Investigation of the molecular basis of three new disorders of brain growth and development identified amongst the Amish

Dr Emma Louise Baple

PhD Thesis

2014

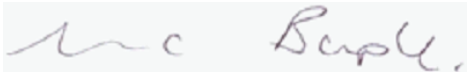
Investigation of the molecular basis of three new disorders of brain growth and development identified amongst the Amish

Submitted by

Dr Emma Louise Baple

to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Medical Studies, September 2014.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

(Signature) 

ACKNOWLEDGMENTS

First and foremost I would like to thank the families who took part in these studies and the Amish community for their support of the wider Windows of Hope project.

I am eternally grateful to my primary supervisor Professor Andrew Crosby, without whom I would never have had all the wonderful experiences this PhD brought me.

I would also like to extend my thanks to Professor Michael Patton who has always supported and encouraged me in whatever scientific or clinical endeavours I have followed. Professor Nigel Brown, Professor Steve Jeffery and Dr Paris Ataliotis based at St George's University of London provided invaluable support and advice throughout my PhD. Since arriving in Exeter, Dr Charles-Shaw Smith joined my supervisory team and I am hugely appreciative of his guidance and encouragement.

None of the work in this thesis would have been possible without the other members of the Windows of Hope project team, Professor Harold Cross, Roselyn Coblentz, Karin Wagner and honorary member Ray Coblentz.

I would like to thank all the members of the Crosby group for their help and support, but in particular Dr Barry Chioza, Dr Ajith Nair, Katy Barwick, Reza Maroofian and Gaurav Harlalka without whom this thesis would never have been completed.

During the course of this PhD I was fortunate to collaborate with many other scientists and clinicians and I would like to thank them all for their contributions. I am particularly grateful to Dr Catherine Green who has helped and encouraged me immeasurably, never giving up on trying to identify the biochemist within me. I would also like to specifically mention Dr Andrew Jackson who supported the MRC clinical fellowship that funded a substantial proportion of this work and welcomed me into his lab. Dr Sahar Mansour was a source of great support and help when describing the clinical phenotypes

described in this thesis and analysis of neuroradiology would not have been possible without Dr Phil Rich.

On a personal level I would like to thank my colleagues Dr Martina Muggenthaler and Dr Olivia Wenger for their friendship and constant helpful advice, my parents and brother who have always been there for me and my husband Peter Savage without whom none of this would have been possible.

The work in this thesis was funded by Newlife foundation for disabled children and the Medical Research Council.

ABSTRACT

Extremes of brain growth have frequently been associated with impaired neurodevelopment and cognition. A significant contribution to our understanding of the processes involved in brain development has been made by the study of single gene disorders which are rare in the general population, but occur with increased frequency in certain endogamous populations. This project stems from findings of a long-running clinical-genetic program, called 'Windows of Hope', based amongst the Amish. The project aims were to clinically characterise three new autosomal recessive disorders of brain growth and development, to define the genes and mutations responsible for each, and to investigate the function of the molecules identified. A final further objective of the project in keeping with the wider aims of the Windows of Hope study was to translate the findings of this research into direct clinical benefits to the families and community involved. A combination of clinical phenotyping, autozygosity and linkage mapping, and functional studies, were used to investigate each disorder, which enabled the identification of the novel disease genes in each case.

Chapter three describes the identification of a hypomorphic mutation in *PCNA* responsible for a novel DNA repair disorder. Thus, although it was considered by many that mutations in PCNA would not be compatible with life due to its crucial role in genomic stability and cell division, this study disproves this notion and describes a molecular 'missing piece' in DNA repair spectrum disorders.

While the relationship between pervasive developmental disorders and megalencephaly is well described, very few single gene disorders associated with this clinical relationship have been identified. Chapter four documents the discovery of two founder mutations in the *KPTN* gene associated with such a phenotype. The functional data shows that the encoded wild type molecule (kaptin) associates with dynamic actin cytoskeletal structures in cultured neurons and that the causative mutations result in loss of function perturbing this interaction, defining kaptin as a new molecule which is crucial for normal human brain development and function.

Chapter five details the investigation of the eponymously named “Hershberger syndrome”, a disorder originally described by McKusick in the Ohio Amish in 1967. Clinical and genetic studies of affected individuals revealed that the syndrome was comprised of two distinct disorders; Aicardi Goutières syndrome due to mutation in *SAMHD1,* and a new condition characterised by profound neurological impairment, cerebellar involvement and nephrosis. This condition, renamed “nephrocerebellar syndrome” was found to be caused by homozygous mutations in two closely linked genes, *WHAMM* and *WDR73.* The genetic and functional data supported the involvement of both molecules in the disease, suggesting that this is a composite phenotype.

