study, 341 moderately severe Alzheimer's disease patients participated in a placebo-controlled study comparing placebo with alpha-tocopherol. The study showed that there was a significant delay in the treatment group reaching the primary endpoints of death, institutionalisation, loss of basic activities of daily living, or severe dementia – compared to that of the placebo group. The American Academy of Neurology has indicated that vitamin E can be considered in an attempt to slow the progression of Alzheimer's disease (Tang-Wai *et al*. (2005)).

There are also other potential medications which are currently undergoing randomised, blinded trials to evaluate their efficacy; these include cholesterol lowering agents, memantine and non-steriodal anti-inflammatory drugs – which have all shown an association to slow or delay disease onset (Tang-Wai *et al*. (2005)).

1.6.2 Dementia with Lewy Bodies

Parkinsonism is a major feature of several dementing diseases whose features are characteristic of Parkinson's disease. This movement disorder is characterised by rigidity and bradykinesia, with rest tremors and gait instability also sometimes being realised. Dementia with Lewy bodies is commonly referred to as a 'parkinson-plus' syndrome, which are neurodegenerative disorders characterised by parkinsonism and at least one other nonparkinsonian neurological manifestation; these can be 'frontal-subcortical' cognitive deficits such as mental slowness, inertia and lack of initiative, forgetfulness, decreased executive functions, visuospatial deficits or mood disturbances. Other 'parkinson-plus' syndromes include progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy and amyotrophic lateral sclerosis/ parkinsonism-dementia complex of Guam.

Dementia with Lewy bodies is by far the most common of these, as it is now the preferred term for a variety of clinical diagnoses, including diffuse Lewy body disease, dementia associated with cortical Lewy bodies, the Lewy body variant of Alzheimer's disease, senile dementia of Lewy body type and Lewy body dementia.

Dementia with Lewy bodies describes a parkinsonian dementia with a widespread distribution of Lewy bodies – eosinophillic cytoplasmic inclusions (refer to section 1.5.3). In addition to the parkinsonian movement disturbance, typical features are visuospatial and executive function cognitive impairments that fluctuate over time (lasting minutes and hours, rather than days), and neuropsychiatric symptoms such as fixed delusional visual hallucinations. The order of emergence of these cognitive disturbances and parkinsonism are variable, though they usually appear within a year of one another. Dementia with Lewy bodies parkinsonism includes rigidity, bradykinesia and disturbances of posture and equilibrium, with the absence of resting tremor being typical. The cognitive fluctuations seen in dementia with Lewy bodies are most commonly realised in marked variations in attention, as well as periods of confusion, inattention or decreased responsiveness.

The neuropsychiatric manifestations of dementia with Lewy bodies are among the most fascinating aspects of the disease, and are extremely helpful in conveying clues for its correct diagnosis. The majority of dementia with Lewy bodies patients experience psychiatric disturbances involving visual hallucinations, which are most commonly fully formed and animate, and often involve deceased relatives, complete strangers or animals.

Depression, apathy, anxiety, insomnia, paranoia and paramnestic phenomena (dreams confused with reality) also frequently occur in dementia with Lewy bodies, which often prove very difficult to manage owing to the characteristic pharmalogical sensitivities exhibited by dementia with Lewy bodies patients. Furthermore, a number of other features, which are variable in occurrence, have also been associated with Dementia with Lewy bodies. These include orthostatic hypotension (low blood pressure when standing), unexplained falls or syncopy (unconsciousness through a fall in blood pressure) and Rapid Eye Movement Sleep Behaviour Disorder (REMSBD) – a disorder characterised by vocalisation and gesticulations with patients acting out their dreams while asleep.

Dementia with Lewy bodies is third most common cause of dementia after Alzheimer's disease and vascular dementia, accounting for about fourteen to twenty percent of patients (McKeith *et al.* (1999)). Two thirds of dementia with Lewy bodies patients are male (Barber *et al.* (2001), Klatka *et al.* (1996)), though it is unclear whether this is due to increased male susceptibility to the disease or reduced male survival. Classical epidemiological studies to ascertain age and sex variation and potential risk factors for dementia with Lewy bodies have yet to be reported.

Dementia with Lewy bodies may realise Lewy bodies that are widely distributed – they can occur in the substantia nigra, locus ceruleus, dorsal motor nucleus of the vagus, nucleus basalis, cholinergic neurons in the basal ganglia, hypothalamus, cerebellar cortex, spinal cord intermediolateral cell column, and autonomic ganglia including submucosal ganglia of lower esophagus. Dependent upon the distribution of the Lewy bodies, dementia with Lewy bodies can be categorised into three types – Type A (involving the brainstem and cortex), Type B (limbic or transitional predominant), or Type C (brainstem predominant). The severity of cognitive impairment in dementia with Lewy bodies has been shown to correlate with Lewy body densities in the frontal and temporal neocortex, as well as Lewy neuritis in the hippocampus (Haroutunian *et al.* (2000), Mattila *et al*. (2000)). Furthermore, visual hallucinations seem to correlate with the presence of Lewy bodies in the parahippocampal and inferior temporal cortices (Harding *et al.* (2002)). The hippocampus also shows significant atrophy and pathology in dementia with Lewy bodies, involving spongiform (sponge-like) and vacuolar changes (Harvey *et al.* (1999)). Moreover, magnetic resonance imaging (MRI) discloses whole brain atrophy with disproportionate atrophy of the temporal lobes (located on the side of the cerebrum), although not to the extent seen in Alzheimer's disease (McKeith *et al.* (1999), Barber *et al.* (2001)). In addition to the characteristic Lewy bodies, the majority of dementia with Lewy bodies patients exhibit senile plaques and neurofibrillary tangles – both characteristic features of Alzheimer's disease pathology (Londos *et al.* (2002), McKeith *et al*. (1998)). Dementia with Lewy bodies brains have also been shown to have decreased concentrations of various neurotransmitters in the putamen and neocortex, including dopamine, serotonin and norepinephrine.

The treatment of dementia with Lewy bodies presents the clinician with several challenges; as although the pathological features have become increasingly well described, there is still controversy in the interpretation of the histological features, i.e. do they indicate a variety of Alzheimer's disease, a variety of Parkinson's disease, a separate distinct entity, a coexistence between the two, or indeed a spectrum disorder? A recent proposal for the reclassification of neurodegenerative disorders into synucleinopathies (or tauopathies) may help clarify its nosological (medical classification) status.

Though this nosological uncertainty must be borne in mid when discussing dementia with Lewy bodies, the treatment of established dementia with Lewy bodies will now be briefly discussed. One main approach in the current treatment of dementia with Lewy bodies, similar to that of Alzheimer's disease, is the use of cholinesterase inhibitors (AChEIs) to counter the severe cholineacetyltransferase deficiency, which is more profound than that seen in Alzheimer's disease. Second generation AChEIs are beginning to be used, and in one randomised trial of 120 patients, significant differences in neuropsychiatric symptoms were found in favour of the treated group over the placebo at 20 weeks. Many similar studies show a comparable benefit of treatment (Byrne *et al*. (2005)). Other putative therapeutic agents aimed at enhancing cholinergic function in dementia with Lewy bodies are muscarinic agonists and nicotinic agonists.

Although dopamine levels are reduced in post-mortem studies in dementia with Lewy bodies, and CSF homovanillic acid levels are reduced in autopsy-confirmed cases, there has been little systematic enquiry into the effects of L-dopa therapy in dementia with Lewy bodies. Early studies however, found little or no L-dopa response in those that were treated, plus acute confusional states (delirium) and other adverse effects which are associated with L-dopa therapy, may all account towards the lack of enthusiasm to use these drugs in dementia with Lewy bodies therapy (Byrne *et al*. (2005)).

Psychiatric symptoms, especially visual hallucinations, are common and troublesome in dementia with Lewy bodies, which are caused by increased sensitivity to neuroleptics. Although neuroleptic sensitivity is not an inevitable consequence of neuroleptic medication in dementia with Lewy bodies, it is certainly common and commonly severe (Byrne *et al*. (2005)). Treatment of psychotic symptoms in dementia with Lewy bodies include the use of

GABAergic agents such as chlormethiazole, which reduce or even obviate visual hallucinations, help in associated sleep disorders and also has neuroprotective effects. (GABA has been suggested as an important transmitter in delirium, and also has an important function in motor control). (Byrne *et al*. (2005)).

1.6.3 Huntington's Disease

Huntington's disease is an autosomal dominant disorder that is clinically characterised by a triad of symptoms and signs:-

1.) A prominent movement disorder, i.e. chorea – rapid, jerky, 'dance-like' movements of the body.

2.) Behavioural and emotional alterations

3.) Cognitive decline.

Huntington's disease is characterised by both extrapyramidal motor abnormalities and an impairment of voluntary movements that effects gait, speech and swallowing. Choreic movements, which are the major symptom of Huntington's disease, are involuntary, abrupt, irregularly timed and randomly distributed movements of the body – with typical accentuation in the fingers and toes (Yang *et al*. (1996)). Choreic severity over time ranges from restlessness, hand fidgeting and an unstable 'dance-like' gait, to a continuous flow of violent movements that are severely disabling and exhausting to the patient. After this first phase of the disease, typically lasting around ten years, the severity of chorea tends to decrease, progressively becoming replaced by bradykinesia and rigidity (core features of parkinsonism) and dystonia (involuntary spasms of muscle contraction that causes abnormal movements/ postures) (Young *et al.* (1986)).

In Huntington's disease, cognitive and intellectual function typically start to slow soon after the chorea begins (Paulsen *et al.* (2001)). As the disease progresses, patients become significantly impaired in an array of cognitive areas including attention and concentration, memory, speech and language, visuospatial skills and frontal executive functions (Snowden *et al.* (2001)). Memory deficits occur in retrieval and procedural learning (Bylsma *et al.* (1991)), and in contrast to Alzheimer's disease, Huntington's disease patients seem to have a

greater difficulty retrieving both recent and remote information, with less difficulty in actually storing memories (Rohrer *et al.* (1999)). Speech and language changes include dysarthria (a disturbance in the strength and coordination of speech muscles), a decreased verbal fluency, and a decreased speech output which ultimately progresses to mutism (Murray & Lenz (2001)). The frontal-executive deficits in Huntington's disease are prominent, and interfere with a patient's ability to perform tasks that require organisation, planning and/ or sequential arrangement of information (Zakzanis (1998)). As the disease advances, all intellectual abilities deteriorate to the point where patients are mute and intellectually devastated in the final stages.

Huntington's disease is equally common in men and women, and effects three to ten people per 100'000 (Folstein (1989)). The disease usually begins in the fourth to fifth decade (in 90-95% of patients), and typically has a 13- to 16-year course (Folstein (1989)).

On autopsy, macroscopic inspection of advanced Huntington's disease patients' brains reveal a weight reduction of 10-20% compared with age matched controls. Much of this weight loss can be attributed to the shrunken neostriatum, the main neuropathological feature in Huntington's disease, with gross atrophy of the putamen and especially the caudate nucleus – which is reduced from a robust structure to a thin ribbon of tissue as little as two or three millimetres thick (Myers (1988)). At the microscopic level, histological changes are concordantly most prominent in the neostriatum, i.e. caudate and putamen, where there is a loss of small spiny neurons, which is hypothesised to have the inhibitory effect on movement mechanisms through a 70-90% reduction in γ-amino butyric acid (GABA) production – an inhibiting neurotransmitter. The striatum also realises shrunken larger neurons, with a conspicuous increase in glial cells (Myers (1988)). It is this striatal pathology which likely underlies the involuntary movements of Huntington's disease chorea and dystonia, as well as the disordered planning, impulsive behaviours, diminished emotional control and some of the other Huntington's disease symptoms (Albin *et al.* (1989)). While the hippocampus also displays moderate neuronal loss, the cerebellum, brainstem and spinal cord are little affected in Huntington's disease (Albin *et al*. (1989)). The extent of the striatal pathology forms the basis of the grading severity of the disease (grades naught to four), which clearly correlates with the clinical progression of Huntington's disease (Myers *et al*. (1988)).

The Huntington's disease gene contains 67 exons (Huntington's Disease Collaborative Research Group (1993)) and encodes a 350 KDa protein called Huntingtin. Huntington's disease has been shown to be caused by an expanded CAG repeat located within the first exon (+36 to 120bp), which is translated into a polyglutamine tract. The number of CAG repeats in the Huntington's disease gene correlates with a younger age of onset and an increased DNA fragmentation in the striatum (Butterworth *et al.* (1998)). It is hypothesised that the abnormal Huntingtin protein aggregates within neurons and glia to form abnormal intracytoplasmic inclusions/ filaments (Lunkes *et al.* (1998)). Furthermore, cysteine aspartate-specific proteases (caspase) may cleave the mutant huntingtin, generating toxic protein fragments that lead to abnormal metabolism, oxidative stress and mitochondrial dysfunction in the cells of the caudate and putamen (Wellington & Hayden (2000)).

Current treatment of Huntington's disease is aimed at the symptomatic control of the motor and psychiatric aspects of the disorder. Neuroleptics such as haloperidol, which is a dopamine antagonist, are used to suppress abnormal movements, though in the later stages of the illness as dopamine receptors are destroyed, this medication will gradually become less useful and may aggravate dystonia, bradykinesia & dysphagia, gait and balance problems (Chua & Chiu (2005)).

The depression in Huntington's disease responds to the same treatments as it does in the general population, but Huntington's disease sufferers may become more sensitive to the side effects such as sedation and anticholinergic-induced cognitive decline. Psychotic symptoms, irritability or behavioural disturbance all respond to neuroleptics (Chua & Chiu (2005)).

Specific treatments which theoretically target the disease process to slow functional decline in disease manifestation, such as antioxidants and other neuroprotective drugs, are under trial. Drugs which may slow progression or improve the chorea whilst being tolerated include antioxidant coenzyme Q, and remacemide, a non-competitive *N*-methyl-D-aspartate (NMDA) receptor antagonist (Chua & Chiu (2005)).

Clinical research determining the early onset and progressive symptoms of Huntington's disease is currently being undertaken by an international collaboration of

researchers in Huntington's disease – the Huntington Study Group. Part of their remit is to research and develop new restorative therapies that rejuvenate or replace malfunctioning neurons in order to restore functions.

1.7 Basic Relevant Molecular Overview

1.7.1 Eukaryotic Transcription Regulation

Regulation of gene expression is fundamental to biological systems. Precise and timely expression of genes is of paramount importance for the proper development of all organisms and correct functioning of all cell types. Control mechanisms have evolved to regulate gene expression according to physiological demands and in response to various stimuli. There are multiple levels of control that determine and influence the expression of a gene into a functional mature protein. Regulatory systems for the control of gene expression exist at various levels, including transcription, precursor-RNA processing, translation of mRNAs, degradation of the mature mRNAs, and degradation of the protein products (Lewin (1997)). However, it is generally accepted that gene expression is predominantly regulated at the level of transcription (Hames (1988)).

Transcriptional activation is a two-step process because most genes exist at some time in an inert state tightly packaged with histones into chromatin (Lewin (1997)). Before transcription can commence this inert structure must be decondensed so transcriptional control sequences are made available to regulatory proteins. The exact mechanisms for decondensing chromatin are not yet fully elucidated but result in the appropriate topology of the DNA helix which will expose the relevant sequences (Hames (1988)). Following decondensation the second activation step takes place which involves the interaction of regulatory proteins known as transcription factors (TFs) with specific DNA sequences (Tjian and Maniatis (1994)).

In the majority of cases, the regulatory sequences immediately adjacent to the encoding region constitute the core promoter region. The core promoter region and additional regulatory sequences situated either upstream or downstream, in close proximity or not to the encoding region, offer binding sites for transcription factors that influence the proper positioning and function of the transcription machinery, and can up-regulate or completely silence the transcription of the particular gene.

Eukaryotic genes are transcribed by one of three polymerases, RNA polymerase I, II or III, with the aid of transcription factors that direct the appropriate polymerase to the start site, unwind the DNA, and perform a range of other tasks without which the polymerase would not be able to function properly.

UCH-L1 belongs to the class of genes transcribed by RNA polymerase II. The typical eukaryotic class II gene includes three kinds of transcriptional control elements: the core promoter, upstream regulatory elements, and distant regulatory elements. Additionally, more recent work has elucidated DNA methylation as another important control factor in mammalian gene transcription.

1.7.1.1 The Core Promoter

The core promoter for every gene is situated immediately adjacent to the transcription start site and is characterised by two key genetic elements: The TATA box and the initiator (Inr) element (Smale and Baltimore (1989)). In experimental *in vitro* systems core promoter sequences were shown to be the minimal sequence elements necessary to nucleate the promoter and recruit RNA polymerase II in order to initiate transcription (Novina *et al.* (1996)).

The TATA box is an adenine-thymine rich (AT rich) stretch of sequence usually situated 25-30 nucleotides upstream from the transcription start site (Breathnach and Chambon (1981)). It forms a binding site for a protein factor, the **T**ATA box **B**inding **P**rotein (TBP). TBP is one of the most highly conserved proteins in eukaryotic evolution and probably one of the most intensively studied (Buratowski (1994)). TBP alone cannot mediate transcription regulation; this activity requires the entire transcription factor D complex (TFIID), which consists of TBP and the TBP-associated factors (Tjian and Maniatis (1994)). In the absence of a TATA box element, components of the TFIID complex can bind either directly or indirectly to the other key core promoter genetic element, the Initiator (Inr) element. The Inr element is defined as sequences spanning a transcription start site that can autonomously function as a promoter (Buratowski (1994)).

1.7.1.2 Upstream Regulatory Elements

The Transcription Initiation Complex (TIC) is made up of a plethora of proteins, which complement and assist RNA polymerase II in starting transcription from any given class II gene promoter. The efficiency and specificity with which the promoter is recognised by the TIC, depends on regulatory sequences further upstream from the core promoter (Lewin (1997)). The core promoter elements (basal elements) and the more upstream elements have different type of functions. The basal elements primarily determine the location of the transcription start site, but can sponsor initiation at a rather low level. They identify the location at which the general transcription factors assemble to form the TIC (Orphanides *et al.* (1996)). The Upstream Regulatory Elements (URE) which lie further upstream influence the frequency of initiation, most likely by interacting directly with the general transcription factors to enhance the efficiency of assembly into an initiation complex (Lewin (1997)). URE and the transcription factors that bind to them may be common and found in a wide variety of promoters, or they may be specific, and particular for transcription in a temporal or timely manner (Hames (1988)) since URE found in any individual promoter differ in number, location and orientation (Lewin (1997)).

1.7.1.3 Distant Regulatory Elements

The mechanisms for regulating transcription require special DNA sequence elements known as enhancers or silencers, which act to respectively increase or decrease the transcriptional levels within certain tissues or cell types. Distant regulatory elements are recognized by specific binding proteins, and as the name implies, can function at considerable distances from the promoter, and are active upstream of, within, or downstream of the gene, regardless of orientation in respect to the promoter (Maniatis *et al.* (1987)). Moreover, these regulatory elements vary in complexity, from simple elements containing one or more binding sites for a single type of binding protein, to enhancers/ silencers containing binding sites for several different proteins (Tjian & Maniatis (1994), Hill & Treisman (1995)).

The main function of transcription enhancer complexes is to recruit the transcription machinery to the promoter through direct protein–protein interactions between the activating transcription factors and one or more components of the transcriptional machinery (Ptashne &

Gann (1997)). Interactions between the transcription machinery and activators are thought to stabilise the binding of general transcription factors to the basal promoter. Thus, enhancers increase the formation of productive pre-transcription-initiation complexes (Pugh (1995)).

Most silencers work together with highly refined enhancers to specify activity (Mori *et al.* (1990)). There are three models for repression: inhibiting DNA binding, blocking activation, and silencing (Renkawitz (1990)); of the three, silencing is the most often utilised mechanism in neuronal-specific genes. Silenced DNA is usually replicated late, suggesting that both DNA and RNA polymerases have a limited access to the silenced sequence (Herschback and Johnson (1993)), thus decreasing the probability of expression. The binding of a repressor protein may directly inhibit the binding of the transcriptional machinery, though such a mechanism would probably only succeed in the case that the silencer overlaps the TATA box itself. Another more likely scenario is that the silencer and its binding protein serve to modify the chromatin structure around the element; the so called 'looping-out' model proposed for distant regulatory elements seems to fit this hypothesis well (Ptashne (1988)). Such 'silencing' probably occurs as the result of the DNA being folded into a form of especially compacted chromatin, obscuring the DNA from the transcriptional machinery (Herschback and Johnson (1993)).

1.7.1.4 DNA Methylation

Recently, methylation of DNA at cytosine bases has been recognised as an important epigenetic modification that plays an important role in the control of gene expression in mammalian cells. One crucial enzyme involved in this process is DNA methyltransferase, which catalyses the transfer of a methyl group from *S*-adenosylmethionine to cytosine residues, forming 5-methylcytosine, a modified base that is found mostly at CpG sites (high density of methylated cytosine residues occurring immediately 5' to G residues) in the genome (Momparler & Bovenzi (2000)). Across the majority of the human genome, these CpG sites are underrepresented and heavily methylated In contrast, there are regions of the genome where CpG sites are found at or above their expected frequency, and a cluster of these CpG sites is referred to as a CpG island (Bird (1986)). Analysis of the spatial relationship between CpG islands and eukaryotic genes has shown that CpG islands are

usually associated with gene promoters and extend further downstream into transcribed regions (Bird (1987)).

Methylation of CpG islands can result in gene silencing (Cedar (1998), Gaston & Fried (1995)), and direct effects of CpG methylation on binding and activity for various transcription factors has been observed (Gaston & Fried (1995)). Furthermore, demethylation is a well-known mechanism for the developmental activation of tissue-specific gene expression (Cedar (1998)).

1.7.2 Genome Variation

The human genome comprises of nuclear (23 chromosomes) and mitochondrial (several to several thousand copies of the circular molecule) genomes, with a total of 3,300 million nucleotides for the haploid cell. Only 30% of the human genome sequence contains genes and 3% encodes for proteins. It is estimated that there are about 35,000 in the human genome, 58% of them have been annotated, and the rest do not have a known function. The average size of a gene is ~27kb with considerable variation.

There are no two individuals (except identical twins) that are exactly the same. This difference in phenotypic traits is the result of sequence variations found within an individual's genes. This 'nucleotide diversity' varies greatly across the genome, and although these variations are found on all chromosomes, their distribution is not uniform.

Two major types of nucleotide sequence variation exist in human genomes; these are single nucleotide polymorphisms (SNPs) and variable number of tandem repeats (VNTRs). SNPs are much more prevalent with over 4 million having so far been discovered, compared to a VNTR number approaching 120,000. Other sequence variations are attributable to insertions/ deletions and rearrangements. SNP classification is dependent on the variant rarer allele's frequency in the population being greater than 1% occurrence. Typically base changes that have a frequency of less than 1% are termed mutations. About half to two thirds of SNPs are shared between two populations, and the rest are ethnic specific. SNPs which are located in genes not only attribute to trait difference between individuals, but it is now becoming apparent that SNP combinations play an important role in disease susceptibility.

Numerous studies have been performed to estimate the sequence variation within the human genome. Screening of a 9.7kb region of the human lipoprotein lipase gene from three diverse populations gave a total of 88 sites showing sequence variation, of which 79 were SNPs and 9 were insertion/ deletion variations (Nickerson *et al*. (1998)). Another study on a 4Mb high density SNP based map around the human apolipoprotein E gene detected a SNP frequency of one SNP per 1.1kb of genomic sequence (Lai *et al*. (1998)). A further study, which took into account all publicly available SNPs spanning the whole genome sequence, indicated a SNP frequency of one every 1.9kb (Sachidanandom *et al*. (2001)). From these three studies a wide range of values are evident, though as briefly mentioned above, the area chosen to evaluate would have had a large impact on the sequence variation realised. The most commonly cited value for the single base variation between two non-related individuals is one SNP every 1kb on average. This figure equates to a 0.1% chance of any base being heterozygous in an individual, and indicates a 99.9% homology among human genomes – the 0.1% variation giving rise to individual phenotypic difference. This 0.1% variation translates into several million sequence variations between two non-related individuals, which results in approximately 100,000 amino acid differences between their proteomes (Brookes (1999)).

Nearly every variable site results from a single historical mutational event; this statement is supported by the fact that the mutation rate at a given site is of the order of 10^{-8} per generation, which is relatively very low contrasted to the number of generations (in the order of 10^{4}) that divides any two humans from a single common ancestor (The International HapMap Consortium (2003)). Each new allele which arises on a particular chromosomal block is initially associated with the other alleles, and this specific set of alleles is referred to as a haplotype.

Coinheritance of SNP alleles on a particular haplotype leads to these alleles being associated in the population, which is referred to as linkage disequilibrium. These associations however decline with distance, as the likelihood of the maternal and paternal chromosomes undergoing recombination (exchanging corresponding segments of DNA) between two SNPs increases with nucleotide length (The International HapMap Consortium (2003)).

1.7.2.1 Hardy-Weinberg Equilibrium

The Hardy-Weinberg equilibrium was named after G. H. Hardy & W. Weinberg, and its underlying principle states that genotype allele frequencies remain constant – or are in equilibrium – from generation to generation, unless specific influences are introduced. Disturbing influences can take the form of mutations, gene flow, random genetic drift, non-random mating, (natural) selection and limited population size.

The Hardy-Weinberg principle can be used as a Punnett square (a cross/ breeding diagram outcome predictor) for populations (Figure 8). It can be used to predict the probability of an offspring's genotype based on the parents' genotype, and vice versa. Furthermore, the Hardy-Weinberg formula can be employed to calculate whether the observed genotype frequency data for a given population is within a normal distribution pattern, by way of calculating the deviation from the expected genotype frequencies through (typically) chi-squared test application.

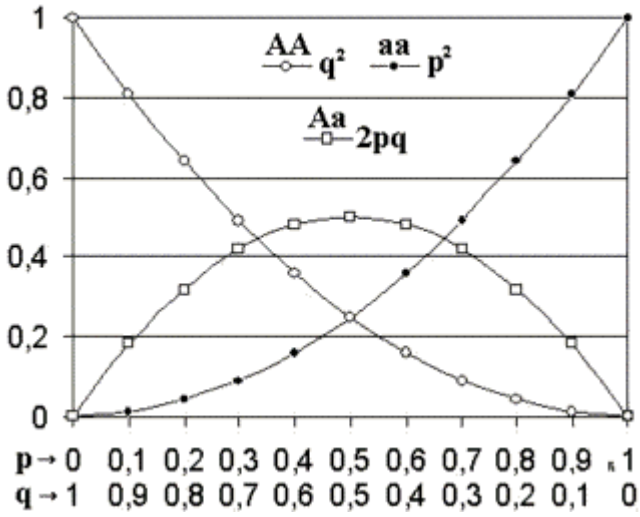


Figure 8 – The Hardy-Weinberg principle for two alleles. The x-axis shows the two allele frequencies *p* and *q*, the y-axis displays the genotype frequencies, and the different glyphs represent the three possible genotypes.
(http://content.answers.com/main/content/wp/en/thumb/b/b7/320px-Hardy-Weinberg.gif).

1.8 *UCH-L1*'s Expression

The cDNA sequence (Day and Thompson (1987)), chromosomal location (4p14) (Edwards *et al*. (1991)) and structure (Leroy *et al*. (1998), Day *et al*. (1990) of *UCH-L1* have all been realised. The gene spans 10kb of genomic DNA and consists of nine exons and eight introns, the smallest of which, exon 2, encodes only four amino acid residues (Day *et al*. (1990)). Below (Figure 9) is a diagrammatic representation of a 1kb region in *UCH-L1*'s promoter region that flanks the 5' end of the UCH-L1 gene; the transcription start site (guanine base) is located 29 nucleotides downstream of the TATA box.

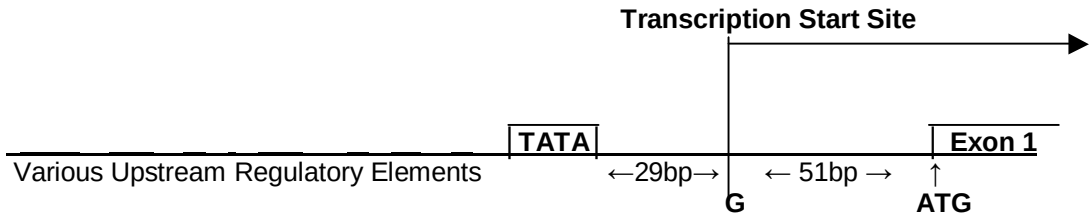


Figure 9 – Basic depiction of *UCH-L1'*s main promoter elements.

Truncation mutagenesis studies (Mann *et al*. (1996)) showed that nucleotides -182 to +51, relative to the transcription start site (-233 to –1 with respect to the translation start site – which is the nucleotide nomenclature reference point of preference within this study), within this region are sufficient to act as a minimal active promoter. Additional 5' sequences were seen to have minor effects on the transcriptional activity of this region, but a truncation to nucleotide –123 (-147) reduced the activity of the promoter to almost background levels. These observations strongly suggest the presence of a critical transcriptional activator within the 59-bp region spanning nucleotides -182 and –123 (-233 and –147 relative to the ATG start site of translation).

Expression of *UCH-L1* is largely confined to cells of the neural and neuroendocrine lineages (Thompson and Day (1988), Wilson *et al*. (1988)). Furthermore, observations link the onset of motor and sensory neuron differentiation with the spatial and temporal expression of *UCH-L1*. Studies in the developing mouse show that UCH-L1 mRNA is expressed in the neural tube at 9 days postcoitum, just before the commencement of neural differentiation. By day 11 of the development UCH-L1 protein accumulated at the ventral horns of the spinal

cords, preceding peak production of motor neurons, and in the dorsal root and cranial nerve ganglia, where it appears before the differentiation of sensory neurons (Schofield *et al*. (1995)). It is also worth noting that this study indicated *UCH-L1* as one of the earliest neural-specific genes to be expressed in the developing nervous system.

UCH-L1's neuronal cell type specificity has been supported by UCH-L1 expression work *in vitro* (Mann *et al*. (1996)). UCH-L1 mRNA was seen to be absent in three different types of nonneuronal cell types, while the UCH-L1 protein was clearly observed in neuroblastoma cell lines using an immunoreactivity strategy. This study also observed that the *UCH-L1* promoter was threefold more active than the SV40 promoter/ enhancer in neuroblastoma cells, and in contrast, was tenfold less active in comparison to the SV40 promoter/ enhancer in nonneuronal cells. This suggests that there are elements within the 1kb promoter that contribute to UCH-L1s neural-specific expression. However, this study also revealed, though at lower activity, that the *UCH-L1* promoter is also active in nonneuronal cells. This linked to the complete absence of UCH-L1 mRNA in nonneuronal cells above, indicates that an additional regulatory element, located outside the 1kb promoter may also be required for achieving neural-restricted expression of UCH-L1, which could then also play a role in its spatial and temporal (CNS) developmental expression.

Such additional regulatory elements could take the form of a 'distant regulatory silencer element', or alternatively a mechanism of gene regulation such as DNA methylation may operate (refer to section 1.6.1.4).

DNA methylation within the 1kb promoter may also play a role in UCH-L1 expression. Analysis of the CpG content of the 5' end of the UCH-L1 gene showed the presence of a complete CpG island region of at least 400bp that spans the transcriptional initiation site (Mann *et al*. (1996)). Interestingly, this region also coincides with the 5' end of the minimal promoter.

Mann *et al*. also elucidated a further 655bp of upstream *UCH-L1* sequence, in addition to the ~340bp that was already known, which then allowed a 1kb human *UCH-L1* promoter sequence to be further analysed with the use of the 'signal scan programme' (HGMP database), which used mammalian, bird, amphibian, insect, plant and yeast

databases to identify known regulatory sequence elements. The results were then aligned with the results also obtained for *Monodelphis domestica* to give a sequence conservation perspective (Figure 10).

The optimal alignment of the human and *Monodelphis UCH-L1* 5' flanking DNA revealed the presence of a perfectly conserved 12bp sequence (PSN) within the 59bp activator region (Figure 10), and its similarity to sequences in the promoters of human and rat synapsin 1 and neurone-specific enolase genes (Sauerwald *et al*. (1990), Oliva *et al*. (1991)), indicates a potentially powerful regulatory element.

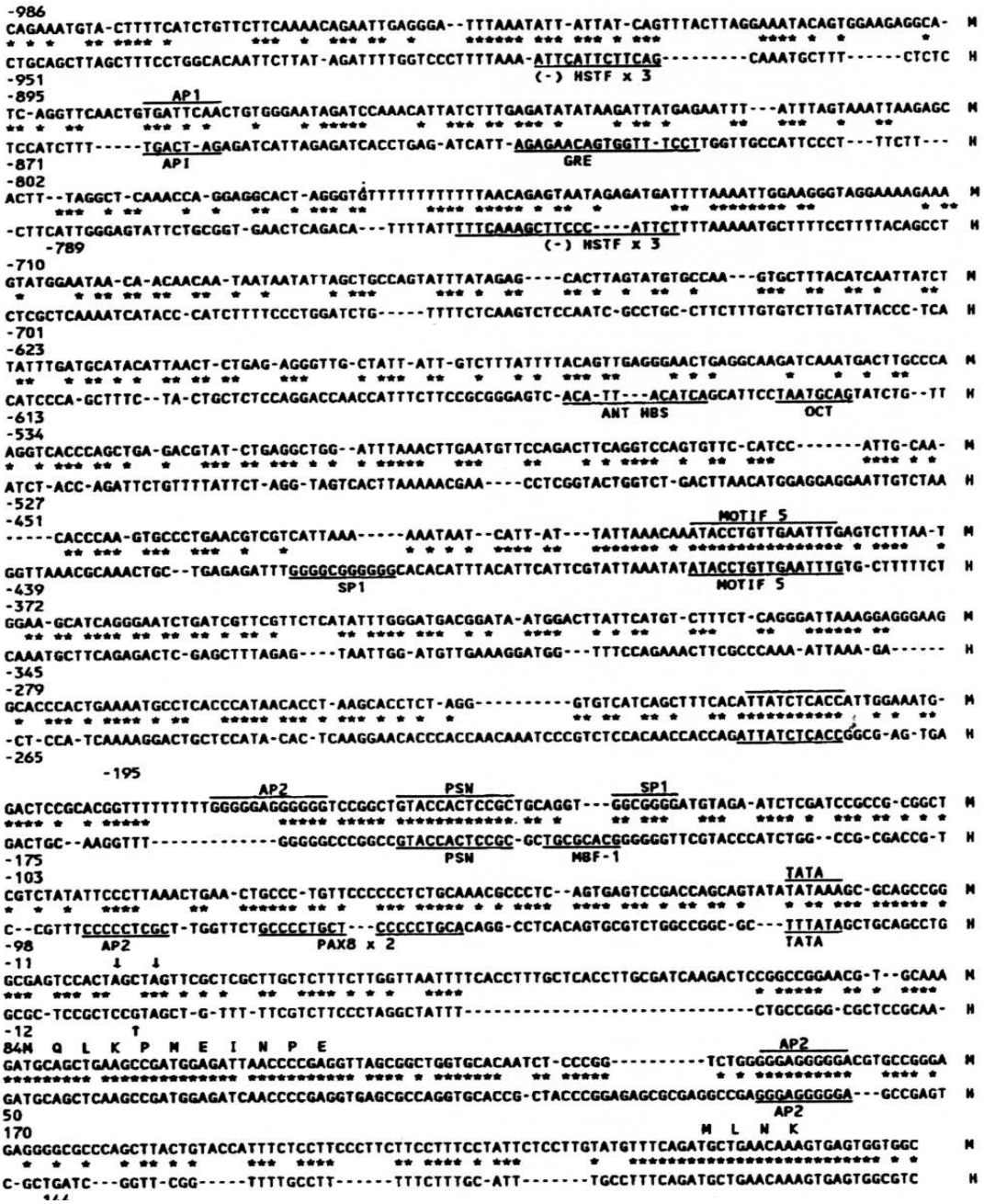


Figure 10 – Comparison of the nucleotide sequences at the 5' ends of the human and *Monodelphis* UCH-L1 genes by optimal sequence alignment. The *Monodelphis* sequence (M) is displayed on the top line, with the human (H) sequence displayed below it. Spacings (-) are used where appropriate to achieve optimal alignment, and those nucleotides seen to be conserved are highlighted (*). Sequences are numbered in the 5' to 3' direction from the major transcription initiation sites, which are numbered +1 and +5 for the *Monodelphis* gene and +1 for the human gene (positions marked by arrows). Sequences for the known transcription factor binding sites TATA, AP1, AP2, SP1, and Oct, the metal binding factor (MBF-1), the heat shock transcription factor (HSTF), the glucocorticoid response element (GRE), the antennapaedia homeobox binding site (ANT HBS), and Pax 8 are shown by the use of lines under the appropriate nucleotides. The PSN and motif 5 sequences conserved between human and *Monodelphis* are highlighted by double lines (one above and one below the sequence), as is a third unnamed conserved motif that contains an inverted GATA factor binding sequence and the ATTA core of the homeobox binding sites (Mann *et al*. (1996)).

1.9 This Study's Importance

This study will be mainly focusing on *UCH-L1*'s promoter sequence variants in the Caucasian population. UCH-L1, as already discussed above (refer to section 1.3), has critical roles as a hydrolyser and a ubiquityl-ligase within the UCS.

What also makes *UCH-L1* such a good candidate gene for further study in reference to neurodegenerative disease susceptibility, is the fact that it has revealed three documented coding region variants associated with neurological disorders: one mutation directly implicated in familial Parkinson's disease with high penetrance (Ile93Met), another mutation possibly with low penetrance (Met124Leu), and a polymorphism (Ser18Tyr) which seems to protect against sporadic Parkinson's disease and Alzheimer's disease in certain populations, whilst also possibly delaying the onset of Huntington's disease.

No work has yet investigated *UCH-L1*'s promoter/ 5' untranslated region, specifically the potential neurodegenerative disease-association of sequence variants within this region, which could alter UCH-L1's expression. Sequence variation within *α-synuclein*'s promoter (complex dinucleotide repeat – refer to section 1.4.1) has already been associated with sporadic Parkinson's disease, and a UCH-L1 pathological route towards Parkinson's disease susceptibility has been established encompassing Rochet & Lansbury's 'critical concentration of α-synuclein' hypothesis (2000). UCH-L1 has an intimate relationship with α-synuclein at the synapse which may even involve its regulation (refer to section 1.3.6), thence unveiling of any common promoter haplotypes which could alter its expression, should be seen as a potentially very important step in uncovering primary risk factors in the pathogenesis of sporadic Parkinson's disease and dementia.

1.10 Objectives

To identify common sequence variants in the UCH-L1 gene, and to examine whether they have functional effects on UCH-L1 expression and/ or function, and to study the sequence variants in relation to neurodegenerative diseases.

2 Materials & Methods

2.1 Polymorphism Identification

2.1.1 DNA Arrays

2.1.1.1 x26 Control DNA Array

Comprised of 26 unrelated healthy genomic DNA samples at a concentration of 7ng/ µl. DNA was extracted from annonymised blood samples (taken with permission from people working in the Human Genetics department in Southampton General Hospital in 2001); thence, no data on age, gender, etc is available. Samples had the following nomenclature: B1-B5; L1-L6, L12, L15, L18-L21 and L23-L30.

2.1.1.2 x64 Control DNA Array

Comprised of 31 unrelated healthy Caucasian genomic DNA samples at a concentration of 7ng/ µl. DNA was extracted from annonymised blood samples (taken with permission from people working in the Human Genetics department in Southampton General Hospital in 2003); thence, no data on age, gender, etc is available. The array was made up of 30 samples in duplicate and 1 in quadruplicate; how this corresponded within the actual 64 well array is outlined below:-

A1 – B2	B1 – A7	C1 – H3	D1 – D6	E1 – B6	F1 – H6	G1 – F2	H1 – H2 G6 – G8
A2 – F4	B2 – A1	C2 – C4	D2 – C6	E2 – E7	F2 – G1	G2 – G3	H2 – H1 G6 – G8
A3 – F8	B3 – C5	C3 – A6	D3 – G4	E3 – G5	F3 – H7	G3 – G2	H3 – C1
A4 – C7	B4 – A5	C4 – C2	D4 – B7	E4 – H5	F4 – A2	G4 – D3	H4 – F7
A5 – B4	B5 – C8	C5 – B3	D5 – F6	E5 – G7	F5 – D7	G5 – E3	H5 – E4
A6 – C3	B6 – E1	C6 – D2	D6 – D1	E6 – E8	F6- D5	G6 – G8 H1 – H2	H6 – F1
A7 – B1	B7 – D4	C7 – A4	D7 – F5	E7 – E2	F7 – H4	G7 – E5	H7 – F3
A8 – H8	B8 – D8	C8 – B5	D8 – B8	E8 – E6	F8 – A3	G8 – G6 H1 – H2	H8 – A8

Table 4 – Indicates the sample duplicity and quadruplicity in the x64 control DNA array.

2.1.1.3 x480 DNA Array with Cognitive Function Data

The sample comprised of 480 Caucasian unrelated men and women, aged 66–75 years, who had been born in the Jessop Hospital for Women, Sheffield, UK. The participants had been traced using the National Health Service Central Register and were still living in Sheffield. The study was approved by the North Sheffield Research Ethics Committee, and all participants gave written informed consent. The investigation conforms with the principles outlined in the Declaration of Helsinki. Participants were interviewed by a fieldworker who administered the cognitive function tests (refer to section 2.5.2). The field worker also took a fasting venous blood sample which was then stored at -80℃. Genomic DNA extracted was from the blood sample and diluted to a final concentration of 7ng/ µl. An aliquot of the DNA samples was used in this study to determine the subjects' respective UCH-L1 genotypes.

2.1.2 Polyacrylamide MADGE Gels

In this study, all PCR products and PCR endonuclease digestions were visualised by way of Microplate Array Diagonal Gel Electrophoresis (MADGE) on horizontal polyacrylamide (H-PAGE) gels (Day *et al.* (1994)).

2.1.2.1 MADGE Background

Microplate Array Diagonal Gel Electrophoresis was invented at the Division of Cardiovascular Genetics, Department of Medicine, University College Medical School in 1994.

The advantages of the MADGE system is that it uses the same 96 well format as is employed in most PCR reactions – a 9mm pitch between wells – which ensures compatibility with existing 96-well arrays and rapid loading with a multichannel pipette. Cubic wells are 2mm and the array is rotated by 71.6° to extend the track length to 26.5 to 50mm (Figure 11) (Gaunt *et al.* (2003)), thus ensuring good resolution of DNA bands and rapid interpretation of results.

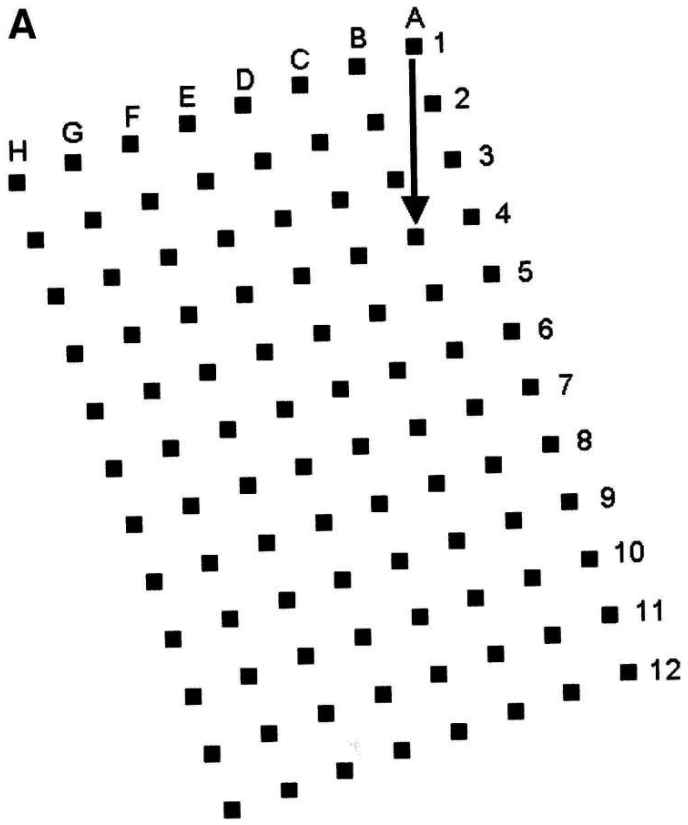


Figure 11 – 96-well MADGE gel showing track 'A1' with direction of electrophoresis (71.6°) (Gaunt *et al.* (2003)).

2.1.2.2 MADGE Gel Preparation

All horizontal polyacrylamide MADGE gels used in this study were 5% gels and were prepared using the following protocol:-

- Glass plate was cleaned using methanol (BDH Laboratory Supplies), and then wiped over lightly with sticky silane (refer to appendices).

- 8.3ml 30% acrylamide (Severn Biotech) was measured into a measuring cylinder, 5ml 10xTBE was added, and was filled to 50ml with deionised water and poured into a beaker.

- 150µl of 25% APS (refer to appendices) and TEMED were then individually added.

- The gel mixture was then poured into the former, and the glass plate was placed over the former and left for 20 minutes to set.

- Once set, the gel was prised apart with the MADGE gel adhering to the glass plate.

- The gel was then stained in an ethidium bromide (Sigma) solution (20µl per 200ml 1xTBE) for 20 minutes.

2.1.2.3 MADGE Gel Loading and Running

- 2µl of MADGE loading dye (refer to appendices) was added to 5µl PCR/ digest product and mixed well.

- 5µl of required samples were loaded into corresponding gel wells.

- Once samples were loaded, the gel was transferred into the 'gel running pack', and anode and cathode connections were secured.

- Gels were typically run between 8-15 minutes at 150 volts (0.02 amps). Time was dependent on size of DNA fragment (larger fragments take longer to run through gel matrix).

- Once gels had been run, they were visualised in a Fluorimager 595 (Molecular Dynamics) using ultra-violet to elucidate required DNA bands.

2.1.2.4 DNA Ladders

In this study, two different DNA ladders were used to decipher the size of DNA bands visualised on polyacrylamide and agarose gels – they were both sourced from Invitrogen and are displayed below:-

2.1.2.4.1 - 100bp DNA Ladder

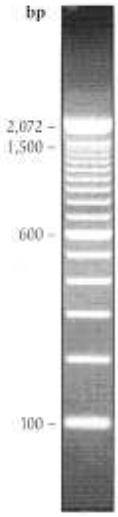


Figure 12 – The 100bp ladder consists of 15 blunt-ended fragments ranging in length from 100 to 1,500bp, at 100-bp increments, and an additional fragment at 2,072bp.

2.1.2.4.2 - 1kb DNA Extension Ladder

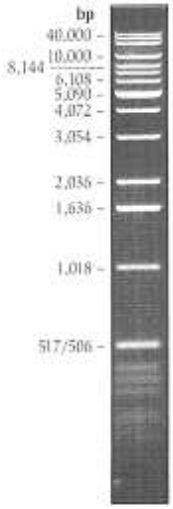


Figure 13 – The 1kb DNA extension ladder consists of 8 fragments ranging in length from 1,018 to 8,144-bp increments, as well as bands of 506bp, 517bp, 1,636bp, 10kb, 15kb, 20kb, and 40kb.

2.1.3 Polymerase Chain Reaction

(to amplify relevant *UCH-L1* DNA for dHPLC analysis)

2.1.3.1 Theory

Polymerase Chain Reaction (PCR) refers to a process for amplifying one or more specific DNA sequences contained in a nucleic acid or mixture of nucleic acids. It utilises two primers complementary to the ends of each specific sequence to be amplified – referred to as the 'sense' and 'anti-sense' primers. Extension of each primer creates a DNA strand including a sequence complementary to the opposite primer.

PCR is a cyclic procedure in which; 1) the relevant DNA strand is denatured; 2) the primers are allowed to anneal with the DNA strand; and 3) the primers are extended, synthesizing DNA complementary to the sequence to be amplified. These steps are repeated many times and result in an exponential amplification of the specific sequence. Each cycle doubles the amount of the specific DNA sequence being amplified (Figure 14).

The PCR method used in this study utilized a thermostable DNA polymerase to effect primer extension. This allowed the steps to be repeated by simple adjustment of the temperature and $MgCl_2$ concentration to affect denaturation, primer annealing, and polymerisation.

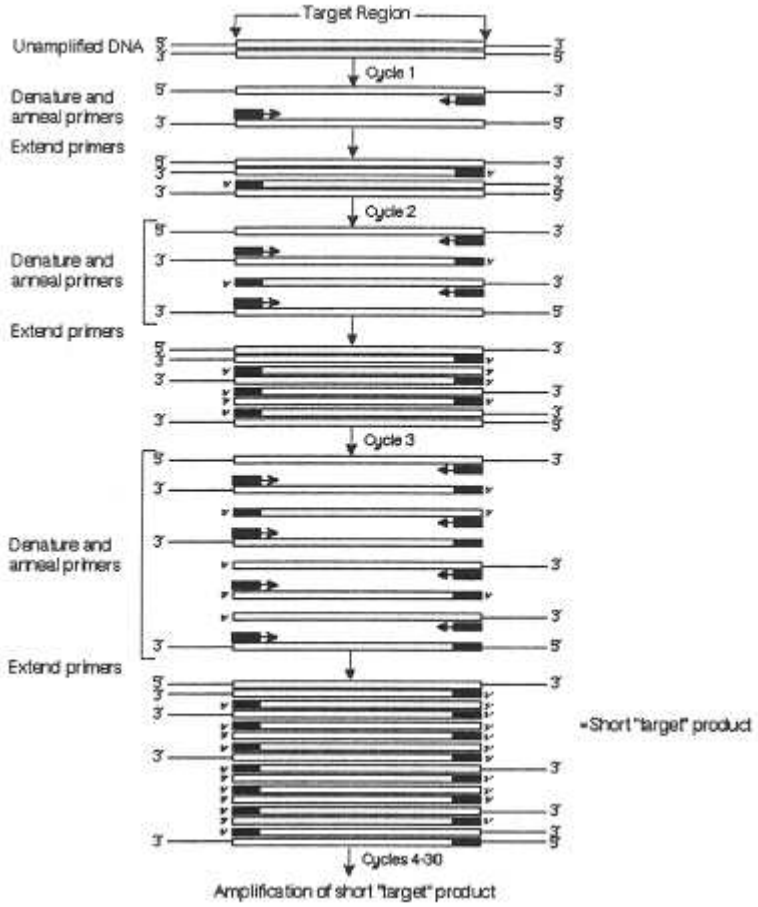


Figure 14 – Basic outline of the PCR reaction (Newton and Graham (1994)).

2.1.3.2 Protocol

The following quantities were used in preparation of all PCR reactions carried out (all reagents used were sourced form Invitrogen):-

- DNA (Genomic ~7ng/ µl/ Plasmid ~0.3µg/ µl) - **1µl**
- PCR Buffer (x 10) (200mM Tris-HCl, 500mM KCl) – **2.5µl**
- dNTPs (8 mM) – **0.625µl**
- $MgCl_2$ (25 mM) – **1.0- 3.0µl** (depending on PCR optimisation results)
- Sense Primer (1µl/ µg) – **0.1µl**
- Anti-sense Primer (1µl/ µg) – **0.1µl**
- DNA *Taq* Polymerase (5 units/ µl) – **0.2µl**
- Deionised Water – **19.475- 17.475µl** (depending on $MgCl_2$ required)

The following PCR reaction programme was employed for all PCR reactions (on a 'PTC DNA Engine Tetrad' (gradient cycler) – MJ Research):-

95℃ for 5 mins

50- 70℃ for 2 mins

72℃ for 2 mins

95℃ for 1 min
50- 70℃ for 1 min
72℃ for 2 mins
x 34

72℃ for 6 mins

END

The specific temperatures used between 50 and 70℃ for each DNA sequence amplified in this study, were determined in PCR optimisation reactions, in which the above reaction programme was employed.

2.1.3.3 Polymerase Chain Reaction + Betaine

2.1.3.3.1 Theory

Betaine improves the amplification of certain DNA sequences by reducing the formation of secondary structure caused by GC-rich regions, and is thus generally applicable to ameliorate the amplification of GC-rich DNA sequences (Henke *et al.* (1997)).

2.1.3.3.2 Protocol

PCR + betaine protocol remained the same as the 'normal' 'PCR protocol (refer to section 2.1.3), except the following quantities were always modified in preparation of the PCR reaction:-

- Addition of Betaine (5M) – **6.5µl** (1.3mM in final 25µl reaction concentration)
- Volume of deionised water was thus also reduced by **6.5µl**

2.1.3.4 PCR Primer Pairs Utilised to Amplify *UCH-L1* Regions

The table below outlines the primer sequences used to amplify the respective *UCH-L1* regions; the fourth column details the size of the resulting amplicon (please refer to appendices for concordant full nucleotide sequence) which were designed within a nucleotide size range between 150-450bp for optimal dHPLC analysis (Rossetti *et al.* (2002)). (Exon 9 was 15bp outside the upper limit of this designated optimal size range, but deemed suitably close for effective analysis).