The identification and functional characterisation of three new genes responsible for abnormalities in brain growth provides an invaluable insight into disease pathogenesis and also identifies molecular pathways important for normal brain development. This is turn enables clinicians to provide a much needed diagnosis for affected individuals and their families as well as the wider community.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3’	…….Three prime
5’	…….Five prime
ANOVA	…….Analysis of variance
APS	…….Ammonium persulphate
Arg	…….Arginine
AT	…….Ataxia telangiectasia
Bp	…….Base pair
BrdU	…….Bromodeoxyuridine
BSA	…….Bovine serum albumin
°C	…….Degrees Celcius
cDNA	…….Coding deoxyribonucleic acid
CS	…….Cockayne syndrome
CSC	…….Clinic for Special Children
CT	…….Computerised tomography
DAPI	…….4'6-diamidino-2-phenylindole
DMEM	…….Dulbecco's modified Eagle's medium
DMSO	…….Dimethylsulphoxide
DNA	…….Deoxyribonucleic acid
dNTP	…….Deoxyribonucleotide triphosphate
ds	…….double stranded

EBV …….Epstein–Barr virus

EDTA …….Ethylene diamine tetraacetic acid

ER …….Endoplasmic reticulum

EVA …….Exome based variant analysis suite

FCS …….Foetal calf serum

Fen1 …….Flap endonuclease 1

FEVA …….Family based exome variant analysis suite

FACS …….Fluorescence-activated cell sorting

γ Ray …….Gamma ray

GFP …….Green fluorescent protein

GG-NER …….Global genome nucleotide excision repair

Gln ……..Glycine

GST …….Glutathione S-transferase

HCL …….Hydrochloric acid

HEPES …….4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid

Het …….Heterozygous

His …….Histidine

hr …….Hour

IDCL …….Interdomain connector loop

Ile …….Isoleucine

IPTG …….Isopropyl-beta-D-thiogalactopyranoside

IQ …….Intelligence quotient

kb …….Kilobases

Kd …….Dissociation constant

KPTN …….Kaptin (Actin-Binding Protein)

L …….Litre

LB …….Luria Broth

Leu …….Leucine

Ig …….Immunoglobulin

Lig1 …….DNA ligase

Lys …….Lysine

NaCl …….Sodium chloride

NaOH …….Sodium hydroxide

NER …….Nucleotide excision repair

OD …….Optical density

μg …….Microgram

μL …….Microlitre

μm …….Micrometre

μM …….Micromolar

m …….Metre

M …….Molar

MeOH …….Methanol

Met …….Methionine

$MgCl_2$ …….Magnesium chloride

min	…….Minute
ml	…….Millilitre
mm	…….Millimetre
mM	…….Millimolar
MRI	…….Magnetic resonance imaging
MTT	…….3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NaCl	…….Sodium chloride
OFC	…….Occipitofontal circumference
PAGE	…….Polyacrylamide gel electrophoresis
PBS	…….Phosphate buffered saline
PBST	…….Phosphate Buffered Saline Tween-20
PCNA	…….Proliferating cell nuclear antigen
PCR	…….Polymerase chain reaction
PFA	…….Paraformaldehyde
PIP	…….PCNA-interaction peptide
Pro	…….Proline
RNA	…….Ribonucleic acid
rpm	…….revolutions per minute
RRS	…….Recovery of RNA synthesis
RT	…….Room temperature
RT-PCR	…….Reverse transcriptase polymerase chain reaction

RU …….Response units

SAMHD1 …….SAM domain and HD domain-containing protein 1

SAP …….Shrimp alkaline phosphatase

SD …….Standard deviation

SDM …….Site directed mutagenesis

SDS …….Sodium dodecyl sulphate

SDS …….Standard deviation score

SDS-PAGE …….sodium dodecyl sulfate - Polyacrylamide gel electrophoresis

Sec …….Second

SEM …….Standard error of the mean

Ser …….Serine

SILAC …….Stable isotope labelling by/with amino acids in cell culture

siRNA …….Small interfering RNA

SNP …….Single nucleotide polymorphism

SPR …….Surface plasmon resonance

TAE …….Tris-acetate-EDTA

TBE …….Tris-Borate-EDTA

TC-NER …….Transcription coupled nucleotide excision repair

TE …….Tris-EDTA

TEMED …….N,N,N^1,N^1-tetramethyl-ethylenediamine

TetR …….Tetracycline- and doxycycline-responsive repressor protein

Tris	…….Tris(hydroxymethyl)aminomethane
TTD	…….Trichothiodystrophy
Tween-20	……. Polysorbate 20
UCSC	…….University of California, Santa Cruz
UDS	…….Unscheduled DNA synthesis
UV	…….Ultraviolet light
V	…….Volt
Val	…….Valine
WDR73	…….WD repeat domain 73
WHAMM	…….WAS protein homolog associated with actin, golgi membranes and microtubules
WHO	…….World health organization
WOH	…….Windows of Hope Amish inherited disease project
WT	…….Wild type
w/v	…….Weight per volume
X-Gal	…….5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
XP	……. Xeroderma pigmentosum

1

CHAPTER ONE

INTRODUCTION

1 INTRODUCTION

1.1 Brain embryological development

The three germ cell layers: ectoderm, mesoderm and endoderm, are established by the end of the second week of human embryonic development. Each layer then gives rise to a particular subset of tissues and organs. It is the ectoderm that gives rise to the nervous system as well as the skin.

Neurulation is the name given to the process of embryonic nervous system formation. The neural plate is formed from a thickening of the dorsal midline ectoderm during the third week of embryonic development, the lateral margins then elevate to create the neural folds. The neural folds subsequently come together and fuse forming the neural tube, which will be fully formed by the end of the 4^{th} week (Figure 1). Some cells from the neural folds subsequently separate to form the neural crests which lie dorsolateral to the neural tube.

The brain is formed from the rostral part of the neural tube, with the caudal part becoming the adult spinal cord. The central cavity of the neural tube will eventually become the spinal canal and the cerebral ventricles. As embryonic development progresses a longitudinal groove, the sulcus limitans develops on the inner surface of the lateral walls of the caudal part of the embryonic brain and inner surface of the lateral walls of the spinal cord. The dorsal and ventral cells become delineated as the alar and basal plates respectively (Figure 2). Cells within the alar plate will generally become part of the sensory system and those of the basal plate will have predominantly motor functions. Around the central part of the canal differentiation occurs to form the grey matter, with the white matter forming as an outer coat.

During the fifth week of embryonic development, the three primary brain vesicles can be identified: the prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain). Further differentiation and division of the primary vesicles creates five secondary brain vesicles. The prosencephalon becomes the telencephalon and the diencephalon and the rhombencephalon divides into the metencephalon and the myelencephalon.

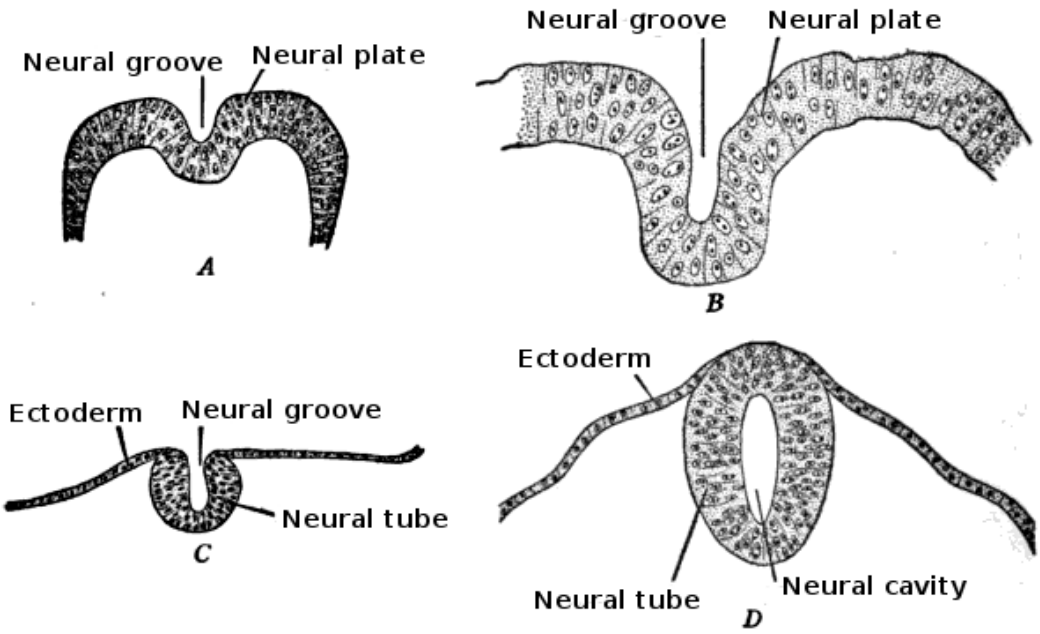


Figure 1: Development of the neural tube. Reproduced from (Ranson, 1920)

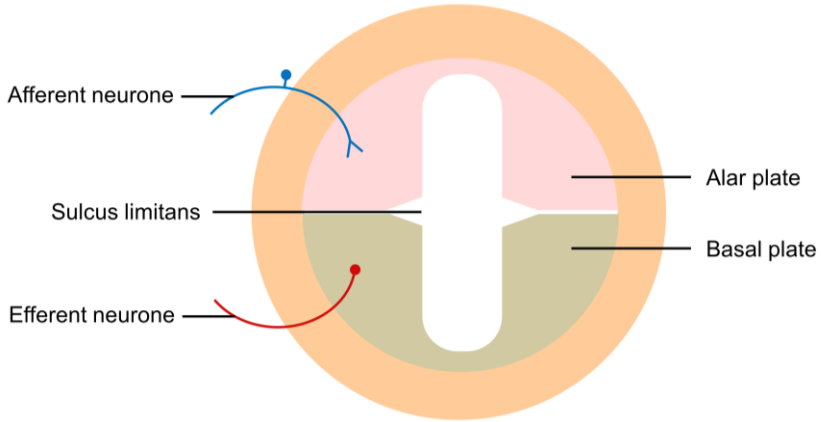


Figure 2: Schematic of the developing neural tube. Modified (Crossman and Neary, 2010)

In the adult brain, the cerebral hemispheres originate from the telencephalon, the thalamus is derived from the diencephalon, the midbrain from the mesencephalon, the pons and cerebellum develop from the metencephalon and the medulla oblongata is derived from the myelencephalon.