Amplified *UCH-L1* Region (and concordant DNA Accession Code)	Name of Primer	Primer Sequence (5' → 3')) Sense (S)/ Anti-sense (AS)	Size of Amplicon (bp)
Promoter Region A (NW_001838903)	P1 P2	CCACTCACTTTGTTCAGCATC (AS) AGGCCTCACAGTGCGTCTG (S)	262
Promoter Region B (NW_001838903)	P3 P4	GGGAAGACGAAAAACAGCTA (AS) CATCAAAAGGACTGCTCCATAC (S)	281
Promoter Region C (NW_001838903)	P5 P6	GTGAGATAATCTGGTGGTTGTGGGA (AS) GTCACTTAAAAACGAACCTCGGTACT (S)	315
Promoter Region D (NW_001838903)	P7 P8	CAGTTTGCGTTTAACCTTAGACAAT (AS) TTTCAAAGCTTCCCATTCTTTTA (S)	328
Promoter Region E (NW_001838903)	P9 P10	AGAAAACAGATCCAGGGAAAAGATG (AS) TCGCTACCTAAGTATTTCTGCAAGC (S)	328
Exons 1 and 2 (X17377)	E1/2a E1/2b	CTCCCCCTGCACAGGCCTCA (S) GTCCCTGCCAGCAGCCGGAA (AS)	353
Exon 3 (AF076269)	E3a E3b	GCTTTGTGCT GTGTCATTGC (S) CTCAAGCTGGGGAGCGGC (AS)	363
Exon 4 (AF060834)	E4a E4b	TGCACTTTCATTCTGAGATG (S) GATGGCTGGCCTCAAAAC (AS)	228
Exons 5 and 6 (AF076270)	E5/6a E5/6b	AGGTTGCTCAGCATGTTCAG (S) CAGTAGAAACACAGATGGC (AS)	364
Exon 7 (AF076271)	E7a E7b	CTTAGTGGGCTTAGAATAGG (S) AAGTGCCCTCATGAGAATAC (AS)	372
Exon 8 (AF076272)	E8a E8b	ATCTAGGCTAGGTAAGCACG (S) TGCTGGCTATACTGGAAGAG (AS)	271
Exon 9 (AF076273)	E9a E9b	GGAGCCTTTCCCTATGTGAC (S) ACCACATCCAAGGTCTTAAC (AS)	465

Table 5 – Lists and details the PCR primer pairs used to amplify the respective UCH-L1 promoter and coding regions (intron distance for *UCH-L1* be found in the appendices).

2.1.4 denaturing High Performance Liquid Chromatography

(to identify *UCH-L1* regions harbouring potential mutations/ polymorphisms)

2.1.4.1 Basic HPLC Theory

High-Performance (or High Pressure) Liquid Chromatography (HPLC) is a form of column chromatography used to separate, identify and quantify. HPLC employs a column that holds chromatographic packing material (the absorbent or stationary phase), a pump that moves the mobile phases (sample being analysed and the solvent that moves it through the column (the most common are methanol and acetonitrile, but the choice is dependent on the nature of the stationary phase and the sample), and a detector that measures the retention times of the molecules.

In HPLC the sample to be analysed (or analyte) is introduced in small volume to the stream of solvent (mobile phase) and is retarded by specific chemical and physical interactions with the stationary phase as it transverses the column length. The nature of the analyte, stationary phase and the mobile phase composition all effect the degree of retardation. One of the fundamental principles of HPLC is the fact that the time it takes for a specific analyte to elute from the mobile phase (retention time) is a unique identifying characteristic of said analyte. The use of pressure in HPLC increases linear velocity through the column, thus giving the component s less time to diffuse, leading to improved resolution in the resulting chromatogram.

Varying the mobile phase composition during the analysis is typically employed and is known as ‘gradient elution’, in which the gradient separates the analyte mixtures as a function of the affinity for the current mobile phase composition relative to the stationary phase (e.g. in a water/ methanol gradient, the more hydrophobic components will elute under relatively high methanol conditions (hydrophobic), whereas the more hydrophilic components will elute relatively low methanol (high water) conditions).

Optimum HPLC methods for a given analyte are those which give the best separation of peaks on the resulting chromatogram.

2.1.4.2 dHPLC Theory

Individuals who are heterozygous in a mutation or polymorphism have a 1:1 ratio of wild-type and mutant DNA. A mixture of hetero- and homoduplexes is formed when the PCR product is hybridised by heating to 95℃ and slowly cool ed. After this treatment, a sample will contain a mixture of hetero- and homoduplexes (Figure 15).

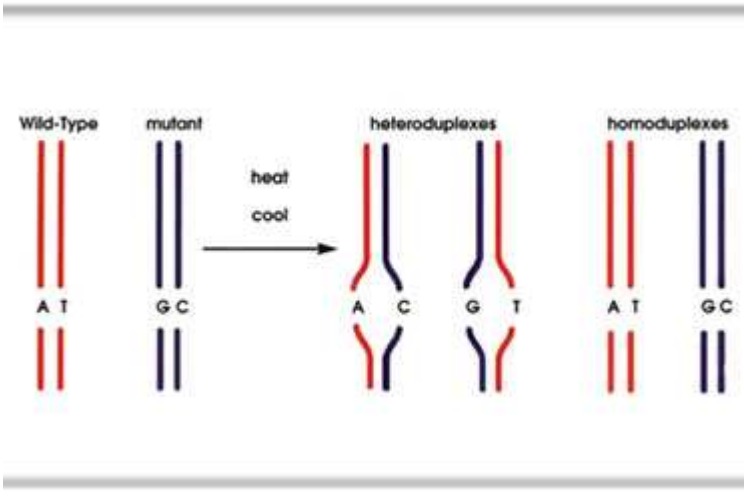


Figure 15 – Creation of a mixture of hetero- and homoduplexes through hybridisation (Taylor et al. (1999)).

Mutations are visualised as a characteristic pattern of peaks corresponding to the mixture of heteroduplexes and homoduplexes formed when wild-type and mutant DNA are hybridised.

Figure 16 (refer below) shows the behaviour of four hybridised species in response to a range of temperature regimes used in an experiment. Under the non-denaturing conditions used for separating the DNA fragments (51℃), all f our species have the same retention time. As the temperature increases to 54℃, the heterodupl ex DNA fragments start to denature in the region either side of the mismatched bases and begin to emerge ahead of the still intact homoduplexes. At 56℃ the homoduplexes start to den ature, with the A-T homoduplex being marginally more denatured than the C-G homoduplex. Optimum separation is seen to be at 57℃.

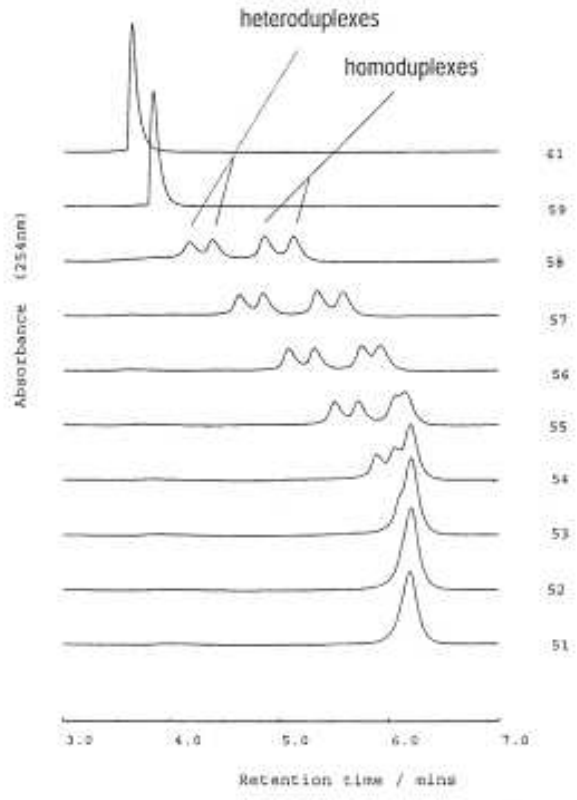


Figure 16 – Temperature-dependent resolution of heteroduplexes from homoduplexes. In the absence of the mutation, only one peak, that of the wild-type homoduplex, would be observed (Taylor *et al.* (1999)).

2.1.4.3 Setting up a dHPLC Method in 'Wavemaker™' (Transgenomic, Inc)

- Required DNA sequence was inserted into the DNA sequence panel.

- A helical fraction vs. temperature graph could then be calculated. From this graph, the optimum temperature for each major melting domain (segment of DNA with a difference in helical fraction vs. temperature characteristic) of the DNA sequence was identified – between 75 and 99% helical fraction.

- Method(s) could then be built for each respective melting domain of each DNA sequence to be dHPLC scanned (entire DNA sequences could be realised within 75 and 99% helical fraction at the same temperature).

- The oven temperature was set to that calculated above.

- The ‘Gradient Template’, which creates the gradient of Buffer A and Buffer B (refer to appendices); critical in the differing elution of heteroduplexes from homoduplexes, and thus their resolution, was then modified in the ‘Gradient Parameters’ panel as follows:-

Drop for Loading – **5%**	- Allows non-retained components such as dNTPs, primers and buffers to pass through the separation cartridge when the PCR products are injected, so they do not influence the separation.
Loading Duration – **0.1 mins**	Sets the time over which the ‘Drop for Loading’ increases (% Buffer B), to the start of the linear gradient.
Gradient Duration – **4 mins**	Length of time over which the % Buffer B increases linearly at a % Buffer B per minute. (Selected by ‘Slope’ option – default of 2% used)
Clean Duration – **0.5 mins**	Length of time at which the % Buffer B remains at 100% to clean high molecular weight DNA such as genomic DNA and contaminants from the cartridge.
Equilibration Duration - **20 mins**	Length of time that the % Buffer B remains at the initial ‘Drop for Loading’ percentage, to equilibrate for the next injection.

- The rest of the gradient parameters were left at the default values, as they were not deemed important in the dHPLC application used.

- Finally the ‘Acquisition Time’ (time of individual dHPLC scan) was set to **6.8 mins**.

- The ‘Wavemaker’ method was then saved and exported for running into ‘HSM’ (main operating software for the dHPLC apparatus – Transgenomic, Inc).

2.1.4.4 dHPLC Protocol

Before the samples could be placed into the 'auto-sampler' for the dHPLC mutation scanning run to commence, the following preparatory steps had to be carried out:-

- Buffers A, B and the Wash Solution (Buffer D) were made up fresh every 7 days. Buffer C (75% acetonitrile (Sigma)) was renewed every 3-4 weeks.

- Pumps A, B, C and D were purged for 3-4 minutes to remove any air build up from the system.

- Autosampler line was washed 5 times to remove any air build up.

- Column equilibration took place for 10 minutes at a flow rate of 0.9ml/ minute.

- 'Rack Parameters' (dimensions of sample well plate) for 96 well plate containing PCR products were checked in 'HSM System Manager' software.

- Sample table was prepared according to PCR products to be scanned and 'Wavemaker' methods required. **8µl** was entered as amount of sample to be injected, and a **3 minute** equilibration time was also entered after each different method to be used in the sample table.

- A 75% acetonitrile (Aldrich) wash to clean the separation cartridge, followed by a low flow (0.05ml/ minute – 'sleep mode') method were always entered into the sample table to end a dHPLC scanning run.

- In the 'Autosampler Set Up' the 'needle down speed' was set to 'Fast', the 'injection method' was set to 'Cut', and the 'lead' and 'rear volumes' were set to 1.0µl.

Once the preparatory steps above had been completed, the samples (in a 96 well plate), which had been heated to 95°C and slowly cooled to allow heteroduplex formation in any samples with a mutation/ polymorphism, were then placed in the autosampler in the pre-specified positions. The dHPLC run was then initialised.

2.1.5 DNA Sequencing

(to characterise specific nature of mutations/ polymorphisms)

2.1.5.1 Underlying Theory

DNA sequencing is the determination of the precise sequence of nucleotides in a sample of DNA. The method used in this study is referred to as the 'dideoxy' or 'chain termination' method.

DNA is synthesised from four deoxynucleotide triphosphates. Figure 17 shows one of them – deoxythymidine triphosphate (dTTP).

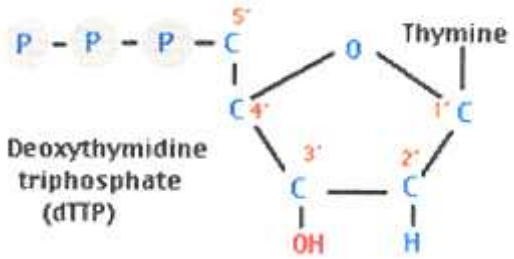


Figure 17 – Chemical structure of deoxythymidine triphosphate.

The 'dideoxy method' gets its name from the critical role played by synthetic nucleotides that lack the OH at the 3' carbon atom (red arrow in Figure 18) – the dideoxynucleotide dideoxythymidine triphosphate (ddTTP) is shown in Figure 18.

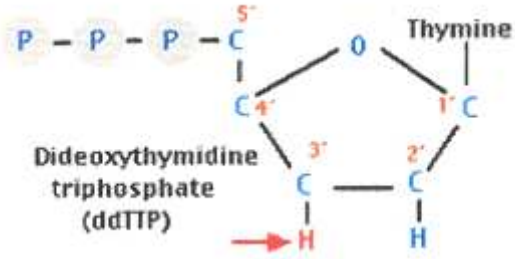


Figure 18 – Chemical structure of dideoxythymidine triphosphate. Lack of a OH group on the 3' carbon atom is indicated with a red arrow.

When a dideoxynucleotide is added to the 'growing' DNA strand, chain elongation stops because there is no 3' OH for the next nucleotide to be added to. For this reason, the 'dideoxy method' is also called the 'chain termination method'.

2.1.5.2 Theory

The DNA to be sequenced is prepared as a single strand. This template DNA is supplied with a mixture of all four normal (deoxy) nucleotides in ample quantity – dATP, dGTP, dCTP and dTTP. A mixture of all four dideoxynucleotides are also supplied in limiting quantities, and each labelled with a fluorescent tag that fluoresces a different colour – ddATP, ddGTP, ddCTP, and ddTTP.

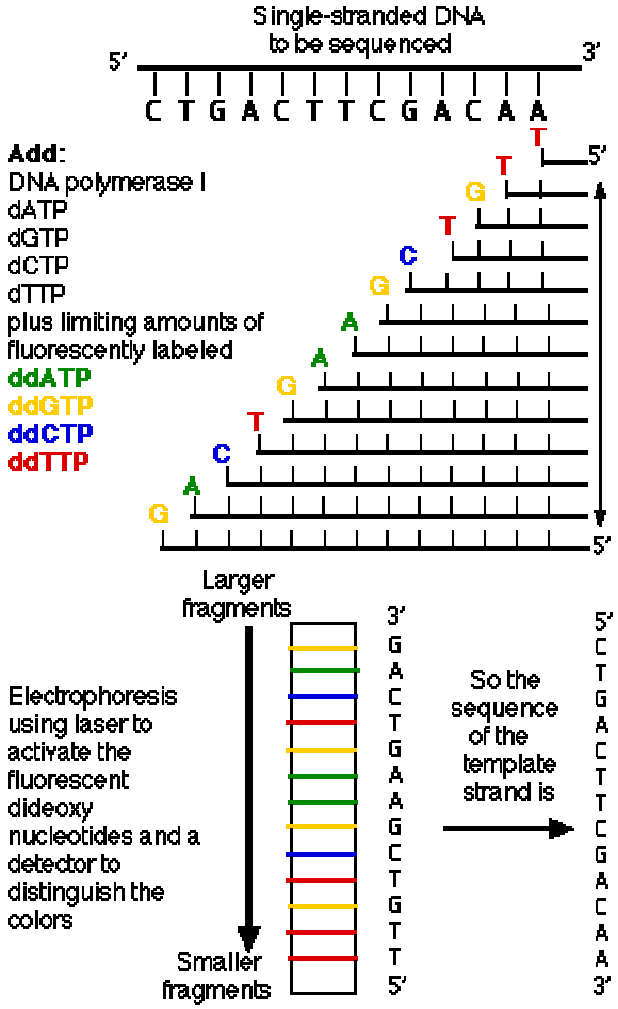


Figure 19 – Pictorial summary of the dideoxy DNA sequencing method.

Because all four normal nucleotides are present, chain elongation proceeds normally until, by chance, DNA polymerase inserts a dideoxy nucleotide (shown as coloured letters in Figure 19) instead of the 'normal' deoxynucleotide (shown as vertical lines in Figure 19). If the ratio of normal nucleotide to the dideoxy versions is high enough, some DNA strands will succeed in adding several hundred nucleotides before insertion of the dideoxy version halts the elongation process.

At the end of the 'incubation period', the fragments are separated by length from longest to shortest on a gel. The resolution is so good that a difference of one nucleotide is enough to separate a particular strand from the next shorter/ longer strand. Each of the four dideoxynucleotides fluoresces a different colour when illuminated by a laser beam, and an automatic scanner provides a printout of the sequence.

2.1.5.3 Shrimp Alkaline Phosphatase/ Exonuclease I Sample Preparation
(PCR Products Only)

Shrimp alkaline phosphatase (SAP) and exonuclease 1 (Exo I) were used to prepare PCR products for sequencing. SAP removes the phosphate groups from the excess dNTPs left over from the PCR reaction, and Exo I digests the single stranded PCR primers into dNTPs.

The following quantities were used in preparation of all SAP/ Exo I reactions carried out on PCR products (all reagents used were sourced form USB):-

- PCR Product (~ 50ng/ µl) - **5µl**
- SAP - **1µl** (1 unit)
- Exo I – **0.1µl** (1 unit)
- SAP Buffer (x10) – **0.9µl**

The following SAP/ Exo I reaction programme was employed for all SAP/ Exo I reactions (on a 'PTC DNA Engine Tetrad' (gradient cycler) – MJ Research):-

37℃ for 30 mins

80℃ for 15 mins

END

2.1.5.4 Sequencing Reaction Protocol

The following quantities were used in preparation of all sequencing reactions carried out (all reagents were sourced from Applied Biosystems):-

- DNA Template – **3.5µl** for PCR products from SAP/ Exo I clean-up reaction,

  1.0µl of plasmid DNA (~0.3µg/ µl)

- Big Dye Terminator Ready Reaction Mix – **4.0µl** (½ unit)
- ½ Sequencing Buffer – **4.0µl** (½ unit)
- Sequencing Primer – **3.2pmol**
- Deionised Water – made up to **20µl**

The following sequencing reaction programme was employed for all sequencing reactions (on a 'PTC DNA Engine Tetrad' (gradient cycler) – MJ Research):-

96℃ for 10 secs
50℃ for 5 secs
60℃ for 4 mins
x 25

END

2.1.5.5 Ethanol Precipitation Protocol

Ethanol allows the precipitation of DNA at low temperatures, which can then be recovered by centrifugation and redissolved in appropriate buffer. This method was employed to clean the single stranded DNA from the sequencing reaction (refer to section 2.1.5.4).

The following protocol was employed for the ethanol precipitation of all sequencing reactions:-

- 16µl deionised water and 64µl ice cold 95% ethanol (Fisher) was added to the 20µl sequencing reaction, which was then vortexed and left at -20℃ for 15 minutes.

- The tubes were then centrifuged at 13,500 revolutions per minute (rpm) for 20 minutes.

- The supernatants were then carefully aspirated and 250µl of 70% ethanol (diluted with deionised water from Fisher stock) was added to the tubes which were then vortexed.

- The tubes were then centrifuged at 13,500rpm for 10 minutes.

- The supernatants were then carefully aspirated, and the samples were placed in a 90℃ heat block for 1 minute. Dried samples were then stored at -20℃.

2.1.5.6 Sequencing Gel Preparation

The following protocol was employed for the preparation of all sequencing gels used in this study:-

- Sequencing plates were cleaned thoroughly using Alconox (Aldrich), and rinsed with deionised water.

- The plates and spacers were then assembled in the gel frame as shown in Figure 20 (gel frame clips were rotated to lock assembly in place).

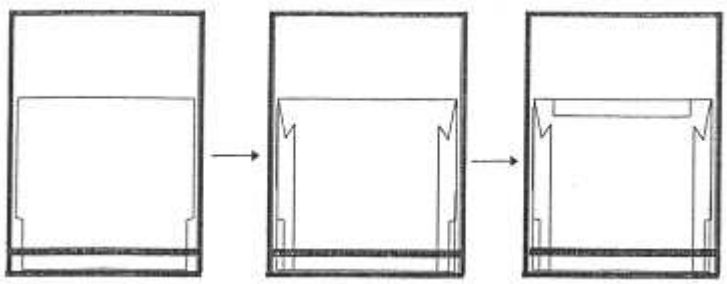


Figure 20 – Diagram outlining the assembly of the sequencing plates and spacers within the gel frame.

- The gel mixture – 30ml 10% 6M Urea (Amresco), 300µl 10% ammonium persulphate solution (APS – refer to appendices) and 30µl 100% N, N, N', N'-Tetramethyl-1-, 2-diaminomethane (TEMED) (Sigma) – was poured, once mixed, from the top centre using a 25ml syringe.

- The continuous edge of the comb was then inserted at the top, and a pressure device was fixed.

- Polymerisation of the gel was allowed to take place - ~2 hours.

- After polymerisation, the outside of the plates were cleaned using deionised water, with particular attention to the scan region (lower eighth).

- Pressure device was then removed, as was the comb, and any loose acrylamide was cleaned off using deionised water and white tissue.

- The comb was then reinserted teeth down, ensuring only 1½mm of the teeth entered the upper gel surface.

2.1.5.6.1 Plate Check (to check scan region cleanliness)

- The lower reservoir was placed in the ABI 377 sequencer (Applied Biosystems), as was the frame/ gel assembly which was clipped in place with four clips. Door was then closed.

- Once the software ('ABI Prism 377 DNA Sequencer Data Collection' – Applied Biosystems) had been initialised, 'Plate Check' was selected.

- A scan window with flat lines in each of the four colours was desired as an indicator of a clean laser scan region. If any peaks were seen, the gel cassette was removed and the scan region was cleaned using deionised water and white tissue until the peaks were no longer visible.

- Once flat lines had been realised the plate check was terminated.

2.1.5.6.2 Pre Run

(to equilibrate gel running conditions prior to sample loading)

- The upper reservoir was locked into place at the top of the gel, and both reservoirs were filled with 1xTBE (refer to appendices; 10xTBE was diluted by a further factor of ten in deionised water). The heat plate was also clipped in place. Door was then closed.

- The desired 'pre run module' was chosen within the software options – 'PR 36E 1200'. Pre run was left for 10 minutes.

2.1.5.7 Sample Preparation/ Loading

The following procedure was carried out to prepare all dry DNA samples from the ethanol precipitation step (above) for DNA sequencing:-

- 6µl of loading buffer – 1 part 25mM ethylenediaminetetraacetic acid disodium salt (EDTA) blue dextran (Applied Biosystems) + 5 parts deionised 100% formamide (Sigma) was added to each DNA pellet. Samples were then vortexed and briefly centrifuged.

- Samples were then denatured by heating them at 96°C for 2 minutes in a heat block. Samples were then immediately placed on ice to prevent re-annealing.

2.1.5.8 Sequence Run

- Once the 10 minute pre run had been completed, the wells were flushed using a syringe.

- 1.5µl of the prepared samples were then loaded into the required wells.

- The upper reservoir lid was then clipped in to avoid buffer evaporation, and the required 'run module' was selected within the software options – 'Seq Run 36E 1200'.

- The required 'sample sheet' was then imported into the run software, and the sequencing run was initialised.

Once the sequencing run had been completed, analysis of the sequencing results took place on a separate computer connected to the ABI 377 sequencer, and employed 'ABI Sequencing Analysis' software (Applied Biosystems) to convert the gel image realised into individual electropherograms displayed in the results section.

2.2 Single Nucleotide Polymorphism Genotyping

(to find common haplotypes within Caucasian population)

2.2.1 Restriction Endonucleases

The table below displays the manufacturer and relevant incubation conditions for all of the restriction endonucleases used in this study (respective buffers were manufactured by the concordant enzyme manufacturer):-

Endonuclease	Manufacturer	Temperature	Duration	Buffer
***Nci* I**	Promega	37℃	16 hours	Buffer 4
***Sgr* AI**	NEB	37℃	16 hours	Buffer 4
***Taq* I**	Promega	65℃	12 hours	Buffer E
***Rsa* I**	Promega	37℃	12 hours	Buffer C
***Fok* I**	NEB	37℃	16 hours	Buffer 4
***Mbo* II**	NEB	37℃	16 hours	Buffer 3
***Kpn* I**	Promega	37℃	12 hours	Buffer J
***Bgl* II**	Promega	37℃	12 hours	Buffer D
***Eco* RI**	NEB	37℃	12 hours	Eco RI Buffer
***Hind* III**	Promega	37℃	12 hours	Buffer E
***Sac* I**	Promega	37℃	12 hours	Buffer 1
***Spe* I**	Promega	37℃	12 hours	Buffer B

Table 6 – Displays all the relevant information for each restriction endonuclease used in this study. – *Abbreviation*: NEB, New England Biolabs.

2.2.2 PCR Product Endonuclease Digestion

The basic quantities indicated below were used in preparation for all endonuclease digests carried out on PCR products in this study (the bovine serum albumin (BSA) and relevant buffers were sourced from the manufacturers of the respective enzyme - detailed in section 2.2.1):-

- DNA – **1.0-3.0µl** (~100ng),
- 10x BSA – **1.0µl**
- 10x relevant buffer – **1.0µl**
- Restriction Endonuclease(s) (10units/µl) – **0.5µl** or **0.4µl** (for two in combination)
- Deionised water – made up to **10µl**

2.2.3 Primer-Introduced Restriction Analysis-Polymerase Chain Reaction

2.2.3.1 Theory

Primer-introduced restriction analysis (PIRA-PCR) was used in an attempt to detect SNPs. It was used to create artificial RFLP, where a mismatch was introduced near the 3' end of the sense or anti-sense primer that was close to the relevant SNP. Computer software (Ke *et al.* (2001)) was used to screen for suitable mismatches, design the required primers and list the relevant restriction enzymes.

2.2.3.2 Protocol

Once the PIRA-PCR strategy had been chosen, the required PIRA-PCR primers were used as normal primers in a PCR reaction (refer to section 2.1.3.2).

2.2.4 Restriction Fragment Length Polymorphism Genotyping Methods

Using restriction analysis software (Lasergene – DNAstar) the first stage in applying this technique was to decipher what changes the observed SNPs have given rise to within the respective region's sequence, in view of endonuclease restriction sites.

The basis to good reliable genotyping using restriction enzymes is to realise an output, i.e. the digested polymorphic versus the digested reference homozygote DNA, which can be easily deciphered between on an agarose/ polyacrylamide gel. Ideally, this would mean the appearance/ disappearance of a DNA digestion band of a molecular weight which is easily observed on a gel.

For this to be the case, in view of the *UCH-L1* regions that have elucidated SNPs (which average around 280bp), restriction endonucleases with more than two cutting sites (in polymorphic or reference homozygote DNA) were discounted as to keep the digestion output simple and easier to decipher.

For three of the five promoter SNPs (C-16T, G-234A and A-307G) and the documented coding region C54A SNP this was accomplished, whereas two of the promoter SNPs - A-24G and A-306G - did not present a viable restriction endonuclease site, and thus required additional DNA manipulation techniques (refer to section 2.2.3) in an attempt to give rise to a feasible RFLP genotyping strategy. Methodologies employed for each of these SNPs are detailed below.

2.2.4.1 C-16T SNP Genotyping

The restriction endonuclease chosen was ***Nci* I**; it realises two restriction sites within promoter region A's sequence, one of which is lost in DNA harbouring the thymine (polymorphic) allele. *Nci* I's restriction site is thus:-

CC↓ N GG
GG N ↑CC

(N denotes any nucleotide).

- Relevant corresponding SNP region:-

CTGCCGGGCGC
C-16T

The enzyme cleaves promoter region A reference homozygote sequences at 90 and 162bp, however the 90bp cut is lost in sequences harbouring the polymorphic T allele.

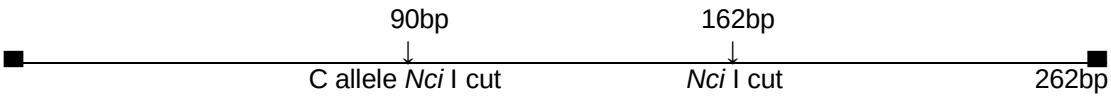


Figure 21 - Linear representation of promoter region A and the *Nci* I restriction sites.

- Thence, DNA harbouring the reference homozygote C allele realises three bands (digested DNA viewed on a gel) of 72, 90 and 100bp. Polymorphic DNA however (containing the T allele), realises only two bands of 100bp and 162bp

Promoter region A/ *Nci* I endonuclease digestion products were then run on a horizontal polyacrylamide gel (refer to section 2.1.2). The critical 162bp band was easily deciphered, and verified using a 100bp DNA ladder (refer to section 2.1.2.4).

2.2.4.2 G-234A SNP Genotyping

The restriction endonuclease chosen was ***Sgr* AI**; it has one restriction site within promoter region B's sequence, which is lost in DNA harbouring the adenine (polymorphic) allele. *Sgr* AI's restriction site is thus:-

CR↓ CCGG YG
GY GGCC ↑RC

(Y denotes a pyrimidine).
(R denotes a purine).

- Relevant corresponding SNP region:-

TCTCACCGGCGAGT
G-16A

The endonuclease cleaves promoter region B reference homozygote sequences at 78bp, however this cut is lost when sequences harbour the polymorphic A allele.

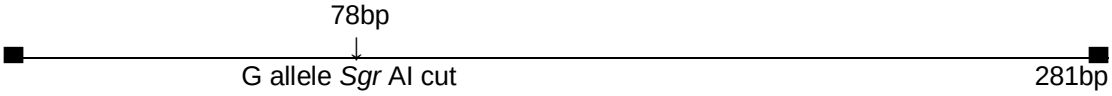


Figure 22 - Linear representation of promoter region B and the *Sgr* AI restriction site.

- Thence, DNA harbouring the reference homozygote G allele realises two bands of 78 and 203bp. Whereas polymorphic DNA (containing the A allele), realises just one band - the undigested 281bp DNA fragment of promoter region B.

Promoter region B/ *Sgr* AI restriction enzyme digestion products were then run on a horizontal polyacrylamide gel. The critical 281bp band was easily deciphered, and verified using a 100bp DNA ladder.

2.2.4.3 A-307G SNP Genotyping

The restriction endonuclease chosen was ***Taq* I**; it realises one cutting site within promoter region C's sequence, and realises a second in DNA harbouring the guanine (polymorphic) allele. *Taq* I's restriction site is thus:-

T↓ CG A
A GC ↑T

- Relevant corresponding SNP region:-

CCATCAAAAG
A-307G

Taq I cleaves promoter region C reference homozygote sequences once at 172bp, however a 244bp cut is also realised in sequences harbouring the polymorphic G allele.

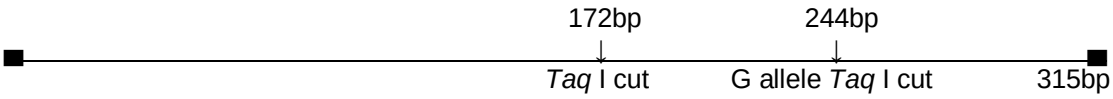


Figure 23 - Linear representation of promoter region C and both *Taq* I's restriction sites.

- Thence, reference homozygote DNA realises two bands of 143 and 172bp. Polymorphic G allele DNA however, realises three bands of 71 and 72bp and another of 172bp.

Promoter region C/ *Taq* I restriction digestion products were then run on a horizontal polyacrylamide gel. The critical 71 and 72bp bands, though indistinguishable, were easily deciphered, and verified using a 100bp DNA ladder.

2.2.4.4 C54A (S18Y) SNP Genotyping

Rsa 1 does not have a restriction site within the amplified reference homozygote exon 3 region, however, a restriction site is created in DNA harbouring the polymorphic adenine allele. *Rsa* 1's restriction site is thus:-

GT↓ AC
GT ↑TG

- Relevant corresponding SNP region:-

GCTGTCCCGG
C54A

The endonuclease cleaves the amplified polymorphic exon 3 sequence at 180bp, no cut is realised in the amplified exon 3 region harbouring the reference homozygote C allele.

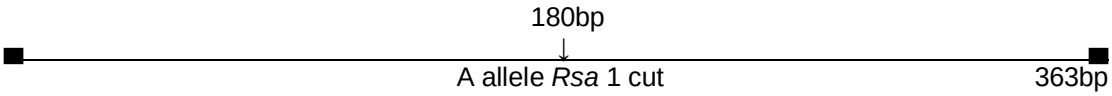


Figure 24 - Linear representation of the amplified exon 3 sequence and the Rsa 1 restriction site.

Thence, reference homozygote C allele DNA realises just one band - the undigested 363bp fragment. Whereas DNA harbouring the polymorphic A allele realise the 180 and 183bp bands.

2.2.4.5 A-24G SNP Genotyping

The A-24G SNP region did not present a viable restriction endonuclease site in its reference homozygote or polymorphic sequence that would have allowed straightforward RFLP genotyping to be carried out.

An alternative option was to introduce a novel restriction site into the 5' UTR SNP region by way of 'primer-introduced restriction analysis (PIRA) PCR' (refer to section 2.2.3).

Using computer software specially constructed for this purpose (Ke *et al.* (2001)), a pair of PCR primers were designed which would amplify the required region, and at the same time incorporate a base mismatch in the required region (brought about by the sense primer)

to give rise to a novel endonuclease restriction site. The actual novel restriction site to be introduced is that of ***Fok* I**.

Primer pair sequences:-

Sense primer (5' → 3'):- TAGCTGTTTTTCGTCTTCCCTAGGAT – (*UCH-L1* -24 M1)

Anti-sense primer (5' → 3'):- TCTCCGGGTAGCTGTGCAC – (*UCH-L1* -24 M2)

- How the primers should adhere to the required *UCH-L1* sequence flanking the A-24G SNP region is displayed:-

TAGCTGTTTTTCGTCTTCCCTAGGAT>
TAGCTGTTTTTCGTCTTCCCTAGGCTATTTCTGCCGGGCGCTCCGCGAAGATGCAGCTC
A-24G
.....AAGCCGATGGAGATCAACCCCGAGGTGAGCGCCAGGTGCACCGCTACCCGGAGA
<CACGTGGCGATGGGCCTCT

- The mismatch brought about by the sense primer is highlighted (above and below) in green, the A-24G SNP is also highlighted.

- The restriction endonuclease site to be created for *Fok* I is thus:--

GGATG…$NNN_{(9)}$..↓
CCTAC……$NNN_{(13)}$..↑

- Relevant corresponding SNP region (after successful mismatch):-

CTAGGATATTT
A-24G

With a successful mismatch introduced, *Fok* I does not cleave the reference sequence, however, a 36bp cut should be realised in sequences harbouring the polymorphic G allele.

Figure 25 - Linear representation of the specified A-24G mismatch PCR region and *Fok* I's restriction site.

- Thence, reference homozygote DNA should realise just one band - the undigested 141bp DNA fragment. Whereas DNA harbouring the polymorphic G allele should realise a 36bp band and another of 105bp.

A-24G mismatch region/ *Fok* I endonuclease digestion products were then run on a ten percent polyacrylamide gel. However, no *Fok* I digestions were realised. Positive controls known to harbour the polymorphic G allele also came out negative. It was thus determined that this mismatch PCR/ endonuclease digestion strategy was failing for an unknown reason (the required mismatch was confirmed through DNA sequencing analysis).

2.2.4.6 A-306G SNP Genotyping

As with the A-24G SNP region, the A-306G SNP region did not present a viable restriction endonuclease site in its reference homozygote or polymorphic sequence which would have allowed straight forward RFLP genotyping to be carried out.

As before, an alternative option was to introduce a novel restriction site into the promoter SNP region by way of 'PIRA-PCR'.

Using the computer software indicated above, a pair of PCR primers were designed which would amplify the required region, and at the same time incorporate a base mismatch in the required region (brought about by the anti-sense primer) to give rise to a novel endonuclease restriction site. The actual novel restriction site to be introduced is that of ***Mbo* II**.

Primer pair sequences:-

Sense primer (5' → 3'):- TCAAATGCTTCAGAGACTCGAGC – (*UCH-L1* -306A/G S)
Anti-sense primer (5'→3'):- TTCCTTGAGTGTATGGAGCAGTTCTT–(*UCH-L1*-306A/GAS)

- How the primers should adhere to the required *UCH-L1* sequence flanking the A-306G SNP region is displayed:-

TCAAATGCTTCAGAGACTCGAGC >
TCAAATGCTTCAGAGACTCGAGCTTTAGAGTAATTGGGATGGTGAAAGGATGGGTTTC
A-306G
CAGAAACTTCGCCCAAAATTAAAGACTCCATCAAAAGGACTGCTCCATACACTCAAGGAA
<TTTATGACGAGGTATGTGAGTTCCTT

- The mismatch brought about by the anti-sense primer is highlighted (above and below) in green, the A-306G SNP is also highlighted.

- The restriction endonuclease site to be created for *Mbo* II is thus:--

GAAGA....$NNN_{(8)}$..↓
CTTCT....$NNN_{(7)}$..↑

- Relevant corresponding SNP region (after successful mismatch):-

TCAAAAAGACT
A-306G

With a successful mismatch introduced, *Mbo* II does not cleave the reference sequence, however, a 12bp cut should be realised in sequences harbouring the polymorphic G allele.

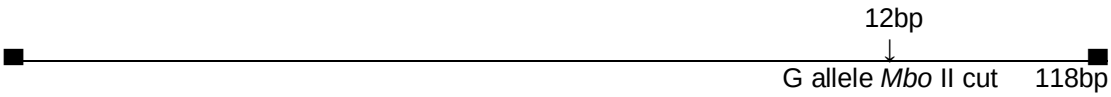


Figure 26 - Linear representation of the specified A-306G mismatch PCR region and *Mbo* II's restriction site.

- Thence, reference homozygote DNA should realise just one band - the undigested 118bp DNA fragment. Whereas DNA harbouring the polymorphic G allele should realise a 12bp band and another of 106bp.

A-306G mismatch region/ *Mbo* II endonuclease digestion products were then run on a ten percent polyacrylamide gel. However, as with the previous strategy, no *Mbo* II digestions were realised. Positive controls known to harbour the polymorphic G allele also came out negative. It was thus determined that this mismatch PCR/ endonuclease digestion strategy

was once again failing for an unknown reason (again, the required mismatch was confirmed through DNA sequencing analysis).

2.3 Cloning Methodologies

2.3.1 Plasmid DNA Vectors

2.3.1.1 pGEM-T Easy Theory

The pGEM-T Easy vector (Promega) (Figure 51) was chosen as it is a convenient system for the cloning of PCR products – in this study the *UCH-L1* promoter haplotypes. The pGEM-T Easy vector has an insertion site which is flanked by single 3'-T overhangs which greatly improve the efficiency of PCR product ligation into the plasmid. This is achieved by the T overhangs preventing recircularisation of the vector, and providing compatible overhangs for PCR products generated by certain thermostable *Taq* polymerases (as used in this study) that add a single deoxyadenosine, in a template-independent fashion, to the 3'-ends of the amplified PCR fragment.

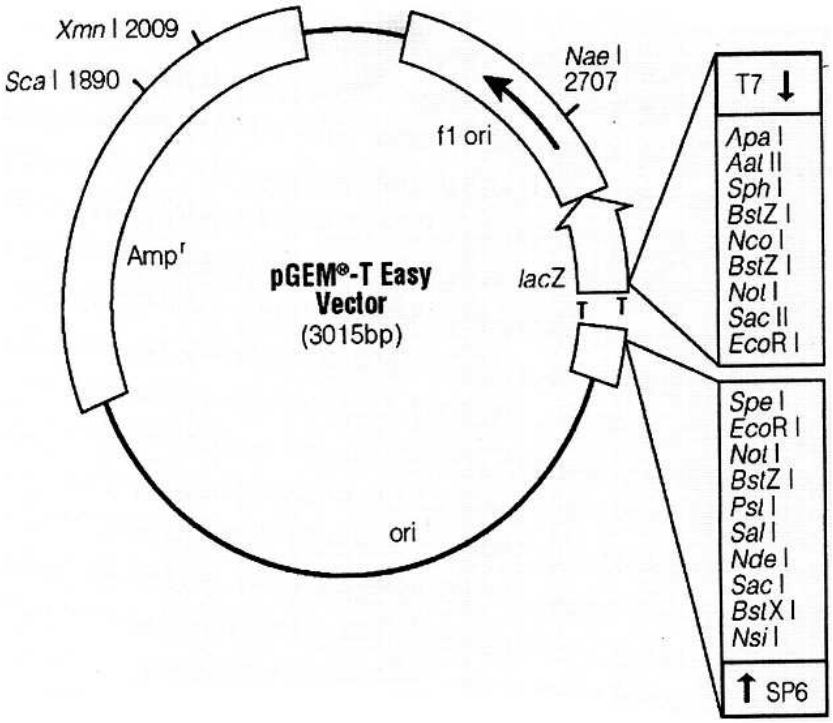


Figure 27 – Schematic representation of the pGEM-T Easy vector showing the gene conferring ampicillin resistance in *E.coli* (Amp^r), the origin of replication in *E.coli* (ori), the origin of replication derived from filamentous phage (f1 ori), the *lac* operator (*lac*Z – start codon), the T7 RNA polymerase transcription site (T7), and the SP6 RNA polymerase promoter (SP6). (www.Promega.com).

2.3.1.2 pGL3-Basic Vector

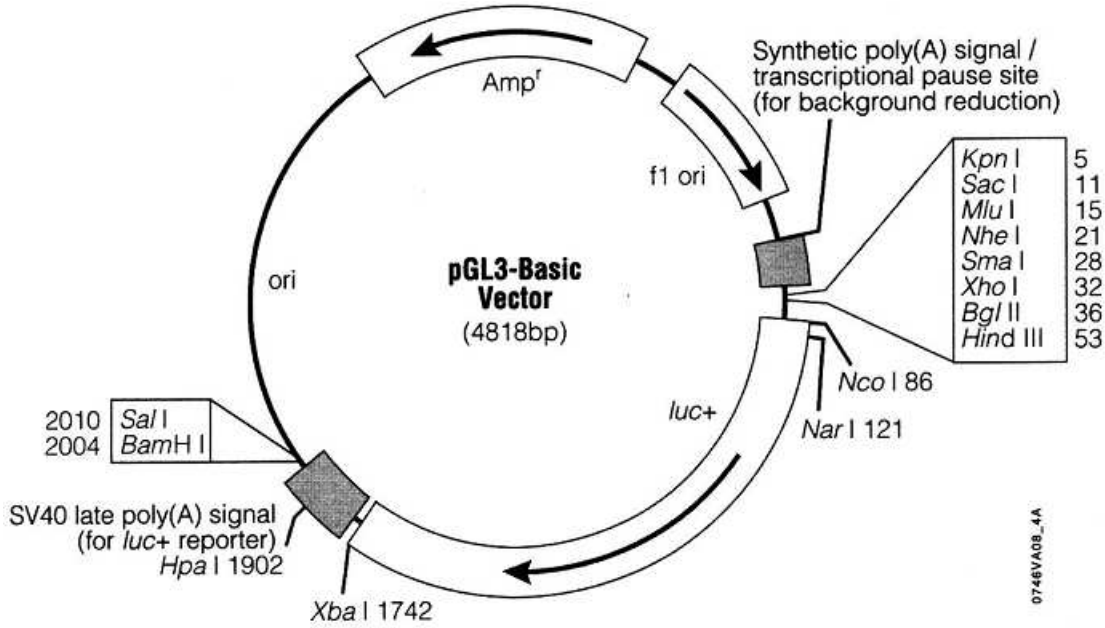


Figure 28 – Schematic representation of the pGL3-Basic vector showing the cDNA encoding the modified firefly luciferase (*luc*+), the gene conferring ampicillin resistance in *E.coli* (Amp^r), the origin of replication in *E.coli* (ori), the origin of replication derived from filamentous phage (f1 ori), and its endonuclease restriction sites; those between nucleotide sequences 1 and 58 make up the 'multiple cloning region'. Other components are described in diagram. (www.Promega.com).

2.3.1.3 pGL3-Eco RI Construction

The pGL3-Eco RI vector used in this study was customly engineered from Promega's pGL3-Basic and pBluescript SK vectors to create a cloning vector with an expanded range of restriction endonuclease sites within the cloning region. The construction strategy is indicated in the figure below:-

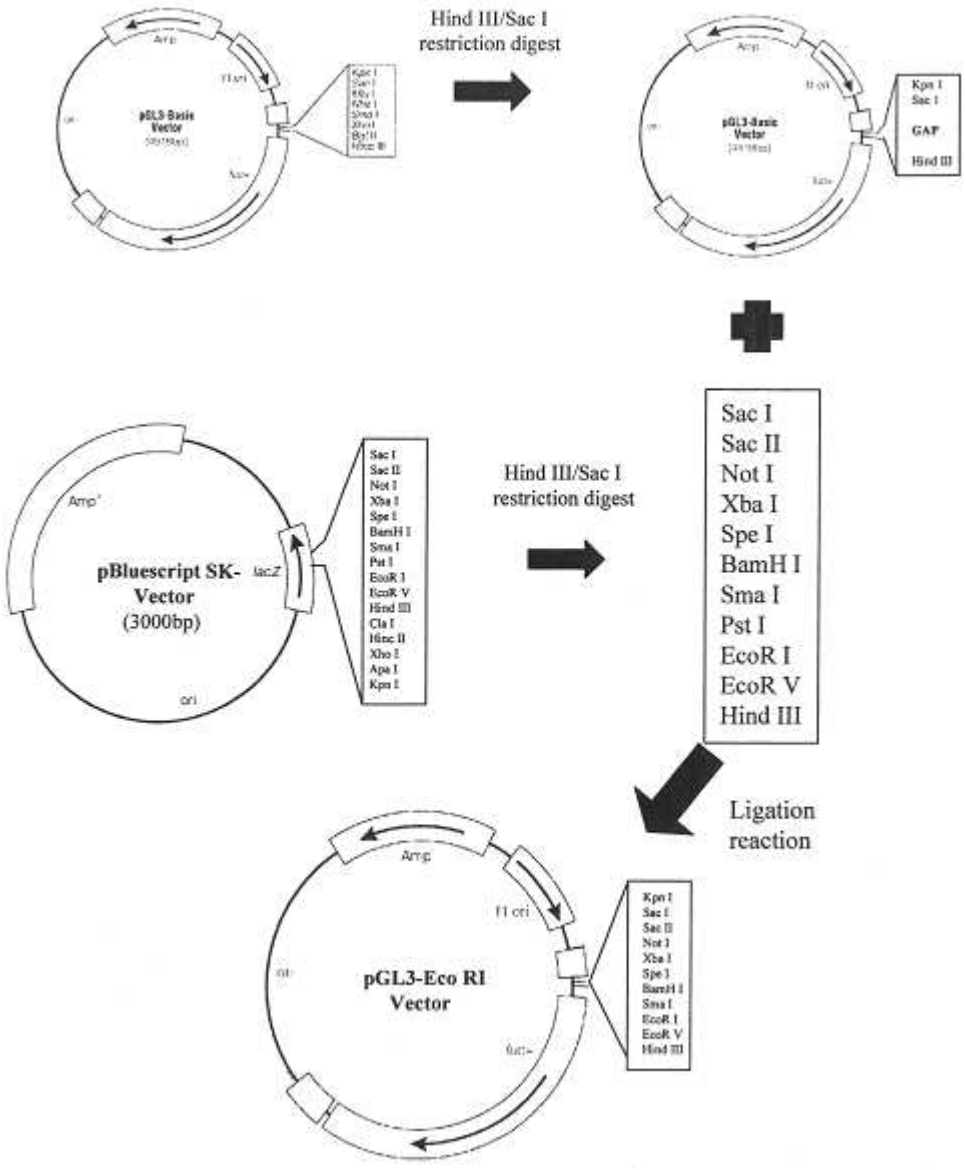


Figure 29 – pGL3-*Eco* RI construction strategy. pGL3-Basic (Promega) was firstly digested with Hind III and Sac 1 restriction endonucleases to remove a segment of the multiple cloning region. The pBluescript SK cloning vector was then also digested with *Hind* III and *Sac* 1. The pBluescript SK's cleaved multiple cloning region was then ligated into the digested pGL3-Basic vector, producing the modified pGL3-Eco RI vector with the expanded multiple cloning site.

2.3.1.4 pGL3-*Eco* RI Modified Vector

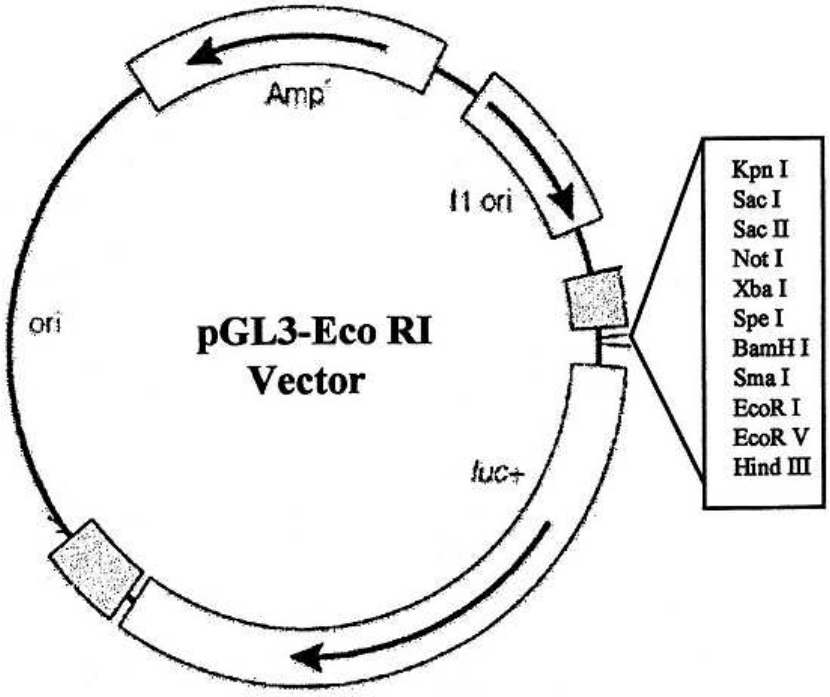


Figure 30 – Schematic representation of the pGL3-Eco RI vector showing the cDNA encoding the modified firefly luciferase (luc+), the gene conferring ampicillin resistance in E.coli (Amp^r), the origin of replication in E.coli (ori), the origin of replication derived from filamentous phage (f1 ori), and its endonuclease restriction sites that make up the modified 'multiple cloning region' (adapted from www.Promega.com)

2.3.2 Agarose Gels

In this study, endonuclease digested minipreped plasmid DNA (including digested and PCR amplified promoter inserts) were visualised on, and if required extracted from, agarose gels. The agarose gels were prepared using the following protocol:-

- 0.5g of agarose powder (GibcoBRL) was added to 50ml of 1xTAE (refer to appendices).

(If the DNA was to be extracted for further use – 'Modified TAE' was used (refer to appendices).

- This was then boiled in a microwave to dissolve the agarose.

- Once the solution had cooled adequately, the solution was poured into the gel former. The well former (comb) was also inserted. The assembly was then left for 20 minutes to set.

2.3.2.1 Gel Loading and Running

- Once set, the comb was removed and the gel was transferred into the 'gel running pack', where it was submerged in 1xTAE.

- 5µl of Orange G loading dye (diluted to x1 from x5 stock solution) (refer to appendices) was added to 10µl of mini/ midipreped DNA and mixed well.

- 12µl of the required samples were then loaded into the agarose gel wells.

- The lid of the gel running pack was then attached securing the anode and cathode connections.

- Gels were typically run at 150 volts for 30 minutes to resolve 1-4 kb bands.

- Once gels had been run, the gel was then soaked in ethidium bromide solution (20µl per 200ml 1xTAE) for 20 minutes.

- After staining, the gels were visualised in the Fluorimager 595 using ultra-violet light to elucidate required DNA bands.