The cerebral hemispheres consist of an inner layer of grey matter, termed the cerebral cortex, with the inner brain mass comprised of white matter. The two hemispheres are separated by a cleft, the great longitudinal fissure. The midbrain, pons and medulla oblongata are collectively referred to as the brainstem. The central cavity within the brain becomes the ventricular system containing cerebral spinal fluid (CSF) (Crossman and Neary, 2010).

1.1.1 Cerebral cortex

The cerebral cortex becomes highly convoluted forming a complex pattern of ridges (gyri) and furrows (sulci) in order to maximise the surface area. Each hemisphere is divided into four lobes: the frontal, parietal, temporal and occipital lobes. The evolutionary origins of the cerebral cortex relate to olfactory function. The phylogenetically old parts of the cortex (referred to as the paleocortex) which include parts of the temporal lobes and hypothalamus, retain an association with the olfactory system and have a primitive three layered cytoarchitecture. These parts of the brain are functionally involved with the emotional aspects of memory and behaviour. The phyogenetically more recent acquisition, the neocortex, comprises the majority of the cerebral cortex, and the neurones within it form a highly organised laminar and radial structure. The adult neocortex is comprised of six layers, within which lie specialised neurones with specific synaptic connections and phenotypes. The fundamental functions of the cerebral cortex are necessary for thought, memory, intellect and conscious awareness. The cerebral cortex also plays a central role in the motor system, as the site where actions are conceived and initiated.

During the first half of human gestation, the neuronal cells that eventually make up the cerebral cortex undergo a process of migration from near the surface of the cerebral ventricles to reach the neocortex. After their last mitotic division, these neuronal cells migrate passed previously towards the pial surface of the cortex to form the cortical plate (Figure 3).

The cortical neurone progenitor cells are confined within a transient pseudoepithelium composed of the narrow ventricular and subventricular zones. These two zones also contain a population of elongated radial glial cells that span the entirety of the embryonic cerebral wall and only exist during the phase of neuronal migration. Each successive generation of neuronal cells migrate past previously formed cells towards the region between the marginal zone and cortical plate. During their migration the neurones use the radial glial cells as a scaffolding to guide their way through the different cortical layers. Radial glial cells can give rise to neurones and eventually astrocytes and oligodendrocytes (Campbell and Gotz, 2002, Rakic, 2007, Gorski et al., 2002). These glial cells typically undergo asymmetrical division producing a post mitotic neurone and a radial glial cell or alternatively a radial glial cell and an intermediate progenitor within the subventricular zone. Intermediate progenitors, like radial glial cells, are a significant neuroprogenic cell population and undergo a symmetrical division to produce two post mitotic neurones (Arnold et al., 2008, Sessa et al., 2010, Sessa et al., 2008).

The radial unit hypothesis of cortical development postulates that the horizontal location of a cortical neurone is determined by the position of its precursor cells in the ventricular zone, and depth is determined by birth order. The consequences of this are that the earliest formed neurones are found in the deeper layers and the latest in the more superficial cortical layers forming columns of interrelated neurones. Consequently the characteristic six layered neocortex is formed from the inside out. Once neurones reach their final laminar positions, they develop synaptic connections with neurones close by as well as more distant neurones in associated regions of the cerebral cortex (Rakic and Lombroso, 1998).

The size of the cortex is critically dependent on the number of neuroprogenitor cells. Any disruption to this cell pool at early stages of cortical development can result in significant differences in cortical size (Sun and Hevner, 2014). The regulation of cell cycle progression plays an important role in the control of neuroprogenitor expansion. For example, knockout of the cytoskeletal associated gene filamin A (*flna*) elongates the cell cycle and reduces the number of neuroprogenitors resulting in a subsequent reduction in cortical size