2.3.3 Plasmid DNA Endonuclease Digestion

The basic quantities indicated below were used in preparation for all endonuclease digests carried out on plasmid DNA in this study (the bovine serum albumin (BSA) and relevant buffers were sourced from the manufacturers of the respective enzyme - detailed in section 2.2.1):-

- DNA – **1.0-10.0µl** (3µg for ligation prep, otherwise ≈300ng),
- 10x BSA – **2.0µl**
- 10x relevant buffer – **2.0µl**
- Restriction Endonuclease(s) – **1.0µl** or **0.8µl** (for two in combination or sequentially)
- Calf Intestinal Alkaline Phosphatase (1unit/µl) (CIAP) – **5.0µl** (if used)
- Deionised water – made up to **20µl**

2.3.4 General DNA Isolation/ Purification Protocols

2.3.4.1 DNA Isolation from Agarose Gels

DNA extraction from agarose gels was carried out using the 'Ultrafree-DA Centrifugal Filter Device' (Millipore).

2.3.4.1.1 Theory

The Ultrafree-DA device from Millipore was designed to recover 100 to 10,000 bp DNA from agarose gel slices in one ten-minute spin. It consists of a pre-assembled sample filter cup with agarose 'Gel Nebulizer', and a microcentrifuge vial. The device utilises gel compression to extract DNA from the agarose. Centrifugal force collapses the gel structure, drives the agarose through a small orifice in the 'Gel Nebulizer' and the resultant gel slurry is sprayed into the sample filter cup. As the agarose is compressed at 5,000xg, DNA is extruded from the gel's pores. The gel matrix is retained by the microporous membrane, and the DNA passes freely through the membrane. DNA can then be recovered in the filtrate vial.

2.3.4.1.2 Protocol

- Required DNA product was first electrophoresed through a modified TAE agarose gel (refer to section 2.3.2).

- Band of interest was located using a long wavelength UV lamp, and the slice of agarose that contained the DNA was carefully excised from the gel (any excess agarose was also then trimmed away).

- The gel slice was then placed into the 'Gel Nebulizer' sample cup assembly device and sealed, so that the cap was attached to the vial.

- The 'Gel Nebulizer' and vial were then centrifuged at 5000x*g* fro 10 minutes.

- DNA was now present in the filtrate. The filter cup and 'Gel Nebulizer' were then discarded, allowing the DNA to be stored in the capped filtrate vial.

2.3.4.2 Miniprep Protocol (QIAgen Spin Miniprep Kit – Qiagen)

The 'QIAgen Spin Miniprep Kit' was used to isolate all the plasmid DNA screened for the required *UCH-L1* promoter inserts. The protocol is described below (all buffers were part of the Qiagen kit):-

- 0.5ml of the overnight culture (above) was centrifuged at 13,500rpm for 45 minutes.

- The bacterial cells were resuspended in 250µl of Buffer P1 (+ RNase A) until no cell clumps were visible.

- 250µl of Buffer P2 was then added. Tubes were inverted 4-6 times to mix.

- 350µl of Buffer N3 was added, and tubes were inverted until the solution became cloudy.

- Tubes were then centrifuged at 13,500 rpm for 10 mins.

- Supernatants from the above step were decanted into the 'QIAprep columns' which were then centrifuged at 13,500rpm for 1 minute. The flowthrough was discarded.

- The QIAprep spin column was then washed by adding 0.5ml Buffer PB, and the centrifuge step at 13,500rpm for 1 minute was repeated. The flowthrough was again discarded.

- A second wash step was carried out by adding 0.75ml Buffer PE, and again centrifuging for 1 minute at 13,500rpm.

- The flowthrough was discarded, and the column was centrifuged for an additional 1 minute to remove any residual wash buffer.

- The plasmid DNA was finally eluted into a clean microcentrifuge tube by adding 50µl Buffer EB to the centre of the QIAprep column, and centrifuging for 1 minute at 13,500rpm after a 1 minute diffusion period.

2.3.4.3 Midiprep Protocol ('PureYield' Plasmid Midiprep System – Promega)

The ''PureYield' Plasmid Midiprep System' was used to isolate all the plasmid DNA that was required for cloning ligation reactions (refer to section 2.3). The protocol, split into two stages, is described below (all solutions and columns were part of the Promega system):-

2.3.4.3.1 Preparation and Lysis of Bacterial Cell Cultures

- Transformed *E.coli* bacterial cells were grown up overnight (≈16 hours) in a 50ml culture.

- The cells were then centrifuged at 10000 x *g* for 10 minutes in order to pellet the cells. The supernatant was then discarded, any excess media was drained on a paper towel.

- The cell pellet was then completely resuspended in 3ml of 'Cell Resuspension Solution'.

- 3ml of 'Cell Lysis Solution' was then added and gently mixed by inverting the tube four times. The tube was then incubated for 3 minutes at room temperature.

- 5ml of 'Neutralization Solution' was then added to the lysed cells, which was then capped and gently inverted four times to mix. The tube was then left in an upright position for 3 minutes to allow a white flocculent precipitate to form.

- The next stage in the midiprep protocol was then carried out:-

2.3.4.3.2 DNA Purification by Centrifugation

- A 'PureYield Clearing Column' was first placed into a new 50ml disposable plastic tube.

- The lysate (from stage 1 above) was then poured in and left to stand for 2 minutes, so as to allow any cellular debris to rise to the top.

- The 'PureYield Clearing Column' was then centrifuged at 3000 x *g* for 5 minutes to filter the lysate through (the centrifugation step was repeated if deemed necessary).

- A 'PureYield Binding Column' was then placed into a new 50ml disposable plastic tube.

- The filtered lysate was then poured in and centrifuged for 3 minutes at 1500 x *g*.

- 5ml of 'Endotoxin Removal Wash Solution' (+ Isopropanol) was then added to the 'PureYield Binding Column' and centrifuged at 1500 x *g* for 3 minutes. The flowthrough was then discarded, and the column reinserted into the tube.

- 20ml of 'Column Wash Solution' (+ ethanol) was then added to the 'PureYield Binding Column', and centrifuged at 1500 x *g* for 5 minutes. The flowthrough was then once again discarded, and the column reinserted into the tube. Centrifugation at 1500 x *g* was then carried out for an additional 10 minutes to ensure complete ethanol removal from the column (the tip of the column was also tapped on a paper towel to remove any final traces).

- The DNA was then eluted by placing the binding column into a new 50ml disposable plastic tube, and adding 600µl of 'Nuclease-Free Water' to the binding membrane of the 'PureYield Binding Column'.

- The 'PureYield Binding Column' was then centrifuged for 5 minutes at 1500 x *g*. The filtrate was then finally transferred into a new 1.5ml tube.

2.3.5 Polymerase Chain Reaction from Bacterial Colony

PCRs could also be set up using neat bacterial colonies from an overnight transformation agar plate, i.e. colonies that had been realised after 37℃ overnight incubation of the transformation culture (refer to 2.3.10) on ampicillin LB agar plates. In this instance, typically used as a rapid detection method of vector uptake during sub-cloning protocol (refer to section 3.2.1), quantities detailed in section 2.1.3.2 remained the same except 1µl was added to the quantity of deionised water used, instead of the usual addition of DNA

The deionised water in this instance, was added to the reaction tube/ well first, and the required bacterial colony was suspended within it. The PCR reaction was then prepared and carried out as indicated previously.

2.3.6 Plasmid DNA Endonuclease Digestion

The basic quantities indicated below were used in preparation for all endonuclease digests carried out on plasmid DNA in this study (the bovine serum albumin (BSA) and relevant buffers were sourced from the manufacturers of the respective enzyme – detailed in section 2.2.1):-

- DNA – **1.0-10.0µl** (3µg for ligation prep, otherwise ≈300ng),
- 10x BSA – **2.0µl**
- 10x relevant buffer – **2.0µl**
- Restriction Endonuclease(s) – **1.0µl** or **0.8µl** (for two in combination or sequentially)
- Calf Intestinal Alkaline Phosphatase (1unit/µl) (CIAP) – **5.0µl** (if used)
- Deionised water – made up to **20µl**

2.3.7 Preparation of Competent Cells – Calcium Chloride Method

The following protocol was used in the preparation of all competent cells used in the cloning section of this study:-

- An overnight culture was grown in Luria-Bertoni (LB) broth from a fresh 'picked' *E.coli* JM109 colony grown overnight on a LB plate.

- 1ml of the overnight culture was diluted in 100ml of fresh LB broth.

- Cells were then grown in a 37℃ shaking incubator until the culture reached an optical density (O.D.) between 0.4-0.6 at a wavelength of 600nm (~2-3 hours).

- Cells were centrifuged at 4000rpm at 4℃ for 10 mins.

- Once centrifuged they were resuspended in 10ml of 0.1M ice-cold calcium chloride ($CaCl_2$) solution, and left on ice for 30 minutes.

- Cells were then again centrifuged at 3000rpm for 5 minutes at 4℃.

- A second resuspension step in 4ml of $CaCl_2$ was then carried out, before leaving the cells on ice for two hours before use.

For Long Term Storage

- After the second resuspension step, 140µl Dimethyl sulfoxide (1%) (DMSO – BDH Laboratory Supplies) was added. Cells were then gently mixed and left on ice for 15 mins.

- A further 140µl DMSO was then added, and the cells were gently mixed, before being aliquoted into microcentrifuge tubes and frozen in liquid nitrogen (for long term storage at -80℃).

2.3.8 Ligations Using the pGEM-T Easy Vector

The following quantities were used in the preparation of all pGEM-T Easy ligation reactions carried out in this study:-

	Standard Reaction	Positive Control	Background Control
2 x Rapid Ligation Buffer -	5µl	5µl	5µl
pGEM-T Easy Vector (50ng/ µl) -	1µl	1µl	1µl
PCR Product -	*µl	-	-
Control Insert DNA (4ng/ µl) -	-	2µl	-
T4 DNA Ligase (3 Weiss units/ µl) -	1µl	1µl	1µl
Deionised Water -	-	1µl	3µl

***** - The amount of PCR product used in the reaction was calculated using the following equation:-

$$\frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio} = \text{ng of insert to be used}$$

- A 3:1 molar to vector ratio was desired.

- ng of PCR product used was quantified on an agarose gel in comparison to a known concentration of DNA. ng of vector was quantified using a Spectrophotometer (Beckman DU-7000) to measure the light absorption of the (suitably diluted) solution at 260nm (absorption at 280nm was also measured so that DNA purity could be ascertained and thus regulated).

- All reactions were then left at 4℃ for 15 hours.

2.3.9 Ligations Using the Modified pGL3-*Eco* RI Vector

The following guideline quantities were used in the preparation of all pGL3-*Eco* RI ligation reactions carried out in this study:-

	Standard Reaction	**Background Control**
5 x Ligation Buffer -	4µl	4µl
PGL3-*Eco*-R1 Vector (≈150ng/ µl) -	2µl	2µl
Promoter Insert -	***µl**	-
T4 DNA Ligase (3 Weiss units/ µl) -	2µl	2µl
Deionised Water -	Up to 20µl	Up to 20µl

***** - The amount of promoter insert used in the reaction was calculated using the following equation:-

$$\frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}} \times \text{insert: vector molar ratio} = \text{ng of insert to be used}$$

- A 3:1 molar to vector ratio was desired.

- ng of PCR product used was quantified on an agarose gel in comparison to a known concentration of DNA. ng of vector was quantified using a Spectrophotometer (Beckman DU-7000) to measure the light absorption of the (suitably diluted) solution at 260nm (absorption at 280nm was also measured so that DNA purity could be ascertained and thus regulated).

- All reactions were then left at 16℃ for 15 hours.

2.3.10 Transformations Using the pGEM-T Easy Vector Ligation System

The following protocol was used to transform all pGEM-T Easy ligation products (quantities in brackets were those used to calculate the transformation efficiency (TE) of the procedure):-

- 2µl (TE - 0.1ng) of each ligation reaction was added to 50µl (TE - 100µl) of thawed *E.coli* JM109 calcium chloride prepared competent cells (refer to section 2.3.7).

- The cells were then gently mixed and placed on ice for 20 minutes.

- The cells were then 'heat-shocked' for 45 seconds in a 42℃ water bath.

- The tubes were immediately returned to ice for 2 mins.

- 950µl (TE - 900µl) of SOC recovery medium (Promega) was added to the transformed cells.

- Incubation then took place in a 37℃ shaking incubato r (150rpm) for 1.5 hours.

- Each transformation culture then had 100µl (TE - 100µl of a 1:10 dilution) plated out (in duplicate) onto LB + ampicillin plates (refer to appendices) which were then incubated for 24 hours at 37°C.

- Those *E.coli* colonies realised on the ampicillin LB agar plates after the 24 hour incubation contained the pGEM-T Easy vector (ampicillin resistance gene), and those colonies corresponding to the 'standard reaction' were thus potential harbourers of the required pGEM-T Easy/ insert recombinant plasmid.

- Those colonies were then screened for the required insert DNA (*UCH-L1* promoter).

- In preparation for the isolation of the pGEM-T Easy plasmid DNA by way of miniprep (refer to section 2.3.4.2), the ampicillin resistant *E.coli* colonies (harbouring the pGEM-T Easy vector) were 'picked off' the plate and grown up overnight at 37℃ in 10mls LB broth + ampicillin (refer to appendices).

N.B. – Performed in parallel during each transformation procedure; 1µl of plasmid PUC 19 DNA (which is a high copy number *E. coli* plasmid cloning vector) was transfected into the same batch of JM109 cells undergoing the same protocol as detailed above, as an extra positive control.

2.3.11 Transformations Using the pGL3 Vector Ligation Products

The following protocol was used to transform all pGL3 ligation products (quantities in brackets were those used to calculate the transformation efficiency (TE) of the procedure):-

- 5µl (TE - 0.1ng) of each ligation reaction was added to 50µl (TE - 100µl) of thawed *E.coli* JM109 calcium chloride prepared competent cells (refer to section 2.3.7).

- The cells were then gently mixed and placed on ice for 5 minutes.

- The cells were then 'heat-shocked' for 45 seconds in a 42℃ water bath.

- The tubes were immediately returned to ice for 2 mins.

- 450µl (TE - 900µl) of SOC recovery medium (Promega) was added to the transformed cells.

- Incubation then took place in a 37℃ shaking incubato r (150rpm) for 1.5 hours.

- The cells were then centrifuged for 10 minutes at 4000rpm (1000 x G), and resuspended in 200µl of SOC recovery medium.

- Each transformation culture then had 100µl (TE - 100µl of a 1:10 dilution) plated out (in duplicate) onto LB + ampicillin plates (refer to appendices) which were then incubated for 24 hours at 37℃.

- Those *E.coli* colonies realised on the ampicillin LB agar plates after the 24 hour incubation contained the pGL3 vector (ampicillin resistance gene), and those colonies corresponding to the 'standard reaction' were thus potential harbourers of the required pGL3/ insert recombinant plasmid.

- Those colonies were then screened for the required insert DNA (*UCH-L1* promoter).

- In preparation for the isolation of the pGL3 plasmid DNA by way of miniprep (refer to section 2.3.4.2), the ampicillin resistant *E.coli* colonies (harbouring the pGL3 vector) were 'picked off'

the plate and grown up overnight at 37℃ in 10mls L B broth + ampicillin (refer to section 6.5.8.2.1).

N.B. – Performed in parallel during each transformation procedure; 1µl of plasmid PUC 19 DNA was transfected into the same batch of JM109 cells undergoing the same protocol as detailed above, as an extra positive control.

2.4 Cell Culture Protocols

2.4.1 Mammalian Cell Lines and Required Growth Media

2.4.1.1 - A2058 (Human Skin Melanoma) (ATCC Number – CRL-11147)

- Description - organism: *Homo sapiens*; organ: skin; disease: melanoma; derived from metastatic site: lymph node; morphology: epithelial; ethnicity: Caucasian; growth properties: adherent.

- Concordant Growth Media - Dulbecco's Modified Eagle's Medium with 4 mM L-Glutamine, 3.7 g/L Sodium Bicarbonate and 1.0 g/L Glucose, 90%; Fetal Bovine Serum, 10%. (ATCC).

2.4.1.2 – MCF-7 (Human Breast Adenocarcinoma) (ATCC Number – HTB-22)

- Description - organism: *Homo sapiens*; organ: mammary gland, breast; cell type: epithelial; disease: adenocarcinoma; morphology: epithelial; ethnicity: Caucasian; growth properties: adherent.

- Concordant Growth Medium - Minimum Essential Eagle’s Medium with 2 mM L-Glutamine and Earle's BSS adjusted to contain 1.5 g/L Sodium Bicarbonate, 0.1 mM non-essential amino acids and 1 mM Sodium Pyruvate and supplemented with 0.01 mg/ml Bovine Insulin, 90%; Fetal Bovine Serum, 10%. (ATCC).

2.4.1.3 – ND-7 (Mouse Neuroblastoma/ Rat Basal Ganglia Neuron Hybrid) (ATCC Number – HTB-22)

- Description - organism: Mus musculus/ *Rattus rattus* hybrid; organ: brain; cell type (rat): dorsal root ganglion neurone; disease (mouse): neuroblastoma; morphology: neuronal; ethnicity: Caucasian; growth properties: adherent.

- Concordant Growth Medium - Dulbecco's Modified Eagle's Medium with 2mM L-Glutamine, 3.7 g/L Sodium Bicarbonate and 4.5 g/L Glucose, 90%; Fetal Bovine Serum, 10%. (ATCC).

2.4.1.4 – HCN-1A (Human Brain Cortical Neuron) (ATCC Number – CRL-11147)

- Description - organism: *Homo sapiens*; organ: brain; cell type: cortical neuron; morphology: neuronal; ethnicity: Caucasian; growth properties: adherent.

- Concordant Growth Medium - Dulbecco's Modified Eagle's Medium with 4mM L-Glutamine, 1.5 g/L Sodium Bicarbonate and 4.5 g/L Glucose, 90%; Fetal Bovine Serum, 10%. (ATCC).

2.4.2 Cell line Maintenance

2.4.2.1 Propagation (From Frozen)

This outlines the basic handling procedure of how cell lines were propagated when received from the supplier in their frozen state.

- The frozen vial was rapidly thawed (approximately 2 minutes) in a 37°C water bath by gentle agitation.

- The vial was removed from the water bath and sprayed with 70% ethanol solution so as to decontaminate the outside of the vial.

- The cell suspension was then centrifuged at 125 x *g* for 7 minutes, and the cryoprotective agent in the supernatant was discarded. The cell pellet was then resuspended in the appropriate amount of relevant fresh growth medium.

- The vial content was then transferred into a culture vessel containing 15-minute pre-incubated growth medium. (The use of pre-incubated growth medium was critical so as to ensure the pH of the medium had equilibrated and was not excessively alkaline.)

- The culture was then incubated at 37°C in a 5% CO_2 air atmosphere.

2.4.2.2 Subculturing Protocol (Passage/ Cell Splitting)

This outlines the basic handling procedure of how cell lines were subcultured when they had reached 90% confluence. The subcultivation ratio was typically 1:3, and took place in 75cm^2 flasks. (All cell lines used in this study were adherent).

- The culture medium was initially discarded.

The cell layer was then briefly rinsed with a 0.25% (w/v) Trypsin-0.53mM EDTA solution (refer to appendices) in order to remove all traces trypsin inhibitor-containing serum.

- 3.0ml of Trypsin-EDTA solution was added to the flask, and then cell layer dispersion was observed under a microscope. Cells typically took 10 minutes to fully disperse.

- 8ml of complete growth medium was then added, and the cells were aspirated by gentle pipetting.

- The cell suspension was then transferred to a centrifuge tube and spun at 125 x *g* for 5 mins.

- The supernatant was then discarded to remove any trypsin-EDTA solution. The cell pellet was resuspended in fresh growth medium, and added to new a culture vessel containing 15-minute pre-incubated growth medium.

- The culture vessel was then placed in a 37°C incubator with a 5% CO_2 air atmosphere.

- (Media renewal typically took place twice per week.)

2.4.3 Vector Transfection into Adherent Mammalian Cells

2.4.3.1 Transfection Related Chemicals

- Serum free medium – animal free protein, without L-glutamine (Sigma – 14591C)

- GeneJuice Transfection agent (Merck Biosciences)

2.4.3.1.1 GeneJuice Theory

Whereas many available transfection reagents are based on cationic lipid formulation, GeneJuice Transfection Reagent is composed of a nontoxic cellular protein and a small

amount of a novel polyamine. GeneJuice Transfection Reagent provides highly efficient DNA transfer in both stable and transient transfection of eukaryotic cells. The unique chemistry provides the advantage of compatibility with both serum-containing and serum-free media, and makes media changes unnecessary.

2.4.3.2 Transfection Preparation

- One day prior to transfection 1-3 x 10^5 cells were plated into 35mm dishes in the required complete growth medium.

- The cells were then incubated overnight at 37°C in a 5% CO_2 air atmosphere.

- Cells were ensured to be 50-80% confluent prior to transfection.

2.4.3.3 Transfection Procedure

- For each 35mm dish that was to be transfected, 100µl of serum free medium (refer to section 2.4.3.1) was added into a sterile tube. To this, 3-6µl of 'GeneJuice' transfection agent (refer to section 2.4.3.1.1) was added drop-wise, and mixed thoroughly by vortexing (actual volume of GeneJuice added was determined through initial optimisation reactions carried out for each cell line).

- Mixtures were incubated at room temperature for 5 minutes.

- 1µg of desired vector DNA was then added to each GeneJuice/ serum free medium mixture and gently mixed in by pipetting.

- The GeneJuice/ DNA mixtures were then incubated at room temperature for a further 15 minutes.

- The entire volume of each GeneJuice/ DNA mixture was then added drop-wise to the prepared 35mm dishes containing the required cells in complete growth medium. The drops were distributed evenly over the entire surface of each dish, and the dish was then gently rocked to ensure even distribution.

- Cells were incubated for 24 hours at 37°C in a 5% CO_2 air atmosphere. (After 8 hours, the transfection mixture was removed from each dish and replaced with complete growth mixture for the remainder of the incubation.)

2.4.4 Dual-Luciferase Reporter Assay (Promega)

2.4.4.1 Dual-Luciferase Reporter Assay Theory

In the Dual-Luciferase Reporter (DLR™) Assay System the activities of firefly (*Phontinus pyralis*) and *Renilla* (*Renilla reniforms*) luciferases are sequentially measured from a single sample. The activity of the firefly luciferase experimental reporter is thus normalised to the activity of the *Renilla* luciferase internal control, thence minimising experimental variability caused by differences in cell viability and/ or transfection efficiency.

Firstly, recombinant firefly luciferase is measured through the addition of Beetle luciferin which is oxidised in the presence of ATP, Mg^{2+} and O_2 (and coenzyme A). Immediately after firefly luciferase quantification, a 'Stop & Glo' reagent is added to quench firefly luciferase luminescence – it also activates the *Renilla* luciferase through the addition of Coelenterazine (in the presence of O_2) – which produces a stabilised luminescent signal that decays slowly.

Thence, the integrated format of the DLR™ Assay allows rapid quantification of the experimental reporter luciferase relative to the internal control *Renilla* luciferase in transfected mammalian cells.

The procedures below outline the protocol used to prepare and perform the Dual-Luciferase Reporter Assays on all of the cell lines investigated in this study. (All buffers, substrates and reagents detailed were supplied by Promega.)

2.4.4.2 Dual-Luciferase Reporter Assay Buffer Preparation (Promega)

- Passive Lysis Buffer

- 4 volumes of Distilled Water to 1 volume of the Buffer of Phosphate Buffered Saline (Promega).

- Luciferase Assay Buffer II

- The lyophilised Luciferase Assay Substrate powder (Promega) was added to the 10ml of supplied Luciferase Assay Buffer II (Promega).

- Stop and Glo Buffer

- 1 volume of Stop and Glo Substrate (x50 concentrate) (Promega) was added to 50 volumes of the supplied Stop and Glo Buffer (Promega).

2.4.4.3 Assay Preparation – Lysis of Cultured Cells

- The initial step in the preparation was to dilute a sufficient quantity of the Passive Lysis Buffer, which was supplied at x 5 concentrate, by adding 4 volumes of distilled water to 1 volume of the buffer. The solution was then mixed well.

- The growth medium was removed from each of the 35mm dishes, and then a sufficient volume of phosphate buffered saline was gently added. The culture vessel was then briefly swirled to remove any detached cells and residual growth medium.

- Once all the rinse solution had been removed, 500µl of diluted (1x) Passive Lysis Buffer was dispensed into each 35mm dish so that the cellular monolayer was completely covered.

- The culture plates were then gently rocked (on a rocking platform) for 15 minutes at room temperature to ensure complete and even coverage of the Passive Lysis Buffer within each 35mm dish.

2.4.4.4 Luciferase Assay Protocol

- Initial preparation of the Luciferase Assay Reagent II was required by resuspending the provided lyophilised Luciferase Assay Substrate in 10ml of the supplied Luciferase Assay Buffer II.

- An adequate volume of Stop and Glo Reagent also required preparation - 100µl required per sample to be assayed. 1 volume of the Stop and Glo Substrate (supplied at x50 concentrate) was added to 50 volumes of Stop and Glo Buffer in a new siliconised polypropylene tube.

- 100µl of Luciferase Assay Reagent II was predispensed into the appropriate number of luminometer tubes required to complete the desired number of dual luciferase assays.

- The luminometer was then programmed to perform a 2-second premeasurement delay, followed by a 10-second measurement period for each reporter assay.

- 20µl of the cell lysate from the required 35mm dish prepared in section 2.4.4.3 above, was then carefully transferred into a luminometer tube containing the Luciferase Assay Reagent. The solution was then mixed by pipetting 2 or 3 times.

- The firefly luciferase activity reading was then measured.

- The sample tube was then removed from the luminometer, and 100µl of Stop and Glo Reagent was then added and vortexed to mix. The sample was replaced in the luminometer, and a second reading initiated.

- The *Renilla* luciferase activity reading was then measured.

- The reaction tube was then discarded, and the next dual luciferase reporter assay was carried out.

2.5 Statistical Analysis (Analysis carried out with 'stata^{tm}' (StataCorp LP)

2.5.1 Statistical Analysis of Luciferase Data

2.5.1.1 Standard Error of Mean Calculations

These were calculated using the following equation:-

$$\text{Standard Error of Mean (SEM)} = \frac{\text{Standard Deviation (SDEV)}}{\text{Square Root } (\sqrt{}) \text{ of the number of observations (n)}}.$$

2.5.1.2 *p*-value Calculations

Two Way Analysis of Variance' (ANOVA) was implemented to examine the relationship between the ratio luminescence values from the luciferase assays outlined above, and the three *UCH-L1* gene promoter haplotypes elucidated in this study – within each of the four cell lines investigated. A *p*-value of <0.05 was considered nominal significance.

2.5.2 Cognitive Function Tests

2.5.2.1 AH4 Cognitive Function Test

The AH4 test is a general intelligence test commonly used to test logical and verbal reasoning within groups. Critically scores tend to decline with age.

2.5.2.2 Mill Hill Cognitive Function Test

The Mill Hill test is a non-verbal intelligence test. Its aim is to record an individual's present recall of acquired information, which is closely related to an individual's store of acquired intellectual skill. Thence, its demands on an individual's present capacity for intellectual activity and rational judgement are reduced to a minimum, which generally causes test scores not to decline with age.

2.5.2.3 Immediate Recall Section of Wechsler Logical Memory Cognitive Test

This section of the Wechsler Logical Memory test is designed to provide a detailed assessment of clinically relevant aspects of immediate memory functioning using both auditory and visual stimulus.

2.5.2.4 Delayed Recall Section of Wechsler Logical Memory Cognitive Test

This section of the Wechsler Logical Memory test is designed to provide a detailed assessment of clinically relevant aspects of delayed memory functioning using both auditory and visual stimulus.

2.5.2.5 Regression of AH4 Test Scores on Mill Hill Test Scores (theory of)

The AH4 test measures fluid intelligence which declines with age, whereas the Mill Hill test measures 'crystallised intelligence' which does not decline with age. Thence, by analysing AH4 test scores in relation to Mill Hill test scores, an indication of cognitive decline can be ascertained. (Negative scores in the raw data (not shown in Section 3.3) meant that an individual had done worse than expected on the AH4 test given their Mill Hill test score, i.e. had declined cognitively with age. Positive scores indicated that an individual had done better than expected, i.e. had not cognitively declined).

2.5.2.6 *p*-value Calculations

One Way Analysis of Variance' (ANOVA) was implemented to examine the relationship between mean scores from the cognitive function tests outlined above, and the five *UCH-L1* single nucleotide polymorphisms investigated in this study. Each of the three mean test scores was analysed in turn in relation to the genotypes of each SNP. A *p*-value of <0.05 was considered nominal significance. A Bonferroni correction for the number of tests would be applied if any of the tests were nominally significant.

(To do a Bonferroni correction for multiple testing, one can multiply the *p*-value by the number of tests performed, e.g. if the *p*-value is 0.08 and eight different variables (such as eight different SNPs) were tested, then the Bonferroni corrected *p*-value would be 0.64).

3 Results

3.1 Results I – Mutation Screening

The overall aim of this section was to elucidate any common mutations or polymorphisms within *UCH-L1*'s promoter, 5'-UTR and exon regions.

dHPLC was initially employed as a rapid detection method to indicate which regions of the UCH-L1 gene harbour any common sequence variations. (As discussed in section 2.1.3.4, for optimal dHPLC analysis, relevant *UCH-L1* nucleotide sequence was divided into regions between 150-450bp in length (with one exception – Exon 9).

Once regions harbouring any common mutations/ polymorphisms had been elucidated through dHPLC, the nature of these sequence variations were established through direct DNA sequencing of the samples that displayed dHPLC heteroduplexes, with reference to the homoduplex samples, to clearly determine the specific nature and location of the sequence variations.

Once these sequence variations had been classified, RFLP genotyping strategies could then be devised (please see section 2.2.4), which allowed concordant genotyping data to be uncovered from the two control DNA arrays (n=31, and n=480). From this data, common haplotypes within the Caucasian population were able to be elucidated, which would form the basis of the functional studies that were carried out in the next stage of the investigation.

3.1.1 dHPLC Analysis

The initial stage of this project involved mutation/ polymorphism scanning of the UCH-L1 gene through dHPLC analysis. Sequence variations within the UCH-L1 gene were identified using genomic DNA from 26 unrelated healthy individuals (refer to section 2.1.1.1).

Once these *UCH-L1* regions had been successfully amplified by Polymerase Chain Reaction (PCR - refer to section 2.1.3) for each of the 26 genomic DNA samples in the array, the samples were then subjected to dHPLC mutation scanning (refer to section 2.1.4.2).

dHPLC scans are only displayed for those regions giving rise to potential sequence mutations/ polymorphisms, i.e. those regions which realised scans exhibiting possible heteroduplexes as well as homoduplexes (refer to section 2.1.4.2). For the relevant regions, one potential heteroduplex positive scan is displayed per region, together with one homoduplex scan for comparison.

(As described in section 2.1.4.3, before any *UCH-L1* region was able to be scanned, a dHPLC method required preparation for the specific PCR amplicon DNA sequence (Wavemaker 3.4/ 4.0). This method dictated the 'oven temperature' and respective buffer concentrations at specific times of acquisition during the homo/ heteroduplex elution within the instrument's separation cartridge).

***UCH-L1* Regions Elucidating dHPLC Heteroduplexes**	
***UCH-L1* Region**	**Corresponding *UCH-L1* Nucleotide Sequence**
Promoter Region A	+158 ↔ -105
Promoter Region B	-31 ↔ -311
Promoter Region C	-235 ↔ -549
Promoter Region D	-475 ↔ -802

Table 7 – Displays the four *UCH-L1* regions in which dHPLC heteroduplexes were elucidated (nucleotide sequences are relative to translation start site, i.e. the adenine nucleotide base of exon 1's ATG start codon corresponds to +1).

3.1.1.1 dHPLC Analysis of Promoter Region A

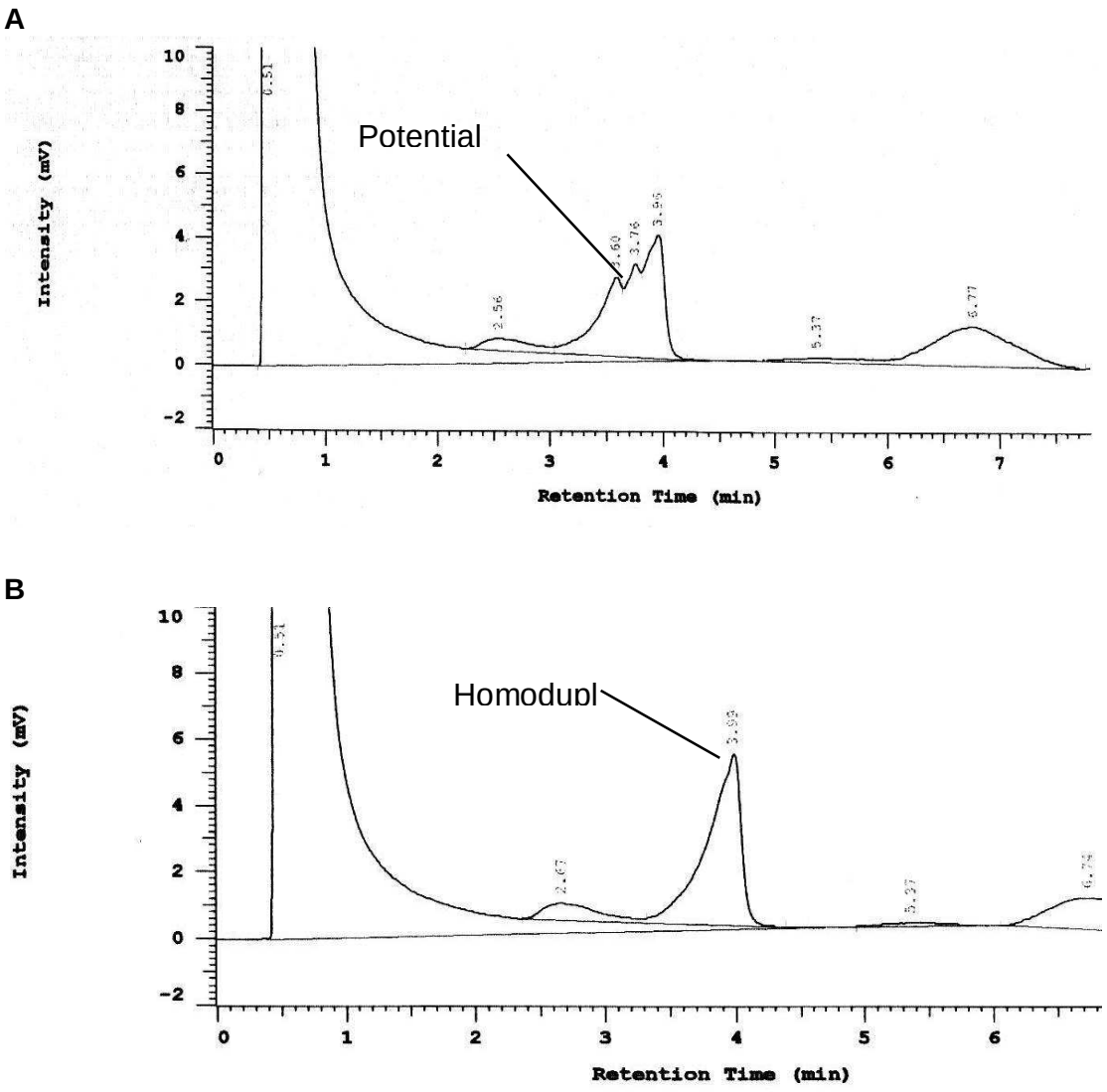


Figure 31 – Comparative dHPLC scans for Promoter Region A (*UCH-L1* nucleotide sequence +158 → -105) showing **(A)** a sample exhibiting potential heteroduplexes, and **(B)** a sample displaying only homoduplex species. Samples were eluted at an 'oven temperature' of 64ºC.

- The y-axis indicates the absorbance intensity in millivolts (mV), whilst the x-axis displays the retention (elution) time in minutes (mins).

- The initial vertical peak at a retention time ≈1 min was the elution of excess PCR primer, the observed shallow peak ≈1 min before hetero/ homoduplex peak elution was non-specific PCR product, and the final peak at ≈7 mins retention time was genomic DNA.

3.1.1.2 dHPLC Analysis of Promoter Region B

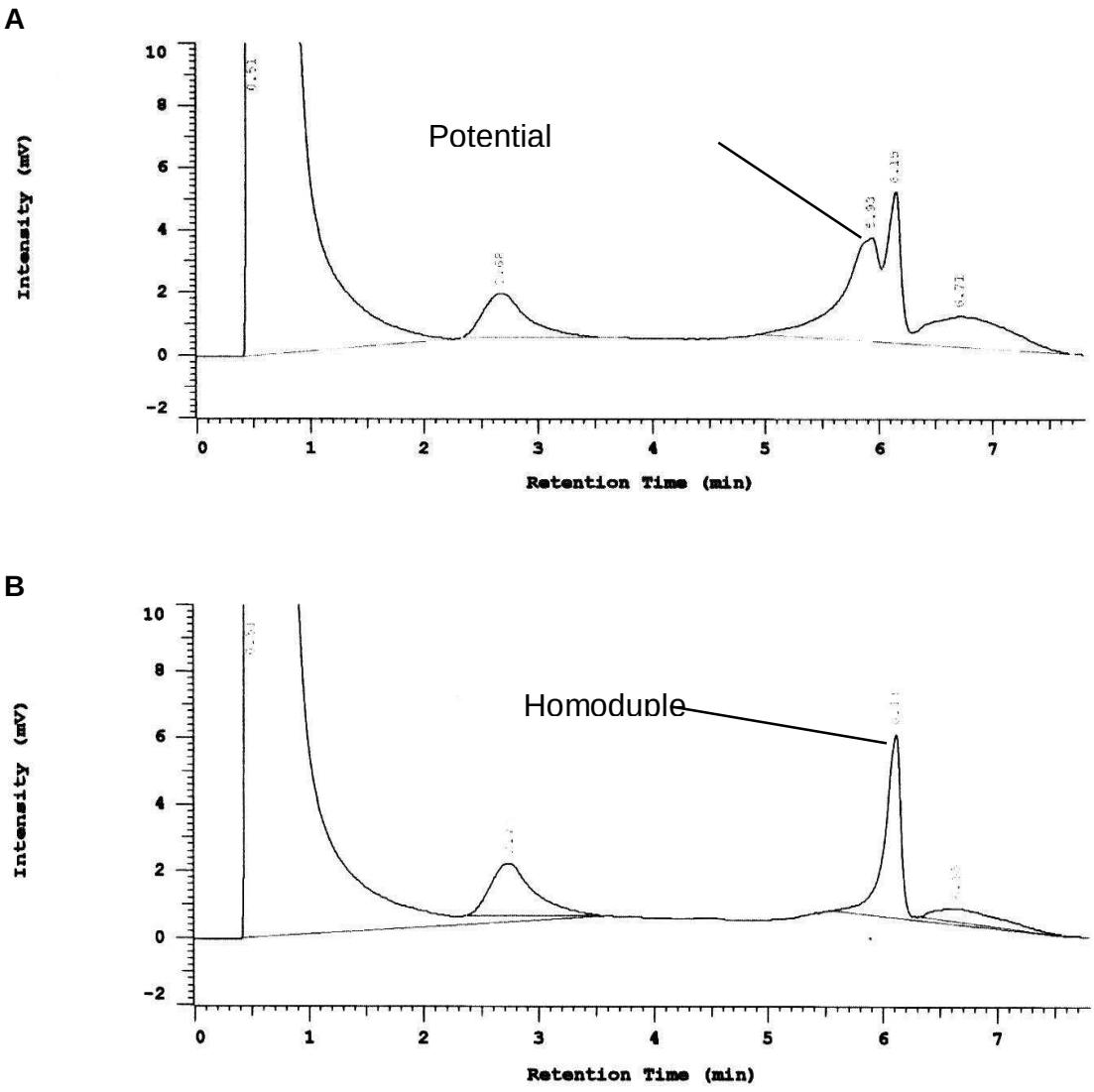


Figure 32 – Comparative dHPLC scans for Promoter Region B (*UCH-L1* nucleotide sequence -31 → -311) showing **(A)** a sample exhibiting potential heteroduplexes, and **(B)** a sample displaying only homoduplex species. Samples were eluted at an 'oven temperature' of 65ºC.

- The y-axis indicates the absorbance intensity in millivolts (mV), whilst the x-axis displays the retention (elution) time in minutes (mins).

- The initial vertical peak at a retention time ≈1 minute was the elution of excess PCR primer, the observed peak at ≈2.75 minutes was non-specific PCR product elution, and the final shallow peak immediately after hetero/ homoduplex peak elution was genomic DNA.

3.1.1.3 dHPLC Analysis of Promoter Region C

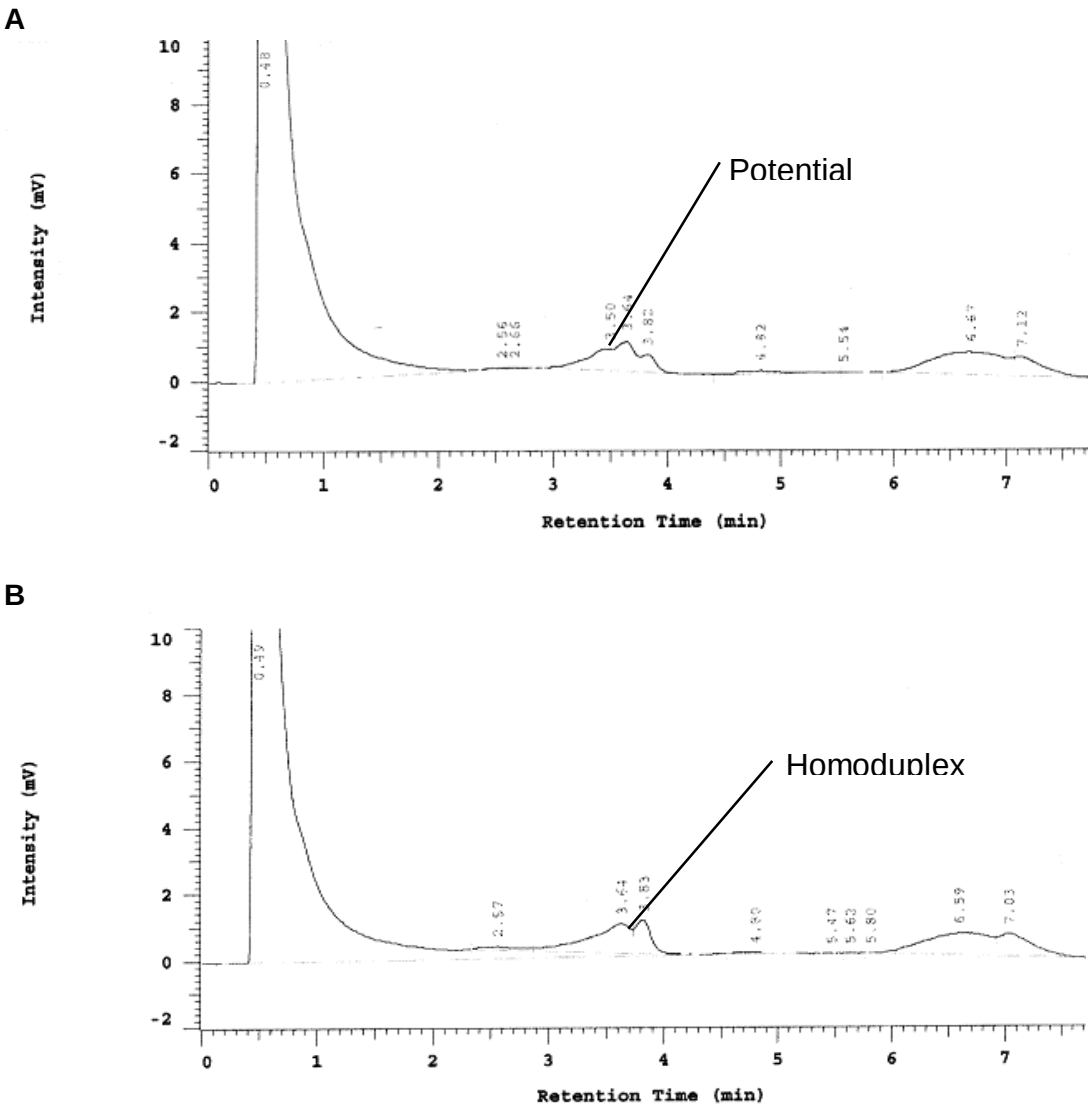


Figure 33 – Comparative dHPLC scans for Promoter Region C (*UCH-L1* nucleotide sequence -235 → -549) showing **(A)** a sample exhibiting potential heteroduplexes, and **(B)** a sample displaying only homoduplex species. Samples were eluted at an 'oven temperature' of 57ºC.

- The y-axis indicates the absorbance intensity in millivolts (mV), whilst the x-axis displays the retention (elution) time in minutes (mins).

- The initial vertical peak at a retention time ≈1 minute was the elution of excess PCR primer, and the final peak(s) at ≈6.75 minutes retention time was genomic DNA elution.

3.1.1.4 dHPLC Analysis of Promoter Region D

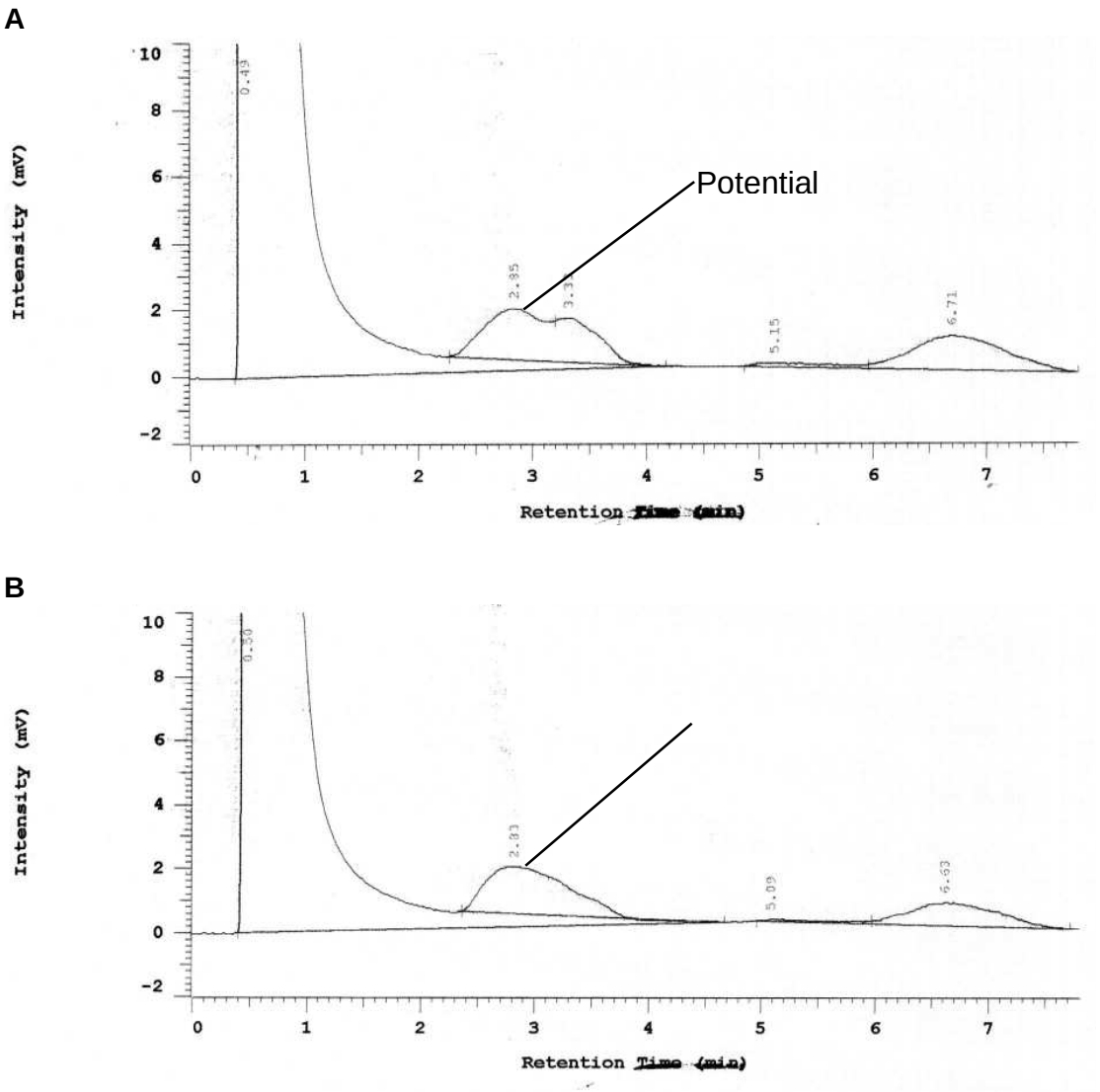


Figure 34 – Comparative dHPLC scans for Promoter Region D (*UCH-L1* nucleotide sequence -475 → -802) showing **(A)** a sample exhibiting potential heteroduplexes, and **(B)** a sample displaying only homoduplex species. Samples were eluted at an 'oven temperature' of 58ºC.

- The y-axis indicates the absorbance intensity in millivolts (mV), whilst the x-axis displays the retention (elution) time in minutes (mins).

- The initial vertical peak at a retention time ≈1 minute was the elution of excess PCR primer, and the final peak at ≈6.75 mins retention time was genomic DNA elution.

3.1.1.5 Interpretation of dHPLC Analysis

Promoter regions A, B, C and D have exhibited potential heteroduplexes, and thus possible sequence mutations or polymorphisms. These samples were then directly sequenced, along with four samples displaying only homoduplex species for nucleotide sequence comparison, to confirm the presence and elucidate the exact nature of these sequence variations.

The following regions did not realise any potential heteroduplexes during dHPLC mutation scanning: promoter region E, exons 1 & 2, exon 4, exons 5 & 6, exon 7, exon 8 and exon 9.

Exon 3 was unable to be mutation scanned using dHPLC as the PCR amplicon attained for this coding region required the use of betaine. Betaine is a substance which cannot be used within Transgenomic's dHPLC apparatus (Transgenomic dHPLC User Manual).

As described in section 1.3.2, *UCH-L1* exon 3 harbours a cytosine to adenine substitution at nucleotide position +54 – corresponding to a serine to tyrosine polymorphism at amino acid position 18 (Ser18Tyr) - which has been hypothesised to confer protection against Parkinson's disease in certain populations. To identify samples which harbour this allele, a viable Restriction Fragment Length Polymorphism (RFLP) genotyping strategy will be utilised, followed by direct DNA sequencing to confirm the presence of this single nucleotide polymorphism (SNP) in this study.

3.1.1.6 Confirming the Presence of Exon 3's Polymorphic +54A allele

A documented RFLP genotyping strategy utilising the restriction endonuclease *Rsa* I (Mellick & Silburn (2000)) was employed to identify samples harbouring the +54A allele (refer to section 2.2.4.4 for details of C54A RFLP genotyping strategy).

UCH-L1 exon 3 region/ *Rsa* 1 restriction enzyme digestion was set up on a set of 16 control DNA samples:-

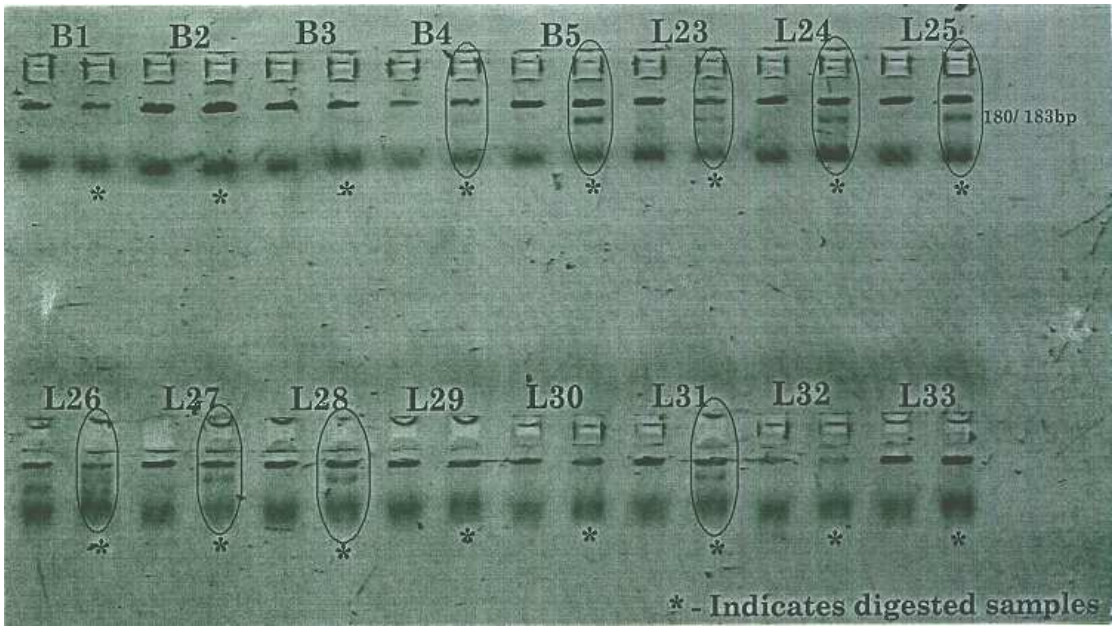


Figure 35 – A 5% polyacrylamide gel (refer to section 2.1.2) displaying *UCH-L1* exon 3 *Rsa* I RFLP analysis of x16 control DNA samples (arbitrarily labelled B1-5,.and L23-33). The nine circled lanes indicate exon 3 PCR products which realised partial *Rsa* I restriction endonuclease digestion, and are thus heterozygous for the polymorphic +54A allele. No homozygotes for the polymorphic A allele were seen to be evident, as no complete *Rsa* I digestion was realised (refer to section 2.2.4.4 for relevant RFLP genotyping strategy details). DNA sample nomenclature are annotated above each brace of polyacrylamide wells; right hand wells are undigested exon 3 PCR controls.

3.1.1.6.1 RFLP of Exon 3's +54A Allele - Interpretation

Nine out of the sixteen samples realised partial *Rsa* I digestion and thus displayed heterozygosity for the polymorphic A allele. In order to confirm the presence and exact nature of the C54A SNP in this study, four of the samples which exhibited heterozygosity were directly DNA sequenced, along with three samples which did not realise cleavage for nucleotide sequence comparison.

3.1.2 Sequencing Results

Sequencing electropherograms are only displayed for those regions giving rise to sequence variations, i.e. SNPs. One sequence electropherogram is displayed per region elucidating the precise nature of the SNP, together with one other electropherogram for comparison purposes which harbours no sequence variations.

- Only sequence regions in the immediate base pair vicinity of realised variations are shown.

- The numbers accompanying each designated sequence SNP, is the corresponding base position in relation to *UCH-L1*'s translation start site (ATG start codon of exon 1).

UCH-L1 **Region**	**Quantity**	**Description**
Promoter Region A	4	A-24G, C-16T (rs9321), G12A, G21A
Promoter Region B	1	G-234A
Promoter Region C	2	A-307G (rs13129604), A-306G
Exon 3	1	C54A (rs5030732)

Table 8 – Summarises the novel *UCH-L1* SNPs elucidated in this section.

- SNP description indicates the sense strand nucleotide change and base pair position relative to the ATG start site of exon 1, i.e. A-24G describes an adenine to guanine SNP at base pair position -24. (C – cytosine, T – thymine).

- If available, concordant 'rs' numbers (NCBI database) are indicated in brackets after the respective SNP.

3.1.2.1 DNA Sequencing of Promoter Region A

As shown in the electropherograms below, promoter region A revealed four novel SNPs - two in the 5' UTR: an adenine to guanine substitution at position -24, and a cytosine to thymine base substitution at position -16; and two located in the coding region of exon 1: both novel guanine to adenine nucleotide substitutions at positions +12 and +21.

- Shown here are the relevant nucleotide sequences (sense and anti-sense) corresponding to the electropherograms displayed for promoter region A in this section. (Sequences were cross referenced with the NCBI website database). The positions of the four observed nucleotide variations are highlighted and denoted in red:-

A-24G C-16T

Sense (5' → 3'):- TTCGTCTTCCCAAGGCTATTTCTGCCGGGCGCTCCGCGAAGA

Anti-sense (3' ← 5'):- AAGCAGAAGGGTTCCGATAAAGACGGCCCGCGAGGCGCTTCT

T-24C G-16A

Anti-sense (5' → 3'):- TCTTCGCGGAGCGCCCGGCAGAAATAGCCTTGGGAAGACGAA

G-16A T-24C

G+12A G+21A

Sense (5' → 3'):- GAAGATGCAGCTCAAGCCGATGGAGATCAACCCCGAGGTG

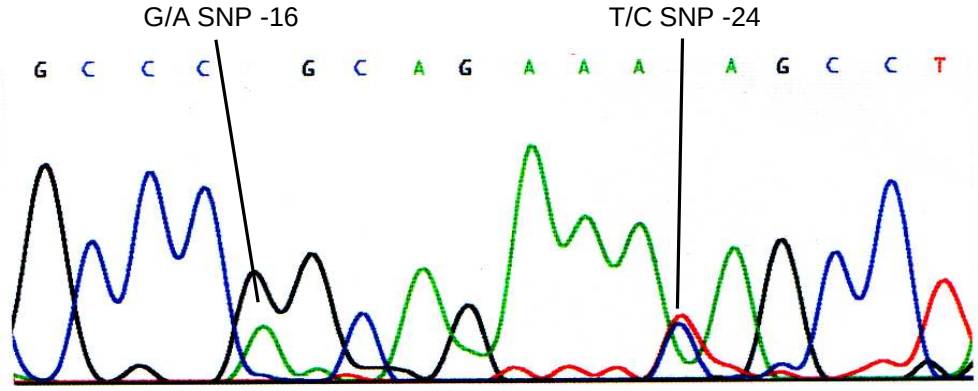
Anti-sense (3' ← 5'):- CTTCTACGTCGAGTTCGGCTACCTCTAGTTGGGGCTCCAC

C+12T C+21T

Anti-sense (5' → 3'):- CACCTCGGGGTTGATCTCCATCGGCTTGAGCTGCATCTTC

C+21T C+12T

A (Anti-sense)



B (Anti-sense)

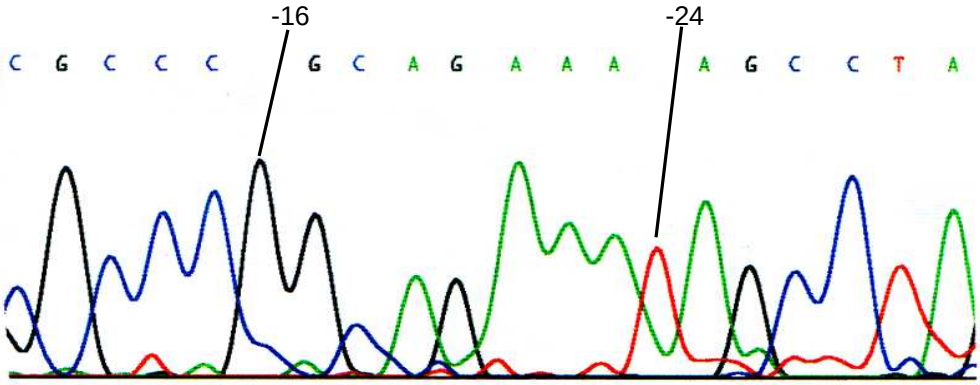


Figure 36 – Comparative DNA sequencing scans for the *UCH-L1* anti-sense nucleotide sequence -11 to -30 of promoter region A (relative to the ATG start codon of exon 1).

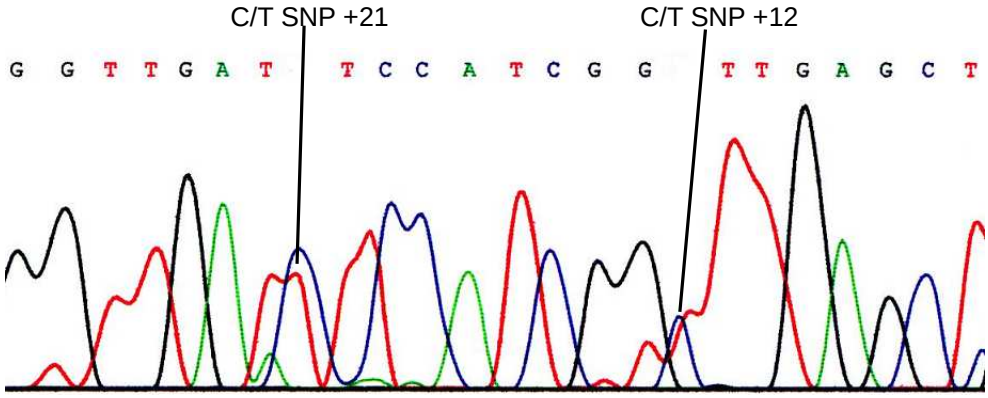
(A) Heterozygous sample exhibiting an anti-sense guanine to adenine SNP at position -16 (G/A SNP -16) and an anti-sense thymine to cytosine SNP at position -24 (T/C SNP -24), corresponding to sense strand SNPs C-16T and A-24G respectively.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair positions are indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

A (Anti-sense)



B (Anti-sense)

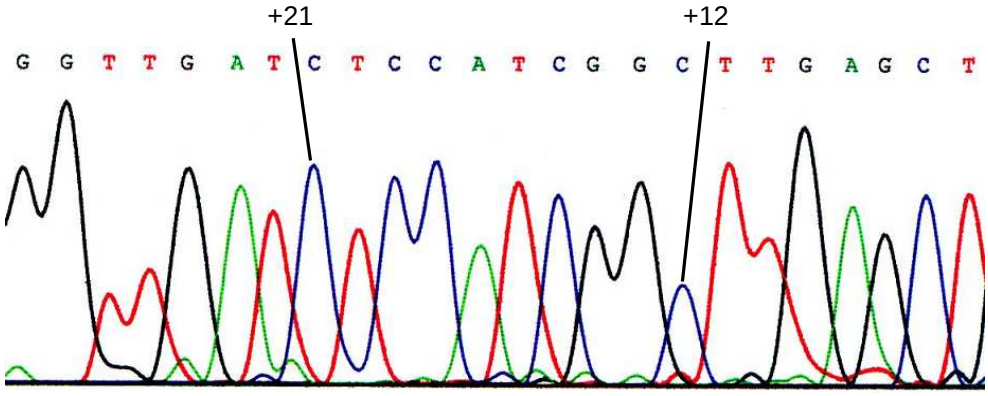


Figure 37 – Comparative DNA sequencing scans for the *UCH-L1* anti-sense nucleotide sequence +28 to +5 of promoter region A (relative to the ATG start codon of exon 1).

(A) Heterozygous sample exhibiting two anti-sense cytosine to thymine SNPs at positions +21 (C/T SNP +21) and +12 (C/T SNP +12), corresponding to sense strand SNPs G+21A and G+12A respectively.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair positions are indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

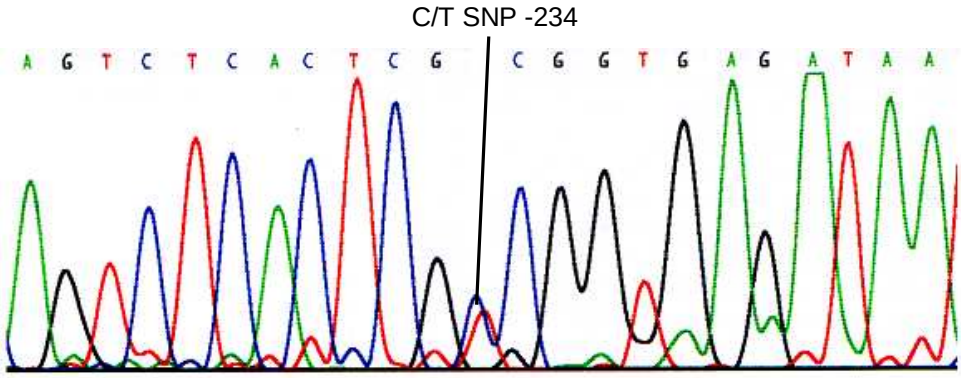
3.1.2.2 Sequencing Analysis of Promoter Region B

As shown in the electropherograms below, promoter region B revealed one novel SNP; a guanine to adenine nucleotide substitution at position -234 (relative to the ATG translation start codon of exon 1).

- Shown here are the relevant nucleotide sequences (sense and anti-sense) corresponding to the electropherograms displayed for promoter region B in this section. (Sequences were cross referenced with the NCBI website database). The position of the observed SNP is highlighted and denoted in red:-

	G-234A
Sense (5' → 3'):-	AACACCAGATTATCTCACCGGCGAGTGAGACTGCAAGGTTT
Anti-sense (3' ← 5'):-	TTGTGGTCTAATAGAGTGGCCGCTCACTCTGACGTTCCAAA
	C-234T
Anti-sense (5' → 3')):-	AAACCTTGCAGTCTCACTCGCCGGTGAGATAATCTGGTGTT
	C-234T

A (Anti-sense)



B (Anti-sense)

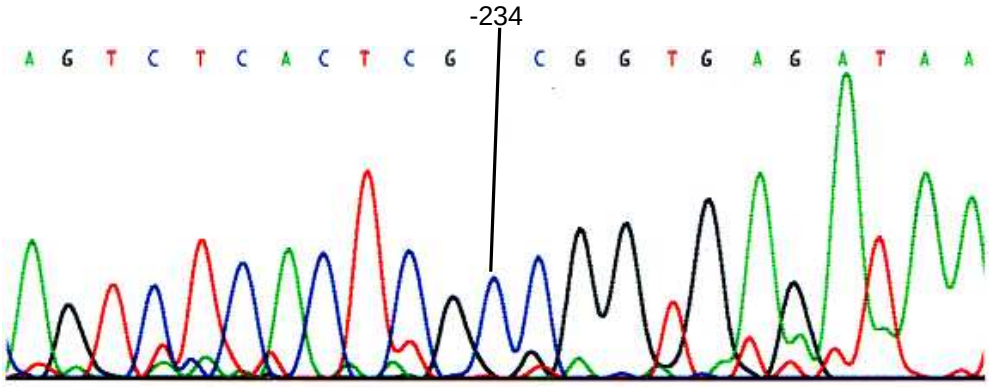


Figure 38 – Comparative DNA sequencing scans for the *UCH-L1* anti-sense nucleotide sequence -223 to -245 of promoter region B (relative to the ATG start codon of exon 1).

(A) Heterozygous sample exhibiting an anti sense cytosine to thymine SNP at position -234 (C/T SNP -234), corresponding to a G-234A sense strand SNP.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair position is indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

3.1.2.3 Sequencing Analysis of Promoter Region C

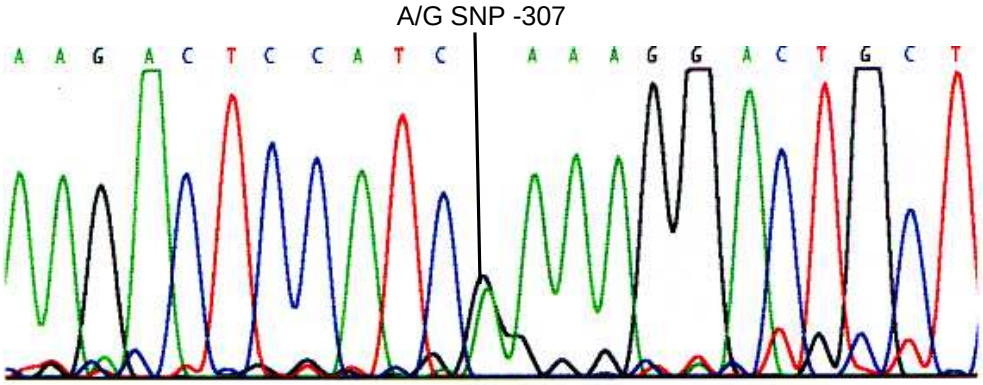
As shown in the electropherograms below, promoter region C revealed two novel SNPs; both adenine to guanine nucleotide substitutions, at positions -307 and -306 (relative to the ATG translation start codon of exon 1).

- Shown here is the relevant nucleotide sequence (sense strand) corresponding to the electropherograms displayed for promoter region C in this section. (Sequences were cross referenced with the NCBI website database). The positions of the two observed SNPs are highlighted and denoted in red:-

(Sense strand (5' → 3')

CCAAAATTAAAGACTCCATCAAAAGGACTGCTCCATACACT
A-307/6G

A (Sense)



B (Sense)

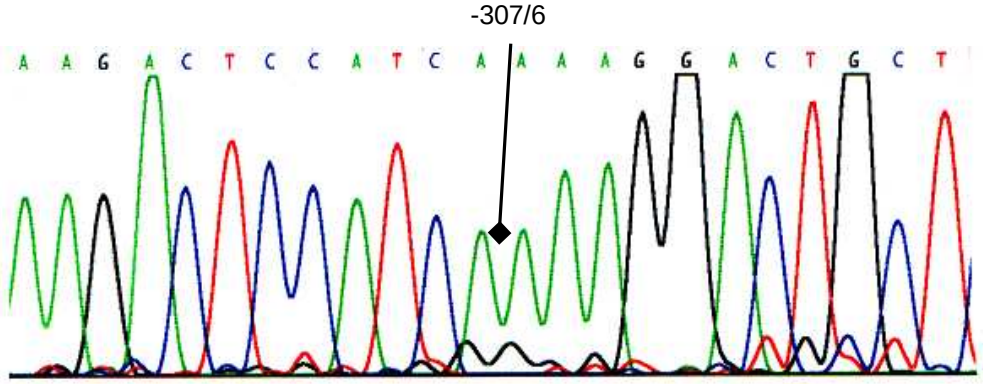


Figure 39 – Comparative DNA sequencing scans for the *UCH-L1* sense nucleotide sequence -318 to -296 of promoter region C (relative to the ATG start codon of exon 1).

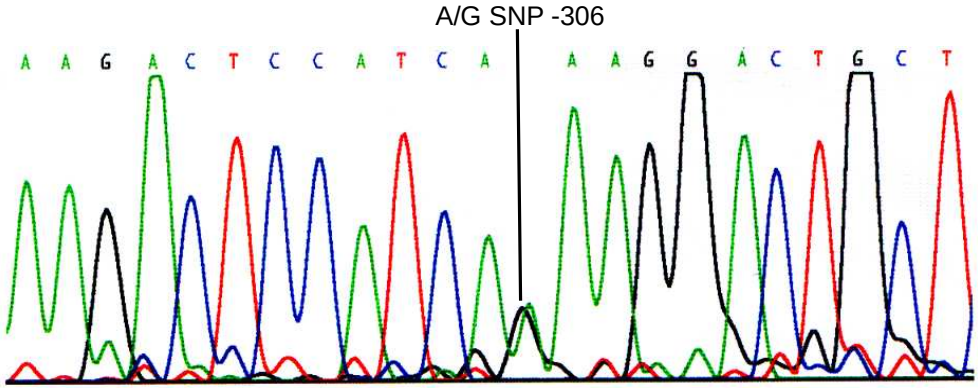
(A) Heterozygous sample exhibiting an adenine to guanine SNP at position -307 (A/G SNP -307); thus a A-307G sense strand SNP.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair positions are indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

A (Sense)



B (Sense)

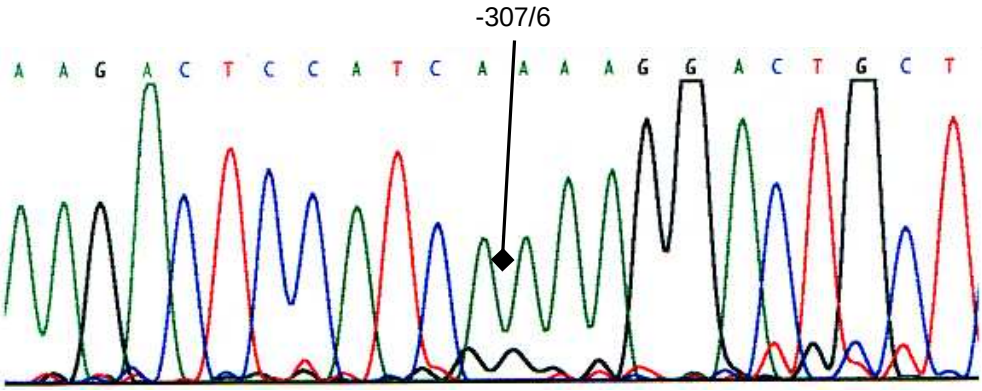


Figure 40 – Comparative DNA sequencing scans for the *UCH-L1* sense nucleotide sequence -318 to -296 of promoter region C (relative to the ATG start codon of exon 1).

(A) Heterozygous sample exhibiting an adenine to guanine SNP at position -306 (A/G SNP -306); thus a A-306G sense strand SNP.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair positions are indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

3.1.2.4 Exon 3 (C54A) Sequencing Results

As shown in the electropherograms below, Exon 3 revealed one previously described SNP (Farrer *et al*. (1999)); a cytosine to adenine nucleotide substitution at position +54 (relative to the ATG translation start codon of exon 1).

- Shown here are the relevant nucleotide sequences (sense and anti-sense) corresponding to the electropherograms displayed for exon 3 in this section. (Sequences were cross referenced with the NCBI website database). The position of the documented C54A SNP is highlighted and denoted in red:-

C+54A

Sense (5' → 3'):- CTCCTCTCCGCAGGTGCTGTCCCGGCTGGGGGTCGCCGGC

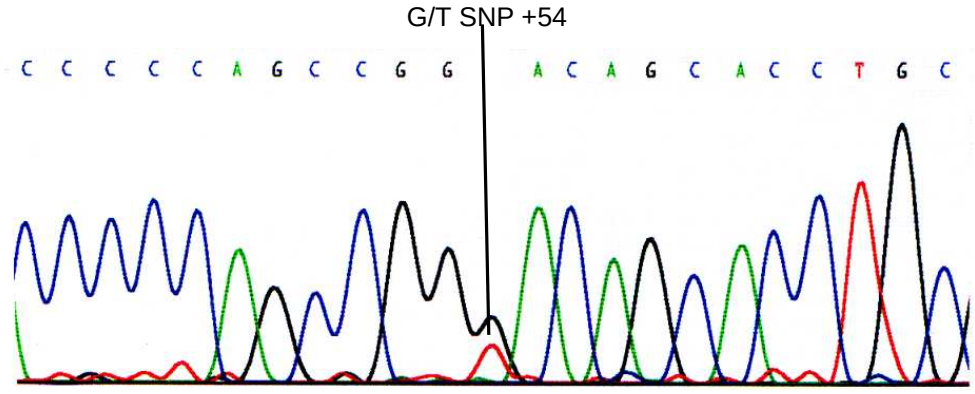
Anti-sense (3' ← 5'):- GAGGAGAGGCGTCCACGACAGGGCCGACCCCCAGCGGCCG

G+54T

Anti-sense (5' → 3'):- GCCGGCGACCCCCAGCCGGGACAGCACCTGCGGAGAGGAG

G+54T

A (Anti-sense)



B (Anti-sense)

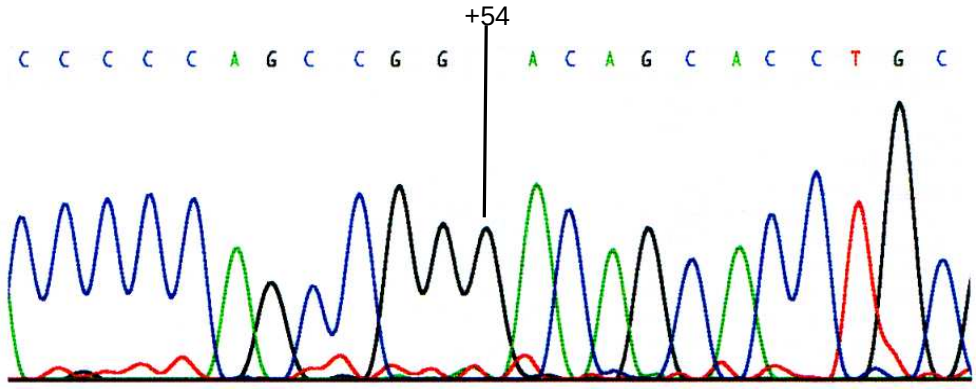


Figure 41 – Comparative DNA sequencing scans for the *UCH-L1* anti-sense nucleotide sequence +65 to +45 of exon 3 (relative to the ATG start codon of exon 1).

(A) Heterozygous sample exhibiting an anti-sense guanine to thymine SNP at position +54 (G/T SNP +54), corresponding to the C+54A documented sense strand SNP.

(B) Homozygous sample harbouring no nucleotide sequence variation (corresponding base pair position is indicated).

- SNPs are identifiable from significant sequencing peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

3.1.2.5 Interpretation of DNA Sequencing

DNA sequencing of the four *UCH-L1* promoter regions exhibiting potential dHPLC heteroduplexes (refer to section 3.1.1) has elucidated a total of seven SNPs - G21A, G12A, C-16T, A-24G, G-234A, A-306G and A-307G (relative to the translation start site). DNA sequencing has also confirmed the position and nature of the documented exon 3 SNP – C54A – whilst corroborating the *Rsa* I RFLP genotyping methodology used to initially identify those samples harbouring the polymorphic A allele.

UCH-L1 promoter region D, though it exhibited potential heteroduplex dHPLC scans (refer to section 3.1.1.4), did not elucidate any nucleotide sequence variations through DNA sequencing. This seems to indicate that the dHPLC scans displaying potential promoter region D heteroduplexes, were perhaps contaminated with non-specific PCR sequences which mimicked heteroduplex peaks.

In addition to the previously documented C54A SNP, two novel coding region SNPs have also been elucidated through DNA sequencing; both are present within *UCH-L1*'s exon 1 nucleotide sequence. However, unlike the C54A SNP, the polymorphic alleles of the G12A and G21A SNPs do not effect a change in UCH-L1's amino acid sequence, i.e. the primary protein structure of UCH-L1 remains unchanged. This is depicted below (polymorphic nucleotide and amino acid changes are indicated in red):-

UCH-L1's exon 1 nucleotide sequence

ATG CAG CTC AAG/A CCG ATG GAG/A ATC AAC

Met Gln Leu Lys/Lys Pro Met Glu/Glu Ile Asn

UCH-L1's corresponding amino acid sequence

Owing to the fact that neither the 12A and 21A polymorphic alleles bring about a variation in UCH-L1's primary structure, led to the cessation of investigative work into these exon 1 SNPs.

The five novel SNPs elucidated within *UCH-L1*'s 5' UTR/ promoter region and the confirmed exon 3 documented SNP, were then genotyped in Caucasian sample arrays to ascertain the most frequent haplotypes governing the distribution of the alleles in the Caucasian population.

3.1.3 Determination of Allele Frequency and *UCH-L1* Promoter Haplotypes

Analysis of the genotyping results attained for the five promoter/ 5' UTR SNPs in this section elucidated three common (greater than two percent) *UCH-L1* promoter/ 5' UTR haplotypes governing their allelic distribution.

Furthermore, *UCH-L1* exon 3's +54A polymorphic allele was not shown to be in linkage disequilibrium with any of the promoter/ 5' UTR polymorphic alleles.

3.1.3.1 RFLP Genotyping Strategies

Determination of the *UCH-L1* promoter/ 5' UTR and exon 3 (+54A) polymorphic allele frequencies was initially carried out in a x64 DNA sample set comprising of 31 unrelated healthy Caucasian DNA samples (refer to section 2.1.1.2).

UCH-L1 promoter/ 5' UTR polymorphic allele frequencies were also determined in a larger x480 DNA sample set comprising of 480 unrelated Caucasian DNA samples, supplemented with cognitive function data (refer to section 2.1.1.3).

Both sample sets had the relevant *UCH-L1* regions amplified by PCR protocol (refer to section 2.1.3), before the relevant RFLP methodology was employed for each SNP (refer to section 2.2.4) to determine their allelic frequencies.

Owing to reasons which have already been briefly discussed later (refer to section 2.2.4.5/ 6), two of the RFLP strategies that relied upon primer-introduced restriction analysis (PIRA – refer to section 2.2.3) – the A-24G and the A-306G RFLP genotyping strategies – did not generate any results. This meant that for each of the DNA sample sets, RFLP methodology had to be supplemented with additional data.

For each successful RFLP strategy, an example gel will be shown to indicate how raw allele frequency data was ascertained.

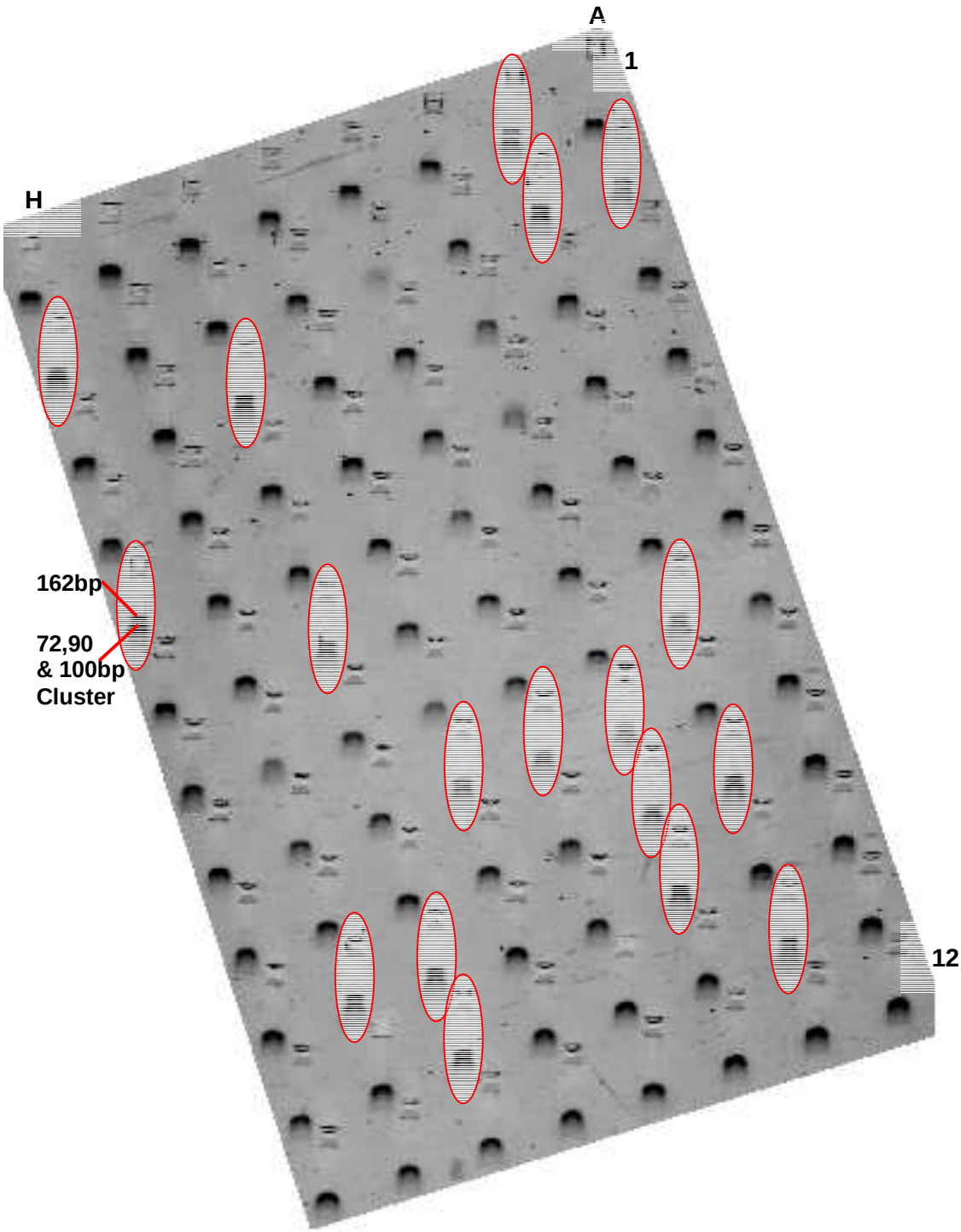


Figure 42 – C-16T SNP RFLP Genotyping Gel Output (visualised under UV light after staining with ethidium bromide). A 5% polyacrylamide gel displaying *UCH-L1* promoter region A *Nci* I RFLP analysis of x96 samples (from x480 DNA sample set). The 18 red circled samples indicate those PCR samples which realised partial double *Nci* I restriction endonuclease digestion, which can be deduced from the presence of both the 162bp DNA band and the 72, 90 & 100bp fragment cluster (e.g. well H5), and are thus heterozygous for the polymorphic -16T allele. No homozygotes for the polymorphic T allele were seen to be evident, as no purely single *Nci* I digestion was realised, i.e. the 72, 90 & 100bp DNA band cluster is visible in all other viable wells, indicating dual *Nci* I digestion, and thence reference homozygote (C allele) DNA. (Refer to section 2.2.4.1 for C-16T RFLP genotyping strategy details). Sample nomenclature runs in well rows from A to H, and in well columns from 1 to 12. Samples D2 and C4 did not elucidate any results owed to sample genomic DNA concentration deficiency.

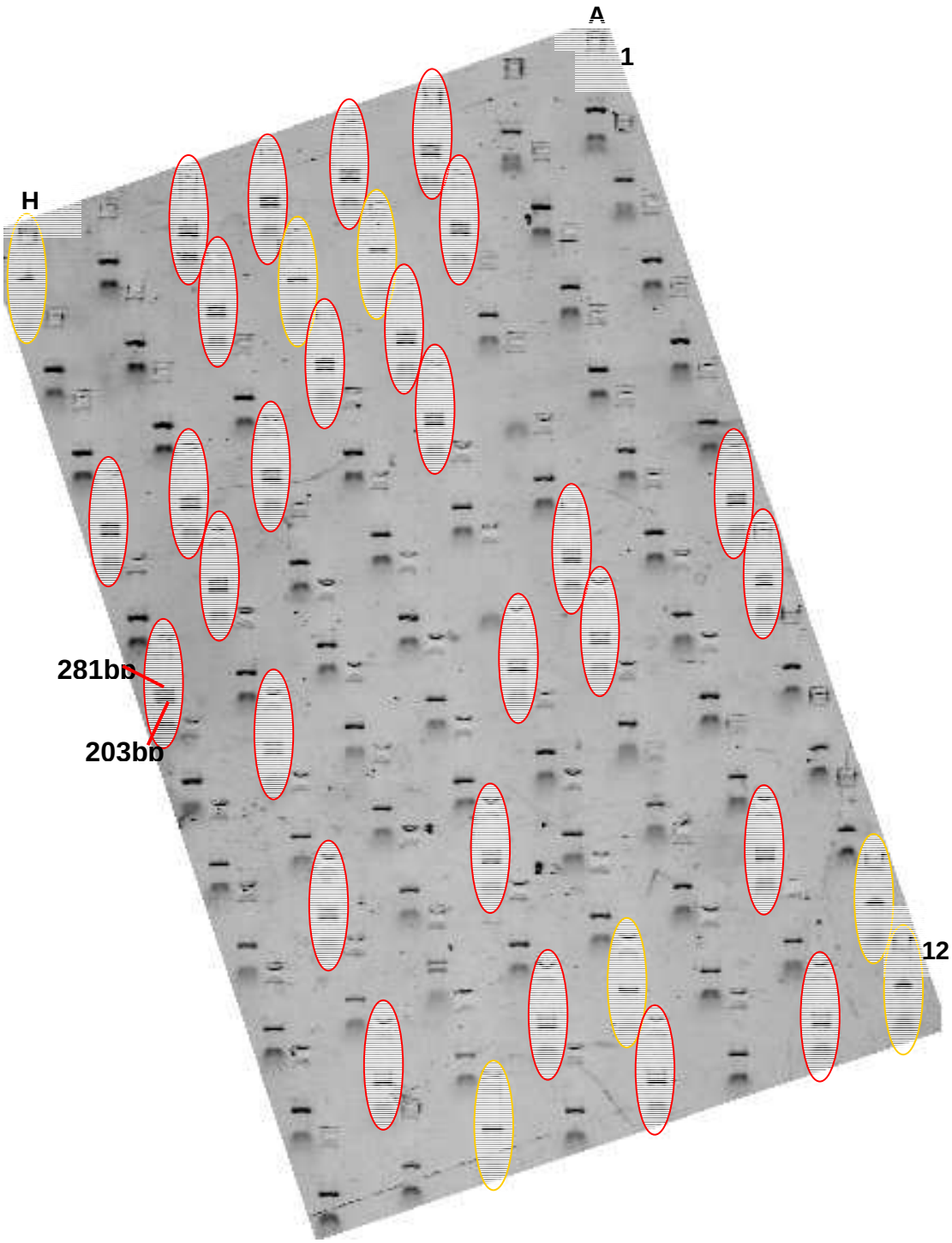


Figure 43 – G-234A SNP RFLP Genotyping Gel Output (visualised under UV light after staining with ethidium bromide). A 5% polyacrylamide gel displaying *UCH-L1* promoter region B *Sgr* AI RFLP analysis of x96 samples (from x480 DNA sample set). The 27 red circled samples indicate those PCR samples which realised partial *Sgr* AI restriction endonuclease digestion, which can be deduced from the presence of both the 281bp (undigested) and 203bp DNA bands (e.g. well H6), and are thus heterozygous for the polymorphic -234A allele. The 7 gold circled digests indicate those samples which realised complete *Sgr* A1 digestion – as no 203bp fragment is evident - and are thus homozygous for the polymorphic A allele. All other viable wells display homozygosity for the reference G allele as no 281bp band is elucidated. (Refer to section 2.2.4.2 for G-234A RFLP genotyping strategy details). Sample nomenclature runs in well rows from A to H, and in well columns from 1 to 12. Samples C4 and D6 did not elucidate any results owed to sample genomic DNA concentration deficiency.

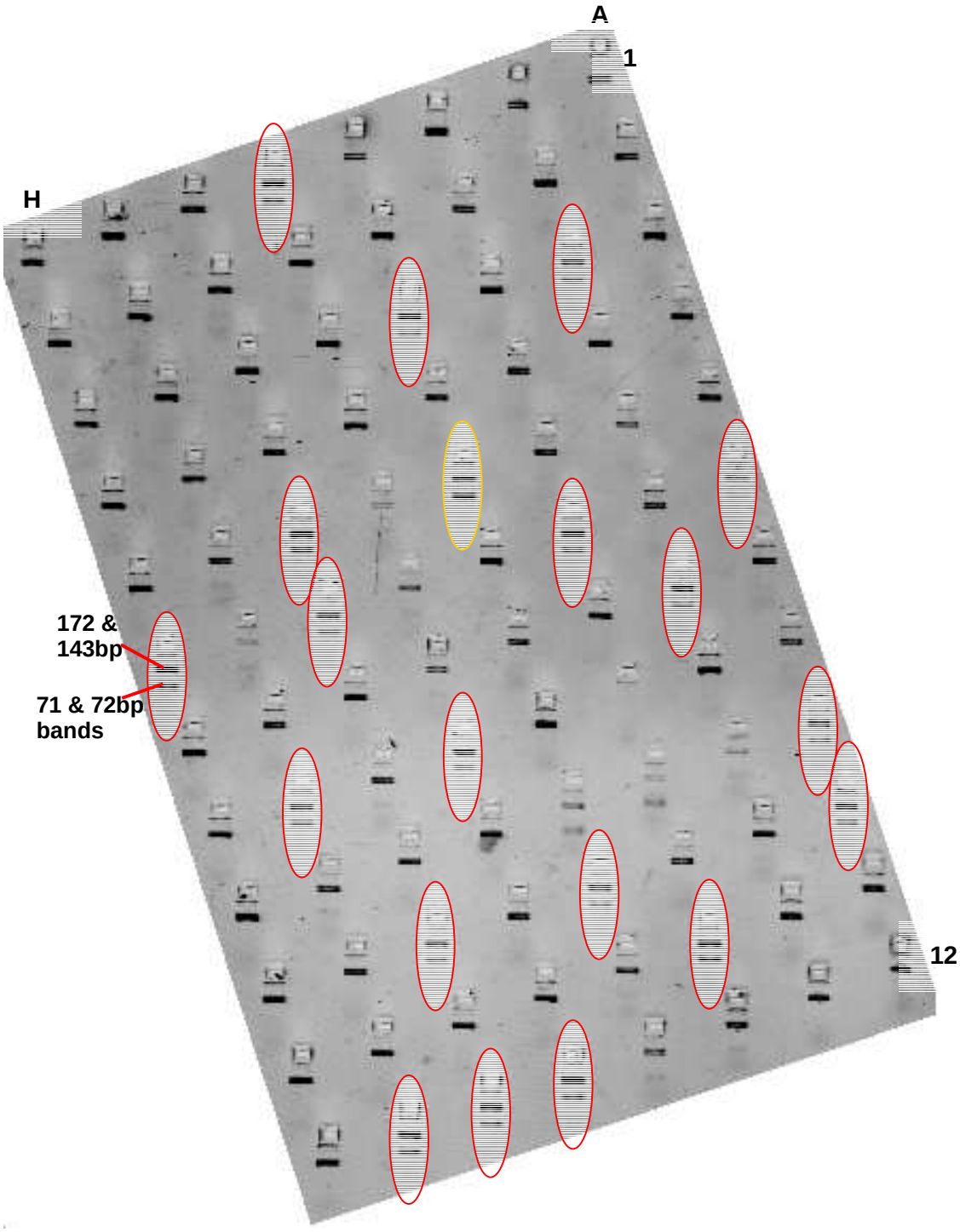


Figure 44 – A-307G SNP RFLP Genotyping Gel Output (visualised under UV light after staining with ethidium bromide). A 5% polyacrylamide gel displaying *UCH-L1* promoter region C *Taq* I RFLP analysis of x96 samples (from x480 DNA sample set). The 19 red circled samples indicate those PCR samples which realised partial double *Taq* I restriction endonuclease digestion, which can be deduced from the presence of both the 172 & 143bp fragments and the 71 & 72bp DNA bands (e.g. well H6), and are thus heterozygous for the polymorphic -307G allele. The gold circled digest indicates a sample which realised complete *Taq* 1 digestion – as only three bands are evident (i.e. no 143bp band) - and is thus homozygous for the polymorphic G allele. All other viable wells display homozygosity for the reference A allele as no 71 & 72bp fragments are seen. (Refer to section 2.2.4.3 for A-307G RFLP genotyping strategy details). Sample nomenclature runs in well rows from A to H, and in well columns from 1 to 12. Samples E5, C8, C9 and D9 did not elucidate any results owed to sample genomic DNA concentration deficiency.

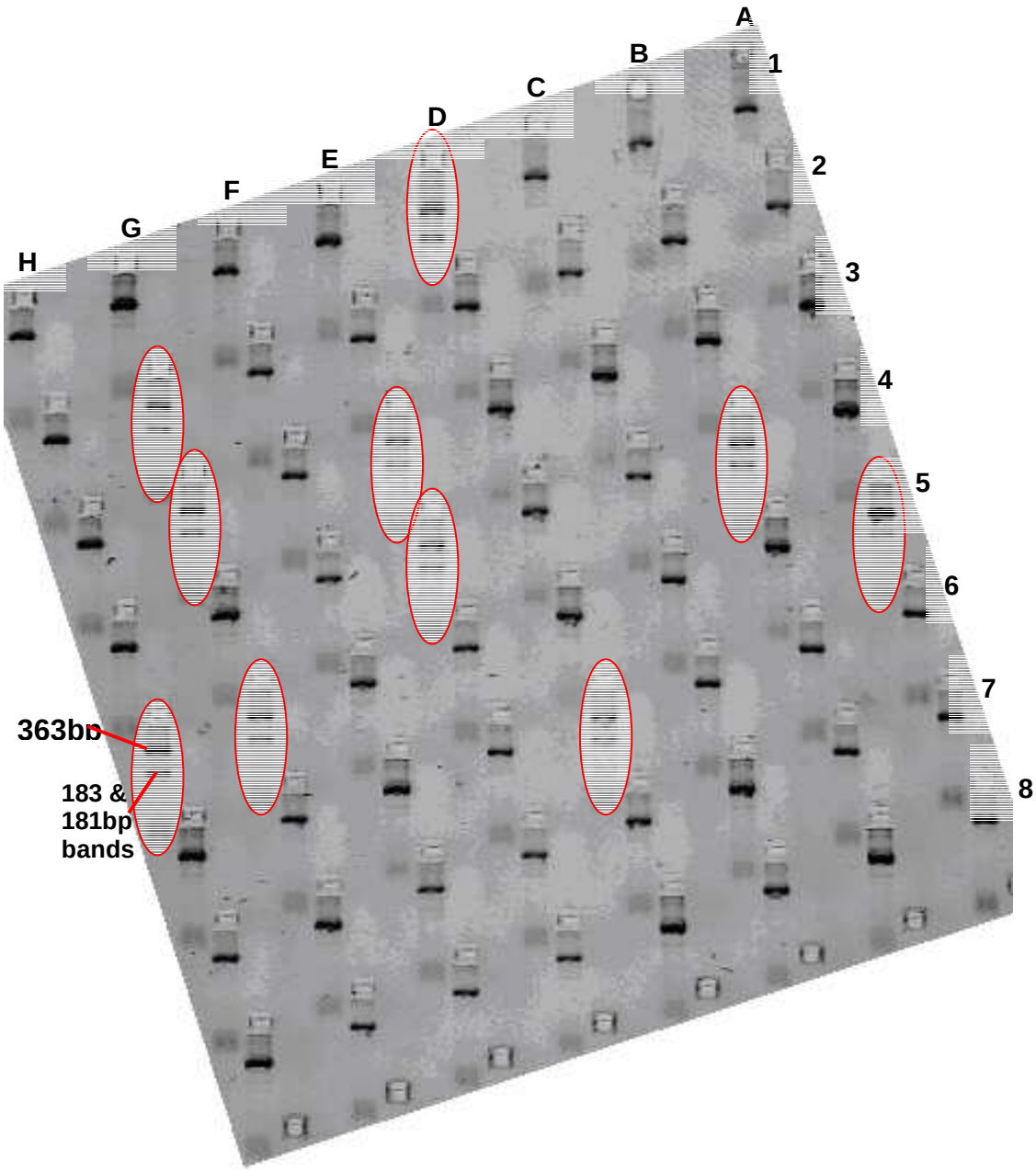


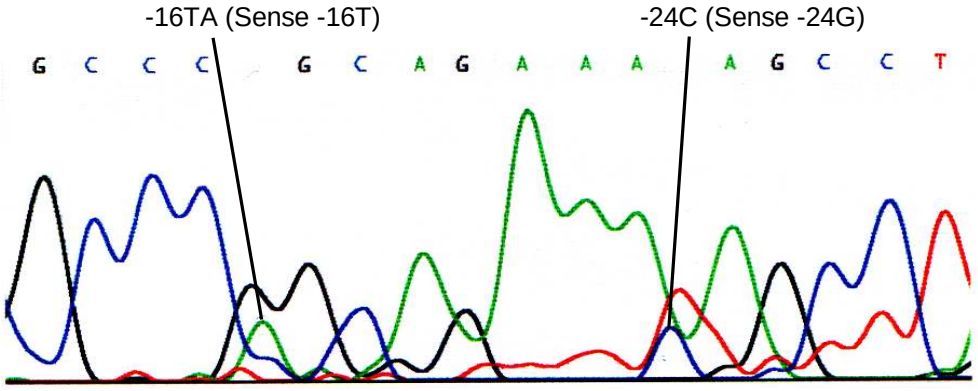
Figure 45 – C54A SNP RFLP Genotyping Gel Output (visualised under UV light after staining with ethidium bromide). A 5% polyacrylamide gel displaying *UCH-L1* exon 3 *Rsa* I RFLP analysis of x64 samples (from x64 DNA sample set). The 10 red circled samples indicate those PCR samples which realised partial *Rsa* I restriction endonuclease digestion, which can be deduced from the presence of both the undigested 363bp DNA band and the digested 180 & 183bp fragments (e.g. well H5), and are thus heterozygous for the polymorphic +54A allele. No homozygotes for the polymorphic A allele were seen to be evident, as no complete *Rsa* I digestion was realised, i.e. the 363bp band is the only visible band in all the remaining lanes, indicating that the remaining samples are homozygous for the reference C allele. (Refer to section 2.2.4.4 for C+54A RFLP genotyping strategy details). Sample nomenclature runs in well rows from A to H, and in well columns from 1 to 8.

As indicated above, owed to the lack of results obtained from both the A-24G and A-307G PIRA-RFLP genotyping strategies, RFLP methodology had to be supplemented. For the x64 DNA sample set, this took the form of supplemental DNA sequencing data.

Direct DNA sequencing was utilised to further investigate the linkage between the -16T and -24G alleles. All the x64 DNA samples that exhibited heterozygosity for the polymorphic -16T allele, through RFLP methodology, were sequenced to elucidate the potential presence of the -24 G allele.

Sequencing of *UCH-L1* promoter region A elucidated the -24G allele in all the DNA samples that had exhibited the polymorphic T allele through RFLP analysis.

A (Anti-sense)



B (Anti-sense)

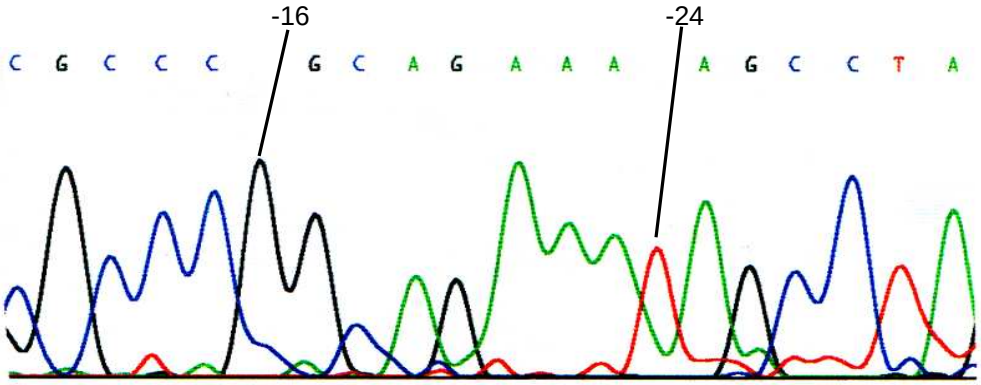


Figure 46 – Comparative DNA sequencing scans for the *UCH-L1* anti-sense nucleotide sequence -11 to -30 of promoter region A (relative to the ATG start codon of exon 1).

(A) Sample that previously exhibited heterozygosity for the polymorphic -16T allele (corresponds to anti-sense -16A allele above) through RFLP methodology, also displays heterozygosity for the anti-sense -24C allele (sense strand -24G allele).

(B) Homozygous sample harbouring no polymorphic alleles (corresponding base pair positions are indicated).

- The polymorphic alleles are identifiable in the heterozygous sample above (A) by way of significant peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

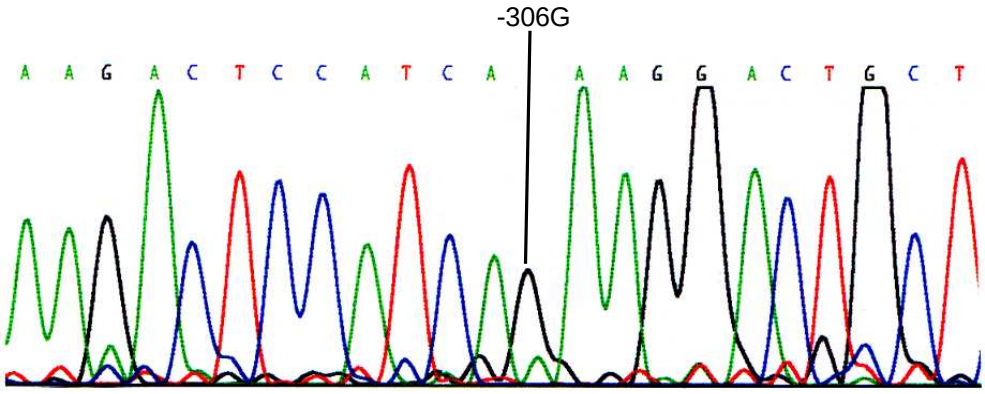
- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

RFLP analysis also indicated that all the samples in the x64 DNA sample set that exhibited heterozygosity for the -16T allele, and thus also the -24G allele (refer above), were correspondingly heterozygous for the -307G allele.

DNA sequencing was employed to investigate the linkage between the -306/ -307 polymorphic G alleles and the -234A allele. All the samples in the x64 DNA sample set that exhibited homozygosity/ heterozygosity for the polymorphic -234A allele through RFLP analysis, were accordingly sequenced to elucidate the potential presence of either the -306 or -307 G allele.