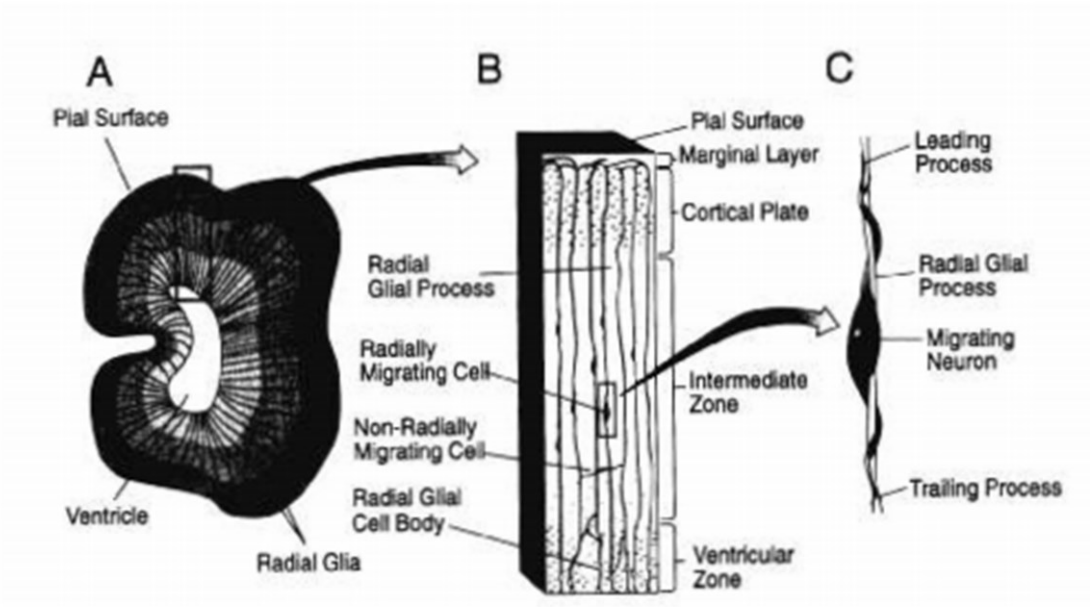


Figure 3: Normal development of the cerebral cortex.

(**A**) Section through the developing primate forebrain showing distribution pattern of radial glial processes that span the fetal cerebral wall from the ventricle to the pial surface. (**B**) Enlargement of the boxed area in (A) to illustrate how neurons migrate from their birth place in the ventricular zone across the intermediate zone to their final destination at the interface between the marginal zone and the developing cortical plate. (**C**) Neuroblasts use the surface of elongated radial glial fibers as a guide during their migration. Reproduced with permission from (Rakic and Lombroso, 1998)

(Lian et al., 2012).

Cilia are important for normal neuroprogenitor protein trafficking and also appear to have a possible role in the cell cycle. Perturbation of normal cilial function as is seen in the emerging group of ciliopathy disorders such as the genetically heterogeneous Joubert syndrome which encompases a range of different brain malformations (Romani et al., 2013). Increased rates of apoptosis in neuronal progenitor cells result in a reduction of cortical size, whereas a decreased rate has the obverse effect. *Brca1* knockout mice have a substantial reduction in cortical size and thickness which results from apoptosis through the activation of p53 at the early stage of neurogenesis (Sun and Hevner, 2014). Caspase genes are an important part of the cellular apoptopic pathway. Knockout of Casp 9 in mice leads to a reduction in apoptosis during early brain development and enlargement and malformation of the brain with significant rates of perinatal mortality (Kuida et al., 1998). Neuroprogenitor cells can be either polarised or unpolarised, undergoing either symmetrical or assymetrical divisions respectively (Sun and Hevner, 2014). Consequently the orientation of mitotic spindles is an important aspect of normal cortical development. The dynein-associated nuclear distribution E homologue 1 (NDE1) plays a central role in microtubule organisation and mitotic spindle position and is highly expressed at the centrosome accumulating on the mitotic spindle of apical neural precursors in early neurogenesis (Alkuraya et al., 2011). Knockout of Nde1 in mice results in abnormal mitotic orientation and progression in cortical neuroprogenitors, leading to microcephaly, most significantly affecting the cerebral cortex (Feng and Walsh, 2004).

1.1.2 Cerebellum

The cerebellum originates from the dorsal aspect of the brain stem and is the largest part of the hindbrain. The cerebellum operates at an unconscious level and is solely involved in motor function. This part of the brain plays a fundamental part in the control and maintenance of balance, muscle tone, the coordination of movements and posture.

The cerebellum consists of two hemispheres which are joined in the midline by the vermis. During embryonic development, the surface becomes highly

convoluted forming transverse folds called folia. Anatomically, the cerebellum is divided into three lobes: the small anterior lobe, larger posterior lobe and the flocculonodular lobe. The outer part of the cortex is comprised of grey matter with an inner core of white matter.

The histological makeup of the cerebellar cortex is consistent throughout. It comprises three layers: an outer molecular layer, intermediate purkinje cell layer and an inner granular layer. The white matter of the cerebellum contains the four cerebellar nuclei (fastigal, globose, emboliform and dentate), which are the primary source of efferent fibres from the cerebellum to other parts of the brain. Clinical signs associated with disruption to normal cerebellar development or functions include dysarthria, cerebellar ataxia, intention tremor and nystagmus.

Phylogenetically the cerebellum is divided into the archicerebellum, paleocerebellum and neocerebellum, with the oldest part in evolutionary terms being the archicerebellum and newest the neocerebellum. The archicerebellum is primarily concerned with normal balance maintenance and comprises the flocculonodular lobe and the fastigal nucleus. The paleocerebellum is central to normal maintenance of muscle tone and is equated with the midline vermis, surrounding paravermis and the globose and emboliform nuclei. The neocerebellum represents the largest part of the cerebellum approximates to the majority of the cerebellar hemispheres and the dentate nuclei, and is responsible for control of muscle coordination (Crossman and Neary, 2010).