Sequencing of promoter region C elucidated the -306G allele (in a homozygous/ heterozygous dependent manner) in all the DNA samples that had exhibited the polymorphic -234A allele through RFLP analysis. The -307G allele was not seen to be present.

A (Sense)



B (Sense)

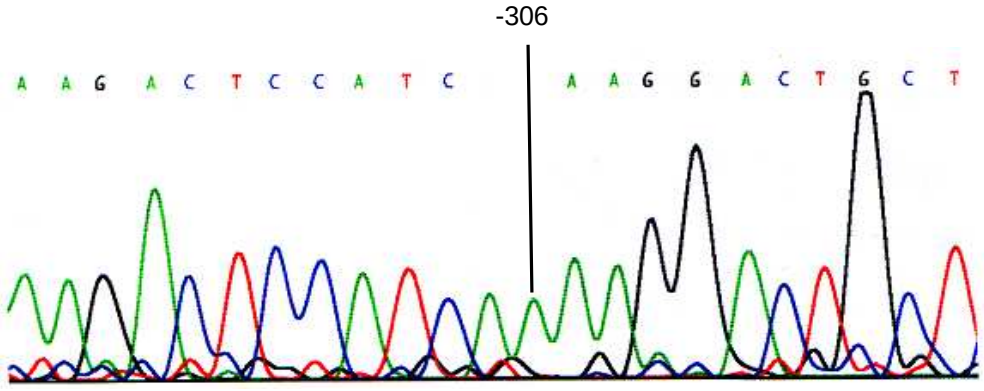


Figure 47 – Comparative DNA sequencing scans for the *UCH-L1* sense nucleotide sequence -318 to -296 of promoter region C (relative to the ATG start codon of exon 1).

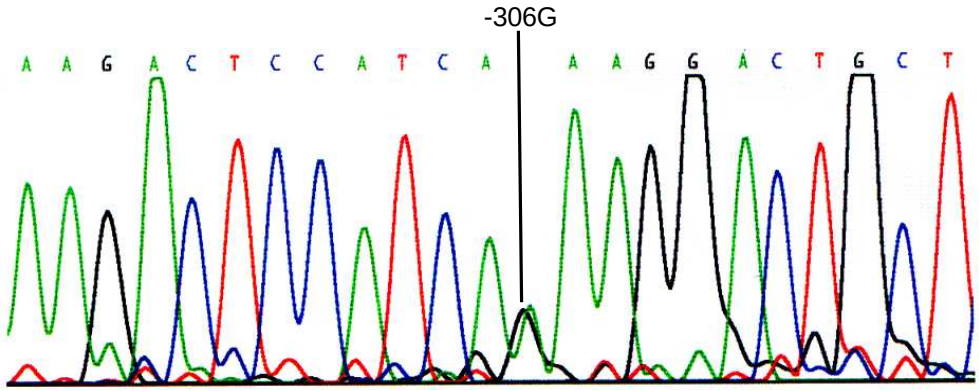
(A) Sample that previously exhibited homozygosity for the polymorphic -234A allele through RFLP analysis, also displays homozygosity for the -306G allele.

(B) Homozygous sample harbouring no polymorphic alleles (corresponding base pair position is indicated).

- The polymorphic allele is identifiable in the homozygous sample above (A) by way of comparison with the non-polymorphic sample (B), which exhibits a different coloured peak (different nucleotide base – see below) at the corresponding base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

A (Sense)



B (Sense)

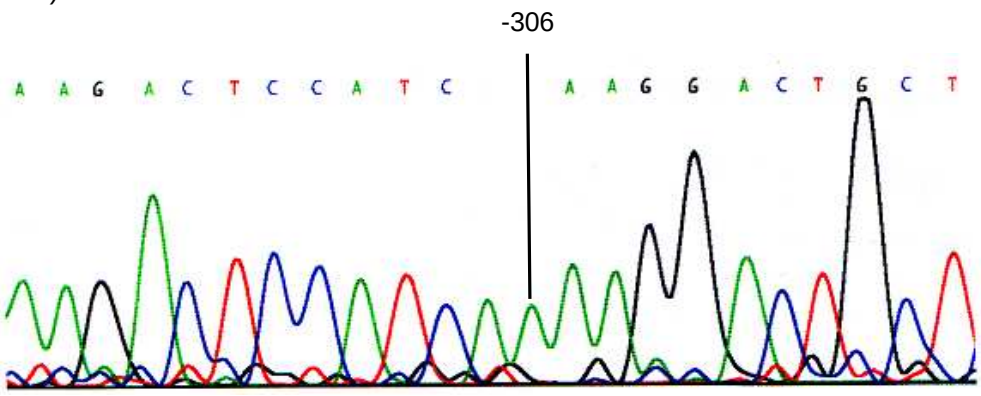


Figure 48 – Comparative DNA sequencing scans for the *UCH-L1* sense nucleotide sequence -318 to -296 of promoter region C (relative to the ATG start codon of exon 1).

(A) Sample that previously exhibited heterozygosity for the polymorphic -234A allele through RFLP analysis, also displays heterozygosity for the -306G allele.

(B) Homozygous sample harbouring no polymorphic alleles (corresponding base pair position is indicated).

- The polymorphic allele is identifiable in the heterozygous sample above (A) by way of significant peaks of different colours (nucleotide bases – see below) being realised at the same base pair position. (Smaller peaks are caused by background non-specific PCR products).

- Nucleotide bases are colour coded accordingly: blue – cytosine(C), black – guanine(G), green – adenine(A), and red – thymine (T).

For the x480 DNA sample set, owed to the lack of results attained from both the A-24G and A-307G PIRA-RFLP genotyping strategies, RFLP methodology was supplemented with allele frequency data supplied by Kbiosciences Ltd.

3.1.3.2 x64 DNA Sample Set (62 Chromosomes) Genotyping Results and Analysis

3.1.3.2.1 Variant Allele Frequency Table

RFLP Analysis

Allele	Frequency
-307 G	8.1%
-234 A	25.8%
-16 T	8.1%
+54 A	8.1%

Sequencing Analysis

Allele	Frequency
-306 G	25.8%
-24 G	8.1%

Table 9 –Displays the frequency of variant alleles obtained from the x 31 DNA sample set through both RFLP and sequencing analysis (refer above). The variant alleles are listed on the left, with their respective allelic frequencies within the sample set indicated on the right.

3.1.3.2.2 Genotype Frequency Table

Nucleotide Position of Allele						
-307	-306	-234	-24	-16	+54	**Genotype Freq.**
A	A	G	A	C	C	36
A	G	A	A	C	C	16
G	A	G	G	T	C	5
A	A	G	A	C	A	5

Table 10 – Collation of the RFLP and sequencing results with respect to the genotypes exhibited by each of the x31 individuals within the x64 DNA sample set. The genotypes elucidated are indicated by A (adenine), G (guanine), C (cytosine) and T (thymine) under the relevant nucleotide position (displayed downstream from left to right). Their frequencies are indicated in the far right hand column. (Variant alleles are denoted in Red).

3.1.3.2.3 Hardy-Weinberg Equilibrium Consistency of Genotype Frequencies

(The Hardy-Weinberg equilibrium concept is introduced in section 1.7.2.1)

	Single Nucleotide Polymorphism			
	A-307G, A-24G, C-16T & C+54A		G-234A & A-306G	
Genotypes	Observed	Expected	Observed	Expected
Homozygotes - Reference	26	26.2	17	17.1
Heterozygotes	5	4.6	12	11.9
Homozygotes - Variant	0	0.2	2	2.1
Variant Allele Frequency	0.08		0.26	
Chi-Squared Test *p* Value	0.8876		0.9982	

(*p* value < 0.05 indicates inconsistency with Hardy-Weinberg Equation).

Table 11 – Exhibits the genotype frequencies for the six *UCH-L1* SNPs investigated for this sample set with the Hardy-Weinberg formula (equation) applied. Both the actual frequencies ('Observed'), and the Hardy-Weinberg Equation theoretically 'Expected' frequencies are displayed. The variant allele frequency is also indicated, with the final row exhibiting the probability (*p* value) of consistency between the 'Observed' and 'Expected' figures, calculated through Chi-Squared statistical analysis.

3.1.3.2.4 Haplotype Frequency Diagrams

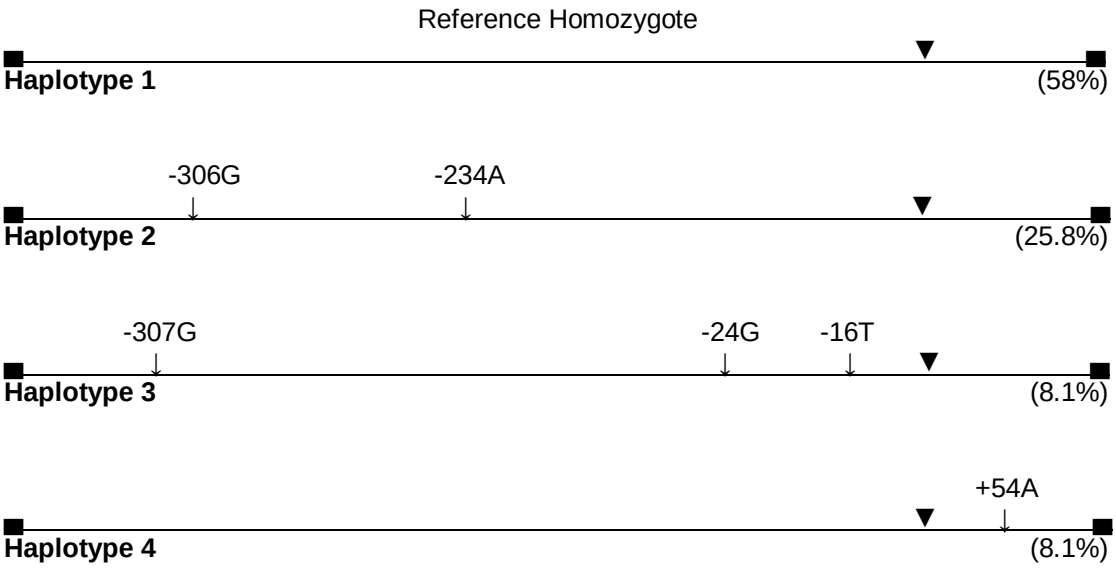


Figure 49 - The four *UCH-L1* haplotypes are diagrammatically represented with reference to the distribution of the six polymorphic alleles, elucidated through RFLP and sequencing genotyping analysis of the x64 sample set (allelic positions not to scale). Their haplotypic frequency within the 62 chromosomes studied is indicated in brackets. (▼ – indicates the translation start site (+1)).

3.1.3.2.5 Interpretation of x64 DNA Sample Set Genotyping Analysis

Through restriction fragment length polymorphism and sequencing analysis the frequencies of the six *UCH-L1* variant alleles were ascertained (Table 9), and once collated with respect to the genotypes exhibited by each of the 31 individuals in the x64 DNA sample (Table 10), haplotype frequencies were able to be determined (Figure 49).

Four distinct haplotypes were elucidated: the reference haplotype (AAGACC) displayed a frequency of 58%, the polymorphic AGAACC haplotype displayed a frequency of 25.8%, whilst both variant GAGGTC and AAGACA haplotypes realised a frequency of 8.1%.

The fact that the genotype frequencies for each of the six *UCH-L1* SNPs investigated were consistent with the Hardy-Weinberg formula lends credence to these results – see Table 9 (the genotype frequencies for each of the SNPs correlated well with those predicted by the Hardy-Weinberg equation).

These results from the 31 individuals (x62 chromosomes) investigated did not elucidate the +54A allele in linkage disequilibrium with any of the promoter/ 5' UTR alleles. Thence, this polymorphism was removed from any proceeding experimental work in this study, as the investigation concentrated on the promoter/ 5' UTR region of *UCH-L1.*

3.1.3.3 x480 DNA Sample Set (960 Chromosomes) Genotyping Results and Analysis

3.1.3.3.1 Variant Allele Frequency Table

Allele	Frequency
-307 G	9.7%
-306 G	17.3%
-234 A	17.6%
-24 G	9.9%
-16 T	9.7%

Table 12 – Displays the frequency of variant alleles obtained from the x 480 DNA sample set. The variant alleles are listed on the left, with their respective allelic frequencies within the sample set indicated on the right.

3.1.3.3.2 Genotype Frequency Table

Nucleotide Position of Allele					
-307	-306	-234	-24	-16	**Genotype Freq.**
A	A	G	A	C	577
A	G	A	A	C	148
G	A	G	G	T	74
G	G	A	G	T	16
A	A	A	A	C	3
A	G	A	G	C	2
G	A	G	G	C	2
A	A	G	A	T	2
G	A	G	A	T	1
A	A	G	G	C	1

(N.B. – 134 Indeterminable alleles)

Table 13 – Collation of the results with respect to the (determinable) genotypes exhibited within the x480 DNA sample set (960 chromosomes). The genotypes elucidated are indicated by A (adenine), G (guanine), C (cytosine) and T (thymine) under the relevant nucleotide position (displayed downstream from left to right). Their frequencies are indicated in the far right hand column. (Variant alleles are denoted in Red).

3.1.3.3.3 Hardy-Weinberg Equilibrium Consistency of Genotype Frequencies

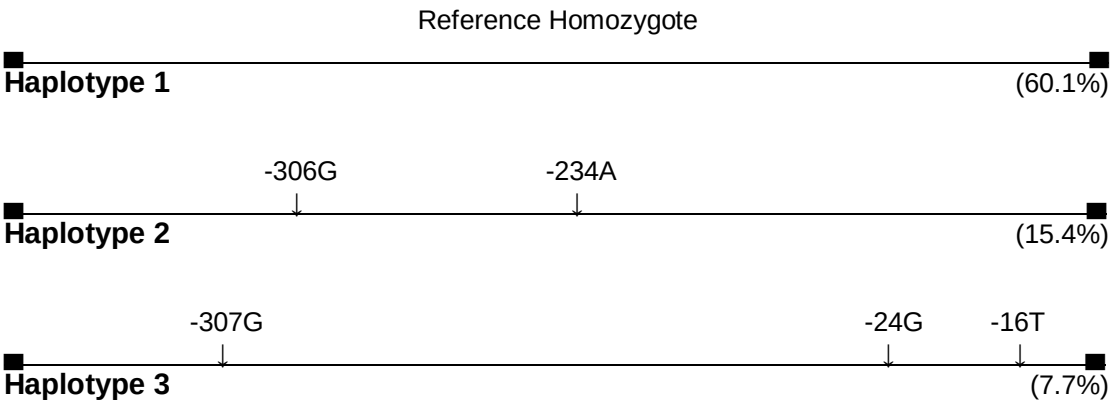
(The Hardy-Weinberg equilibrium concept is introduced in section 1.7.2.1)

	Single Nucleotide Polymorphism									
	A-307G		A-306G		G-234A		A-24G		C-16T	
Genotypes	O	E	O	E	O	E	O	E	O	E
Homozygotes (Reference)	376.	377.6	275	275.2	317	312.3	367	368.9	376	377.6
Heterozygotes	87	83.7	134	133.6	129	138.4	89	85.2	87	83.7
Homozygotes (Variant)	3	4.6	16	16.2	20	15.3	3	4.9	3	4.6
Variant Allele Frequency	0.10		0.20		0.18		0.10		0.10	
Chi-Squared Test *p* Value	0.6993		0.9979		0.3446		0.6285		0.6993	

(*p* value < 0.05 indicates inconsistency with Hardy-Weinberg Equation).

Table14 – Exhibits the genotype frequencies for the five *UCH-L1* SNPs investigated for this sample set with the Hardy-Weinberg formula (equation) applied. Both the actual frequencies ('O' for 'Observed'), and the Hardy-Weinberg Equation theoretically 'Expected' ('E') frequencies are displayed. The variant allele frequency is also indicated, with the final row exhibiting the probability (*p* value) of consistency between the 'Observed' and 'Expected' figures, calculated through Chi-Squared statistical analysis.

3.1.3.3.4 Haplotype Frequency Diagrams



(For continuation of less common *UCH-L1* promoter haplotypes, please continue overleaf)

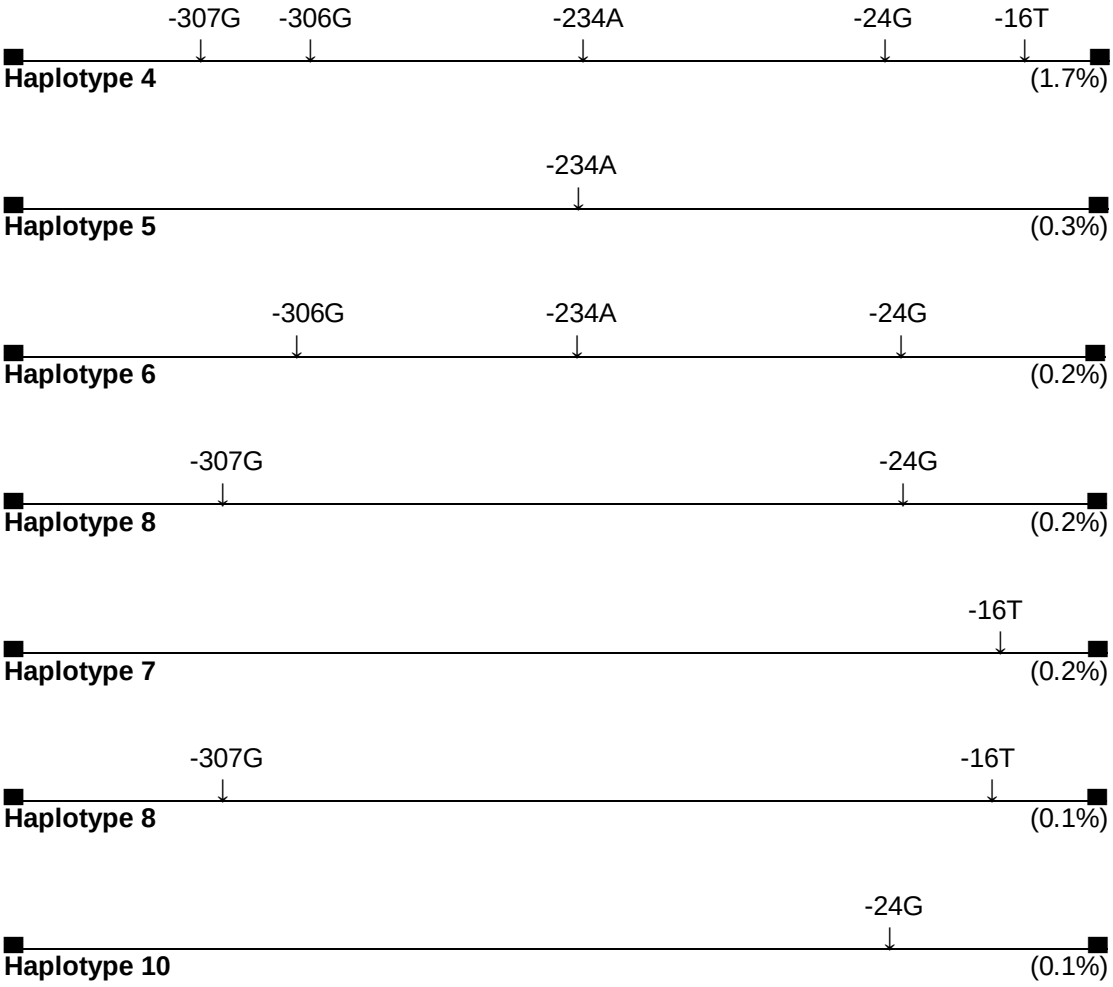


Figure 50 - The ten *UCH-L1* promoter haplotypes are indicated with reference to the distribution of the five promoter/ 5' UTR polymorphic alleles elucidated in this study (allelic positions not to scale). Their haplotypic frequency within the 960 chromosomes studied in this sample set is indicated in brackets.

3.1.3.3.5 Interpretation of x480 DNA Sample Set Genotyping Analysis

Through restriction fragment length polymorphism and supplemental allele frequency data, the frequencies of the five *UCH-L1* promoter polymorphic alleles were ascertained (Table 12), and once collated with respect to the genotypes exhibited by each of the 480 individuals (x480 DNA sample set) (Table 13), haplotype frequencies were able to be determined (Figure 50).

Ten haplotypes were elucidated; three displayed a frequency of greater than 5%, with seven others displaying a total frequency of 2.8% (Figure 50). With reference to the three haplotypes elucidated from the x64 DNA sample set (refer to section 3.1.3.2), a good correlation was observed with the three most common haplotypes uncovered from the x480 DNA sample set. The AAGAC promoter haplotype being the most common (60.1%), the AGAAC haplotype being the next frequent (15.4%), and the GAGGT haplotype exhibiting a frequency of 7.7%. This promoter haplotype order, with respect to frequency, was the same in both DNA sample sets investigated.

As previously discussed, this study's main focus was on potential changes in expression that common *UCH-L1* Caucasian promoter haplotypes could bring about on UCH-L1 levels within mammalian cells. Thus, only those haplotypes with a frequency of over 5% were deemed relevant to the next stage of this study, which was to insert the promoter haplotypes within suitable vectors, to allow expression studies within mammalian cell lines to be carried out.

N.B. – *UCH-L1* promoter haplotype nomenclature:-

AAGAC – *UCH-L1* reference haplotype.

AGAAC – Haplotype exhibiting –306G and –234A polymorphic alleles.

GAGGT – Haplotype exhibiting –307G, -24G and –16T polymorphic alleles.

3.1.4 Discussion of Mutation Screening

From the mutation screening carried out above, three common promoter haplotypes have emerged which govern the distribution of five novel SNPs within the Caucasian population.

As with all biallelic polymorphisms of Mendelian inheritance, it was important to ascertain whether the genotype distribution of the promoter/ 5'-UTR SNP alleles uncovered in this study (upon which the next stage of this study was to be based) were consistent with the Hardy-Weinberg principle – the qualifying law in population genetics to ensure a population is panmictic and in evolutionary stasis. The Hardy-Weinberg law was used to test the null hypothesis in both DNA sample sets (refer to sections 3.1.3.2.3 and 3.1.3.3.3), and both populations displayed a high level of consistency – thus ensuring that downstream analysis was sound.

What the mutation screening in this section also elucidates is that the previously cited S18Y (C54A) variant allele is not in linkage disequilibrium with any of the promoter variant alleles uncovered in this study. In a recent large, case-control study (*n*=3023) (Healy *et al.* (2006)), this previously reported polymorphism was seen to have no significant association with sporadic Parkinson's disease (previous studies had reported the S18Y polymorphism as having a protective effect against Parkinson's disease – please refer to section 1.3.2 for further discussion). This lends credence to the fact that this polymorphism will not be studied further in this investigation, and also sustains the objective to concentrate the focus of this study on the promoter haplotypes, the polymorphic variants of which, are not in haplotypic linkage with the S18Y variant allele.

Within this same study (Healy *et al.* (2006)), the authors also provide frequency data for an array of SNPs within the UCH-L1 gene (resequencing data from 128 control chromosomes), only one of which relates to the promoter polymorphisms elucidated in this study – designated C-16T in this study (rs9321, clone position (NT_006238) 960944) – they report a frequency of 10.0% (0.32 HW), which supports the allele frequency observed in this study within both control arrays of 9.7 and 8.1%. This SNP also has population diversity information within the NCBI database (Entrez SNP (dbSNP)); frequency from two populations

have been uploaded – one reports a 15% frequency of the variant allele in a Japanese control array (*n*=100), and the second reports a frequency of 4% for the polymorphic T allele in a mixed ethnicity control array (*n*=92) – though owing to both populations not being 100% Caucasian, it is difficult to draw any direct conclusion from this data. One other SNP that has been elucidated in this study within *UCH-L1*'s promoter region also has been designated on the NCBI database – A-307G: rs13129604 – but has no population diversity frequency data associated with it.

The five polymorphisms that this study elucidates within the promoter were analysed for their predicted effect on transcription factor binding. The 'TFSEARCH' program was utilised (at http://www.rwcp.or.jp/lab/pdappl/papia.html (Akiyama, Y. (1995))), which searches highly correlated sequence fragments versus 'TFMATRIX' transcription factor binding site profile database in 'TRANSFAC' databases developed at GBF – Braunschweigh, Germany, and then carries out a simple correlation calculation with required inputted sequence data – its use has been designated suitable in publications (Heinemeyer *et al*. (1998)).

Any significant alterations to predicted transcription factor binding, in contrast to the reference alleles (Wt alleles), that the TFSEARCH software uncovered for the five *UCH-L1* promoter SNPs elucidated in this study, are outlined in Table 15 below.

Owed to the predictive nature of the TFSEARCH software, it is difficult to draw any specific conclusions from the analysis carried out, although Table 15 does describe significant changes in predicted transcription factor binding (compared to that of the reference alleles) for three of the variant alleles – with at least one harboured on each of the two *UCH-L1* polymorphic promoter haplotypes. This supports the necessity to study these haplotypic sequence variants further in functional analyses, which were to be carried out in the next stage of this project.

Variant Allele(s)	**Change in Predicted Transcription Factor Binding from that of Reference Allele**	**Species** (if applicable)
-307G	No Predicted Change	-
-306G	Novel GATA-binding factor-1	*Mus musculus* (Mouse)
-234A	No Predicted Change	-
-24G/ -16T	Novel SRY site (Sex-Determining region Y gene product)	*Homo sapiens* (Human)
	Lost 2x CdxA sites	*Gallus gallus* (Chick)

Table 15 – Outlines the predicted changes in transcription factor binding that the five promoter *UCH-L1* SNP variant alleles realised from that of the reference alleles utilising the 'TFSEARCH' software (Akiyama, Y. (1995)) (at default settings, for Vertebrate classification). – The –24G and –16T polymorphic alleles were analysed together owed to their close proximity, and the fact that they reside on the same polymorphic haplotypes.

3.2 Results II – Functional Studies

The ultimate aim of this phase of the investigation was to incorporate the three common (those realised at a frequency above 5%) *UCH-L1* promoter haplotypes identified in section 3.1 into a suitable mammalian reporting vector, which would then allow any differences in transcriptional activity between the *UCH-L1* promoter haplotypes to be quantitatively measured within mammalian cells, including human cortical neuronal cells.

The mammalian reporter vector chosen for this purpose was the 'pGL3 Luciferase Reporter Vector (Promega) (Figure 28). The pGL3 luciferase reporter vector provides a basis for the quantitative analysis of *cis*-acting factors, i.e. promoters that regulate mammalian gene expression. The vector critically contains a modified coding region for firefly (*Photinus pyralis*) luciferase that has been optimised for monitoring transcriptional activity of inserted upstream promoters in transformed eukaryotic cells.

Once all three haplotypes had been incorporated into the pGL3 vector with 100% sequence fidelity and in the correct orientation (5' to 3' upstream of the luc+ gene), any differences in their transcriptional activity was investigated using a Dual-Luciferase Reporter Assay System (Promega – refer to section 2.4.4.1) upon the mammalian cells that had been transfected with the three pGL3/ *UCH-L1* promoter constructs.

The dual-luciferase reporter assays were carried out for all three pGL3/ *UCH-L1* promoter haplotypes (plus a control pGL3-*Eco* R1 vector with no (promoter) insert to normalise for baseline activity values) in four different mammalian cell types:-

A2058 – Human Skin Melanoma

MCF-7 – Human Breast Adenocarcinoma

ND-7 – Mouse Neuroblastoma/ Rat Basal Ganglia Neuron Hybrid

HCN-1A – Human Brain Cortical Neuron

The HCN-1A human cortical neuronal cell line was chosen as the end point to this, the 'Functional Studies' section of this investigation, as it represented most closely the cells in which *UCH-L1*'s promoter would be most active *in vivo*, i.e. the intranuclear protein and

transcription factor profiles would be relatively comparable. Though non-human, the ND-7 hybrid cell line also represented an excellent mammalian cell line to study interhaplotypic transcriptional variation within the *UCH-L1* promoter, as it had partial origins within the basal ganglia (rat) – a region of the brain which is associated with increased *UCH-L1* mRNA (Leroy *et al*. (1998)) and Parkinson's disease pathogenesis. Furthermore, its non-human origin would allow potential interspecies conclusions to be drawn.

The A2058 cells were chosen as the cells to undertake the initial luciferase experiments with, as not only were they adherent and of Caucasian ethnicity, as with the HCN-1A cell line, but there invasiveness would guarantee rapid and efficient growth. Any interhaplotypic variation uncovered in these non-neural cells would also allow inferences to be made into the causative proteins/ transcription factors being neural specific or indeed more ubiquitous in cells throughout the rest of the body. The MCF-7 cell line was chosen for similar reasons to the A2058 cell line, though its different origin did have the potential to allow expansion of any inter cell line inferences. Additionally, the MCF-7 mode of adherent growth was more similar (compared to A2058) to that of the neuronal cells.

3.2.1 Construction of pGL3/ *UCH-L1* Promoter Vectors

The remainder of this section will describe the strategies that were employed to incorporate the three *UCH-L1* promoter haplotypes into the pGL3 Luciferase Reporter Vector.

3.2.1.1 Initial Cloning Strategy – direct incorporation of haplotypes into pGL3

The initial strategy was to engineer – through PIRA-PCR (refer to section 2.2.3) – two different restriction sites, that were also present within pGL3's 'multiple cloning region' either side of the three *UCH-L1* promoter haplotypes sequences, to facilitate direct ligation into pGL3's cloning region in the correct orientation (after concordant endonuclease excision).

The restriction sites selected were *Kpn* I and *Bgl* II (Figure 51), which importantly were not realised within the pGL3 or the *UCH-L1* sequences (haplotypic as well as reference).

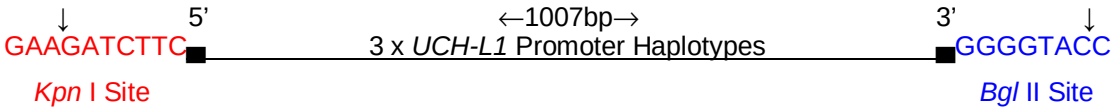


Figure 51 – Diagrammatic representation of required *UCH-L1* promoter sequences (↓ = specific 'cutting' site of restriction endonuclease).

Primer Design Considerations

Anti-sense Primer:-

- 5' end had to begin upstream of ATG translation start site, as only 5' UTR/ promoter sequences were required (ATG start codon was present in pGL3's *luc*+ gene).
- 3' end had to end downstream of the C-16T SNP to encompass its interhaplotypic variation.

Sense Primer:-

- 3' end had to encompass/ be upstream of all identified regulatory sequence elements (with reference to Mann *et al.* (1996)).

All three UCH-L1 promoter/ 5'UTR haplotypes were successfully amplified by PCR and the respective nucleotide sequence fidelities were confirmed through DNA sequencing. Kpn I/ Bgl II double restriction digest was then carried out on the three UCH-L1 PCR amplified haplotypes and the pGL3 vector, to allow ligation reactions to be employed to allow promoter insertion into the vector.

Interpretation

Through reasons that remain unclear, despite many modifications in time, temperature and quantities being made to each segment of the cloning strategy (post initial sequence confirmation), including *Kpn* I/ *Bgl* II restriction digestion, the ligation reaction and *E.coli* transformation and recovery, no pGL3/ *UCH-L1* promoter constructs were identified. This particular cloning strategy was thus suspended in search of a more successful approach.

3.2.1.2 Revised Cloning Strategy – pGL3 via pGEM-T Easy holding vector

3.2.1.2.1 Incorporation of Haplotypes into pGEM-T Easy

The basic outline of this first phase of the revised cloning strategy was to incorporate the individual *UCH-L1* promoter haplotypes into an initial transient holding vector - pGEM-T Easy (Promega) (Figure 27) – before incorporating the promoter sequence into pGL3. The pGEM-T Easy system was chosen as PCR products could be inserted directly into the vector without any post-modification (refer to section 2.3.1.1). It was also chosen as it contained several restriction sites flanking the PCR insertion (cloning) site which were not realised within the *UCH-L1* promoter haplotype sequences (refer to appendices for 'Lasergene' output of non-cutting endonucleases), which were then able to be used to remove the *UCH-L1* promoter sequences with a single restriction digest protocol.

As for the 'Initial Cloning Strategy' (refer to section 3.2.1.1), the same region of the UCH-L1 promoter was amplified by PCR (the primers were only modified to remove the restriction sites and further reduce any DNA secondary structure). Once amplified through PCR and the respective nucleotide sequence fidelity confirmed, all three *UCH-L1* promoter haplotypes were inserted into pGEM-T Easy through specific pGEM-T Easy ligation reactions (see section 2.3.8). Their incorporation was then confirmed through *E.coli* subcloning protocol and subsequent *Eco* RI restriction digest of the relevant plasmid DNA to display the pGEM-T Easy vector and *UCH-L1* promoter inserts.

Results

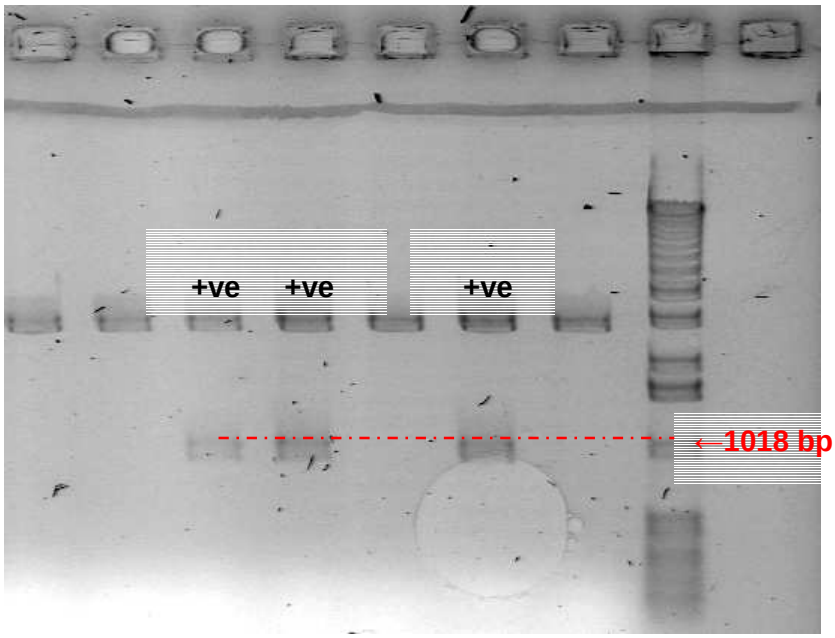


Figure 52 – Agarose get identifying three pGEM-T Easy constructs harbouring the *UCH-L1* promoter inserts (+ve). The upper bands are the linear pGEM-T Easy vectors. The lower bands are the *Eco* RI 'cleaved' promoter inserts. The 1018 bp band of the DNA ladder confirmed the inserts were of the required size (~1 Kb) (denoted and marked in red).

*(Sequence fidelity of all three of the *UCH-L1* promoter inserts were confirmed through DNA sequencing).

Interpretation

Now that all the reference AAGAC, and the two polymorphic – AGAAC & GAGGT – haplotypes had been successfully incorporated into pGEM-T Easy, all three common *UCH-L1* promoter haplotypes could now be excised and incorporated into pGL3.

3.2.1.2.2 Transfer from pGEM-T Easy into pGL3-*Eco* RI

Confirmation of the *UCH-L1* promoter sequences in pGEM-T Easy allowed phase two of the revised cloning strategy to be undertaken. The basic outline of this phase was to transfer the *UCH-L1* promoter inserts from pGEM-T Easy to a modified pGL3 vector – pGL3-Eco RI (refer to section 2.3.1.3 for an outline of its construction) (Figure 29).

The pGL3-*Eco* RI vector was specifically chosen for phase two of this cloning strategy owing its 'custom' *Eco* RI site (from which its name derives) that had been engineered into its 'multiple cloning site' (Figure 30). The *Eco* RI restriction site allowed the potential for the three *UCH-L1* inserts to be transferred from pGEM-T Easy into pGL3-*Eco* RI through the use of the *Eco* RI and *Spe* I endonucleases in a sequential fashion (Figure 53)

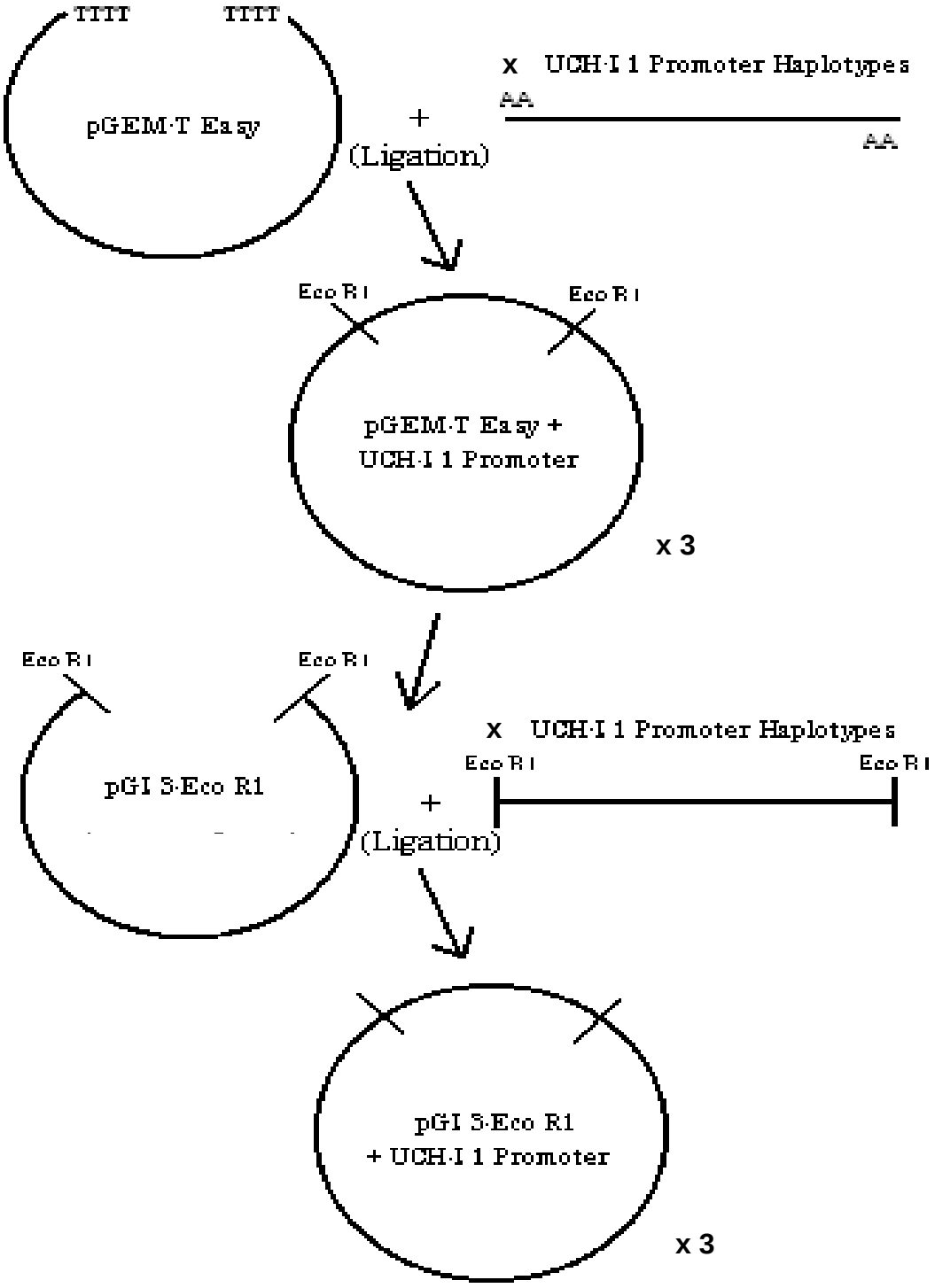


Figure 53 – Basic outline of the ultimately successful 'Revised Cloning Strategy'.

Orientation of the *UCH-L1* promoter inserts in pGEM-T Easy were of critical importance. As seen in Figure 27 (section 2.3.1.1), pGEM-T Easy harbours *Eco* RI restriction sites either side of its 'PCR insertion site', and on the 'SP6 branch' of its 'multiple cloning region' the *Spe* I restriction site is located closer to the 'PCR insertion site' than *Eco* RI. This then allowed, if done sequentially (i.e. not at the same time), the *UCH-L1* promoter inserts in pGEM-T Easy to be extracted with 'sticky ends' for *Eco* RI at one end and *Spe* I at the other. Though for this to be done correctly *Spe* I would have to be utilised first, then followed by *Eco* RI.

pGL3-*Eco* RI contains these two restriction sites within its modified 'multiple cloning region', however *Spe* I is 5' and *Eco* RI is 3' in respect to the luc+ gene (the correct control of which is fundamental to the next stage of this investigation). This meant that pGEM-T Easy/ *UCH-L1* inserts had to be identified in the 5' to 3' orientation in respect to the *lac* operator, so that they could then be transferred to the pGL3-*Eco* RI vector in the correct orientation. (Referring back to the sequencing electropherograms for the pGEM-T Easy/ *UCH-L1* promoter constructs obtained in section 3.2.1.2.1, one construct for each haplotype that incorporated the *UCH-L1* insert in the correct orientation (*lac*Z 5' to 3') was identified).

Spe I followed by *Eco* RI restriction endonuclease digestion of these pGEM-T Easy constructs and of the pGL3/ Eco RI vector was then carried out, and three pGL3-Eco RI/ *UCH-L1* promoter constructs were then formed through three separate ligation reactions using the purified digestion products, followed by transformations and selective ampicillin agar plating as normal.

Results

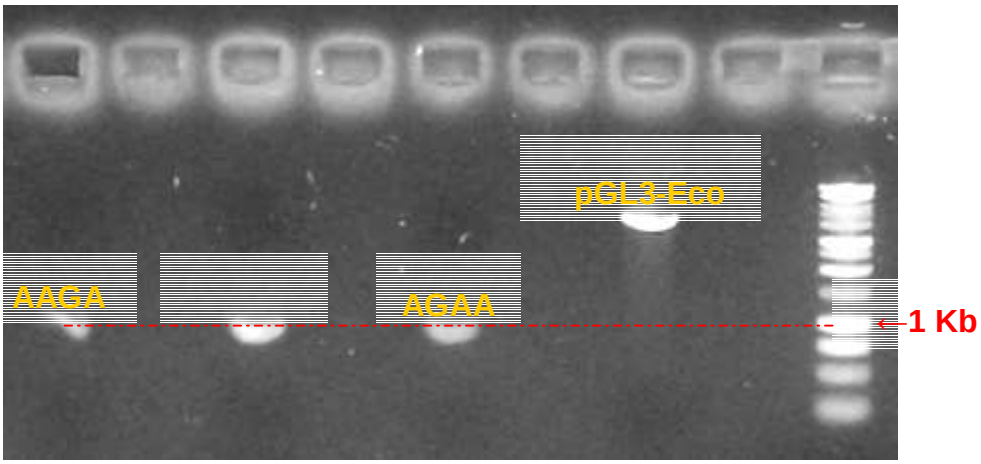


Figure 54 – 'Modified TAE agarose gel' showing the pGL3-Eco RI vector and the three *UCH-L1* promoter haplotype inserts after *Spe* I → *Eco* RI endonuclease digestion, in preparation for the respective ligation reactions. The 1 Kb band of the DNA ladder (refer to section 2.1.2.4) confirmed the inserts were of the required size (~1 Kb) (denoted and marked in red).

This cloning route finally proved successful with good colony growth being realised for each *UCH-L1* haplotype. Colonies were then grown up and minipreped, with *Spe* I/ *Eco* RI digest revealing all three haplotypes required (Figure 55).

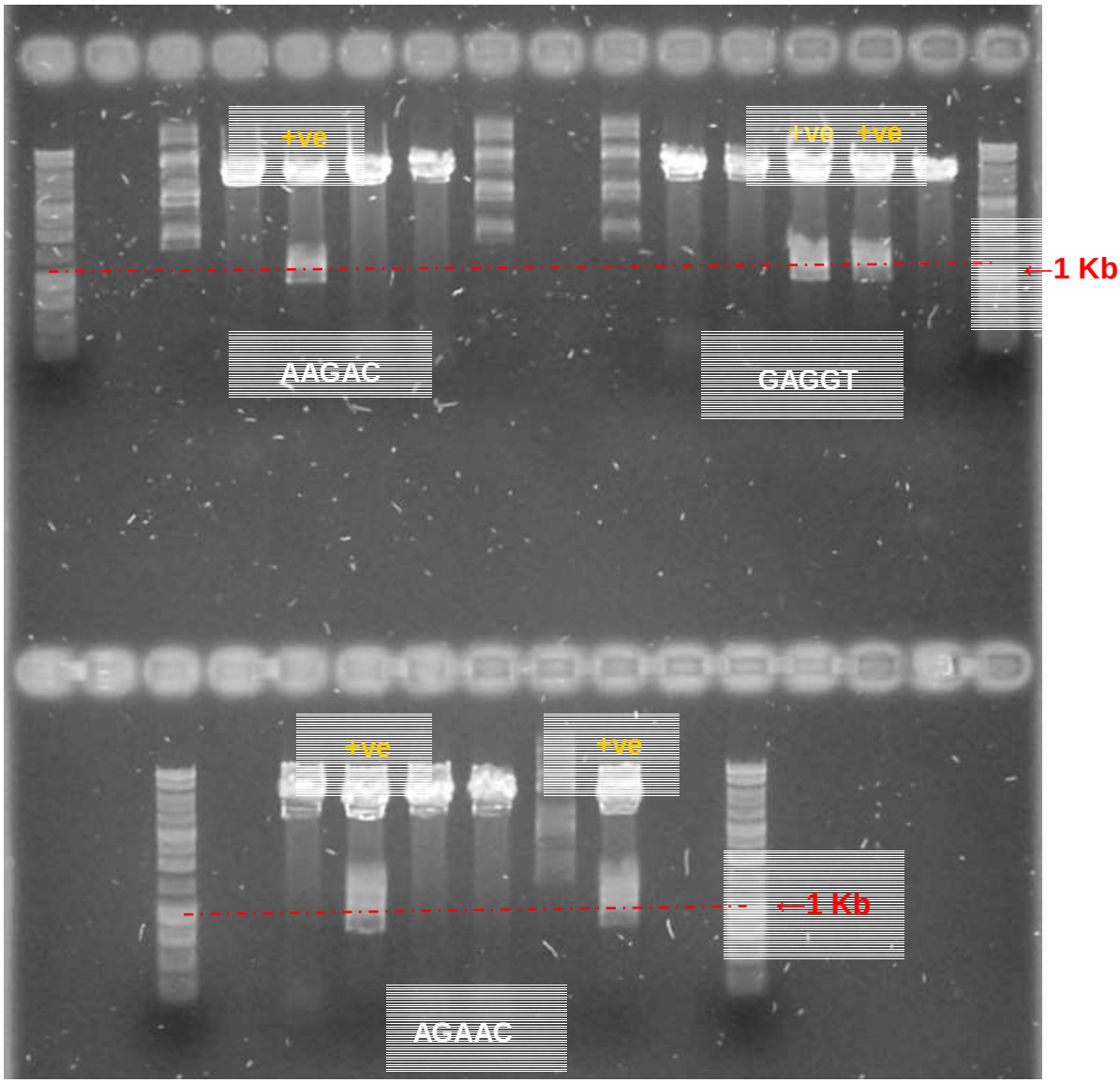


Figure 55 - Agarose gel identifying three pGL3-Eco RI constructs harbouring the *UCH-L1* promoter inserts (+ve) for all three promoter haplotypes (denoted in white). The upper bands are the linear pGL3-Eco RI vectors. The lower bands are the *Spe* I/ *Eco* RI 'cleaved' promoter inserts. The 1 Kb band of the DNA ladder confirmed the inserts were of the required size (~1 Kb) (denoted and marked in red).

- One insert for each *UCH-L1* haplotype represented within the pGL3-Eco RI/ *UCH-L1* promoter constructs elucidated were DNA sequenced to confirm their respective 100% sequence fidelity (electropherograms not shown).

Interpretation

All three *UCH-L1* promoter haplotypes had been successfully incorporated into the modified pGL3-*Eco* RI vector, which allowed the next stage of this study to be initialised; in which the transcriptional activities of the *UCH-L1* promoters could now be investigated within mammalian cell lines to elucidate any potential differences between the haplotypes.

3.2.2 Transcriptional Activity of the *UCH-L1* Promoter Haplotypes

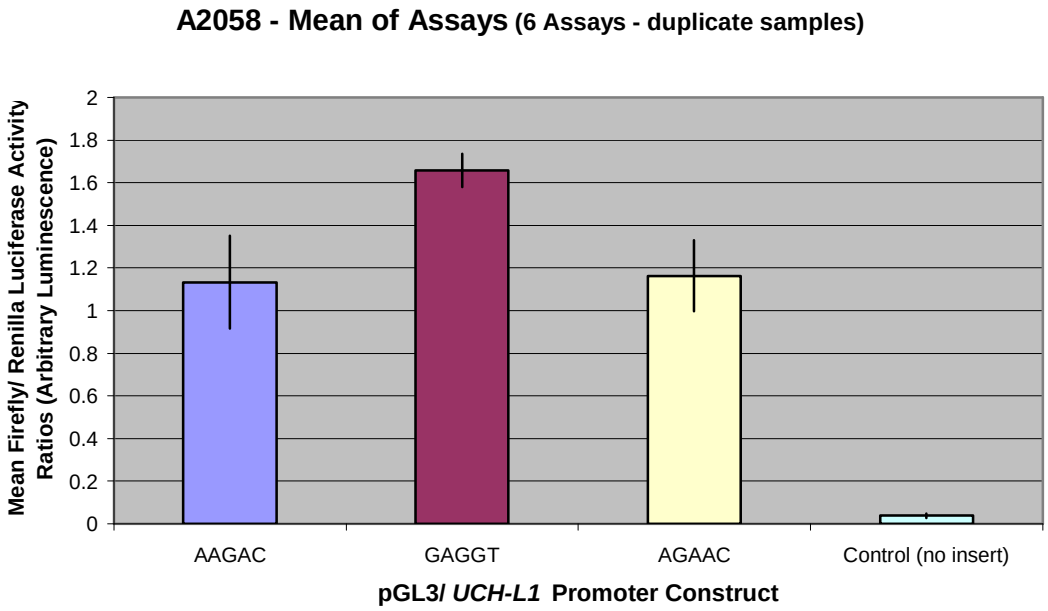
- Each respective cell line was prepared individually for the dual-luciferase assays as described in the relevant subsections (for each cell line). The general preparation for each cell line studied involved all of the following steps in order: culture propagation, subculturing, transfection and lysis - before the dual-luciferase assay could be carried out.

As described in section 2.4.4.1, each dual-luciferase reaction carried out on each pGL3/ *UCH-L1* promoter haplotype replicate in a particular cell line realised two measurements: one for the firefly luciferase luminescence (promoter transcriptional activity dependent), followed by a *Renilla* luciferase luminescence measurement (control reporter). The initial 'experimental' measurement was then divided by the control measurement to normalise the transcriptional luciferase activity relative to the 'internal' control's luciferase activity, to realise a ratio value that effectively eliminates any potential experimental variability, i.e. differences in cell viability, transfection efficacy, etc.

In all the human cell lines tested the pGL3/ *UCH-L1* GAGGT promoter haplotype construct showed a statistically significant increase in mean firefly/ renilla luciferase activity ratios compared to the pGL3/ *UCH-L1* AAGAC and AGAAC promoter haplotype constructs. No significant variation in activity ratios was realised between the pGL3/ *UCH-L1* promoter constructs in the mouse neuroblastoma/ rat basal ganglia neuron hybrid cell line.

3.2.2.1 A2058 Dual-Luciferase Reporter Assays

The mean ratios from the six dual-luciferase assays carried out in duplicate on the A2058 cell line above were averaged. The values obtained are displayed in graph and numerical form below (together with the standard error of mean(s) (S.E.M)):-



Bar Graph Data:-

Mean of Assays	AAGAC	GAGGT	AGAAC	(no insert)
Mean Ratio Value	1.133	1.658	1.163	0.037
S.E.M	0.218	0.079	0.168	0.009

Probabilities of Differences Between the pGL3/ *UCH-L1* promoter haplotype Constructs:-

Between **AAGAC**, **GAGGT** and **AGAAC** Haplotypes	***p* = 0.000602**
Between **AAGAC** and **GAGGT** Haplotypes	***p* = 0.000757**
Between **GAGGT** and **AGAAC** Haplotypes	***p* = 0.000469**
Between **AAGAC** and **AGAAC** Haplotypes	*p* = 0.839171

Figure 56 – A bar graph displaying the mean firefly/ renilla luciferase activity ratios realised for the three pGL3/ *UCH-L1* promoter constructs and a control pGL3 vector with no insert, in the A2058 human skin melanoma cell line. Concordant graph data is also detailed below the

graph, as are the respective *p*-values – which were calculated by way of 'Two Way Analysis of Variance' (ANOVA) of all luciferase data.

- The mean ratio, standard error of mean and *p*-values were derived form all the data from 6 separate assays (in which measurements were taken in duplicate).
- The x-axis indicates the mean firefly/ renilla luciferase activity ratios in an arbitrary luminescence scale. The y-axis indicates the pGL3 construct.

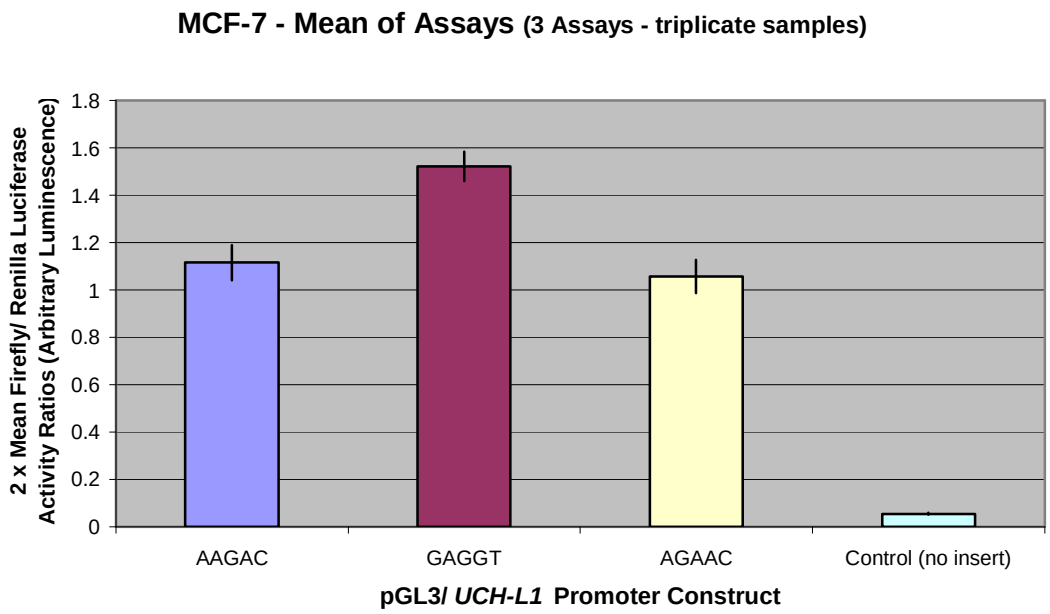
- The standard error of means are graphically indicated as (+/-) error lines at the top of each separate bar.

- Prior to cell lysis and dual-luciferase reporter assay methodology – 1-3x10^5 cells were grown to 50-80% confluence at 37℃ (5% $C0_2$) in full growth medium; 1µg of vector together with 'GeneJuice' transfection agent were then added, and cells were incubated for 24 hours at 37℃ (5% $C0_2$).

N.B. – Raw data from each of the dual-luciferase assays carried out for the A2058 cell line is detailed in the relevant section of the appendix.

3.2.2.2 MCF-7 Dual-Luciferase Reporter Assays

The mean ratios from the three dual-luciferase assays carried out in triplicate on the MCF-7 cell line above were averaged. The values obtained are displayed in graph and numerical form below (together with the standard error of mean(s) (S.E.M)):-



Bar Graph Data:-

Mean of Assays	AAGAC	GAGGT	AGAAC	(no insert)
2x Mean Ratio Value	1.116	1.523	1.056	0.054
2 x S.E.M	0.074	0.062	0.071	0.007

Probabilities of Differences Between the pGL3/ *UCH-L1* promoter haplotype Constructs:-

Between **AAGAC**, **GAGGT** and **AGAAC** Haplotypes	***p* = 0.000000307**
Between **AAGAC** and **GAGGT** Haplotypes	***p* = 0.0000152**
Between **GAGGT** and **AGAAC** Haplotypes	***p* = 0.00000321**
Between **AAGAC** and **AGAAC** Haplotypes	*p* = 0.341209

Figure 57 – A bar graph displaying the mean firefly/ renilla luciferase activity ratios realised for the three pGL3/ *UCH-L1* promoter constructs and a control pGL3 vector with no insert, in the MCF-7 human breast adenocarcinoma cell line. Concordant graph data is also detailed

below the graph, as are the respective *p*-values – which were calculated by way of 'Two Way Analysis of Variance' (ANOVA) of all luciferase data.

- The mean ratio, standard error of mean and *p*-values were derived form all the data from 3 separate assays (in which measurements were taken in triplicate).
- The x-axis indicates 2 x the mean firefly/ renilla luciferase activity ratios in an arbitrary luminescence scale. The y-axis indicates the pGL3 construct.

-The standard error of means are graphically indicated as (+/-) error lines at the top of each separate bar.

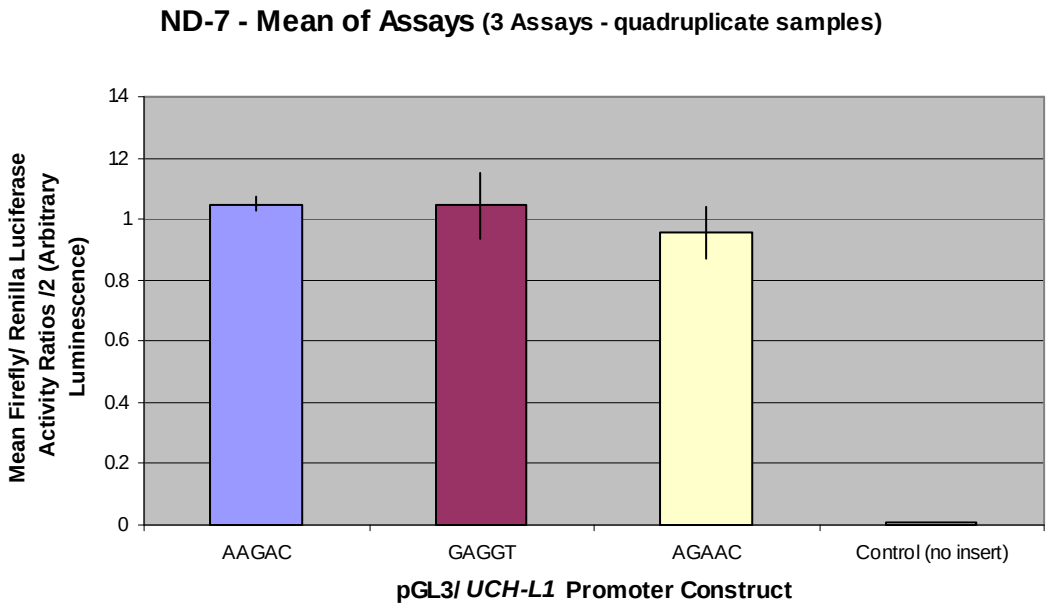
- Arbitrary luminescence values for each pGL3 construct were multiplied by a factor of 2 for clearer graph comparison purposes between the cell lines.

- Prior to cell lysis and dual-luciferase reporter assay methodology – 1-3x10^5 cells were grown to 50-80% confluence at 37℃ (5% $C0_2$) in full growth medium; 1µg of vector together with 'GeneJuice' transfection agent were then added, and cells were incubated for 24 hours at 37℃ (5% $C0_2$).

N.B. – Raw data from each of the dual-luciferase assays carried out for the MCF-7 cell line is detailed in the relevant section of the appendix.

3.2.2.3 ND-7 Dual-Luciferase Reporter Assays

The mean ratios from the three dual-luciferase assays carried out in quadruplicate on the ND-7 cell line above were averaged. The values obtained are displayed in graph and numerical form below (together with the standard error of mean(s) (S.E.M)):-



Bar Graph Data:-

Mean of Assays	AAGAC	GAGGT	AGAAC	(no insert)
Mean Ratio Value /2	1.049	1.044	0.957	0.009
S.E.M /2	0.022	0.106	0.085	-

Probabilities of Differences Between the pGL3/ *UCH-L1* promoter haplotype Constructs:-

Between **AAGAC**, **GAGGT** and **AGAAC** Haplotypes	$p = 0.608964$
Between **AAGAC** and **GAGGT** Haplotypes	$p = 0.956315$
Between **GAGGT** and **AGAAC** Haplotypes	$p = 0.497216$
Between **AAGAC** and **AGAAC** Haplotypes	$p = 0.244074$

Figure 58 – A bar graph displaying the mean firefly/ renilla luciferase activity ratios realised for the three pGL3/ *UCH-L1* promoter constructs and a control pGL3 vector with no insert, in the ND-7 mouse neuroblastoma/ rat basal ganglia neuron hybrid cell line. Concordant graph

data is also detailed below the graph, as are the respective *p*-values – which were calculated by way of 'Two Way Analysis of Variance' (ANOVA) of all luciferase data.

- The mean ratio, standard error of mean and *p*-values were derived form all the data from 3 separate assays (in which measurements were taken in quadruplicate).
- The x-axis indicates the mean firefly/ renilla luciferase activity ratios divided by 2 in an arbitrary luminescence scale. The y-axis indicates the pGL3 construct.

-The standard error of means are graphically indicated as (+/-) error lines at the top of each separate bar.

- Arbitrary luminescence values for each pGL3 construct were divided by a factor of 2 for clearer graph comparison purposes between the cell lines.

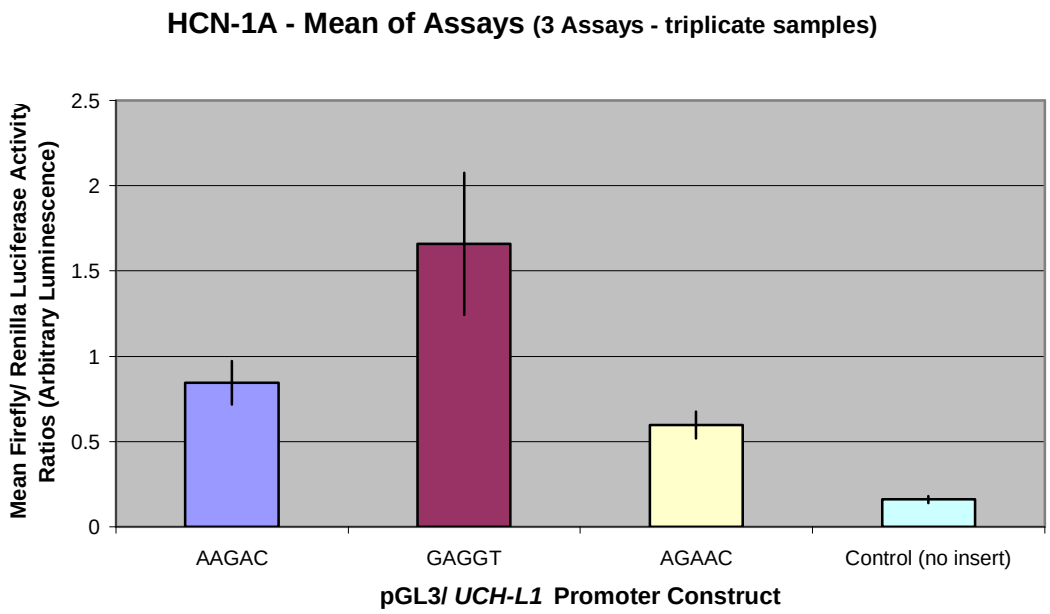
- Prior to cell lysis and dual-luciferase reporter assay methodology – $1\text{-}3x10^5$ cells were grown to 50-80% confluence at 37℃ (5% $C0_2$) in full growth medium; 1µg of vector together with 'GeneJuice' transfection agent were then added, and cells were incubated for 24 hours at 37℃ (5% $C0_2$).

N.B. – Raw data from each of the dual-luciferase assays carried out for the ND-7 cell line is detailed in the relevant section of the appendix.

.

3.2.2.4 HCN-1A Dual-Luciferase Reporter Assays

The mean ratios from the three dual-luciferase assays carried out in triplicate on the HCN-1A cell line above were averaged. The values obtained are displayed in graph and numerical form below (together with the standard error of mean(s) (S.E.M)):-



Bar Graph Data:-

Mean of Assays	AAGAC	GAGGT	AGAAC	(no insert)
Mean Ratio Value	0.846	1.660	0.597	0.160
S.E.M	0.127	0.418	0.079	0.021

Probabilities of Differences Between the pGL3/ *UCH-L1* promoter haplotype Constructs:-

Between **AAGAC**, **GAGGT** and **AGAAC** Haplotypes	***p* = 0.005316**
Between **AAGAC** and **GAGGT** Haplotypes	***p* = 0.040501**
Between **GAGGT** and **AGAAC** Haplotypes	***p* = 0.009974**
Between **AAGAC** and **AGAAC** Haplotypes	*p* = 0.059379

Figure 59 – A bar graph displaying the mean firefly/ renilla luciferase activity ratios realised for the three pGL3/ *UCH-L1* promoter constructs and a control pGL3 vector with no insert, in the HCN-1A human brain cortical neuron cell line. Concordant graph data is also detailed

below the graph, as are the respective *p*-values – which were calculated by way of 'Two Way Analysis of Variance' (ANOVA) of all luciferase data.

- The mean ratio, standard error of mean and *p*-values were derived form all the data from 3 separate assays (in which measurements were taken in triplicate).
- The x-axis indicates the mean firefly/ renilla luciferase activity ratios in an arbitrary luminescence scale. The y-axis indicates the pGL3 construct.

-The standard error of means are graphically indicated as (+/-) error lines at the top of each separate bar.

- Prior to cell lysis and dual-luciferase reporter assay methodology – 1-3x10^5 cells were grown to 50-80% confluence at 37℃ (5% $C0_2$) in full growth medium; 1μg of vector together with 'GeneJuice' transfection agent were then added, and cells were incubated for 24 hours at 37℃ (5% $C0_2$).

N.B. – Raw data from each of the dual-luciferase assays carried out for the HCN-1A cell line is detailed in the relevant section of the appendix.

3.2.2.5 Interpretation of Luciferase Reporter Assays

From the four dual-luciferase reporter assays carried out, it can be seen that in ND-7 cells (mouse neuroblastoma/ rat basal ganglia neuron hybrid) no statistically significant variation in mean firefly/ renilla luciferase activity ratios between the promoter haplotypes was realised (Figure 58). However, in all three of the human call lines investigated, the GAGGT promoter haplotype realised a significant increase in mean luciferase activity ratios (Figures 56, 57 & 59), compared with the other two *UCH-L1* haplotypes.

The increase in mean GAGGT luciferase activity ratios varied between cell types – a 44.4% increase was realised in the A2058 (Human Skin Melanoma) cells, whilst in the MCF-7 (Human Breast Carcinoma) cells a 40.2% increase was exhibited, and the HCN-1A (Human Brain Cortical Neuron) cells displayed a 129.9% increase for the GAGGT promoter haplotype – relative to the average of the mean activity ratios obtained for the AAGAC and AGAAC haplotype constructs.

3.2.3 Discussion of Functional Studies

The dual-luciferase reporter assays carried out in section 3.2.2 have thence elucidated a significant increase in the transcriptional activity of GAGGT polymorphic *UCH-L1* promoter in human cells; Figure 60 illustrates these increases relative to the mean of the other two haplotypes in all the cell lines tested.

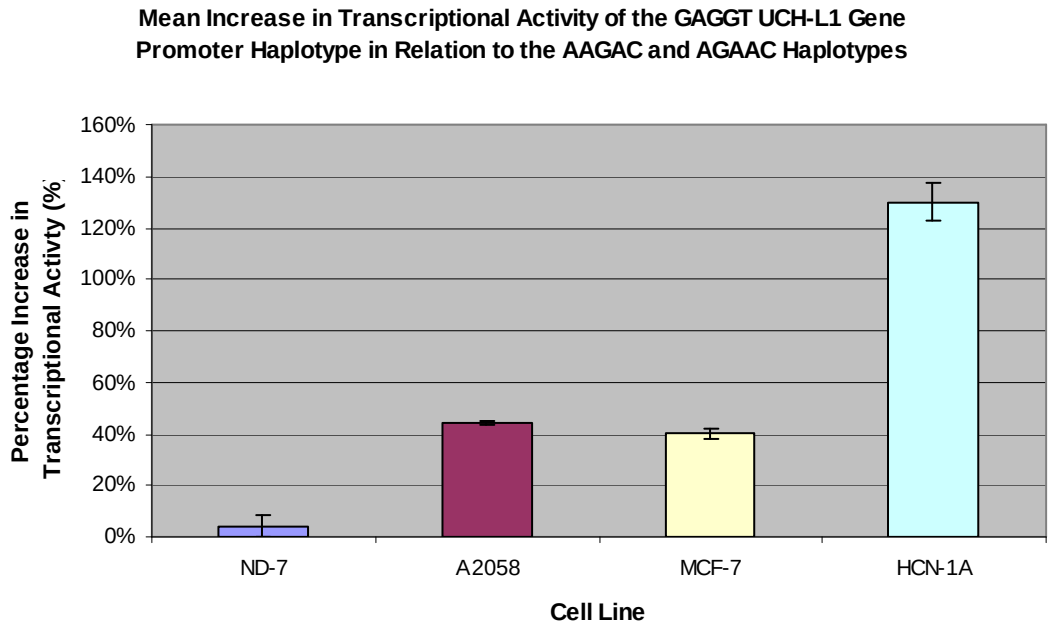


Figure 60 – A bar graph displaying the mean increase in firefly/ renilla luciferase mean activity ratios (which is directly related to transcriptional activity of promoter insert, of the GAGGT UCH-L1 gene promoter haplotype, in relation to the AAGAC (reference homozygote) and AGAAC gene promoter haplotypes, in the four mammalian cell lines assayed.

These results suggest that the –16T, -24G and –307G polymorphic alleles significantly effect the *UCH-L1* promoter's transcriptional activity in human cells *in vitro*, especially in neuronal cells (please refer to section 4 for further discussion with regards to the differences in transcriptional activity elucidated between the promoter haplotypes and cell lines tested).

To support the use of the Dual-Luciferase Reporter Assay (Promega), and thence any downstream conclusions/ inferences made, a recent functional study on the parkin gene promoter (UCS related neuron-specific gene – refer to section 1.4.2) with reference to

sporadic Parkinson's disease, which also employed the dual-luciferase system, will now be discussed (West *et al.* (2002)).

In this study, the group investigated two SNPs within the *parkin* core promoter and assayed their respective promoter activities using the dual luciferase system. The results elucidated one allele (-258G) that reduced luciferase activity by the order of 25% in human neuroblastoma cells, compared to that of the reference allele (-258T). The group then subsequently indicate, in a large population-based series of cases (*n*=319) and controls (*n*=196), that this polymorphism was associated with sporadic Parkinson's disease. Further supporting functional work (EMSA) showed that the polymorphic G allele did not bind protein from the human substantia nigra as well as the reference T allele, indicating a potential transcription factor binding deficiency.

This study has shown that interhaplotypic variations in the transcriptional activity of promoters, elucidated through the use of the dual-luciferase reporter assay, can certainly help uncover polymorphic alleles that are associated with disease. With particular applicability to *UCH-L1*, the target gene in this work was also a neuron-specific enzyme intricately involved in the UCS, and was found to have association with a neurodegenerative disease (Parkinson's disease – lending considerable credence to the experimental rationale in this investigation.

The next stage in this investigation will attempt to uncover any potential neurodegenerative disease association of the *UCH-L1* promoter SNP alleles in an elderly population with associated cognitive function data.

3.3 Results III – Population Studies

The overall aim of this final result's section was to identify if any of the reference or polymorphic SNP alleles elucidated in this study were associated with any variation in cognitive function that could be attributed to a neurodegenerative phenotype.

The x480 MRC DNA sample set genotyped for the five *UCH-L1* promoter/ 5'UTR SNPs that were elucidated in this study, had supplemental cognitive function data on the respective individuals whose DNA it included. This allowed the collated genotyping results for each of the SNPs to be correlated with the cognitive function data held for each individual, to identify – through simple statistical analysis – if any variation in cognitive function was associated.

The cohort comprised of 480 unrelated Caucasian men and women aged between 66-75 years old, and the concordant cognitive function tests were administered in the field through interview. Four cognitive function tests were administered to each individual (refer to section 2.5.2 for further details on each of the specific tests), though two tests which were of particular/ increased value to this investigation were the AH4 cognitive function test which measured fluid intelligence - which declines with age - and the Mill Hill cognitive function test which measured 'crystallised intelligence' - which importantly does not decline with age; as a simple regression of the AH4 scores onto those of the Mill Hill test would allow an indication of cognitive decline to be ascertained.

Alzheimer's disease, Huntington's disease and dementia with Lewy bodies have all been associated with *UCH-L1* SNP alleles (section 1.3) and all realise cognitive decline as a main disabling feature (Tang-Wai *et al*. (2005), Chua & Chiu (2005), Bryne (2005)). Parkinson's disease has been most strongly associated with *UCH-L1* (section 1.3), and although not regarded as a major clinical symptom, cognitive decline/ dysfunction in Parkinson's disease patients is high, and has recently received much more attention in the field as a feature of the disease (Pai & Chan (2001)). Thence, any variation uncovered in cognitive decline that is associated with any of the *UCH-L1* promoter SNPs elucidated in this study, must be seen as a potential genetic risk factor in neurodegeneration, especially when

the very high risk age range (66-75 years old) of the cohort is borne in mind, in relation to age of onset in the vast majority of neurodegenerative disease cases.

The cognitive function data comprises of scores attained by each individual in the four separate cognitive function tests, plus the AH4/ Mill Hill regression analysis. For each test/ analysis in cognitive function, the mean test scores for each genotype of a particular allele (i.e. reference homozygote, heterozygote and variant homozygote) was statistically analysed by two-way Analysis of Variance (ANOVA) to uncover any significant differences between them, which would indicate associated cognitive function variation. If any nominally significant differences were uncovered in any of the tests through ANOVA, a post-hoc test that would be applied is a Bonferroni correction (refer to section 2.5.2.6). Furthermore, confounders – such as smoking for example – would not be taken into account in the first instance; if nominally significant/ significant results were elucidated from the initial analysis, further multi-variate tests would be applied to take such confounders into consideration (age was not a confounder in this analysis, as it was taken into account by way of the AH4/ Mill Hill regression analysis described above).

Statistical analysis revealed no significant correlative differences between the cognitive function mean test scores and any of the *UCH-L1* allele genotypes.

Statistical analysis in this section was carried out by Dr Catherine Gale (MRC Epidemiological Centre, Southampton General Hospital).

Statistical Analysis of AH4 Cognitive Function Test Scores from x480 DNA Array

SNP	Genotype	Mean Score	S.D.	*n*	*p*
-16	TT	22.33	1.53	3	0.9966
	CT	21.89	9.69	93	
	CC	21.88	9.62	347	
-24	GG	22.33	1.53	3	0.9257
	AG	21.52	9.79	92	
	AA	21.96	9.60	339	
-234	AA	21.60	10.45	20	0.4943
	GA	22.72	8.89	148	
	GG	21.57	9.95	276	
-306	GG	21.31	10.87	16	0.6727
	AG	22.40	8.87	124	
	AA	21.49	9.76	258	
-307	GG	22.33	1.53	3	0.9198
	AG	21.61	9.84	93	
	AA	22.07	9.65	347	

Table 16 – Table displaying a summary of the statistical analysis (from x480 DNA array) for the AH4 cognitive function test scores in relation to each *UCH-L1* allele investigated in this study, and their respective genotypes.

- *UCH-L1* alleles are denoted in the first column by the nucleotide (SNP) position in relation to the gene's translation start site. For each allele, the reference homozygote, heterozygote and variant homozygote Genotypes are indicated in the second column, with the respective mean AH4 cognitive function test score (Mean Score), standard deviations (S.D.) there in, and the frequency (*n*) of each genotype displayed in a row through columns three, four and five in the table. The final column in the table elucidates the probability (*p*) that any of the respective genotypes for that allele correlate with a significant difference in AH4 cognitive function test scores.

Statistical Analysis of Mill Hill Cognitive Function Test Scores from x480 DNA Array

SNP	Genotype	Mean Score	S.D.	*n*	*p*
-16	TT	18.00	4.00	3	0.6158
	CT	18.12	4.67	93	
	CC	18.69	5.16	350	
-24	GG	18.00	4.00	3	0.4084
	AG	17.95	5.03	92	
	AA	18.74	5.14	341	
-234	AA	19.30	5.69	20	0.1360
	GA	19.13	4.58	150	
	GG	18.17	5.27	277	
-306	GG	18.75	6.26	16	0.2445
	AG	19.16	4.68	125	
	AA	18.22	5.31	259	
-307	GG	18.00	4.00	9	0.3928
	AG	17.92	4.98	93	
	AA	18.73	5.14	350	

Table 17 – Table displaying a summary of the statistical analysis (from x480 DNA array) for the Mill Hill cognitive function test scores in relation to each *UCH-L1* allele investigated in this study, and their respective genotypes.

- *UCH-L1* alleles are denoted in the first column by the nucleotide (SNP) position in relation to the gene's translation start site. For each allele, the reference homozygote, heterozygote and variant homozygote Genotypes are indicated in the second column, with the respective mean Mill Hill cognitive function test score (Mean Score), standard deviations (S.D.) there in, and the frequency (*n*) of each genotype displayed in a row through columns three, four and five in the table. The final column in the table elucidates the probability (*p*) that any of the respective genotypes for that allele correlate with a significant difference in Mill Hill cognitive function test scores.

Statistical Analysis of Scores from Immediate Recall Section of Wechsler Logical Memory Cognitive Function Test in x480 DNA Array

SNP	Genotype	Mean Score	S.D.	*n*	*p*
-16	TT	16.67	9.50	3	0.3852
	CT	22.31	6.83	89	
	CC	21.69	7.61	350	
-24	GG	16.67	9.50	3	0.4112
	AG	22.24	7.34	88	
	AA	21.69	7.50	341	
-234	AA	22.95	9.13	20	0.3845
	GA	22.34	6.83	149	
	GG	21.42	7.73	274	
-306	GG	23.25	9.13	16	0.7236
	AG	21.74	6.74	124	
	AA	21.70	7.76	256	
-307	GG	16.67	9.50	3	0.4681
	AG	22.09	7.07	89	
	AA	21.80	7.63	350	

Table 18 – Table displaying a summary of the statistical analysis (from x480 DNA array) for the immediate recall section of Wechsler logical memory cognitive function test scores in relation to each *UCH-L1* allele investigated in this study, and their respective genotypes.

- *UCH-L1* alleles are denoted in the first column by the nucleotide (SNP) position in relation to the gene's translation start site. For each allele, the reference homozygote, heterozygote and variant homozygote Genotypes are indicated in the second column, with the respective mean cognitive function test score (Mean Score), standard deviations (S.D.) there in, and the frequency (*n*) of each genotype displayed in a row through columns three, four and five in the table. The final column in the table elucidates the probability (*p*) that any of the respective genotypes for that allele correlate with a significant difference in cognitive function test scores.

Statistical Analysis of Scores form Delayed Recall Section of Wechsler Logical Memory Cognitive Function Test in x480 DNA Array

SNP	Genotype	Mean Score	S.D.	*n*	*p*
-16	TT	14.67	6.66	3	0.8530
	CT	17.19	7.20	91	
	CC	17.12	7.73	350	
-24	GG	14.67	6.66	3	0.8443
	AG	17.23	7.56	90	
	AA	17.09	7.58	341	
-234	AA	18.70	9.27	20	0.0717
	GA	18.11	7.10	149	
	GG	16.49	7.73	276	
-306	GG	18.75	8.99	16	0.4742
	AG	17.37	6.89	124	
	AA	16.72	7.78	258	
-307	GG	14.67	6.66	3	0.8388
	AG	17.03	7.39	91	
	AA	17.20	7.73	350	

Table 19 – Table displaying a summary of the statistical analysis (from x480 DNA array) for the delayed recall section of Wechsler logical memory cognitive function test scores in relation to each *UCH-L1* allele investigated in this study, and their respective genotypes.

- *UCH-L1* alleles are denoted in the first column by the nucleotide (SNP) position in relation to the gene's translation start site. For each allele, the reference homozygote, heterozygote and variant homozygote Genotypes are indicated in the second column, with the respective mean cognitive function test score (Mean Score), standard deviations (S.D.) there in, and the frequency (*n*) of each genotype displayed in a row through columns three, four and five in the table. The final column in the table elucidates the probability (*p*) that any of the respective genotypes for that allele correlate with a significant difference in cognitive function test scores.

Statistical Analysis on the Residuals of a Regression of AH4 Test Scores on Mill Hill Test Scores from x480 DNA Array

SNP	Genotype	Mean	S.D.	*n*	*p*
-16	TT	0.2048	0.609	9	0.6532
	CT	0.1209	1.04	93	
	CC	0.0205	0.962	346	
-24	GG	0.2048	0.609	9	0.7598
	AG	0.0999	1.03	92	
	AA	0.0220	0.960	338	
-234	AA	0.1358	1.05	20	0.6618
	GA	0.0661	0.985	147	
	GG	0.0669	0.961	276	
-306	GG	0.0762	1.04	16	0.8673
	AG	0.0170	1.01	123	
	AA	0.0457	0.932	258	
-307	GG	0.2048	0.609	3	0.7716
	AG	0.1166	1.05	93	
	AA	0.0400	0.960	346	

Table 20 – Table displaying a summary of the statistical analysis (from x480 DNA array) for the residuals of a regression of AH4 Test Scores on Mill Hill cognitive function test scores, in relation to each *UCH-L1* allele investigated in this study, and their respective genotypes.

- *UCH-L1* alleles are denoted in the first column by the nucleotide (SNP) position in relation to the gene's translation start site. For each allele, the reference homozygote, heterozygote and variant homozygote Genotypes are indicated in the second column, with the respective mean regression score (Mean), standard deviations (S.D.) there in, and the frequency (*n*) of each genotype displayed in a row through columns three, four and five in the table. The final column in the table elucidates the probability (*p*) that any of the respective genotypes for that allele correlate with a significant difference in the regression of cognitive function test scores.

Interpretation of the Statistical Analysis

Statistical analysis of the mean test scores from each of (or indeed part of) the cognitive function tests which had been carried out on the individuals whose DNA made up the x480 DNA array (and the regression work there of), indicated no significant relationships with respect to the five *UCH-L1* promoter SNPs and the concordant reference/ variant homozygote and heterozygote genotypes.

3.3.1 Population Genetics Discussion

Thence, none of the *UCH-L1* promoter SNP alleles investigated, in the statistical based population study above, was seen to be associated with any variation/ decline in cognitive function. Decline in cognitive function is a major feature of the dementias, and although cognitive dysfunction is also a feature of Parkinson's disease, the motor abnormalities – which have their pathophysiological basis in the dopaminergic neurons of the substantia nigra – are still the defining characteristics of Parkinson's disease. Furthermore, of all the neurodegenerative diseases which *UCH-L1* is associated, Parkinson's disease certainly has the strongest association in the literature – with particular reference to the Ile93Met mutation (Leroy *et al.* (1998b)) and the protective Ser18Tyr polymorphic allele (Maraganore *et al*. (1999), Gasser *et al*. (1999), Wintermeyer *et al*. (2000), Zhang *et al*. (2000)). Thence, despite these seemingly negative statistical results within this cohort, *UCH-L1* still remains a good candidate gene to investigate a potential genetic risk factor for Parkinson's disease.

Using data which is available from the 'International HapMap Project' (www.hapmap.org), the linkage disequilibrium (LD) blocks that span the *UCH-L1* gene in Caucasians could be ascertained. Analysis of these LD blocks indicates that there is some LD across the whole gene, though there is also some division after *UCH-L1*'s exon 2. Thence, the five promoter SNPs elucidated in this study are in a LD block with exons 1 and 2, but show less LD with exons 3-9, meaning that the promoter SNPs would not mark any SNPs downstream of exon 2 very well, which was seen to be the case with the previously reported exon 3 C54A SNP (Ser18Tyr polymorphism) (refer to section 3.1 above).

With particular reference to Parkinson's disease, SNP association data from whole genome studies are also available from the National Institute of Neurological Disorders and Stroke (NINDS) Repository Parkinson's Disease Collection (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?id=phs000003), which allowed the *UCH-L1* region to be analysed for SNPs that were associated with the disease. From the analysis of the data (using Ensembl – www.ensembl.org), the one *UCH-L1* SNP (rs6848261) that was most strongly associated with Parkinson's disease in the *UCH-L1* expanded region was only very weakly associated, was 15.7Mb downstream, and was thence not on the same LD block as the promoter SNPs elucidated in this study.

Data with regards to the five promoter SNPs elucidated is still lacking in these SNP databases available in the public domain; only the C-16T and A-307G SNPs are present on the dbSNP database, and the associated data in relatively scarce (see section 3.1.4). Thence, the analysis of these SNPs as Parkinson's disease risk alleles cannot be completed with the data that is currently available. However, the *UCH-L1* promoter SNPs elucidated have been investigated with reference to the more general area of neurodegeneration and dementia, with focus on the main cognitive decline phenotype, and the results of these analyses have now focused future work on these SNPs toward the motor dysfunction phenotypes associated with Parkinson's disease – to augment the significant transcriptional effects that have been found *in vitro* and elucidate the concordant effects *in vivo*.

4 Discussion

The main objective of this study was to identify common sequence variants of the Ubiquitin Carboxyl-Terminal Hydrolase-L1 (UCH-L1) gene promoter in the Caucasian population and to examine whether they had potential functional effects on UCH-L1's expression. In order to initially identify common promoter sequence variations, and any potential coding region polymorphism association there of, denaturing High Performance Liquid Chromatography (dHPLC) was employed to scan the entire *UCH-L1* promoter/ 5' UnTranslated Region (UTR) and coding region (exons) in two DNA sample sets composed of unrelated healthy Caucasian individuals.

dHPLC scan heteroduplexes identified regions of the promoter/ 5' UTR and exon 1 as harbouring potential sequence variations. DNA sequencing analysis of these regions elucidated seven novel Single Nucleotide Polymorphisms (SNPs) in the UCH-L1 gene – five in the promoter/ 5' UTR and two in exon 1. Exon 1's guanine to adenine substitutions at nucleotide position 12 and 21 (in relation to the translation start site) were not deemed necessary for further study owed to no consequential change in UCH-L1's amino acid sequence being realised. From work previously done by Mann *et al*. (1996), it could be seen that four of the five promoter/ 5' UTR SNPs uncovered were located within highly conserved regions of the *UCH-L1* sequence – sequence comparison with an evolutionary distant species: *Monodelphis domestica* (Figure 61). Direct DNA sequencing of *UCH-L1* exon 3 also elucidated the well documented C54A SNP, whose polymorphic A allele codes for the Ser18Tyr protective UCH-L1 variant (refer to section 1.3.2) – this will be discussed further below.

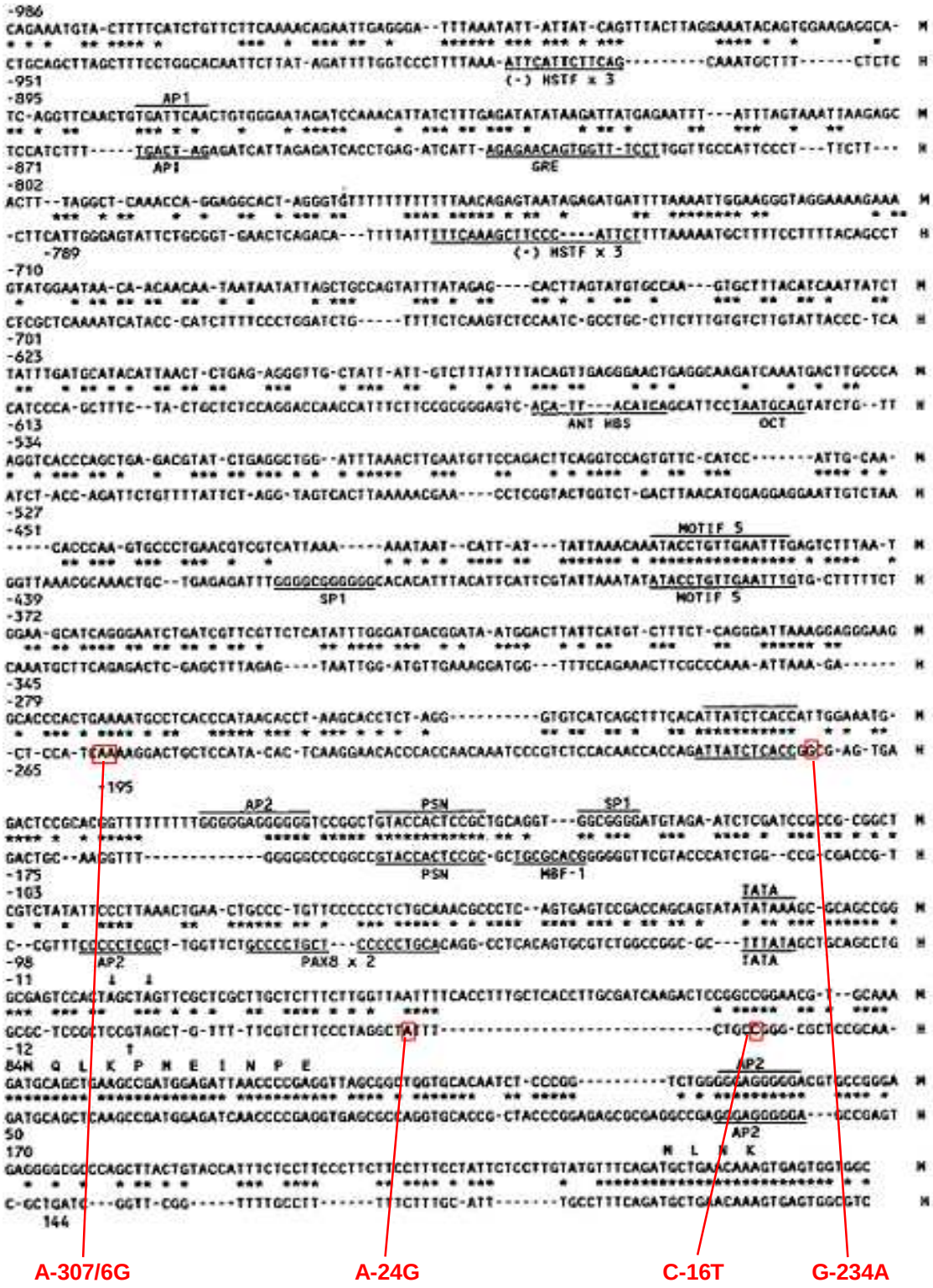


Figure 61 – Comparison of the nucleotide sequences at the 5' end of the human (H) and *Monodelphis domestica* (M) UCH-L1 genes by optimal sequence alignment. Those nucleotides seen to be conserved are highlighted (*), and all relevant promoter architecture is also shown (Mann *et al.* (1996) – refer to Figure 10 for further description). Adapted above to show the positions of the five novel promoter/ 5' UTR SNPs elucidated in this study (positions in relation to ATG translation start site).

The adenine to guanine substitution at positions -307 and -306 (A-307G, A-306G) were located in a highly conserved tetra-adenine sequence, and similarly the two 5' UTR SNPs – an adenine to guanine substitution at position -24 (A-24G) and a cytosine to thymine nucleotide change at position -16 (C-16T) – were also realised within short highly conserved sequences. The guanine to adenine SNP at position -234 (G-234A), though not conserved across species, was positioned one base pair upstream of a 59bp region of critical importance to the promoter's activity (between nucleotides -233 and -147 – refer to section 1.8 for further explanation), and one base pair downstream of a highly conserved transcriptional motif that contains an inverted GATA factor binding sequence, and the ATTA core of the homeobox binding sites (Mann *et al*. (1996)), which indicates a high potential for transcription factor binding. The positioning of the five novel SNPs elucidated in *UCH-L1*'s promoter/ 5' UTR indicated the potential to affect a variation in transcriptional activity, and were thus deemed suitable for further investigation. The three SNPs - A-307G, A-306G and G-234A - that were located in *UCH-L1*'s promoter region (upstream of the transcription start site), have the potential to effect transcription by effecting the formation of the Transcription Initiation Complex (see Section 1.7.1.2) by altering transcription factor recruitment to at least two upstream regulatory elements. The G-234A SNP could also have an impact on the critical 59bp 'minimal active promoter' identified by Mann *et al*. (1996). The two 5' UTR SNPs – A-24G and C-16T – could effect the assembly of components of the TFIID complex (see Section 1.7.1.1), which would also have an impact on *UCH-L1* transcription rates.

The next phase of the study required information on the haplotypic distribution of the five novel SNPs, so that the common Caucasian haplotypes incorporating them could ultimately be investigated for any differences in transcriptional activity realised between them.

Genotyping implemented through Restriction Fragment Length Polymorphism (RFLP) theory and direct DNA sequencing of the two DNA sample sets of unrelated healthy Caucasian individuals, identified two UCH-L1 gene haplotypes governing the distribution of the five novel promoter/ 5' UTR SNPs elucidated in this study. Thence, it was able to be concluded that three common UCH-L1 gene promoter haplotypes exist in the Caucasian population (Figure 62).

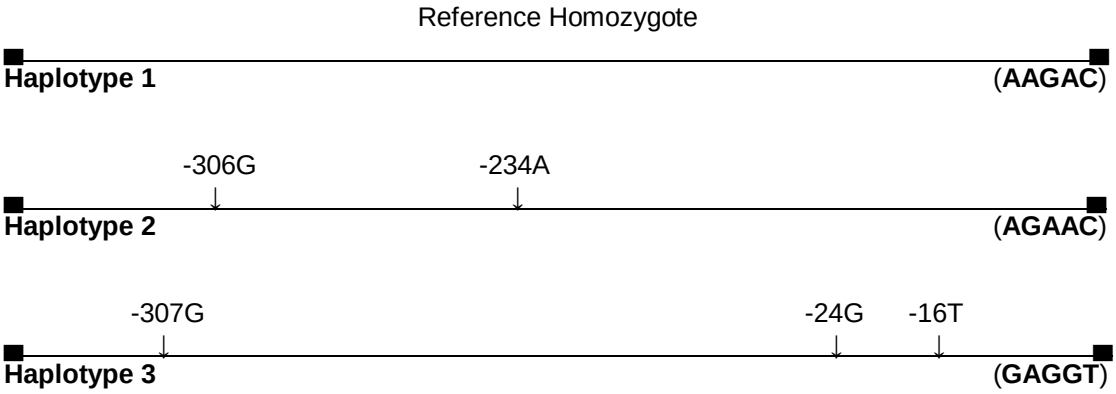


Figure 62 - The three *UCH-L1* promoter haplotypes are diagrammatically represented with reference to the distribution of the five novel promoter/ 5' UTR polymorphic alleles elucidated from the genotyping analysis in this study (allelic positions not to scale). Haplotype designations are indicated in brackets underneath and right.

Initial genotyping results in the smaller sample set (x62 chromosomes) were also confirmed in the larger Caucasian DNA sample set of 480 individuals (960 chromosomes), in which the three promoter haplotypes elucidated above (Figure 62), were the only haplotypes to be realised at frequencies above 5%. In this study, 5% was designated the threshold to determine a 'common' haplotype with the potential to confer genetic susceptibility for multifactorial neurodegenerative disease prevalence (refer to section 1.6) in the Caucasian population. The respective haplotype frequencies were: AAGAC (reference homozygote) – 60.1%, AGAAC – 15.4%, and GAGGT – 7.7% (seven other less common promoter haplotypes realised a combined frequency of 2.8%).

As discussed in section 3.3.1, there is linkage disequilibrium across the promoter region, though there is an indication from the data available in the public domain (HapMap), that after exon 2 there is a division in LD blocks, which support the results obtained in section 3.1 that the previously reported C54A SNP (protective Ser18**Tyr** allele) is not in linkage disequilibrium with the promoter SNPs associated.

Once the common Caucasian *UCH-L1* promoter haplotypes had been identified, they were incorporated into the mammalian cloning vector pGL3 (slightly modified to increase cloning strategy options) via pGEM-T Easy as a 'holding vector'. Through (dual) luciferase

reporter assays conducted on transfected mammalian cells, the pGL3 luciferase reporter vector allowed the transcriptional activity of the three inserted *UCH-L1* promoter/ 5' UTR haplotype sequences to be quantitatively compared.

The transcriptional activity of the three *UCH-L1* promoter haplotype inserts were analysed in four different mammalian cell lines – three of human origin. The only non-human cell line investigated was a mouse neuroblastoma/ rat basal ganglia neuron hybrid – ND-7 – which realised no significant differences in the transcriptional activities of the three promoter haplotypes. Assays conducted in human cell lines however, elucidated a significant difference in the transcriptional activity of the GAGGT *UCH-L1* promoter (Figure 60). In the A2058 human skin melanoma cell line and the MCF-7 human breast adenocarcinoma cell line, the GAGGT haplotype realised a mean increase in transcriptional activity of 44% and 40% respectively, compared to the AAGAC and AGAAC haplotypes. This increase in transcriptional activity of the GAGGT haplotype was even more pronounced in human cortical neuronal cells (HCN-1A cell line), which unveiled a 130% increase compared to the other *UCH-L1* promoters. Statistical analysis (employing two-way ANOVA) consistently supported these findings as significant.

These results clearly demonstrate that the -16T, -24G and -307G polymorphic alleles have a significant impact on the *UCH-L1* promoter's transcriptional activity in human cells *in vitro*. Furthermore, the effect of the GAGGT haplotype on transcriptional activity became more pronounced the closer the cell type investigated corresponded to the cells in which UCH-L1 is expressed *in vivo*, i.e. human neuronal cells. No effect was observed in the rat/ mouse ND-7 hybrid cell line, human skin (A2058) and breast cancer (MCF-7) cell lines realised comparable effects of around a forty percent increase in transcriptional activity, whilst the GAGGT promoter haplotype brought about an increase in transcriptional activity around three times this in the human neuronal cell line - HCN-1A.

The marked increase in transcriptional activity brought about by the GAGGT haplotype's three constituent polymorphic alleles in neuronal cells would seem to suggest that their effect on transcriptional activity is brought about by *cis*-acting sequence motif alteration, as *trans*-acting gene specific transcriptional binding factors often have a cell-type specific expression pattern which could well account for this large variation in transcriptional activity

between the cell lines (Weinzier (1999)). This would be especially true for *UCH-L1*, whose protein has been observed to be completely absent in non-neuronal cell lines due to regulation of expression at the transcriptional level (Mann *et al.* (1996)). Moreover, if the SNPs had affected the secondary structure or DNA methylation state of the *UCH-L1* promoter, a more general, cell line non-specific transcriptional effect would have been realised, and should thus be discounted as the underlying mechanism of the observed inter-cell-type transcriptional variation. (It should be noted that this analysis in view of DNA methylation was in particular reference to the C-16T and C-24G SNPs, as the A-307G SNP is not located in the CpG island region identified by Mann *et al*. (1996) – refer to section 1.8).

Thence, it is proposed that the -16T, -24G and -307G polymorphic alleles effect the transcriptional activity of the *UCH-L1* promoter by positively effecting the efficiency of the Transcription Initiation Complex (TIC) assembly and/ or its initiation by modifying (or indeed creating new) specific RNA/ protein interactions in view of transcription binding factor recruitment.

In reference to *UCH-L1*'s well documented association to Parkinson's disease, together with the novel role I have proposed for UCH-L1 at the synapse with α-synuclein (refer to section 1.3.6 for a discussion on UCH-L1 and review of recent literature), an increase in *UCH-L1* transcription could be a potential route in the disease pathogenesis – through proportionally increasing UCH-L1 expression in neuronal cells. UCH-L1's hypothesised role as an α-synuclein regulator (through its E3-like ligase activity) in the 'SNARE protein cycle' of neurotransmitter release/ vesicle recycling, suggests a potential pathogenic route for Parkinson's disease; an increase in UCH-L1 expression would cause an increase in α-synuclein ubiquitination/ activation throughout life, which could ultimately lead to the 'critical concentration of α-synuclein' being exceeded (Rochet and Lansbury (2000)) when protein turnover and degradation pathways become less efficient in later life, giving rise to the resultant protein aggregation and related cytoplasmic stresses which seem to be the acute pathophysiological cause of Parkinson's disease.

This can basically be seen as the inverse of the biochemical route hypothesised for UCH-L1's Ser 18Tyr variant's protective effect (in certain populations), with reduced ubiquitination/ activation of α-synuclein leading to a reduced build up of associated protein

aggregates in later life. Concordantly with this observation, *UCH-L1*'s 54A polymorphic allele which codes for the Ser18Tyr variant was also genotyped in this study through RFLP analysis, and though it was detected at a frequency of 8.1% (within a DNA sample set of 31 unrelated healthy Caucasian individuals), it was not seen to be distributed with any of the *UCH-L1* promoter haplotypes harbouring the novel SNPs elucidated in this study, i.e. the -16T, -24G and -307G alleles were not haplotypically associated with the 'protective' 54A polymorphic allele.

A further objective of this study was to investigate any possible association the polymorphic alleles potentially had in relation to neurodegenerative symptoms. This was achieved through analysing the data obtained from the RFLP genotyping that was carried out on the DNA sample set composed of 480 individuals that had supplemental cognitive function data on each individual. Such analysis allowed each polymorphic allele to be analysed with respect to the scores attained by the individuals in four separate tests, each examining specific areas of cognitive function. Furthermore, score comparisons from two of the cognitive function tests – the AH4 and Mill Hill tests – allowed any potential cognitive decline (with age) associated with any of the promoter polymorphic alleles to be uncovered. Cognitive decline with age is a symptom heavily associated with dementia, i.e. Alzheimer's disease, Dementia with Lewy Bodies and Huntington's disease – all of which have been associated with UCH-L1 (see Section 1.6). Ergo, this phase of the study attempted to uncover any evidence that the *UCH-L1* promoter SNPs were associated with dementia pathogenesis. Ultimately, none of the five novel *UCH-L1* promoter/ 5' UTR SNPs elucidated in this study were seen to be associated with any variation in cognitive function. These results suggest that the increased transcriptional activity observed for the GAGGT promoter haplotype does not seem to have any obvious, highly penetrant detrimental effects in reference to cognitive function/ decline with age, which indicates that this haplotype may not play a pathogenic role in dementia onset.

The cohort investigated in this study, in an attempt to elucidate any cognitive variation with respect to the *UCH-L1* promoter SNPs, varied from those cohorts that initially identified *UCH-L1* as an important gene in neurodegenerative diseases (refer to section 1.3), by way of not having disease phenotypes. The rationale for this was to potentially uncover a more

general association with neurodegenerative disease that would be signified by an association with cognitive decline; thence potentially uncovering a broader link with dementia. The cohort was large (n = 480) in comparison to much of the work that has already been done on *UCH-L1* with regards to more specific neurodegenerative disease investigation, and thence should be treated as a useful insight into the potential neuropathogenic role that UCH-L1 may play in the Caucasian population, as this study has certainly refocused *UCH-L1* towards the more motor dysfunctional side of neurodegeneration – with which it has been most consistently associated since Leroy *et al.*'s work in 1998. One would have certainly expected a well established genetic risk factor for a common dementia to have been associated with a variation in cognitive decline in the cohort studied - ApoE4 as a genetic risk factor for Alzheimer's for instance.

One of the limitations of this study was that the *UCH-L1* promoter SNPs were not evaluated as Parkinson's disease risk alleles, and although this would have provided a much clearer picture into *UCH-L1*'s potential pathophysiological role in the brain, especially bearing in mind the increased transcriptional activity uncovered in one of the promoter haplotypes, the fact that its association with cognitive decline has been investigated should still be regarded as very useful insight into one of the most abundant proteins in the brain (1-5% of total soluble protein).

The rationale that the GAGGT *UCH-L1* promoter haplotype could have a pathogenic association with Parkinson's disease and not dementia, can be explained by the fact that although the proposed pathogenic transcriptional effect (which leads to a hypothesised increase in UCH-L1 expression) would be felt throughout the neuronal cells of the brain, it would be more pronounced in the neurons of the substantia nigra – where *UCH-L1* is already transcribed at higher levels (Leroy *et al.* (1992)). Thence, owed to a lifetime of increased α-synuclein ubiquitination/ activation at the synapse, any damaging neuronal effects of α-synuclein/ protein aggregation brought about by this hypothesised pathogenic model in later life, would be realised in the neurons of the substantia nigra before they made an impact anywhere else in the brain – thus effecting the control of voluntary movement (which this region of the brain facilitates) in the first instance - as seen in Parkinson's disease.