1.2 Abnormalities of brain growth

Human brain growth and development involves a precisely controlled cascade of molecular and cellular events that involve genetic, epigenetic and environmental factors. The simplest definition of growth is an increase in size. One of the most important differences between organisms is their brain size. The expansion in size of the cortical surface of the brain that has occurred during evolution to humans is particularly remarkable. During the 23 million years since humans and macaque monkey diverged from a common ancestor, there has been around a 15 fold increase in the brain's cortical surface. Even more significantly, there is a more than 1000 fold difference in cortical size

between mice and humans (Rakic, 1995). Disorders of brain growth and development underlie a significant proportion of intellectual disability and autistic spectrum disorders (Mirzaa and Paciorkowski, 2014).

1.3 Occipitofrontal circumference

When age and sex are considered, head circumference, or occipitofrontal circumference (OFC), correlates well with intracranial volume and with cognitive function. OFC measurements can be a helpful indicator of underlying brain pathology to the clinician, however currently there remains a shortage of population specific occipitofrontal circumference data. Nevertheless, occipitofrontal circumference reference charts remain a useful clinical tool (Stevenson et al., 1997).

Until the late 1990's the Tanner-Whitehouse charts were the standard centile charts used to assess head circumference in the United Kingdom. To update the British growth reference, anthropometric data for weight, height, body mass index and head circumference from 17 distinct surveys representative of England, Scotland and Wales (37,700 children, age range 23 weeks gestation to 23 years) were analysed by maximum penalised likelihood using the LMS (lambda-mu-sigma) method. The British 1990 growth reference provides centiles for weight, height and body mass index from birth to 23 years. Measurements can be converted to normally distributed standard deviation scores. The reference sample comprised Caucasian British children with none excluded for health reasons making it a growth reference and not a growth standard. Included are 6444 head circumference measurements for boys and 4917 for girls, aged between 23 weeks gestation to 18 years (males) or 17 (females). The Edinburgh Growth Study, a longitudinal survey of children residing in Edinburgh followed from birth to age 20, was also included to provide additional head circumference reference data, the growth charts that resulted from this endeavour are referred to as the UK90 reference charts. (Cole et al., 1998).

In 1992, Bushby et al also suggested that the use of the Tanner-Whitehouse charts for adult head circumference was inappropriate particularly for males,

given that reference data was only available up to the age of 16 and head growth would continue after this time, moreover others had previously concluded that head circumference and brain weight were correlated with height in adulthood as well as childhood (Bushby et al., 1992). They produced centile charts for head circumference against height based on the measurements of 354 adults in two British centres (Newcastle upon Tyne and Cardiff), 159 males and 195 females attending dental clinics were measured, those with conditions likely to affect cranial anatomy were excluded. The centile charts constructed from these measurements indicate that adult head circumference is related to height. The mean head circumference of a male of average height being above the 97th centile for a 16 year old on the Tanner-Whitehouse charts. Unfortunately this study has never been repeated and the number used to construct the centile charts was small.

In 2007 the World Health Organisation released growth charts intended as an international tool based on data produced from the Multicentre Growth Reference Study undertaken between 1997 and 2003. The study included children from 6 countries: Brazil, Ghana, India, Norway, Oman, and the United States. All the children included were healthy and environmental factors like breast feeding and exposure to smoking were tightly controlled. The resulting occipitofrontal charts only go to age 5 years and are a growth standard as opposed to a growth reference (Rollins et al., 2010). Moving to a different chart, such as UK90 after age 5 years creates difficulties in monitoring children's head growth as the two tools are not directly comparable.

Natale and Rajagopalan compared data from the World Health Organisation's Multicentre Growth Reference Study with data from studies performed in 55 countries or ethnic groups. They found that the mean head circumference varied widely between different national and ethnic groups. In many groups, means were consistently 0.5–1 standard deviation above the Multicentre Growth Reference Study mean. Head size in breastfed children at any age examined was far closer to local norms than to the Multicentre Growth Reference Study means. The differences between national or ethnic group head circumference means were large enough that using the WHO charts would put many children at risk for misdiagnosis of macrocephaly or

microcephaly. Their findings indicate that the use of a single international standard for head circumference is not justified (Natale and Rajagopalan, 2014).

Head circumference is not often measured routinely beyond infancy, and then usually only for monitoring of head growth. The Tanner-Whitehouse and UK90 references show striking differences from each other beyond this age range, with the UK90 reference on average more than one standard deviation higher. The two curves differ significantly around the time of puberty. This means that neither of these references are robust for diagnosing micro or macrocephaly after the age of 2 years or for assessing head growth after the onset of puberty.

1.4 Microcephaly

Microcephaly is defined as a head circumference greater than or equal to three standard deviations (SDS) below the mean. Microcephaly can be either congenital or postnatal in onset and is either the consequence of impaired brain growth during early development, or a destructive neurological process. In itself microcephaly is a relatively non-specific finding, with a number of both environmental and genetic aetiologies including single gene disorders, chromosome anomalies, metabolic conditions, teratogen exposure, congenital infection and birth asphyxia (Murray and Jackson, 2012).