This thesis has outlined the pathogenic models in which both coding region and promoter/ 5' UTR sequence variations could potentially give rise to the cytoplasmic stresses upon the neurons of the substantia nigra, which in turn bring about the symptoms seen in Parkinson's disease. The effect of coding region polymorphisms are easy to track through the processes of transcription and translation, as the change in UCH-L1's amino acid sequence allows one a straightforward insight. The effects of polymorphic variations within the promoter region/ 5' UTR of a gene however, are not so easy to determine *in vivo* owed to the different levels of post-transcriptional control available to a cell. The pathogenic model for the –307G, -24G and –16T polymorphic promoter/ 5' UTR alleles described in this thesis, relies upon the increase in transcription – which they certainly bring about *in vitro* – to be mirrored with an increase in the protein's translation *in vivo*.

Transcriptional control is by far the most basic and thus effective form of control a cell can have on the production of a particular protein – and is concordantly also the most common (Mata *et al.* (2005)) – for without the mRNA transcript no protein can be synthesised, and inversely, more transcript generally gives rise to more protein (Beyer *et al.* (2004)). However, eukaryotic cells have evolved a variety of methods that can regulate gene expression post transcription-rate regulation. These all centre around the processing, export, localisation, turnover, stability and translation of the transcript mRNAs by various combinations of RNA-binding proteins, which add substantial complexity to the control of gene expression (Mata *et al*. (2005)).

Regulation of mRNA stability is one widespread and common method of controlling transcript levels via several different exonucleolytic or endonucleolytic pathways (Parker and Sang (2004)). Decay rates can be specified by control or AU-rich elements that are usually located within the 3' UnTranslated Regions (UTR) and are recognised by various RNA-binding proteins (Parker and Sang (2004)). Though it is possible that *UCH-L1*'s 3' UTR region could contain a control element that would effect its mRNA pre-translation stability, work done by Yang *et al*. (2003) does seem to suggest that transcripts encoding vital proteins – such as UCH-L1 – would have long half lives; making it markedly stable, when its relatively short nucleotide length is borne in mind.

Another level of post-transcriptional control takes place during translation, specifically the multi-step, rate-limiting process of translation initiation. Transcript-specific regulation involves RNA-binding proteins that associate with particular structural features in target transcript UTRs (Mata *et al*. (2005)) – in much the same way as outlined for mRNA decay. *UCH-L1*'s 3' UTR may well harbour such control elements, and is certainly worth further investigation, though its relatively short UTR nucleotide length does make the chances of any critical control mechanism being present in this capacity more unlikely. The same rationale would also be applied to other control processes involving RNA-binding proteins that could alter *UCH-L1* mRNA through its potential processing, export or localisation. Length of mRNA also effects more 'global' regulation of translation, by inversely effecting the density of ribosomes, i.e. it has been shown that long transcripts seem to have reduced ribosome density, which seems to be caused by less efficient translation initiation – though the reason for this remains unclear (Arava *et al.* (2005)). Ergo, this finding seems to suggest that shorter transcripts – such as *UCH-L1* mRNA – would be less likely to be effected by translation initiation limitations owed to a reduction in ribosomal density.

Though post-transcriptional regulation is an established multi-level route whereby cells can control the expression of certain genes, a wide-ranging recent report does still suggest that there is a significant whole genome correlation between absolute transcript abundance, and translational efficiency in steady state conditions (Beyer *et al.* (2004)).

It is my contention that although post transcriptional regulation of *UCH-L1* should not to be discounted, neither should it be overstated in view of theoretically negating any increase in the UCH-L1 protein.

In summary, it is my contention that, the increased transcriptional activity of *UCH-L1*'s GAGGT promoter haplotype ascertained from the (dual) luciferase expression analysis in human cell lines, interpreted with reference to the novel pathogenic model of Parkinson's disease proposed in this thesis, certainly warrants further investigation as a potential genetic risk factor for sporadic Parkinson's disease.

5 Future Work

In view of the potential pathogenic route towards Parkinson's disease elucidated in this study, with reference to the increased transcriptional activity brought about by the -16T, -24G and -307G polymorphic alleles upon *UCH-L1*'s promoter, potential association of these SNPs with Parkinson's disease must be investigated. The methodology employed could take the form of the RFLP genotyping strategies outlined in this investigation within DNA sample sets of sporadic Parkinson's disease suffers.

This study seemed to indicate that the increased transcriptional activity of the *UCH-L1* GAGGT promoter haplotype was caused by one or more alterations in the binding of transcription factors, which positively effected the efficiency of the Transcription Initiation Complex (TIC) assembly/ initiation. Investigative work should now be carried out to confirm any change in *UCH-L1* promoter/ protein binding which would support this hypothesis. An Electrophoretic Mobility Shift Assay (EMSA) could not only uncover any variations in protein binding between the haplotypes, it could also specifically examine any changes in protein binding realised between the reference homozygote and polymorphic alleles for each individual SNP. DNase 1 Footprinting could also be similarly used to elucidate any differences in DNA-protein interaction between the haplotypes.

Furthermore, whether or not an increase in *UCH-L1* mRNA transcript brings about concordant increases in UCH-L1 protein within neuronal cells has not yet been confirmed, as post-transcriptional regulation could somehow stabilise intracellular UCH-L1 at a certain concentration - buffering any increase in transcriptional rate. The influence of post-transcriptional gene regulation could be established through post mortem neuronal tissue analysis of UCH-L1 protein levels in those individuals harbouring the GAGGT promoter haplotype, compared to those individuals carrying the AGAAC or reference homozygote *UCH-L1* haplotypes. This would thus confirm that the increase in transcriptional activity brought about by the GAGGT haplotype, gives rise to an increase in UCH-L1 protein *in vivo.*

Though the investigation discussed above would be a very effective route to directly determine the role of post-transcriptional regulation for UCH-L1 *in vivo*, the reality of securing

the methodology's required resources may prove difficult. Thence, simpler, less resource consuming *in vitro* methodologies must be discussed. UCH-L1 protein expression analysis from neuronal tissue culture could provide an effective *in vitro* route to determine whether the GAGGT haplotype brings about an increase in UCH-L1 protein compared to the two other promoter haplotypes elucidated in this study.

The presence of any 3'UTR control elements in post-transcriptional regulation (most likely region in which these control elements would be present – refer to section 4 for a discussion) of UCH-L1 could also be investigated. *UCH-L1* sequence could be cross-referenced for known post-transcriptional control elements, and if present, *UCH-L1* sequence mutants could be constructed that lack these elements, which could then be investigated to determine whether any change in protein expression results, compared to that of the reference homozygote 3'UTR sequence.

Moreover, similar to the approach and methodology used in this thesis, further SNP screening work could take place in *UCH-L1*'s 3'UTR, in which any polymorphic alleles elucidated could be genotyped within Parkinson's disease DNA sample sets, to establish any association. This, in addition to the RFLP genotyping work for the –16T, -24G and –307G polymorphic alleles within Parkinson's disease sample sets already set out in this section, would give a more complete overview of any potential association of the UCH-L1 gene in Parkinson's disease.

With the transcription of *UCH-L1* still in mind, the fact that its 13kb transcript is present in the substantia nigra at higher levels than else where in the brain (Leroy *et al.* (1992)), indicates that the neurons of the substantia nigra would be a good investigative target to uncover the route by which increased *UCH-L1* transcription is actually brought about *in vivo*. Elucidation of such *in vivo* methods of increasing *UCH-L1*'s transcription rate, would not only provide a clearer understanding of *UCH-L1* transcriptional regulation, but it may also impart valuable information for any potential future putative therapy that involved targeting neuronal UCH-L1 levels in the control of neurodegeneration in Parkinson's disease (see below).

The true role of UCH-L1 in neurons is far from clear. Further *in vitro* work in neuronal cells is required to confirm its potential role as an E3-like 'ubiquitinator'/ activator of α-synuclein at the synapse. What would be of immediate interest, in view of the hypothesised pathogenic model of Parkinson's disease onset put forward in this study, is whether synaptic α-synuclein levels proportionally rise as UCH-L1 expression levels are increased in human neuronal cells.

Furthermore, with reference to the hypothesised basis put forward in this thesis (refer to section 1.3.6) of the documented protective effect brought about by the Ser18Tyr UCH-L1 mutant in certain populations, it would also be important to establish whether carriers of the polymorphic allele do actually realise a decreased level of SNARE complex associated neural activity throughout life.

From what has been discussed and put forward in this thesis regarding Parkinson's disease pathogenesis involving UCH-L1, a theoretically valid and potentially effective target for Parkinson's disease therapy becomes apparent. It has been shown that a polymorphic variation in *UCH-L1*'s coding region can protect against Parkinson's disease in certain populations (Maraganore *et al.* (1999), Zhang *et al.* (2000), Satoh *et al.* (2001) & (Wintermeyer *et al.* (2000) – refer to section 1.3.2), and this seems to be associated with the polymorphic protein's reduced ligase activity (Liu *et al.* (2002)). Thence, any therapeutic strategy that can effect a similar change in UCH-L1's ligase enzyme kinetics, would have a good theoretical basis of providing a similar protective effect to an individual. Potential targets for such a therapeutic strategy, could take the form of repressing *UCH-L1*'s transcription – if indeed elevated UCH-L1 levels are found to be a major risk factor. Though, as discussed above, the specifics of neuronal *UCH-L1* transcriptional regulation need to be more fully elucidated before this can be attempted. Therapeutic strategies could also potentially target UCH-L1's as yet unknown post-transcriptional regulatory mechanisms (if any are indeed present), i.e. investigations outlined above could uncover 3'UTR translational control elements within the UCH-L1 gene, which could thus be biochemically targeted.

However, before such potential therapies can be administered, genetic predisposition to Parkinson's disease must be identified and confirmed, so that those individuals at risk can

be quickly and easily identified through genetic screening; the outcome of any further research investigating the association of UCH-L1's GAGGT gene promoter haplotype will be eagerly anticipated.

6 Appendix

6.1 *UCH-L1* Sequence Management

UCH-L1's promoter and exons were partitioned into regions which could be amplified by way of PCR to then be scanned using denaturing high performance liquid chromatography (dHPLC) – (refer to section 2.3).

UCH-L1's sequence was obtained from the NCBI website database. The sequence was cross referenced with relevant publications; promoter region (Mann *et al*. (1996)), exon (and intron) structure (Leroy *et al*. (1998), Day *et al*. (1990)).

The promoter sequence of *UCH-L1* was sub sectioned into five regions, *UCH-L1*'s coding sequences (9 exons) were subdivided into seven distinct regions:-

……..**(-1023)**GCCTATTTTTAAAAAATACTCAACCTTCTTGCTTCCTTCGCTACCTAAG
P10 >TCGCTACCTAAG

TATTTCTGCAAGCCCACTTTGTTCTGCAGCTTAGCTTTCCTGGCACAATTCTTATAGATTT
TATTTCTGCAAGC**>**

TGGTCCCTTTTAAAATTCATTCTTCAGCAAATGCTTTCTCTCTCCATCTTTTGACTAGAGAT
CATTAGAGATCACCTGAGATCATTAGAGAACAGTGGTTTCCTTGGTTGCCATTCCCTTTC
TTCTTCATTGGGAGTATTCTGCGGTGAACTCAGACATTTTATTTTTCAAAGCTTCCCATTC
P8>TTTCAAAGCTTCCCATTC

TTTTAAAAATGCTTTTCCTTTTACAGCCTCTCGCTCAAAATCATACCCATCTTTTCCCTGG
TTTTA**>** **<**GTAGAAAAGGGACC

ATCTGTTTTCTCAAGTCTCCAATCGCCTGCCTTCTTTGTGTCTTGTATTACCCTCACATCC
TAGACAAAAGA<**P9**

CCCAGCTTTCTACTGCTCTCCCAGGACCAACCATTTCTTCCGCGGGAGTCACATTACATC
AGCATTCCTAATGCAGTATCTGTTATCTACCAGATTCTGTTTTATTCTAGGTAGTCACTTA
P6>GTCACTTA

AAAACGAACCTCGGTACTGGTCTGACTTAACATGGAGGAGGAATTGTCTAAGGTTAAAC
AAAACGAACCTCGGTACT**>** **<**TAACAGATTCCAATTTG

GCAAACTGCTGAGAGATTTGGGGCGGGGGGCACACATTTACATTCATTCGTATTAAATAT
CGTTTGAC**<P7**

ATACCTGTTGAATTTGTGCTTTTTCTCAAATGCTTCAGAGACTCGAGCTTTAGAGTAATTG
GGATGGTGAAAGGATGGGTTTCCAGAAACTTCGCCCAAAATTAAAGACTCCATCAAAAG
P4>CATCAAAAG

GACTGCTCCATACACTCAAGGAACACCCACCAACAAATCCCGTCTCCACAACCACCAGA
GACTGCTCCATAC**>** AGGTGTTGGTGGTCT

TTATCTCACCGGCGAGTGAGACTGCAAGGTTTGGGGGCCCGGCCGTACCACTCCGCGC
AATAGAGTG**<P5**

TGCGCACGGGGGGTTCGTACCCATCTGGCCGCGACCGTCCGTTTCCCCCTCGCTTGGT
TCTGCCCCTGCTCCCCCTGCACAGGCCTCACAGTGCGTCTGGCCGGCGCTT**TATA**GCT
P2>AGGCCTCACAGTGCGTCTG**>**
E1/2a>CTCCCCCTGCACAGGCCTCA**>**

GCAGCCTGGGCGGCTCC**G**CTAGCTGTTTTTCGTCTTCCCTAGGCTATTTCTGCCGGGCG
<ATCGACAAAAAGCAGAAGGG**<P3** **A-24G** **C-16T**

Exon 1>

CTCCGCGAAG**ATG**CAGCTCAAGCCGATGGAGATCAACCCCGAGGTGAGCGCCAGGTGC
ACCGCTACCCGGAGAGCGCGAGGCCGAGGGAGGGGGAGCCGAGTCGCTGATCGGTTC

Exon 2>

GGTTTTGCCTTTTTCTTTGCATTTGCCTTTCAGATGCTGAACAAAGTGAGTGGCGTCTCG
<CTACGACTTGT TTCACTCACC**<P1**

CGCCGTCTCTGGCCCCCTCCCCCGCGAGCGCCGAGGCGGGGGCGCCCACCGGTTCCG
<AAGGC

GCTGCTGGCAGGGACCAAGCCGCCCGCTGCGAGCACCGGAGACGGCCGGGCTGGGG
CGACGACCGTCCCTG**<E1/2b**

CGTGGGCTGGGCGCAGCACAGACTCGGCTGCACGGGCTTCGCGGGCGCCACGTGTGG
GCCGCGCTTTGTGCTGTGTCATTGCGCCGGCCCGGGTGGGGGTGGCAGGGCGGGACT
E3a>GCTTTGTGCTGTGTCATTGC**>**

GGGGCTCCTCCCAGGCTCGGGTGCGGGCGCGGAGGGCGCGCGCCTCCTGGCCCCGC
CCCCTGGCAGGTGCCCGCGACCCGCGTGTCCCCGTGCGCCTGGCCGCCTTGTCTCCTC

Exon 3>

TCCGCAGGTGCTGTCCCGGCTGGGGGTCGCCGGCCAGTGGCGCTTCGTGGACGTGCT
C54A (S18Y)

GGGGCTGGAAGAGGAGTCTCTGGGCTCGGTGCCAGCGCCTGCCTGCGCGCTGCTGCT
GCTGTTTCCCCTCACGGCCCAGGTAGGGCGTGGGGCCCAGGATGCGCCGGCCGCCGG
CAGTGCACGCCGCTCCCCAGCTTGAGTCCTCGGGGTAGTTGGTAGAACTCATGTGCTG
<CGGCGAGGGGTCGAACTC**<E3b** **Exon 4>**

CCATCTGTTCTTTGCACTTTCATTCTGAGATGTAAAAACGCTTTTTACATTCGCAGCATGA
E4a>TGCACTTTCATTCTGAGATG**>**

GAACTTCAGGAAAAAGCAGATTGAAGAGCTGAAGGGACAAGAAGTTAGTCCTAAAGTGT
ACTTCATGAAGCAGACCATTGGGAATTCCTGTGGCACAATCGGACTTATTCACGCAGTG
GCCAATAATCAAGACAAACTGGGATTTGGTAGGTGTGGGTTTTGAGGCCAGCCATCCTA
C277G (I93M) <CAAAACTCCGGTCGGTAG**<E4b**

AGCTTGAACTTGAAACATGGAGTTCAGAAACAGCTGTTTATCCACAACCCTGGAGGCAGT
ATTAAAGATTCAGGTTGCTCAGCATGTTCAGCAAAGGCTTAAGTCAACAATAAATATGTAC
E5/6a>AGGTTGCTCAGCATGTTCAG**>** **Exon 5>**

CCACTTGTATTATTTTACCTATACTAACACATCCATTTTTTTTTTAAGAGGATGGATCAGTT
CTGAAACAGTTTCTTTCTGAAACAGAGAAAATGTCCCCTGAAGACAGAGCAAAATGCTTT
GAAAAGAATGAGGTAAGAGAACTTACAGAGCATGGCCTTTAAATAACTCTAGAGATTTTT

Exon 6>

GTGCTAATTATTTTCTTTTTTCCGCAGGCCATACAGGCAGCCCATGATGCCGTGGCACAG
GAAGGCCAATGTCGGGTAAATGCAAATACAAATCGGAGCCAGGCTGCCTGGGTGCCAT
<CGGTA

CTGTGTTTCTACTGAAATTGTGCAGGAATCTCTTACTGGAACCTCATAGAGTTGTTCTGA
GACACAAAGATGAC**<E5/6b**

CAGTTAACAGTATATTTACCTTAGTGGGCTTAGAATAGGGCTTAATGTAAGACATACATTA
E7a>CTTAGTGGGCTTAGAATAGG**>**

AATATTAGCTATAATTTCTAAAAATCAAGTCAGTTCAAGCACATTTCACTTGAATTGCAAG

Exon 7>

ATAATTTTTAAAATACAGCTTACACTCATTTTCAAAAATTTCTTGACTTTCTTTAGGTAGAT
GACAAGGTGAATTTCCATTTTATTCTGTTTAACAACGTGGATGGCCACCTCTATGAACTTG
GTATGTTTTACTCCATTTTTGGAACCCAGTGTAGTTTCATGTGTTCTTTCAGACTGAATTT
CTCTTGATATATTGTGTGACTTTATGGCACTTGGCATATCATTGTTTATAAAGCCACAATA
ACAAAGTATTCTCATGAGGGCACTTAACCCCTTATCTGTGGGTTTGGCAGTGGTTTTTGG
<CATAAGAGTACTCCCGTGAA**<E7b**

AAGAATCTAAAACTTCCATCTAGGCTAGGTAAGCACGGTAGCCAGAAAACATGCAGAGA
E8a>ATCTAGGCTAGGTAAGCACG**>**

Exon 8>

AAATTGACATGCCTGGCTTCTTTGTTACAGATGGACGAATGCCTTTTCCGGTGAACCATG
GCGCCAGTTCAGAGGACACCCTGCTGAAGGTCATCTTTGGAATGCATCTCTTCTTAATGT
GCCTCACAATTCTTTGGCTAAATTTTCTATTCTAAAGTGCTAACACTCTTCCAGTATAGCC
<GAGAAGGTCATATCGG

AGCATCAGTTGCCTGGGTTAATAGCAGTCTTTAGGGTGAACTTTGAGCATTAATAGACCT
TCGT**<E8b**

TGGAGCCTTTCCCTATGTGACTTTCATTTTGAGCTCTTGCTGTTTGGATTTTAATGACATT
>GGAGCCTTTCCCTATGTGAC**>E9a**

Exon 9>

TCTCCTTTCCAGGACGCTGCCAAGGTCTGCAGAGAATTCACCGAGCGTGAGCAAGGAGA
AGTCCGCTTCTCTGCCGTGGCTCTCTGCAAGGCAGCC**TAA**TGCTCTGTGGGAGGGACTT
TGCTGATTTCCCCTCTTCCCTTCAACATGAAAATATATACCCCCCCATGCAGTCTAAAATG
CTTCAGTACTTGTGAAACACAGCTGTTCTTCTGTTCTGCAGACACGCCTTCCCCTCAGCC
ACACCCAGGCACTTAAGCACAAGCAGAGTGCACAGCTGTCCACTGGGCCATTGTGGTGT
GAGCTTCAGATGGTGAAGCATTCTCCCCAGTGTATGTCTTGTATCCGATATCTAACGCTT
TAAATGGCTACTTTGGTTTCTGTCTGTAAGTTAAGACCTTGGATGTGGTTTAATTGTTTGT
<CAATTCTGGAACCTACACCA**<E9b**

CCTCAAAAGGAATAAAACTTTTCTGCTGATAAGATAGCCACAGCTGATTCTCATTTTCTTT
TACCCTCTCCTCAATATGTCAGG……..

Figure 63 – Displays *UCH-L1*'s promoter and coding regions with PCR primer sequences adhered (primer sequences are indicated directly below the relevant *UCH-L1* sequence – sense primers in >blue>, and anti-sense primers in <green< (refer to section 2.1). *UCH-L1* exons are underlined and labelled above the 5' end. Sequences not relevant to *UCH-L1's* promoter or coding regions, i.e. intronic regions, are not represented. Major sites within the sequence are indicated in larger, bold font; *UCH-L1*'s TATA box, transcription start site (a guanine residue), ATG start codon (Exon 1) and TAA stop codon (Exon 9). Also, documented *UCH-L1* single nucleotide polymorphisms (SNPs) are highlighted in red, with their sequence position and base change displayed immediately below.

- Intronic nucleotide distances for *UCH-L1* are as follows: intron 1 – 105bp, intron 2 – 463bp, intron 3 – 2700bp, intron 4 – 950bp, intron 5 - 75bp, intron 6 – 1300bp, intron 7 800bp and intron 8 - >200bp.

6.2 Cloning Strategy Preparation

6.2.1 *UCH-L1* Restriction Endonuclease Sequence Information

The restriction endonucleases that did not realise cutting sites within *UCH-L1*'s promoter sequence – taking into account all sequence changes brought about by the haplotypic variance – are listed in Figure 48 below:-

Aar I	*Aat* II	*Acc* 65I	*Acc* I	*Acl* I	*Afe* I	*Afl* II
Afl III	*Age* I	*Ahd* I	*Alo* I	*Alo* I'	*Asc* I	*Ase* I
Bae I	*Bae* I'	*Bam* HI	*Ban* I	*Bbe* I	*Bbv* CI	*Bcg* I
Bcg I'	*Bci* VI	*Bcl* I	*Bfr* BI	***Bgl* II**	*Bmr* I	*Bpl* I
Bpl I'	*Bpm* I	*Bpu* 10I	*Bsa* AI	*Bsa* HI	*Bsa* I	*Bsa* WI
Bsi WI	*Bsp* E I	*Bsp* HI	*Bsp* MI	*Bsr* BI	*Bsr* DI	*Bsr* GI
Bss HII	*Bss* S I	*Bst* API	*Bst* BI	*Bst* EII	*Bst* XI	*Bst*Z17I
Bsu 36I	*Btr* I	*Bts* I	*Cla* I	*Dra* III	*Drd* I	*Ear* I
Eci I	*Eco* I CR I	*Eco* NI	*Eco* RI	*Eco* RV	*Fal* I	*Fal* I'
Fse I	*Fsp* A I	*Hinc* II	*Hpa* I	*Hpy* 99I	*Hpy*CH4IV	*Kas* I
***Kpn* I**	*Mfe* I	***Mlu* I**	*Msc* I	*Msl* I	*Nar* I	*Nco* I
Nde I	*Not* I	*Nru* I	*Nsi* I	*Nsp* I	*Oli* I	*Pac* I
Pci I	*Pfl* M I	*Pme* I	*Pml* I	*Ppi* I	*Ppi* I'	*Ppu* 10I
Ppu MI	*Psh* A I	*Psi* I	*Psr* I	*Psr* I'	*Pvu* I	*Pvu* II
Rsr II	***Sac* I**	*Sal* I	*San* DI	*Sap* I	*Sbf* I	*Sca* I
Sex AI	*Sfi* I	*Sfo* I	*Sgf* I	***Sma* I**	*Sna* BI	*Spe* I
Sph I	*Srf* I	*Ssp* I	*Swa* I	*Taq* II	*Taq* II'	*Tat* I
Tsc I	*Tth* 111I	*Xba* I	*Xma* I	*Xmn* I	*Zra* I	

Figure 64 - Output obtained from 'Lasergene – DNAstar' restriction analysis software. (The restriction enzymes highlighted in **bold** were those that were deemed suitable to be utilised).

6.2.2 *UCH-L1* Promoter Cloning Primer Positions

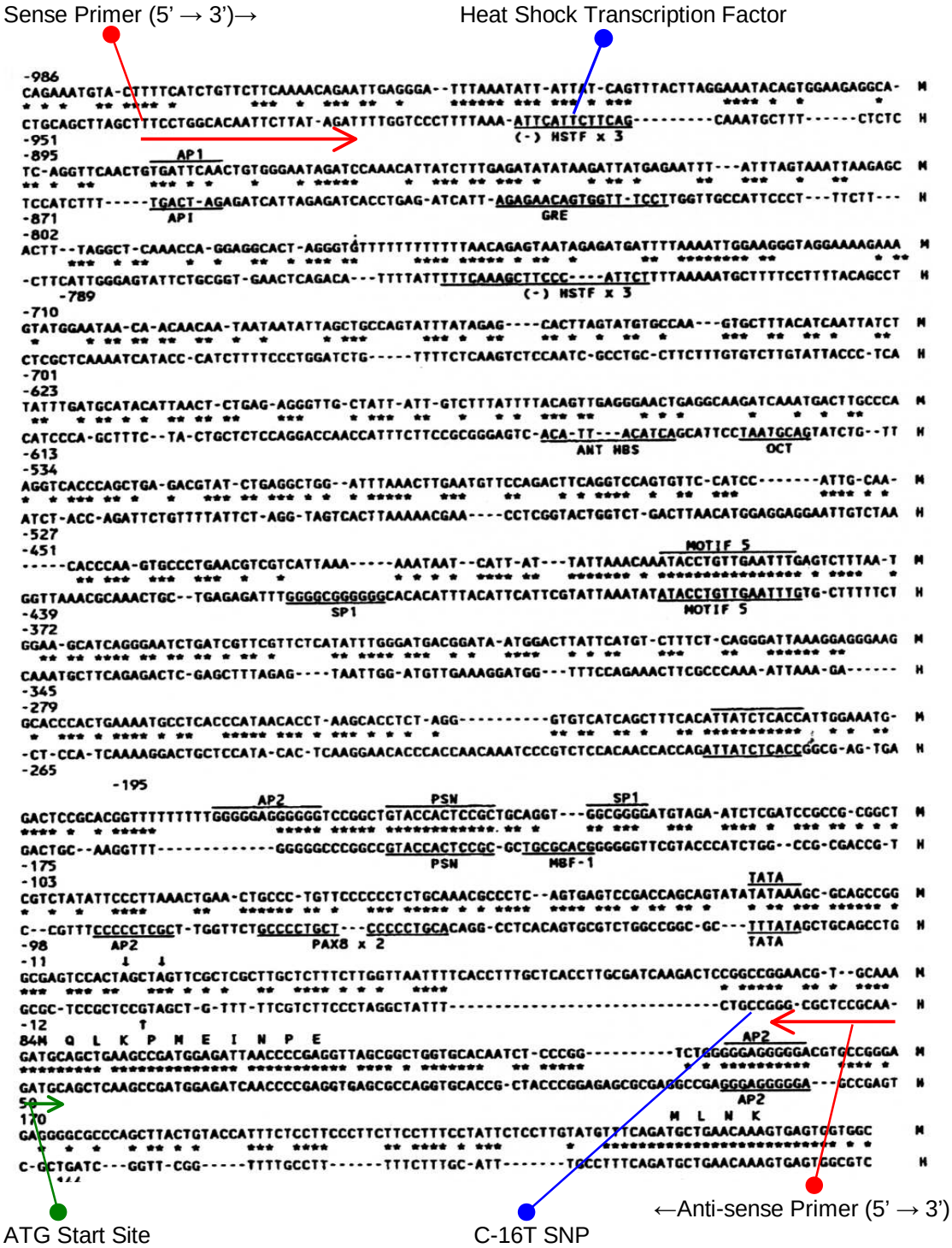


Figure 65 - Comparison of the human (H) *UCH-L1* promoter sequence with that of *Monodelphis domestica* (M), displaying all relevant promoter architecture (Mann *et al.* (1996) – see also Figure 7). Adapted above to show the location of the sense and anti-sense primers required for intact *UCH-L1* promoter amplification. Other regions denoted indicate important sequences relevant in primer design.

6.3 Initial Cloning Strategy Results

- All three *UCH-L1* promoter/ 5' UTR haplotype PCR bands were successfully amplified:-

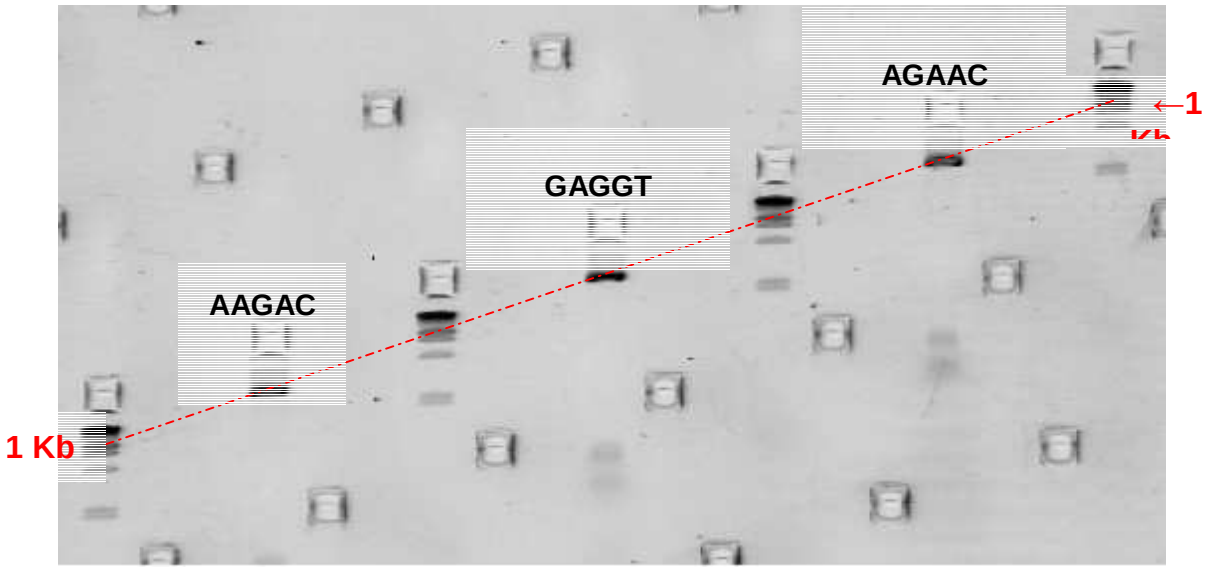


Figure 66 – MADGE gel clearly showing all three haplotype PCR bands at the required size (~1 Kb). Alternate wells display 1 Kb DNA ladders (refer to section 3.4.1.2), with the 1 Kb band clearly marked and denoted in red.

*(The *KPN* I and *Bgl* II restriction sites were also confirmed through DNA sequencing (electropherograms not shown)).

6.4 Raw Luciferase Data

6.4.1 A2058 Luciferase Data

LAR II/ Ranilla Ratios (per assay)

	1	2	3	4	5	6
AAGAC	0.16869	0.409532	0.464812	1.600454	1.168695	2.559046
AAGAC	1.191219	0.953563	0.635797	1.406574	0.990509	2.052191
GAGGT	0.375141	0.643423	0.651809	2.387029	1.637049	3.966622
GAGGT	0.422891	0.807864	0.562145	2.290851	2.115007	4.035088
AGAAC	0.371139	0.59361	0.387204	1.101707	1.352162	3.167936
AGAAC	0.224587	0.600362	0.452258	2.041433	0.871158	2.795326

Means

	1	2	3	4	5	6
AAGAC	0.680	0.682	0.550	1.504	1.080	2.306
GAGGT	0.399	0.726	0.607	2.339	1.876	4.001
AGAAC	0.298	0.597	0.420	1.572	1.112	2.982
Control (no insert)	0.022	0.053	0.030	0.071	0.013	0.032

Standard Deviation

	1	2	3	4	5	6
AAGAC	0.723037	0.384688	0.120904	0.137094	0.125997	0.3584
GAGGT	0.033765	0.116277	0.063402	0.068008	0.337967	0.048413
AGAAC	0.103628	0.004774	0.046	0.664486	0.340122	0.263475
Control (no insert)	0.006248	0.005623	0.020086	0.018077	0.002272	0.002045

SEM

AAGAC	0.511	0.272	0.085	0.097	0.089	0.253
GAGGT	0.024	0.082	0.045	0.048	0.239	0.034
AGAAC	0.073	0.003	0.033	0.470	0.241	0.186
Control (no insert)	0.006	0.006	0.020	0.018	0.002	0.002

6.4.2 MCF-7 Luciferase Data

LAR II/ Ranilla Ratios (per assay)

	1	2	3
AAGAC	0.589581	0.736922	0.387264
AAGAC	0.47863	0.619924	0.51341
AAGAC	0.524761	0.648434	0.521836
GAGGT	0.851022	0.89262	0.577915
GAGGT	0.787264	0.987256	0.578177
GAGGT	0.689539	0.876482	0.612583
AGAAC	0.53595	0.605134	0.347481
AGAAC	0.675409	0.635024	0.407564
AGAAC	0.609459	0.502971	0.434887

Means

	1	2	3	**Av. (x2)**
AAGAC	0.531	0.668	0.474	**1.116**
GAGGT	0.776	0.919	0.590	**1.523**
AGAAC	0.607	0.581	0.397	**1.056**
Control (no insert)	0.028	0.027	0.027	**0.054**

Standard Deviation

	1	2	3
AAGAC	0.055737	0.061007	0.07538
GAGGT	0.081335	0.059844	0.01994
AGAAC	0.069764	0.069244	0.044715
Control (no insert)	0.004952	0.002101	0.00327

SEM

				Av. (x2)
AAGAC	0.032	0.035	0.044	**0.074**
GAGGT	0.047	0.035	0.012	**0.062**
AGAAC	0.040	0.040	0.026	**0.071**
Control (no insert)	0.005	0.002	0.003	**0.007**

6.4.3 ND-7 Luciferase Data

LAR II/ Ranilla Ratios (per assay)

	1	2	3
AAGAC	5.013398	0.371552	0.9562
AAGAC	4.738495	0.377671	0.828825
AAGAC	4.992474	0.407731	0.995058
AAGAC	5.160542	0.39973	0.943772
GAGGT	4.292822	0.347971	0.672486
GAGGT	4.014011	0.384565	0.696142
GAGGT	6.158737	0.339166	0.78304
GAGGT	6.282653	0.3341	0.744643
AGAAC	5.420359	0.282063	0.711644
AGAAC	5.731897	0.33256	0.717097
AGAAC	4.272887	0.259336	0.560606
AGAAC	3.866999	0.269649	0.545323

Means

	1	2	3	**Av./2**
AAGAC	4.976	0.389	0.931	**1.049**
GAGGT	5.187	0.351	0.724	**1.044**
AGAAC	4.823	0.286	0.634	**0.957**
Control (no insert)	0.045	0.002	0.010	**0.009**

Sdev

	1	2	3
AAGAC	0.175247	0.017307	0.071511
GAGGT	1.200026	0.022808	0.049469
AGAAC	0.894333	0.032464	0.093423
Control (no insert)			

SEM

				Av./2
AAGAC	0.088	0.009	0.036	**0.022**
GAGGT	0.600	0.011	0.025	**0.106**
AGAAC	0.447	0.016	0.047	**0.085**

6.4.4 HCN-1A Luciferase Data

LAR II/ Ranilla Ratios (per assay)

	1	2	3
AAGAC	0.145119	0.481069	1.587689
AAGAC	0.068872	0.678304	2.346939
AAGAC	0.132839	0.733788	1.43576
GAGGT	0.261019	0.4375	5.898601
GAGGT	0.515489	0.948546	2.7008
GAGGT	0.2241	0.852	3.098295
AGAAC	0.113986	0.604265	0.934354
AGAAC	0.117722	0.541436	1.474378
AGAAC	0.131004	0.644022	0.814103

Means

AAGAC	0.116	0.631	1.790	**0.846**
GAGGT	0.334	0.746	3.899	**1.660**
AGAAC	0.121	0.597	1.074	**0.597**
Control (no insert)	0.035	0.147	0.296	**0.160**

Standard Deviation

AAGAC	0.040939	0.13282	0.488158
GAGGT	0.158654	0.271508	1.742874
AGAAC	0.008944	0.051723	0.351674
Control (no insert)	0.024303	0.012954	0.026479

SEM

AAGAC	0.024	0.077	0.282	**0.127**
GAGGT	0.092	0.157	1.006	**0.418**
AGAAC	0.005	0.030	0.203	**0.079**
Control (no insert)	0.024	0.013	0.026	**0.021**

6.5 Stock Solutions/ Preparations

6.5.1 PCR

- PCR Buffer (x 10, 200mM Tris-HCl, 500mM KCl) (Invitrogen)
- dNTPs (8 mM) (Invitrogen)
- $MgCl_2$ (25 mM) (Invitrogen)
- DNA *Taq* Polymerase (5 units/ µl) (Invitrogen)

- Betaine (5M) - 5.86g Betaine (Sigma) in 10ml deionised water.

6.5.2 Polyacrylamide Gel Related

6.5.2.1 Gel Plate Preparation

- Methanol (BDH Laboratory Supplies)

- **Sticky Silane** – 495ml Ethanol (Fisher)
  - 2.5ml Glacial acetic acid (BDH Laboratory Supplies)
  - 2.5ml Methacryloxypropylthimethoxy-silane (Fisher)

6.5.2.2 Gel Preparation

30% Acrylamide (Severn Biotech)

- **TBE (10x)** - 108g Tris Base (Sigma)
  - 55g Orthaboric acid (BDH Laboratory Supplies)
  - 93g EDTA (BDH Laboratory Supplies)

  (Made up to 1000ml with deionised water.)

- Ammonium Persulphate Solution 25% (25% APS)

- 2.5g Ammonium persulphate powder (Fisher) in 10ml deionised water.
- Ethidium bromide (Sigma)

6.5.2.3 Gel Running

- MADGE Loading Dye - 980µl Deionised formamide (Sigma)

- 200µl **0.5M Ethylenediaminetetraacetic acid disodium salt**
  - 186.12g EDTA (BDH Laboratory Supplies) in 1000ml diH_20
- 0.025% Xlene cyanol ff (Sigma) – as required.
- 0.025% Bromophenol blue (Sigma) – as required.

(Made up to 10ml with deionised water.)

6.5.3 dHPLC Buffers

- Buffer A (0.1M Triethylammonium acetate (TEAA) - 50ml TEAA (2M – Transgenomic)
- 250µl Acetonitrile (Aldrich)
- Made up to 1000ml with HPLC diH_20

- Buffer B (0.1M TEAA and 250ml acetonitrile (25%)) - 50ml TEAA (2M – Transgenomic)
- 250ml Acetonitrile (Aldrich)
- Made up to 1000ml with HPLC diH_20

- Buffer C (75% Acetonitrile cleaning solution) - 750ml Acetonitrile (Aldrich)
- Made up to 1000ml with HPLC diH_20

- Buffer D (Syringe wash solution – 8% Acetonitrile) - 80ml Acetonitrile (Aldrich)
- Made up to 1000ml with HPLC diH_20

6.5.4 DNA Sequencing Related

- Shrimp Alkaline Phosphatase (SAP) (Promega)
- SAP Buffer (Promega)
- Exonuclease I (Exo I) (Promega)
- Big Dye Terminator Ready Reaction Mix (ABI)
- Sequencing Buffer (ABI)
- 95% Ethanol (Fisher)
- 70% Ethanol (diluted with deionised water from Fisher stock)
- Alconox (Aldrich)

- Additional Chemicals Used in the Preparation of the Gel Mixture:

- 10% 6M Urea (Amresco)

- Ammonium Persulphate Solution 10% (10% APS)

- 1.0g Ammonium persulphate powder (Fisher) in 10ml deionised water.
- N, N, N', N'-Tetramethyl-1-, 2-diaminomethane (TEMED) (Sigma)

- Sequencing Loading Buffer

- 1 part 25mM Ethylenediaminetetraacetic acid disodium salt (EDTA) (BDH Lab Supplies)
- 5 parts deionised Formamide (Sigma)
- Blue Dextran (Applied Biosystems) – as required

- Other Restriction Digest Related Chemicals:

- 10 x Bovine Serum Albumin (BSA) (Invitrogen)
- Calf Intestinal Alkaline Phosphatase (CIAP) (Promega)

6.5.5 Agarose Gel Related

6.5.5.1 Gel Preparation

- Agarose powder (GibcoBRL)

- TAE (1x) - 242g Tris base (Sigma)
- 57.1ml Glacial acetic acid (BDH Laboratory Supplies)
- 100ml 0.5M EDTA (see above)

(Made up to 1000ml with deionised water.)

- Modified TAE (1x) - 48.25g Tris base (Sigma)
- 11.42ml Glacial acetic acid (BDH Laboratory Supplies)
- 0.372g EDTA (BDH Laboratory Supplies)

(Made up to 500ml with deionised water)

6.5.5.2 Gel Running

- Orange G Loading Dye – 5ml Glycerol (BDH Laboratory Supplies)
- 1ml 0.5M EDTA (see above)
- 1ml 2% Orange G (Sigma)
- 0.1ml 10% Lauryl sulfate sodium salt (SDS) (BDH Lab. Supplies)

(Made up to 10ml with deionised water.)

6.5.6 Miniprep Protocol Buffers - (QIAgen Spin Miniprep Kit)

- Buffer P1 (+ RNase A) (Qiagen)
- Buffer P2 (Qiagen)
- Buffer N3 (Qiagen)
- Buffer PB (Qiagen)
- Buffer PE (Qiagen)
- Buffer EB (Qiagen)

6.5.7 Midiprep Protocol Solutions

('PureYield' Plasmid Midiprep System – Promega)

- Cell Resuspension Solution (Promega)
- Cell Lysis Solution (Promega)
- Neutralization Solution (Promega)
- Endotoxin Removal Wash Solution (+ Isopropanol) (Promega)
- Column Wash Solution (+ ethanol) (Promega)
- Nuclease-Free Water (Promega)

6.5.8 Bacterial Sub-Cloning

6.5.8.1 Bacterial Strain

- *E.coli* – JM109 (Original Glycerol Stock – Promega)

6.5.8.2 Luria-Bertoni + Ampicillin Growth Medium

In this study, two types of Luria-Bertoni (LB) + ampicillin growth medium was used for the selected growth of *E.coli* colonies harbouring the respective Promega cloning vectors (see above) which conveyed ampicillin resistance – agar plates and broth (both the agar and broth powder were sourced from Gibco BRL).

6.5.8.2.1 Luria-Bertoni + Ampicillin Agar Plates

All LB + ampicillin agar plates used in this study were prepared as indicated below:-

- 30g LB agar powder was dissolved in 1000ml deionised water (this contained 10g tryptone, 5g yeast extract, 5g sodium chloride and 10g agar).

- This was then autoclaved at 121℃ for 15 minutes t o sterilise.

- Once the solution had adequately cooled (> 50℃) 1ml of ampicillin (1g/ 10ml) was added.

- The agar was then dispensed into agar plates (~ 40/ 1000ml).

6.5.8.2.2 Luria-Bertoni + Ampicillin Broth

All LB + ampicillin broth used in this study was prepared as indicated below:-

- 20g LB broth powder was dissolved in 1000ml deionised water (this contained 10g tryptone, 5g yeast extract and 5g sodium chloride).

- This was then autoclaved at 121℃ for 15 minutes t o sterilise.

- Once the solution had adequately cooled (> 50°C) 1ml of ampicillin (1g/ 10ml) was added

6.5.9 Ligations

- 0.1M calcium chloride ($CaCl_2$) (BDH Laboratory Supplies)
- 1Dimethyl sulfoxide (DMSO) (BDH Laboratory Supplies)
- PUC 19 Control DNA (Promega)

6.5.9.1 pGEM-T Easy Vector Ligation Related Materials

- Rapid Ligation Buffer (Promega)
- pGEM-T Easy Vector (Promega)
- Control Insert DNA (Promega)
- T4 DNA Ligase (Promega)

6.5.9.2 pGL3-*Eco* RI Vector Ligation Related Materials

- Ligation Buffer (Promega)
- pGL3-*Eco*-RI (Modified from Promega – see above)
- Control Insert DNA (Promega)
- T4 DNA Ligase (Promega)

6.5.9.3 Additional Post Ligation Media

- SOC recovery medium (Promega)

6.5.10 Mammalian Cell Culture

6.5.10.1 Solutions Related to Propagation Methodology (Passaging)

- 0.25% (w/v) Trypsin-0.53mM EDTA solution (Sigma)
- 70% Ethanol (diluted with deionised water from 95% Fisher stock)

6.5.10.2 Transfection Related Chemicals

- Serum free medium – animal free protein, without L-glutamine (Sigma)

GeneJuice (Merck Biosciences) – Theory:

Whereas many available transfection reagents are based on cationic lipid formulation, GeneJuice Transfection Reagent is composed of a nontoxic cellular protein and a small amount of a novel polyamine. GeneJuice Transfection Reagent provides highly efficient DNA transfer in both stable and transient transfection of eukaryotic cells. The unique chemistry provides the advantage of compatibility with both serum-containing and serum-free media, and makes media changes unnecessary.

6.5.10.3 Dual Luciferase Reporter Assay Buffer Preparation (Promega)

- Passive Lysis Buffer

- 4 volumes of Distilled Water to 1 volume of the Buffer of Phosphate Buffered Saline (Promega).

- Luciferase Assay Buffer II

- The lyophilised Luciferase Assay Substrate powder (Promega) was added to the 10ml of supplied Luciferase Assay Buffer II (Promega).

- Stop and Glo Buffer

- 1 volume of Stop and Glo Substrate (x50 concentrate) (Promega) was added to 50 volumes of the supplied Stop and Glo Buffer (Promega).

7 References

Albin, R. L., Young, A. B. & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. Trends in Neuroscience. **12**: 366-75.

Amerik, A. Y. & Hochstrasser, M. (2004). Mechanism and function of deubiquitinating enzymes. Biochim. Biophys. Acta. **1695:** 189–207.

Arava, Y., Boas, F. E., Brown, P. O., Herschlag, D. (2005). Dissecting eukaryotic translation and its control by ribosome density mapping. Nucleic Acids Res. **33**: 2421-2432.

Ardley, H. C & Robinson, P. A. (2005). E3 Ubiquitin Ligases. In: Essays in Biochemistry – The Ubiquitin-Proteasome System. Eds. Mayer, R. J. & Layfield, R. pp. 15-30 (Chapter 2). Portland Press, London.

Argiles, J. M., Lopez-Soriano, F. J. & Pallares-Trujillo, J. (1998). Ubiquitin and Disease. Eds. Argiles, J. M., Lopez-Soriano, F. J. & Pallares-Trujillo, J. R. G. Landes Company, Texas.

Armstrong, M., Daly, A. K., Cholerton, S., Bateman, D. N. & Idle, J. R. (1992). Mutant debrisoquine hydroxylation genes in Parkinson's disease. Lancet **339**: 1017-1018.

Barbeau, A., Roy, M., Bernier, G., Campanella, G. & Paris, S. (1987). Ecogenetics of Parkinson's disease: prevalence and environmental aspects of rural areas. Canadian Journal Neurological Science. **14**: 36-41.

Barber, R., McKeith, I. G., Ballard, C., Gholkar, A. & O'Brien, J. T. (2001). A comparison of medial and temporal lobe atrophy in dementia with Lewy bodies and Alzheimer's disease: magnetic resonance imaging volumetric study. Dementia and Geriatric Cognitive Disorders. **12**: 198-205.

Barber, R., Panikkar, A. & McKeith, I. G. (2001). Dementia with Lewy bodies: diagnosis and management. International Journal of Geriatric Psychiatry. **62**: 16-21.

Barden, J. A., Cottee, L. J. & Bennett, M. R. (1999). Vesicle-associated proteins and P2X receptor clusters at single sympathetic varicosities in mouse vas deferens. Journal of Neurocytology. **28**: 469-480.

Bear, M. F., Connors, B. W. & Paradiso, M. A. (1996). Neuroscience: Exploring the Brain. Williams and Wilkins.

Ben-Neriah, Y. (2002). Regulatory functions of ubiquitination in the immune system. Nature Immunology.**31**: 20-26.

Bennett, M. C., Bishop, J. F., Leng, Y., Chock, P. B., Chase, T. N. & Mouradian, M. M. (1999). Degradation of α-synuclein by proteasome. Journal of Biological Chemistry. **274**: 33855-58.

Betarbet, R., Sherer, T. B., MacKenzie, G., Garcia-Osuna, M., Panov, A. V. & Greenamyre, J. T. (2000). Chronic systemic pesticide exposure reproduces features of Parkinson's disease. Nature Neuroscience. **3** (12): 1301-06.

Beyer, A., Hollunder, J., Nasheuer, H. P. & Wilhelm, T. (2004). Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. Molecular and Cellular Proteomics. **3**: 1083-1092.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature. **32**: 209-212.

Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends in Genetics. **3**: 342–347.

Blandini, F., Nappi, G., Tassorelli, C. & Martignoni, E. (2000). Functional changes of the basal ganglia circuitry in Parkinson's disease. Progress in Neurobiology. **62**: 63–88.

Borodovsky, A., Kessler, B. M., Casagrande, R., Overkleeft, H. S., Wilkinson, K. D. & Ploegh, H. L. (2001). A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. EMBO Journal. **20**: 5187-196.

Braun, B. C., Glickman, M., Kraft, R., Dahlmann, B., Kloetzel, P. M., Finley, D. & Schmidt, M. (1999). The base of the proteasome regulatory particle exhibits chaperone-like activity. Nature. Cell Biology. **1**: 221-6.

Breathnach, R. & Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Annual Reviews in Biochemistry. **50**: 349-383.

Brookes, A. J. (1999). The essence of SNPs. Gene. **234**, (2): 177-186.

Brookmeyer, R., Gray, S. & Kawas, C. (1998). Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. American Journal of Public Health. **88**: 1337-1342.

Buratowski, S. (1994). The basics of basal transcription by RNA polymerase II. Cell. **77**:1-3.

Burn, D. J., Mark, M. H., Playford, E. D., Maraganore, D. M., Zimmerman, T. R., Duvoisin, R. C., Harding, A. E., Marsden, C. D. Brooks, D. J. (1992). Parkinson's disease in twins studied with 18F-dopa and positron emission tomography. Neurology. **42**: 1894-900.

Burnett, B., Li, F. & Pittman, R. N. (2003). The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. Human Molecular Genetics. **12**: 3195–3205.

Butterworth, N. J., Williams, L., Bullock, J. Y., Love, D. R., Faull, R. L. & Dragunow, M. (1998). Trinucleotide (CAG) repeat length is positively correlated with the degree of DNA fragmentation in Huntington's disease striatum. Neuroscience. **87**: 49-53.

Bylsma, F. W., Rebok, G. W. & Brandt, J. (1991). Long-term retention of implicit learning in Huntington's disease. Neuropsychologia. **29**: 1213-1221.

Byrne, J. E. (2005). The treatment of dementia with Lewy bodies. In: Dementia. Eds. Burns, A., O'Brien, J. & Ames, D. pp. 648-653 (Chapter 50). Edward Arnold, London. 3rd Edition.

Cedar, H. (1988). DNA methylation and gene activity. Cell. **53**: 3-4.

Chandara, S., Gallardo, G., Fernandez-Chacon, R., Schluter, O. M. & Sudhof, T. C. (2005). α-Synuclein cooperates with CSPα in preventing neurodegeneration. Cell. **123**: 383-396.

Chua, P. & Chiu, E. (2005). Huntington's disease. In: Dementia. Eds. Burns, A., O'Brien, J. & Ames, D. pp. 754-760 (Chapter 62). Edward Arnold, London. 3rd Edition.

Chung, K. K., Zhang, Y., Lim, K. L., Tanaka, Y., Huang, H., Gao, J., Ross, C. A., Dawson, V. L. & Dawson, T. M. (2001). Parkin ubiquinates the α-synuclein-interacting protein, synphillin-1: implications for Lewy-body formation in Parkinson's disease. Nature Medicine. **7**: 1144-1150.

Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. Cell. **79**: 13-21.

Ciechanover, A., Orian, A. & Schwartz, A. L. (2000). Journal of Cellular Biochemistry **34**: 40-51 (Supplement).

Clough, C. G., Chaudhuri, K. R. & Sethi, K. D. (2003). Fast Facts – Indispensable Guides to Clinical Practice: Parkinson's Disease. Health Press, Oxford.

Conway, K. A., Rochet, J. –C., Bieganski, R. M. & Lansbury, P. T. (2001). Kinetic stability of the α-synuclein protofibril by a dopamine-α-synuclein adduct. Science. **294**: 1346-49.

Cordato, D. J. & Chan, D. K. Y. (2004). Genetics and Parkinson's disease. Journal of Clinical Neuroscience. **11** (2): 119-123.

Dahlmann, B. (2005). Proteasomes. In: Essays in Biochemistry – The Ubiquitin-Proteasome System. Eds. Mayer, R. J. & Layfield, R. pp. 31-48 (Chapter 3). Portland Press, London.

Day, I. N. M. & Thompson, R. J. (1987). Molecular cloning of cDNA coding for human PGP 9.5 protein – A novel cytoplasmic marker for neurones and neuroendocrine cells. FEBS Letters. **210** (2): 157-160.

Day, I. N. M., Hinks, L. J. & Thompson, R. J. (1990). The structure of the human gene encoding protein gene product 9.5 (PGP9.5), a neuron-specific ubiquitin C-terminal hydrolase. Journal of Biochemistry. **268**: 521-524.

Day, I. N. M. & Humphries, S. E. (1994). Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. Analytical Biochemistry. **222** (2): 389-95.

DeStefano, A. L., Golbe, L. I., Mark, M., Lazzarini, A. M., Maher, N. E., Saint-Hilaire, M. H., Feldman, R. G., Guttman, M., Watts, R. L., Suchowersky, O., Lafontaine, A. L., Labelle, N., Lew, M. F., Waters, C. H., Growdon, J. H., Singer, C., Currie, L. J., Wooten, G. F., Vieregge, P., Pramstaller, P. P., Klein, C., Hubble, J. P., Stacy, M., Montgomery, E., MacDonald, M. E., Gusella, J. F. & Myers, R. H. (2001). Genome-wide scan for Parkinson's Disease: The GenePD Study. Neurology. **57**:1124-1126.

Dickson, D. W. (1999). Neuropathologic differentiation of progressive supranuclear palsy and corticobasal degeneration. Journal of Neurology. **246** (S2): 6-15.

Doran J. F., Jackson, P., Kynoch, P. A. M. & Thompson, R. J. (1983). Isolation of PGP 9.5, a new human neurone-specific protein detected by high-resolution two-dimensional electrophoresis. Journal of Neurochemistry. **40** (6): 1542-47.

Dunnet, S. B. & Bjorklund, A. (1999). Prospects for new restorative and neuroprotective treatments in Parkinson's disease. Nature. **399**: A32-39.

Eddins, M. J. & Pickart, C. M. (2005). Ubiquitin-conjugating Enzymes. In: Protein Degradation – Ubiquitin and The Chemistry of Life (Vol. 1). Eds. Mayer, R. J., Ciechanover, A. & Rechsteiner, M. pp. 102-134 (Chapter 5). Wiley-VCH, Weinheim.

Edwards, Y. H., Fox, M. F., Povey, S., Hinks, L. J., Thompson, R. J. & Day, I. N. M. (1991). The gene for human neurone specific ubiquitin C-terminal hydrolase (UCH-L1, PGP9.5) maps to chromosome 4p14. Annuls of Human Genetics. **55**: 273-78.

Elbaz, A., Levecque, C., Clavel, J., Vidal, J. –S., Richard, F., Correze, J. –R., Delemotte, B., Amouyel, P., Alperovitch, A., Chartier-Harlin, M. –C. & Tzourio, C. (2003). S18Y polymorphism in the UCH-L1 gene and Parkinson's disease: evidence for an age-dependent relationship. Movement Disorders. **18** (2): 130-37.

Engelender, S., Kaminsky, Z., Guo, X., Sharp, A. H., Amaravi, R. K., Kleiderlein, J. J., Margolis, R. L., Troncoso, J. C., Lanahan, A. A., Worley, P. F., Dawson, V. L., Dawson, T, M. Ross, C. A. (1999). Synphilin-1 associates with α-synuclein and promoted the formation of cytosolic inclusions. Nature Genetics. **22**: 110-14.

Evans, P. C., Smith, T.S., Lai, M. J., Williams, M. G., Burke, D. F., Heyninck, K., Kreike, M. M., Beyaert, R., Blundell, T. L. & Kilshaw, P. J. (2003). A novel type of deubiquitinating enzyme. Journal of Biological Chemistry. **278**: 23180-86.

Fall, P. –A., Fredrikson, M., Axelson, O. & Granérus, A. –K. (1999). Nutritional and occupational factors influencing the risk of Parkinson's disease: A case-control study in southeastern Sweden. Movement Disorders. **14**: 28-37.

Farrer, M., Gwinn-Hardy, K., Muenter, M., DeVrieze, F. W., Crook, R., Perez-Tur, J., Lincoln, S., Maraganore, D., Adler, C., Newman, S., MacElwee, K., McCarthy, P., Miller, C., Waters, C. & Hardy, J. (1999). A chromosome 4p haplotype segregating with Parkinson's disease and postural tremor. Human Molecular Genetics. **8**: 81-85.

Farrer, M., Destee, A, Becquet, E, De Vrieze, F. W-., Mouroux, V., Richard, F., Defebvre, L., Lincoln, S., Hardy, J., Amouyel, P. & Chartier-Harlin, M. –C. (2000). Linkage exclusion in French families with probable Parkinson's disease. Movement Disorders. **15** (6): 1075-83.

Folstein, S. E. (1989). Huntington's disease: a disorder of families. Johns Hopkins University Press, Baltimore.

Fon, E. (2006). *Unpublished data.* Centre for Neuronal Survival, Montreal Neurological Institute and Hospital, 3801 University Street, Montreal, Quebec. H3A 2B4.

Funayama, M., Hasegawa, K., Kowa, H., Saito, M., Tsuji, S. & Obata, F. (2002). A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Annals of Neurology. **51**: 296-301.

Gao, S., Hendrie, H. C., Hall, K. S. & Hui, S. (1998). The relationships between age, se, and the incidence of dementia and Alzheimer's disease. Archives of General Psychiatry. **55**: 809-815.

Gasser, T. (2001). Genetics of Parkinson's disease. Journal of Neurology. **248**: 833-40.

Gasser, T., Pilz, P., Omasmeier, D. & Miller-Myhsok, B. (2000). Ubiquitin C-terminal hydrolase L1 in sporadic Parkinson's disease. Movement Disorders. **15**: 201 (Supplement).

Gasser, T., Muller-Myshok, B., Wszolek, Z. K., Oehlmann, R., Calne, D. B., Bonifati, V., Bereznai, B., Fabrizio, E., Vieregge, P. & Horstmann, R. D. (1998). A susceptibility locus for Parkinson's disease maps to chromosome 2p13. Nature Genetics. **18**: 262-65.

Gaston, K. & Fried, M. (1995). CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Research. **23** (6): 901-909.

Gaunt, T. R., Hinks, L. J., Rassoulian, H. & Day, I. N. M. (2003). Manual 768 or 384 well microplate gel 'dry' electrophoresis for PCR checking and SNP genotyping. Nucleic Acids Research. **31**: 9 (e48).

Gelb, D. J., Oliver, E. & Gilman, S. (1999). Diagnostic Criteria for Parkinson Disease. Archives of Neurology. **56**: 33-39.

Giasson, B. I., Duda, J. E., Murray, I. V. J., Chen, Q., Souza, J. M., Hurtig, H. I., Ischiropoulos, H., Trojanowski, J. Q.& Lee, V. M. –Y. (2000). Oxidative damage linked to neurodegeneration by selective α-synuclein nitration in synucleinopathy lesions. Science. **290**: 985-989.

Gibb, W. R. G. & Lees, A. J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. Journal of Neurology, Neurosurgery and Psychiatry. **51**: 745–752.

Glickman, M. H. (2000). Getting in and out of the proteasome. Seminars in Cell and Developmental Biology. **11**: 149-58.

Goedert, M., Spillantini, M. G. & Davies, S. W. (1998). Filamentous nerve cell inclusions in neurodegenerative diseases. Current Opinion in Neurobiology. **8**: 619-32.

Goodrich, J. S., Clouse, K. N. & Schupbach, T. (2004). Hrb27C, Sqd and Otu cooperatively regulate gurken RNA localisation and mediate nurse cell chromosome dispersion in Drosophila oogenesis. Development. **131**: 1949-58.

Gorrell, J. M., DiMonte, D. & Graham, D. (1996). The role of the environment in Parkinson's disease. Environmental Health Perspective. **104**: 652.

Groll, M., Bochtler, M., Brandstetter, H., Clausen, T. & Huber, R. (2005). Molecular machines for protein degradation. Chembiochem. **6**(2): 222-56.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. Nature. **386**: 463−471.

Gwinn-Hardy, K., Chen, J. Y., Liu, H. C., Liu, T. Y., Boss, M., Seltzer, W., Adam, A., Singleton, A., Koroshetz, W., Waters, C., Hardy, J. & Farrer, M. (2000). Spinocerebellar ataxia type 2 with Parkinsonism in ethnic Chinese. Neurology. **55**: 800-5.

Gwinn-Hardy, K., Singleton, A., O'Suilleabhain, P., Boss, M., Nicholl, D., Adam, A., Hussey, J., Critchley, P., Hardy, J. & Farrer, M. (2001). Spinocerebellar ataxia type 3 phenotypically resembling parkinson disease in a black family. Archives of Neurology. **58**: 296-9.

Hames, B. D. & Glover, D. M. (1988). Transcription and Splicing. IRL Press. 2nd Edition.

Harding, A. J., Broe, G. A. & Halliday, G. M. (2002). Visual hallucinations in Lewy body disease relate to Lewy bodies in the temporal lobe. Brain. **125**: 391-403.

Harhangi, B. S., Farrer, M. J., Lincoln, S., Bonifati, V., Meco, G., De Michele, G., Brice, A., Dürr, A., Martinez, M., Gasser, T., Bereznai, B., Vaughan, J. R., Wood, N. W., Hardy, J., Oostra' B. A. & Breteler, M. M. B. (1999). The Ile93Met in the ubiquitin carboxy-termianl-hydrolase-L1 gene is not observed in European cases with familial Parkinson's disease. Neuroscience Letters. **270** (1): 1-4.

Haroutunian, V., Serby, M., Purohit, D. P., Perl, D. P., Marin, D., Lantz, M., Mohs, R. C. &. Davis K. L. (2000). Contribution of Lewy body inclusions to dementia with patients with and without Alzheimer's disease neuropathological conditions. Archives of Neurolgy. **57**: 1145-1150.

Harvey, G. T., Hughes, J., McKeith, I. G., Briel, R., Ballard, C., Gholkar, A., Scheltens, P., Perry, R. H., Ince, P. & O'Brien, J. T. (1999). Magnetic resonance imaging differences between dementia with Lewy bodies and Alzheimer's disease: a pilot study. Psychological Medicine. **29**: 181-187.

Hashimoto, M., Takeda, A., Hsu, L. J., Takenouchi, T. & Masliah, E. (1999). Role of cytochrome *c* as a stimulator of α-synuclein aggregation in Lewy body disease. Journal of Biological Chemistry. **274**: 28849-28852.

Heal, D. G., Abou-Sleiman, P. M., Casas, J. P., Ahmadi, K. R., Lynch, T., Gandhi, S., Muqit, M. M. K., Foltynie, T., Barker, R., Bhatia, K. P., Quinn, N. P., Lees, A. J., Gibson, M. J., Holton, J. L., Revesz, T., Goldstein, D. B. & Wood, N. W. (2006). *UCH-L1* is not a Parkinson's disease susceptibility gene. Annals of Neurology. **59**: (4) 627-33.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A., Kel, O., Ignatieva, A., Ananko, E., Podkolodnaya, O., Kolpakov, F., Podkolodny, N. & Kolchanov, N. (1998). Databases on Transcriptional Regulation: TRANSFAC, TRRD, and COMPEL. Nucleic Acids Research. **26**: 364-70.

Henke, W., Herdel, K., Schnorr, D. & Loening, S. A. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Research. 25 (19): 3957-58.

Hershko, A., Heller, H., Elias, S. and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system: resolution, affinity purification and role in protein breakdown. Journal of Biological Chemistry. **258**: 8206-8214.

Hershko, A. & Ciechanover, A. (1992).The ubiquitin system for protein degradation. Annual Reviews of Biochemistry. **61**: 761-807.

Hicks, A. A., Petursson, H., Jonsson, T., Stefánsson, H., Jóhannsdóttir, H. S., Sainz, J., Frigge, M. L., Kong, A., Gulcher, J. R., Stefánsson, K. & Sveinbjörnsdóttir, S. (2001). A susceptibility gene for late-onset idiopathic Parkinson's disease successfully mapped. American Journal of Human Genetics. **69**: (123) 200 (Supplement).

Hill, C. S. & Treisman, R. (1995). Transcriptional regulation by extracellular signals: mechanisms and specificity. Cell, **80**: 199–211.

Hochstrasser, M. (1996). Protein degradation or regulation: Ub the judge. Cell. **84**: 813-15.

Hoppe, T. (2005). Multiubiquitylation by E4 enzymes: one size doesn't fit all. Trends in Biochemical Science. **30**: 183-187.

Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J. W., Gu, W., Cohen, R. E. & Shi, Y. (2002). Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. Cell. **111**: 1041-1054.

Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. **72**: 971-83.

Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., Hackett, J., Adamson. J., Lincoln, S., Dickson, D., Davies, P., Petersen, R. C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J. M., Nowotny, P., Che, L. K., Norton, J., Morris, J. C., Reed, L. A., Trojanowski, J., Basun, H., Lannfelt, L., Neystat, M., Fahn, S., Dark, F., Tannenberg, T., Dodd, P. R., Hayward, N., Kwok, J. B., Schofield, P. R., Andreadis, A., Snowden, J., Craufurd, D., Neary, D., Owen, F., Oostra, B. A., Hardy, J., Goate, A., van Swieten, J., Mann, D., Lynch, T. & Heutink, P. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature. **393**: 702-5.

Imai, Y., Soda, M., Inoue, H., Hattori, N., Mizuno, Y. & Takahashi, R. (2001). An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of parkin. Cell. **105**: 891-902.

Irizarry, M. C., Growdon, W., Gomez-Isla, T., Newell, K., George, J. M., Clayton, D. F. & Hyman, B. T. (1998). Nigral and cortical Lewy bodies and dystrophic nigral neurites in Parkinson's disease and cortical Lewy body disease contain α-synuclein immunoreactivity. Journal of Neuropathological and Experimental Neurology. **57**: 334-337.

Jellinger, K. A. (2001). The Pathology of PD. In: Parkinson's Disease: Advances in Neurology (Vol. 86). Eds. Calne, D. & Calne, S. M. pp. 55-72 (Chapter 6). Lippincott, Williams & Wilkins, Philadelphia.

Jenner, P. & Olanow, C. W. (1998). Understanding cell death in Parkinson's disease. Annals of Neurology. **44**: S72-84.

Jenner, P. & Olanow, C. W. (1998). Understanding cell death in Parkinson's disease. In: Beyond the Decade of the Brain (Vol. 3). Eds. Olanow, C. W. & Jenner, P. pp.145-166, Wells Medical Limited, Royal Tunbridge Wells.

Johnston, S. C., Larsen, C. N., Cook, W. J., Wilkinson, K. D. & Hill, C. P. (1997). Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8áAáresolution. The EMBO Journal. **16** (13): 3787-3796.

Johnston, S. C., Riddle, S. M., Cohen, R. E. & Hill, C. P. (1999). Structural basis for the specificity of ubiquitin C-terminal hydrolases. The EMBO Journal. **18** (14): 3877-3887.

Kato, M., Miyazawa, K., Kitamura, N. (2000). A Deubiquitinating Enzyme UBPY Interacts with the Src Homology 3 Domain of Hrs-binding Protein via a Novel Binding Motif PX(V/I)(D/N)RXXKP. Journal of Biological Chemistry**. 275**: 37481-37487.

Kawamata, H., McLean, P. J., Sharma, N. & Hyman, B. T. (2001). Interaction of α-synuclein and synphillin-1: effect of Parkinson's disease associated mutations. Journal of Neurochemistry. **77**: 929-34.

Ke, X., Collins, A. & Shu, Y. (2001) PIRA PCR designer for restriction analysis of single nucleotide polymorphisms. Bioinformatics. **17** (9):838-39.

Kitada, T., Asakawa, S., Hittori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. & Shimizu, N. (1998). Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. Nature. **392** (6676): 605-8.

Klatka, L. A., Louis, E. D. & Schiffer, R. B. (1996). Psychiatric features in diffuse Lewy body disease: a clinicopathological study using Alzheimer's disease and Parkinson's disease comparison groups. Neurology. **47**: 1148-1152.

Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H. D., Mayer, T. U. & Jentsch, S. (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly**.** Cell. **96**: 635–644.

Korell, M. & Tanner, C. M. (2005). Epidemiology of PD: An Overview. In: Parkinson's Disease. Eds. Ebadi, M. & Pfeiffer, R. F. pp. 39-50 (Chapter 5). CRC Press, Florida.

Kramer, M. (1999). Annals of Neurology. **46**: 179-82.

Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, S., Kosel, S., Przuntek, H., Epplen, J. T., Schols, L. & Riess, O. (1998). Als39Pro mutation in the gene encoding α-synuclein in Parkinson's disease. Nature Genetics. **18**: 106-108.

Kruger, R., Vieira-Saecker, A. M., Kuhn, W., Berg, D., Muller,, T., Kuhnl, N., Fuchs, G. A., Storch, A., Hungs, M., Woitalla, D., Przuntek, H., Epplen, J. T., Schols, L. and Riess, O. (1999). Increased susceptibility to sporadic Parkinson's disease by a certain combined α-synuclein/ apolipoprotein E genotype. Annals of Neurology. **45**: 611-17.

Kulkull, W. A. & Ganguli, M. (2000). Epidemiology of dementia: concepts and overview. In: Neurologic Clinics. Ed. DeKosky, S. pp. 923-49 (Volume18). W. B. Saunders, Philadelphia.

Lai, E., Riley, J., Purvis, I. & Roses, A. (1998). A 4-Mb High-Density Single Nucleotide Polymorphism-Based Map around Human APOE*1. Genomics. **54** (1): 31-38.

Lam, Y. A., Xu, W., DeMartino, G. N. & Cohen, R. E. (1997). Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. Nature. **385**: 737-740.

Lang, A. E. & Lozano, A. M. (1998). Parkinson's disease. New England Journal of Medicine. **339**: 1044-1053.

Lang, A. E. & Lozano, A. M. (1998). Parkinson's disease: Medical progress part II. New England Journal of Medicine. **339**: 1130-1143.

Langston, J. W., Ballard, P., Tetrud, J. W. *et al.* (1983). Chronic parkinsonism in humans due to a product of a meperidine-analog synthesis Science. **219**: 979-80.

Larsen, C N., Krantz, B. A. & Wilkinson, K. D.(1998). Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases. Biochemistry. **37**: 3358-68.

Lee, M., Hyun, D. –H., Halliwell, B. & Jenner, P. (2001). Effect of the overexpression of wild-type or mutant α-synuclein on cell susceptibility to insult. Journal of Neurochemistry. **76**: 998-1009.

Lee, V. Y. M., Goedert, M. & Trojanowski, J. Q. (2001). Neurodegenerative tauopathies. Annual Review of Neuroscience. **24**: 1121-59.

Leroy, E., Boyer, R. & Polymeropoulos, H. (1998a). Intron-exon structure of ubiquitin C-terminal Hydrolase-L1. DNA Research. **5**: 397-400.

Leroy, E., Boyer, R., Auburger, G., Leube, B., Ulm, G., Mezey, E., Harta, G., Brownstein, M. J., Jonnalagada, S., Chernova, T., Dehejia, A., Lavedan, C., Gasser, T., Steinbach, P. J., Wilkinson, K. D. & Polymeropoulos, M. H. (1998b). The ubiquitin pathway in Parkinson's disease. Nature. **395** (6701): 451-2.

Levecque, C., Destee, A., Mouroux, V., Becquet, E., Defebvre, L., Amouyel, P. & Chartier-Harlin, M. C. (2001). No genetic association of the ubiquitin carboxy-terminal hydrolase-L1 gene S18Y polymorphism with familial Parkinson's disease. Journal of Neural Transmitters. **108**: 979-84.

Lewin, B. (1997). Genes VI. Oxford University Press.

Lim, K. L., Chew, K. C. M., Tan, J. M. M., Wang, C., Chung, K. K. K., Zhang, Y., Tanaka, Y., Smith, W. L., Engelender, S., Ross, C. A., Dawson, V. L. & Dawson, T. M. (2005). Parkin mediates non-classical, proteasomal-independent, ubiquitination of Synphilin-1: Implications for Lewy Body Formation. Journal of Neuroscience. **25**: 2002-2009.

Lincoln, S., Vaughan, J., Wood, N., Baker, M., Adamson, J., Gwinn-Hardy, K., Lynch, T., Hardy, J. & Farrer, M. (1999). Low frequency of pathogenic mutations in the ubiquitin carboxy-terminal hydrolase gene in familial Parkinson's disease. Neuroreport. **10** (2) 427-29.

Londos, E., Passant, U., Risberg, J., Gustafson, L. & Brun, A. (2002). Contributions of other brain pathologies in dementia with Lewy bodies. Dementia and Geriatric Cognitive Disorders. **13**: 130-148.

Lucking, C. B., Dürr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B. S., Meco, G., Denefle, P., Wood, N. W., Agid, Y. & Brice, A. (2000). Association between early-onset Parkinson's disease and mutations in the *parkin* gene. New England Journal of Medicine. **342**: 1560-67.

Lunkes, A., Trottier, Y. & Mandel, J. L. (1998). Pathological mechanisms in Huntington's disease and other polyglutamine expansion diseases. Essays in Biochemistry. **33**: 149-163.

Lyons, K. E., Pahwa, R. & Troster, A. I. (1997). A Comparison of Parkinson's Disease Symptoms and Self-reported Functioning and Well Being. Parkinsonism & Related Disorders. **3**: 207-209.

Maher, N. E., Golbe, L. I., Lazzarini, A. M., Mark, M. H., Currie, L. J., Wooten, G. F., Saint-Hilaire, M., Wilk, J. B., Volcjak, J., Maher, J. E., Feldman, R. G., Guttman, M., Leew, M., Schuman, S., Suchowersky, O., Lafontaine, A. L., Labelle, N., Vieregge, P., Pramstaller, P. P., Klein, C., Hubble, J., Reider, C., Growdon, J., Watts, R., Montgomery, E., Baker, K., Singer, C., Stacy, M. & Myers, R. H. (2002). Epidemiological study of 203 sibling pairs with Parkinson's disease: the GenePD study. Neurology. **58**: 79-84.

Makarova, K. S., Aravind, L. & Koonin, E. V. (2000). A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae. Trends in Biochemistry Science. **25**: 50-52.

Manago, Y., Kanahori, Y., Shimada, A., Sato, A., Amano, T., Sato, Y., Setsuie, R., Sakurai, M., Aoki, S., Wang, Y. L., Osaka, H., Wada, K. & Noda, M. (2005). Potentiation of ATP-induced currents due to the activation of P2X receptors by ubiquitin carboxy-terminal hydrolase L1. Journal of Neurochemistry. **92**: 1061-1072.

Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. Science **236**: 1237–1245.

Mann, D. A., Trowern, A. R., Lavender, L., Whittaker, P. A. & Thompson, R. J. (1996) Identification of evolutionary conserved regulatory sequences in the 5' untranscribed region of the neural-specific ubiquitin C-terminal hydrolase (PGP9.5) gene. Journal of Neurochemistry. **66**: 35-46.

Maraganore, D. M., Farrer, M. J., Hardy, J. A., Lincoln, S. J., McDonnell, S. K. & Rocca, W. A. (1999). Case-control study of the ubiquitin carboxyl-terminal hydrolase L1 gene in Parkinson's disease. Neurology. **53**: 1858-860.

Marmor, M. D. & Yarden, Y. (2004). Role of protein ubiquitylation. in regulating endocytosis of receptor tyrosine kinases. Oncogene, **23**: 2057-2070.

Martin, E. R., Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H., Goetz, C. G., Small, G. W., Masterman, D., Mastaglia, F., Laing, N. G., Stajich, Ribble, R. C., Booze, M. W., Rogala, A., Hauser, M. A., Zhang, F., Gibson, R. A., Middleton, L. T., Roses, A. D., Haines, J. L., Scott, B. L., Pericak-Vance, M. A. & Vance, J. M. (2001). Association of single-nucleotide polymorphisms of the *tau* gene with late-onset Parkinson's disease. Journal of the American Medical Association. **286**: 2245-50.

Mattila, P. M., Rinne, J. O., Helenius, H., Dickson, D. W. & Roytta, M. (2000). α-Synuclein immunoreactive cortical Lewy bodies are associated with cognitive impairment in Parkinson's disease. Acta Neuropathology. **100**: 285-290.

Mayer, A. N. & Wilkinson, K. D. (1989). Detection, resolution and nomenclature of multiple ubiquitin carboxyl-terminal esterases from bovine calf thymus. Biochemistry. **28**: 166-172.

Maytal-Kivity, V., Reis, N., Hofmann, K. & Glickman, M. H. (2002). MPN+, a putative catalytic motif found in a subset of MPN domain proteins from eukaryotes and prokaryotes, is critical for Rpn11 function. BMC Biochem. **3**: 28.

McKeith, I. G., Ince, P., Jaros, E. B., Fairbairn, A., Ballard, C., Grace, J., Morris, C. M. & Perry, R. H. (1998). What are the relations between Lewy body disease and AD? Journal of Neural Transmission. Supplementum. **54**: 107-116.

McKeith, I. G., O'Brien, J. T. & Ballard, C. (1999). Diagnosing dementia with Lewy bodies. Lancet. **354**: 1227-1228.

McNaught, K. S. & Jenner, P. (2001). Proteasomal function is impaired in the substantia nigra in Parkinson's disease. Neuroscience Letters. **297**: 191-94.

McNaught, K. S., Olanow, C. W., Halliwell, B., Isacson, O. & Jenner, P. (2001). Failure of the ubiquitin-proteasome system in Parkinson's disease. Nature Review of Neuroscience. **2**: 589-94.

Mellick, G. D. & Silburn, P. A. (2000). The ubiquitin carboxy-terminal hydrolase-L1 gene S18Y polymorphism does not confer protection against idiopathic Parkinson's disease. Neuroscience Letters. **293** (2): 127-30.

Mezey, E., Dehejia, A., Harta, G., Papp, M. I., Polymeropoulos, M. H. & Brownstein, M. J. (1998). α-Synuclein in neurodegenerative disorders: murderer or accomplice? National Medicine. **4**: 755-57.

Misaghi, S., Galardy, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L. & Gaudet, R. (2005). Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. Journal of Biological Chemistry. **280**: 1512-1520.

Momose, Y., Murata, M., Kobayashi, K., Tachikawa, M., Nakabayashi, Y., Kanazawa, I. & Toda, T. (2001) Association studies of multiple candidate genes for Parkinson's disease using single nucleotide polymorphisms. Annals of Neurology. 51 (1): 133-36.

Momparler, R. L. & Bovenzi, V. (2000). DNA methylation and cancer. Journal Cell Physiology. **183**: 145.

Morris, J. R. & Solomon, E. (2004). BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. Human Molecular Genetics. **13**: 807-817.

Mori, N., Stein, R. & Anderson, D. J. (1990). A cell type-preferred silencer element that controls the neural-specific expression of the SCG10 gene. Neuron. **4**: 583-94.

Murray, L. L. & Lenz, L. P. (2001). Productive syntax abilities in Huntington's and Parkinson's diseases. Brain and Cognition. **46**: 213-219.

Myers, R. H., Vonsattel, J. P., Stevens, T. J., Cupples, L. A., Richardson, E. P., Martin, J. B. & Bird, E. D. (1988). Clinical and neuropathologic assessment of severity in Huntington's disease. Neurology. **38**: 341-7.

Naito, S., Mochizuki, H., Yasuda, T., Mizuno, Y., Furusaka, M., Ikeda, S., Adachi, T., Shimizu, H. M., Suzuki, J., Fujiwara, S., Okada, T., Nishikawa, K., Aoki, S. & Wada, K. (2006). Characterization of multimetric variants of ubiquitin carboxyl-terminal hydrolase L1 in water by small-angle neutron scattering. Biochemical and Biophysical Research Communications. **339**: 717-725.

Nanao, M. H., Tcherniuk, S. O., Chroboczek, J., Dideberg, O., Dessen, A. & Balakirev, M. Y. (2004). Crystal structure of human otubain 2. EMBO Journal. **8**: 783-88 (Reports 5).

Naze, P., Vuillaume, I., Destee, A., Pasquier, F. & Sablonniere, B. (2002). Mutation analysis and association studies of the ubiquitin carboxy-terminal hydrolase L1 gene in Huntington's disease. Neuroscience Letters. **328**: 1-4.

Newton, C. R. & Graham, A. (1994). PCR. BIOS Scientific Publishers.

Nicastro, G., Menon, R. P., Masino, L., Knowles, P. P., McDonald, N. Q. & Pastore, A. (2005). The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. Proceedings of the National Academy of Science - USA. **102**: 10493-10498.

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengård, J., Salomaa, V., Vartiainen, E., Boerwinkle' E. & Sing, C. F.(1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nature Genetics. **19** (3): 233-240.

Nijman, S. M. B., Luna-Vargas, M. P. A., Velds, A., Brummelkamp, T. R., Dirac, A. M. G., Sixma, T. K. & Bernards, R. (2005). A Genomic and Functional Inventory of Deubiquitinating Enzymes. Cell, **123**: 773-786.

Nishikawa, K., Li, H., Kawamura, R., Osaka, H., Wang, Y. –L., Hara, Y., Hirokawa, T., Monago, Y., Amano, T., Noda, M., Aoki, S. & Wada, K. (2003). Alterations of structure and hydrolase activity of parkinsonism-associated human ubiquitin carboxyl-terminal hydrolase L1 variants. Biochemical and Biophysical Research Communications. **304**: 176-83.

Oliva, D., Cali, L., Feo, S. & Giallongo, A. (1991). Complete structure of the human gene encoding neuron-specific enolase. Genomics. **10**: 157-165.

Orphanides, G., Lagrange, T. & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Genes & Development. **10:** 2657-2683.

Osaka, H., Wang, Y-L., Takada, K., Takizawa, S., Setsuie, R., Li, H., Sato, Y., Nishikawa, K., Sun, Y-J., Sakurai, M., Harada, T., Hara, Y., Kimura, I., Chiba, S., Namikawa, K., Kiyama, H., Noda, M., Aoki, S. & Wada, K. (2003). Ubiquitin carboxy-terminal hydrolase L1 binds to and stabilizes monoubiquitin in neuron. Human Molecular Genetics. **12**: 1945-1958.

Osawa, Y., Wang, Y. L., Osaka, H., Aoki, S. & Wada, K. (2001). Cloning, expression, and mapping of a mouse gene, *UCH-L4*, highly homologous to human and mouse *UCH-L3*. Biochemical and Biophysical Research Communications. **283** (3): 627-633.

Pagano, M. (1997). Cell cycle regulation by the ubiquitin pathway. The FASEB Journal. **11** (13): 1067-1075.

Pai, M.-C. & Chan S.-H. (2001). Education and cognitive decline in Parkinson's disease: a study of 102 patients. Acta Neurologica Scandinavica. **103** (4): 243-47.

Parker, R. & Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nature Structural & Molecular Biology. **11**: 121-7.

Paulsen, J. S., Zhao, H., Stout, J. C., Brinkman, R. R., Guttman, M., Ross, C. A., Como, P., Manning, C., Hayden, M. R., Shoulson, I. & the Huntington Study Group. (2001). Clinical markers of early disease in persons near onset of Huntington's disease. Neurology. **57**: 658-662.

Peng, X. R., Jia, Z., Zhang, Y., Ware, J. Trimble, W. S. (2002). The septin CDCrel-1 is dispensable for normal development and neurotransmitter release. Molecular Cell Biology. **22**: 378-87.

Piccini, P., Morrish, P. K., Turjanski, N., Sawle, G. V., Burn, D. J., Weeks, R. A., Mark, M. H., Maraganore, D. M., Lees, A. J. & Brooks, D. J. (1997). Dopaminergic function in familial Parkinson's disease: a clinical and 18F-dopa positron emission tomography study. Annals of Neurology. **41**: 222-29.

Piccini, P., Burn, D. J., Ceravolo, R., Maraganore, D. & Brooks, D. J. (1999). The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. Annals of Neurology. **45**: 577-82.

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. Annual Review of Biochemistry. **70**: 503-33.

Pickart, C. M. (2004). Back to the Future with Ubiquitin. Cell. **116**: 181-190.

Pickart, C. M., & Fushman, D. (2004). Polyubiquitin chains: polymeric protein signals. Curr. Op. Chem. Biol. **8**: 610-616.

Pickart, C. M. & Rose, I. A. (1985). Ubiquitin carboxyl-terminal hydrolase acts on ubiquitin carboxyl- terminal amides. Journal of Biological Chemistry. **260**: 7903-7910.

Polymeropoulos, M. H., Higgins, J. J., Golbe, L. I. *et al.* (1996). Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. Science. **274** (5290): 1197-99.

Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E. S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W. G., Lazzarini, A. M., Duvoisin, R. C., Di Iorio, G., Golbe, L. I. & Nussbaum, R. L. (1997). Mutation in the *α-synuclein* gene identified in families with Parkinson's disease. Science. **276** (5321): 2045-47.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature. **335**: 683 - 689.

Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. Nature. **386**: 569.

Puente, X. S. & López-Otín, C. (2004). A genomic analysis of rat proteases and protease inhibitors. Genome Research. **14**: 609-622.

Pugh, B. F. (1996). Mechanisms of transcription complex assembly. Current Opinion in Cell Biology. **8**: 303-311

Rajput, A. H., Offord, K. P., Beard, C. M. & Kurland, L. T. (1984). Epidemiology of parkinsonism: incidence, classification and mortality. Annals of Neurology. **16**: 278–82.

Rechsteiner, M. (2005). The 26S Proteasome. In: Protein Degradation – Ubiquitin and The Chemistry of Life (Vol. 1). Eds. Mayer, R. J., Ciechanover, A. & Rechsteiner, M. pp. 220-247 (Chapter 9). Wiley-VCH, Weinheim.

Reiss, O., Jakes, R. & Krugger, R. (1998). Molecular Medicine Today. **4**: 438-444.

Renkawitz, R. (1990). Transcriptional repression in eukaryotes. Trends in Genetics. **6**: 192-97.

Ribeiro, C. S., Carneiro, K., Ross, C. A., Menezes, J. R. & Engelender, S. (2002). Synphilin-1 is developmentally localized to synaptic terminals, and its association with synaptic vesicles is modulated by α-synuclein. Journal of Biological Chemistry. **277**: 23927–23933.

Rochet, J. C, & Lansbury, P. T. (2000). Amyloid fibrillogenesis: themes and variations. Current Opinion in Structural Biology. **10**: 60-68.

Rohrer, D., Salmon, D. P., Wixted, J. T. & Paulsen, J. S. (1999). The disparate effects of Alzheimer's disease and Huntington's disease on semantic memory. Neuropsychology. **13**: 381-388.

Rosati, G., Granieri, E., Pinna, L., Aiello, I., Tola, R., De Bastiani, P., Pirisi, A. & Devoto, M. C. (1980). The risk of Parkinson disease in Mediterranean people. Neurology. **30**: 250-55.

Rossetti, S., Chauveau, D., Walker, D., Saggar-Malik, A., Winearls, C. G., Torres, V. E. & Harris, P. C. (2002). A complete mutation screen of the ADPKD genes by DHPLC. Kidney International. **61**: 1588-99.

Sachidanandom, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., *et al*. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. **409**: 928-933.

Saeki, Y., Tayama, Y., Toh-e, A. & Yokosawa, H. (2004). Definitive evidence for Ufd2-catalyzed elongation of the ubiquitin chain through Lys48 linkage. Biochemical and Biophysical Research Communication. **320**: 840-845.

Saigoh, K., Wang, Y. –L., Suh, J. –G., Yamanishi, T., Sakai, Y., Kiyosawa, H., Harada, T., Ichihara, N., Wakana, S., Kikuchi, T. & Wada, K. (1999). Intragenic deletion in the gene encoding ubiquitin carboxy-terminal hydrolase in gad mice. Nature Genetics. **23**: 47-51.

Sarkis, C. & Mallet, J. (2005). Gene therapy. In: Neurodegenerative Diseases – Neurobiology, Pathogenesis & Therapeutics. Eds. Beal, M. F., Lang, A. E. & Ludolph, A. pp. 329 (Chapter 22). Cambridge University Press, Cambridge.

Satoh, J. & Kuroda, Y. (2001). A polymorphic variation of seine to tyrosine at codon 18 in the ubiquitin C-terminal hydrolase-L1 gene is associated with a reduced risk of sporadic Parkinson's disease in a Japanese population. Journal of Neurological Sciences. **189**: 113-117.

Satoh J.-I.; Kuroda Y. (2001). Ubiquitin C-terminal hydrolase-L1 (PGP9.5) expression in human neural cell lines following induction of neuronal differentiation and exposure to cytokines, neurotrophic factors or heat stress. Neuropathology & Applied Neurobiology. **27** (2): 95-104.

Sauerwald, A., Hoesche, C., Oschwald, R. & Kilimann, M. W. (1990). The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-less, phylogenetically conserved sequence with cell type-specific promoter function. Journal of Biological Chemistry. **265** (25): 14932-14937.

Savettieri, G., De Marco, E. V., Civitelli, D., Salemi, G., Nicoletti, G., Annesi, G., Cirò Candiano, I. C. & A. Quattrone, A. (2001). Lack of association between ubiquitin carboxy-terminal hydrolase L1 gene polymorphism and Parkinson's disease. Neurology. **57**: 560-561.

Sawle, G. (1999). Parkinsonism: Parkinsons's disease. In: Movement disorders in clinical practice. Ed. Sawle, G. pp. 7-33. Isis Medical Media Ltd, Oxford.

Schofield, J. N., Day, I. N. M., Thompson, R. J. & Edwards, Y. H. (1995). PGP9.5, a ubiquitin C-terminal hydrolase; pattern of mRNA and protein expression during neural development in the mouse. Developmental Brain Research. **85** (2): 229-238.

Schubert, U., Antón, L. C., Gibbs, J., Norbury, Yewdell, J. W. & Bennink, J. R. (2000). A large fraction of newly synthesized proteins are rapidly degraded by proteasomes. Nature. **404**: 770-774.

Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H., Goetz, C. G., Small, G. W., Masterman, D., Mastaglia, F., Laing, N. G., Stajich, J. M., Slotterbeck, B., Booze, M. W., Ribble, R. C., Rampersaud, E., West, S. G., Gibson, R. A., Middleton, L. T., Roses, A. D., Haines, J. L., Scott, B. L., Vance, J. M. & Pericak-Vance, M. A. (2001). Complete genomic screen in Parkinson's disease: evidence for multiple genes. Journal of the American Medical Association. **286**: 2239-44.

Seidler, A., Hellenbrand, W., Robra, B. P., Vieregge, P., Nischan, P., Joerg, J., Oertel, W. H., Ulm, G. & Schneider, E. (1996). Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. Neurology. **46** (5): 1275-1284.

Semple, C. A. M. (2003). The comparative proteomics of ubiquitination in mouse. Genome Research. **13**: 1389-1394.

Schindelin, H. (2005). Evolutionary Origin of the Activation Step During Ubiquitin-dependent Protein Degradation. In: Protein Degradation – Ubiquitin and The Chemistry of Life (Vol. 1). Eds. Mayer, R. J., Ciechanover, A. & Rechsteiner, M. pp. 21-43 (Chapter 3). Wiley-VCH, Weinheim.

Schlesinger, M. J. & Bond, U. (1987). Oxford Survey on Eukaryotic Genes. **4**. 77-91.

Selkoe, D. J. (2000). The genetics and molecular pathology of Alzheimer’s disease: roles of amyloid and the presenilins. In: Neurologic Clinics. Ed. DeKosky, S. pp. 903-21 (Volume 18). W. B. Saunders, Philadelphia.

Semchuk, K. M., Love, E. J. & Lee, R. G. (1991). Parkinson's disease and exposure to rural environmental factors: a population based case-control study. Canadian Journal of Neurological Sciences. **18**: 279-86.

Semchuk, K. M., Love, E. J. & Lee, R. G. (1992), Parkinson's disease and exposure to agricultural work and pesticide chemicals. Neurology. **42** (7): 1328-1335.

Shimura, H., Hattori, N., Kubo, S.-I., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K. & Suzuki, T. (2000). Familial Parkinson's disease gene product, parkin, is a ubiquitin-protein ligase. Nature Genetics. **25**: 302-5.

Shimura, H., Schlossmacher, M. G., Hattori, N., Frosch, M. P., Trockenbacher, A., Schneider, R., Mizuno, Y., Kosik, K. S. & Selkoe, D. J. (2001). Ubiquitination of a new form of α-synuclein by parkin from human brain: implications for Parkinson's disease. Science. **293**: 263-69.

Shoffner, J. M., Watts, R. L., Juncos, J. L., Torroni, A. & Wallace, D. (1991). Mitochondrial oxidative phosphorylation defects in Parkinson's disease. Annals of Neurology. **30** (3): 332-339.

Skipper, L & Farrer, M. (2002). Parkinson's Genetics: Molecular Insights for the New Millennium. Neurotoxicology. **23** (4): 503-514 (12).

Smale, S. T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica Biophysica Acta. **1351**(1-2): 73-88

Small, G. W.; Rabins, P. V., Barry, P. P., Buckholtz, N. S., DeKosky, S. T., Ferris, S. H., Finkel, S. I., Gwyther, L. P., Khachaturian, Z. S., Lebowitz, B. D., McRae, T. D., Morris, J. C., Oakley, F., Schneider, L. S., Streim, E., Sunderland, T., Teri, L. A. & Tune, L. E. (1997). Diagnosis and treatment of Alzheimer's disease and related disorders: consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society. The Journal of the American Medical Association. **278**: 16.

Smith, C. A. D., Gough, A. C., Leigh, P. N., Summers, B. A., Harding, A. E., Maraganore, D. M., Sturman, S. G., Schapiro, A. H. & Williams, A. C. (1992). Debrisoquie hydroxlase gene polymorphism and susceptibility to Parkinson's disease. Lancet. **339**: 1375-77.

Snider, J., Lazarevic, D. & Wolfenden, R. (2002). Catalysis by Entropic Effects: The Action of Cytidine Deaminase on 5,6-Dihydrocytidine. Biochemistry. **41**: 3925-3930.

Snowden, J., Craufurd, D., Griffiths, H., Thompson, J. & Neary, D. (2001). Longitudinal evaluation of cognitive disorder in Huntington's disease. Journal of the International Neuropsychological Society. **7**: 33-44.

Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. (1998). alpha-synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies**.** Proceedings of the National Academy of Science. **95**: 6469-6473.

Spillantini, M. G., Schmidt, M. L., Lee, V. M. –Y., Trojanowski, J. Q., Jakes, R. & Goedert, M. (1997). α-Synuclein in Lewy bodies. Nature. **388**: 839-40.

Stefanis, L., Larsen, K. E., Rideout, H. J., Sulzer, D. & Greene, L. A. (2001). Expression of A53T mutant but not wild-type α-synuclein in PC12 cells induces alterations of the ubiquitin-dependent degradation system, loss of dopamine release, and autophagic cell death. Journal of Neuroscience. **21**: 9549-60.

Sveinbjornsdottir, S., Hicks, A. A., Jonsson, T., Petursson, H., Guomundsson, G., Frigge, M. L., Kong, A., Gulcher, J. R. & Stefansson, K. (2000). Familial aggregation of Parkinson's disease in Iceland. New England Journal of Medicine. **343**: 1765-770.

Tan, E. K., Khajavi, M., Thornby, J. I., Nagamitsu, S., Jankovic, J. & Ashizawa, T. (2000). Variability and validity of polymorphism association studies in Parkinson's disease. Neurology. **55** (4): 533-538.

Tang-Wai, D. F., Josephs, K. A. & Petersen, R. C. (2005). Alzheimer's disease - overview. In: Neurodegenerative Diseases – Neurobiology, Pathogenesis & Therapeutics. Eds. Beal, M. F., Lang, A. E. & Ludolph, A. pp. 329 (Chapter 22). Cambridge University Press, Cambridge.

Taylor, P., Munson, K. & Gjerde, D. (1999) Detection of mutations and polymorphisms on the WAVE™ DNA fragment analysis system. Transgenomic Application Note. **101**.

The International HapMap Consortium. (2003). The International HapMap Project. Nature. **426**: 789-96.

Thompson, R. J. & Day, I. N. M (1988). Protein gene product 9.5: a new neuronal and endocrine marker, in *Neuronal and Glial Proteins*, Vol. 2 (Marangos, P. C., Campbell, I. C. & Cohen, R. M., eds), pp. 209-228. Academic Press, San Diego.

Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. EMBO J., **19**: 94–102.

Tjian, R. & Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. Cell. **77** (1): 5-8.

Tompkins, M. M., Basgall, E. J., Zamrini, E. & Hill, W. D. (1997). Apoptotic-like changes in Lewy-body-associated disorders and normal aging in substantia nigral neurons. American Journal of Pathology. **150**: 119-131.

Touchman, J. W., Dehejia, A., Chiba-Falek, O., Cabin, D. E., Schwartz, J. R., Orrison, B. M., Polymeropoulos, M. H. & Nussbaum, R. L. (2001). Human and mouse α-synuclein genes: comparative genome sequence analysis and identification of a novel gene regulatory element. Genome Research. **11**: 78-86.

Twelves, D., Perkins, K., & Counsell, C. (2003). Systematic review of incidence studies of Parkinson's disease. Movement Disorders. **18**: 19-31.

Ulrich, H. D. (2002). Natural Substrates of the Proteasome and Their Recognition by the Ubiquitin System. In: The Proteasome Ubiquitin Protein Degradation Pathway. Eds. Zwickl, P. & Baumeister, W. pp. 137-174 (chapter 6). Springer-Verlag, Berlin.

Valente, E. M., Valente, E. M., Bentivoglio, A. R., Dixon, P. H., Ferraris, A., Ialongo, T., Frontali, M., Albanese, A. & Wood, N. W. (2001). Localisation of a novel locus for autosomal recessive early-onset parkinsonism, *PARK 6*, on human chromosome 1p35-p36. American Journal of Human Genetics. **68**: 895-900.

van Duijn, C. M., Dekker, M. C., Bonifati, V., Galjaard, R. J., Houwing-Duistermaat, J. J., Snijders, P. J., Testers, L., Breedveld, G. J., et al. (2001). *Park 7*, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. American Journal of Human Genetics. **69**: 629-34.

van Duinen, S. G., Lammers, G. J., Maat-Schieman, M. L. & Roos, R. A. (1999). Numerous and widespread α-synuclein-negative Lewy bodies in an asymptomatic patient. Acta Neuropathology (Berlin). **97**: 533-539.

van Slegtenhorst, M., Lewis, J. & Hutton, M. (2000). The molecular genetics of the tauopathies. Experimental Gerontology. **35:** 461-71.

Wakabayashi, K., Engelender, S., Yoshimoto, M., Tsuji, S., Ross, C. R. & Takahashi, H. (2000). Synphillin-1 is present in Lewy bodies in Parkinson's disease. Annals of Neurology. **47**: 521-523.

Wang, J., Zhao, C. –Y., Si, Y. –M., Liu, Z. –L., Chen, B. & Yu, L. (2002) *ACT* and *UCH-L1* polymorphisms in Parkinson's disease and age of onset. Movement Disorders. **17** (4): 767-71.

Weinzierl, R. O. J. (1999). Gene-Specific Transcription Factors. In: Mechanisms of Gene Expression. Ed. Weinzierl, R. O. J. pp. 76-138 (Chapter 3). Imperial College Press, London.

Weissman, A.M. 2001. Themes and variations on ubiquitylation. Nature Reviews Molecular Cell Biology. **2**:169–178.

Welchman, R. L., Gordon, C. & Mayer, R. J. (2005). Ubiquitin and ubiquitin-like proteins as multifunctional signals. Nature Reviews Molecular Cell Biology. **6**: 599–609.

Wellington, C. L. & Hayden, M. R. (2000). Caspases and neurodegeneration: on the cutting edge of new therapeutic approaches. Clinical Genetics. **57**: 1-10.

West, A. B., Maraganore, D., Crook, J., Lesnick, T., Lockhart, P. J., Wilkes, K. M., Kapatos, G., Hardy, J. A. & Farrer, M. J. (2002). Functional association of the *parkin* gene promoter with idiopathic Parkinson's disease. Human Molecular Genetics. **11** (22): 2787-92.

Wicks, S. J., Haros, K., Maillard, M., Song, L., Cohen, R. E., Dijke, P. T. & Chantry, A. (2005). The deubiquitinating enzyme UCH37 interacts with Smads and regulates TGF-beta signalling. Oncogene. **24**: 8080–8084.

Wilkinson, K. D., Lee, K., Deshpande, S., Duerrksen-Hughes, P., Boss, J. M. & Pohl, J. (1989). The neuron-specific PGP9.5 is a ubiquitin C-terminal hydrolase. Science. **246**: 670-73.

Wilkinson, K. D., Deshpande, S. & Larsen, C. N. (1992). Comparisons of neuronal (PGP 9.5) and non-neuronal ubiquitin C-terminal hydrolases. Biochemical Society Transactions. 20: 631-37.

Wilkinson, K. D., Tashayev, V. L., O'Connor, L. B., Larsen, C. N., Kasperek, E. & Pickart, C. M. (1995). Metabolism of the polyubiquitin degradation signal. Structure, mechanism, and role of isopeptidase T. Biochemistry. **34**: 14535-14546.

Wilson, P. O. G., Barber, P. C., Hamid, Q. A., Power, B. J. & Polak, J. M. (1988). The immunolocalisation of protein gene product 9.5 using rabbit polyclonal and mouse monoclonal antibodies. British Journal of Experimental Pathology. **69**: 91.

Wilson, S. M., Bhattacharyya, B., Rachel, R. A., Coppola, V., Tessarollo, L., Householder, D. B., Fletcher, C. F., Miller, R. J., Copeland, N. G. & Jenkins, N. A. (2002). Synaptic defects in ataxia mice result from a mutation in Usp14, a ubiquitin-specific protease. Nature Genetics. **32**: 420-425.

Wintermeyer, P., Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Berg, D., Becker, G., Leroy, E., Polymeropoulos, M., Berger, K., Przuntek, H., Schols, L., Epplen, J. T. & Olaf, R. (2000). Mutation analysis and association studies of the UCH-L1 gene in German Parkinson's disease patients. Neuroreport. **11**: 2079-82.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M. & Darnell, J. E. Jr. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Research. **13**:1863-1872.

Yichin, L., Fallen, L., Lashuel, H. A., Liu, Z. & Lansbury, P. T. (2002). The UCH-L1 gene encodes two opposing enzymatic activities that affect α-synuclein degradation and Parkinson's disease susceptibility. Cell. **111**: 209-218.

Young, A. B., Shoulson, I., Penney, J. B., Starosta-Rubinstein, S., Gomez, F., Travers, H., Ramos-Arroyo, M. A., Snodgrass, R., Bonilla, E., Moreno, H. & Wexler, N. S. (1986).

Huntington’s disease in Venezuela: neurologic features and functional decline. Neurology. **36**: 244-9.

Zackzanis, K. K. (1998). The subcortical dementia of Huntington’s disease. Journal of Clinical Experimental Neuropsychology. **20**: 565-578.

Zhang, J., Hattori, N., Giladi, N. & Mizuno, G. (2000). Failure to find mutations in ubiquitin carboxy-terminal hydrolase L1 gene in familial Parkinson's disease. Parkinsonism & Related Disorders. **6** (4): 199-200.

Zhang, J., Hattori, N., Leroy, E., Morris, H. R., Kubo, S. –I., Kobayashi, T., Wood, N. W., Polymeropoulos, M. H. & Mizuno, Y. (2000). Association between a polymorphism of ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) gene and sporadic Parkinson's disease. Parkinsonism and Related Disorders. **6**: (4) 195-97.