1.4.1 Congenital Microcephaly

Microcephaly is frequently associated with structural brain malformation or other developmental anomalies. However, rarely a significant and specific reduction in brain size with grossly preserved macroscopic appearance is seen in the absence of additional features. Typically this presentation is associated with autosomal recessive inheritance. This phenotype has been termed primary microcephaly (MCPH). Despite the marked reduction in cerebral volume to sometimes as little as a third of normal, intellect is often relatively preserved with most individuals only mild/moderately impaired (Murray and Jackson, 2012, Kaindl et al., 2010, Thornton and Woods, 2009).

To date a total of 11 genes have been shown to be associated with the primary microcephaly phenotype (Verloes et al., 1993). The majority of the protein products of these genes have related functions and play an important role in the mitotic phase of the cell cycle localising to the centrosomes and mitotic spindle poles. It is thus proposed that primary microcephaly is a disorder of neurogenic mitosis, with a reduction in neuroprogenitor cells resulting in fewer mature neurones and a reduced cerebral volume (Murray and Jackson, 2012). The centrosome is one of the main microtubule organising centres in the cell and centrosome duplication and maturation are intimately linked to cell-cycle progression (Thornton and Woods, 2009). Elucidation of the molecular causes of primary microcephaly have revealed a wide spectrum of phenotypes associated with the causative genes. *WDR62,* for example, also results in a variety of structural brain disorders including lissencephaly, cerebellar hypoplasia and hypoplasia of the corpus callosum (Yu et al., 2010). The most profound reduction in brain size is seen in patients homozygous for loss of function mutations in the centrosomal protein NDE1 with brain size less than 10% of normal. Affected individuals are profoundly neurologically impaired and cortical lamination patterns are also abnormal with microlissencephaly seen on neuroimaging (Guven et al., 2012).

Interestingly, two of the genes associated with primary microcephaly, *CEP152* (Kalay et al., 2011, Guernsey et al., 2010) and *CENPJ* (Bond et al., 2005), have also been associated with the microcephalic primordial dwarfism family of disorders, a group of conditions with global growth failure. Other genes associated with microcephalic primordial dwarfism shown to cause defects in centrosomal and spindle microtubule function include: *PCNT*, *CENPE* and *POC1A* (Griffith et al., 2008, Shaheen et al., 2012, Mirzaa et al., 2014). However other genes implicated in microcephalic primordial dwarfism have more diverse functions and include disorders of genome replication associated with (*ORC1, ORC4, ORC6, CDT1, and CDC6*) (Bicknell et al., 2011b, Bicknell et al., 2011a), DNA damage response (*ATR, Lig4*) (O'Driscoll et al., 2003, Murray et al., 2014), growth hormone axis (*IGF1*)(Woods et al., 1996) and mRNA splicing (*RNU4atac*) (Edery et al., 2011).

Postnatal onset of microcephaly comprises a more diverse set of disorders including Angelman syndrome, Rett syndrome and Christianson syndrome. In some cases specific biological pathways in which defects lead to postnatal microcephaly can be identified for example the nucleotide excision DNA repair pathways discussed in more detail in chapter 3. (Seltzer and Paciorkowski 2014).

1.5 Megalencephaly

The definition of Megalencephaly is a brain size of greater than 2 standard deviations of the age and sex related mean. Megalencephaly is divided into developmental/anatomical types and metabolic types (Olney, 2007) Neurometabolic syndromes such as cerebral organic acid disorders including Glutaric aciduria type I, disorders of lysine metabolism, lysosomal storage disorders, mucopolysacharidoses and leukoencephalopathies cause megalencephaly as a result of accumulation of metabolic substrates, cellular hypertrophy or oedema (Mirzaa and Poduri, 2014).

Anatomical megalencephaly can be an isolated feature or associated with a more generalised overgrowth phenotype such as in Soto's syndrome, Weaver syndrome, Simpson Golabi Behmel (Olney, 2007) and DNMT3A related overgrowth (Tatton-Brown et al., 2014).

The principle pathway implicated in isolated megalencephaly is PI3K-AKT-mTOR, which includes PTEN related disorders. Molecules within this pathway have important roles to play in the regulation of both the cell cycle and apoptosis. Dysregulation can cause significant increase in brain size, typically 4-5 SDS above the mean, but occasionally as much as 8 SDS. Mutations perturbing the PI3K-AKT-mTOR pathway can be germline as with *PTEN* and *STRADA* (Parker et al., 2013), but are predominantly somatic and mosaic. Post zygotic activating mutations in *PIK3CA, AKT3* and *mTOR* have all been shown to result in unilateral hemispheric enlargement termed hemimegalencephaly (Lee et al., 2012, Poduri et al., 2012). Hemimegalencephaly is usually accompanied by profound neurological impairement and severe early onset intractable seizures (Mirzaa and Poduri, 2014).

1.6 Classification of intellectual disability

The definitions of intellectual disability (previously termed mental retardation) and developmental delay are rather complex and have evolved over time. The reasons for this evolution relate to changes in public and professional acceptability of certain terminology, for example mental retardation. In addition there have been significant advances in our understanding of the underlying pathophysiology of these disorders much of which relates to the rapid progress made in the field of neurogenetics, neurobiology, neuroimaging and epidemiology (APA, 2013).

The classification of mental disorders, known as psychiatric nosology is a key aspect of psychiatry. The responsibility for classification of neurodevelopmental disorders like intellectual disability has thus typically lain with this field. The two most widely recognised classification systems for mental disorders are ICD and DSM.

The history of the International Statistical Classification of Diseases and Related Health Problems (ICD) dates back over a century and has been the responsibility of the World Health Organisation since it was founded in 1948, the current iteration is ICD-10, although this is under revision and ICD-11 is expected to be released in 2017.

ICD used as the global health information standard for statistics relating to morbidity and mortality, which can then be used to identify global health trends. It is increasingly used as a diagnostic tool in healthcare and ICD-10 has been translated into 43 languages and is used by over 100 countries worldwide. Unlike the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM), it is free to access through the worldwide web. Around 70% of health expenditures worldwide are allocated using statistics collected through use of the ICD (www.who.int/classifications/icd/factsheet/en/).

The history of the DSM dates back to 1844, when the American Psychiatric Association published a statistical classification of institutionalised mental patients, the first predecessor to DSM. The aim of this was to improve communication between institutions about the different types of disorders they

saw. The DSM evolved to be a diagnostic classification system describing the full range of mental health disorders to guide diagnosis and inform management strategies, the recently revised 5th edition DSM-V was released towards the end of 2013 (APA, 2013). One of the most important differences between the ICD and DSM is that the ICD is produced by a global agency with a constitutional public health mission, whereas the DSM is produced a single national professional psychiatric association. Both the American Psychiatric Association and the World Health Organisation would like to see the harmonisation of DSM–V and ICD–11. Small differences in wording, number of symptoms necessary for diagnosis as well as exclusion criteria can lead to significant diagnostic discordance. The WHO has formed a DSM–ICD Harmonisation Coordinating Group, which comprises members of the DSM–V Task Force and the International Advisory Group for the Revision of ICD–10 with the aim being to facilitate uniformity between the two classification schemes (First, 2009).

The diagnosis of intellectual disability (intellectual developmental disorder) is revised from the DSM-IV diagnosis of mental retardation and will be similarly revised in ICD-11 to intellectual developmental disorders. Currently within ICD-10 it remains under Chapter V – Mental and behavioural disorders, subsection; Mental retardation (F70-79), where is classified as mild, moderate, severe or profound with subdivisions to identify behavioural impairment. ICD-10 defines mental retardation as “A condition of arrested or incomplete development of the mind, which is especially characterised by impairment of skills manifested during the developmental period, skills which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities. Retardation can occur with or without any other mental or physical condition.” The primary assessment tool suggested is a standardised Intelligence test, although it is also recommended that social adaption assessments can also be used to aid diagnosis.

Within DSM-5 the importance of using both clinical assessment and standardised testing of intelligence when diagnosing intellectual disability is emphasised. The classification of severity of impairment is now based on adaptive functioning rather than Intelligence quotient (IQ) test scores alone. The overarching aim of this revision was to avoid an over emphasis of IQ scores and

to ensure that a person's functional ability was the primary consideration by removing them from the diagnostic criteria but retaining them within the text. The American Psychiatric Association believe that these changes will permit a more comprehensive assessment across conceptual, social, and practical domains and result in a diagnosis that accurately reflects the impairment of an individual's ability to function in day to day life resulting from their deficit in general mental abilities. Intellectual disability is listed in DSM-V under the chapter heading: Neurodevelopmental disorders, which they define as " a group of conditions with onset in the developmental period". Three criteria must be met for a diagnosis to be made including; (A) deficits in general intellectual functioning, (B) deficits in everyday adaptive functioning when compared to age, gender and socioculturally matched peers and (C) onset of these deficits in the developmental period. Severity is graded as with the ICD into mild, moderate, severe and profound. In summary, the significant changes within DSM-V address the naming of a disorder, diagnostic criteria and impact on an individual's ability to function (APA, 2013).

1.6.1 Developmental delay

Developmental delay can be either specific or global. A specific developmental delay is when one area of development is affected. Significant delay is defined as performance of two standard deviations or more below the mean on age-appropriate, standardised norm-referenced developmental screening or assessment tests. Global developmental delay is generally defined as a significant delay in two or more developmental domains: gross and fine motor; speech and language; cognition; personal and social development; or activities of daily living. The extent of delay can be classified as mild if functional age is 33% below chronological age, moderate if functional age is 34–66% of chronological age, and severe if functional age is 66% below chronological age (McDonald et al., 2006, Silove et al., 2013, Shevell et al., 2000). Developmental assessment tools commonly used to assess the level of developmental functioning include the Griffiths Mental Development Scales and the Denver II Developmental screening test (Frankenburg et al., 1992). DSM-V lists global developmental delay as a diagnosis within neurodevelopmental disorders, but states clearly that is a term reserved for individuals < 5 years old. The DSM-V

classification of neurodevelopmental disorders is summarised in Figure 4 (APA, 2013).

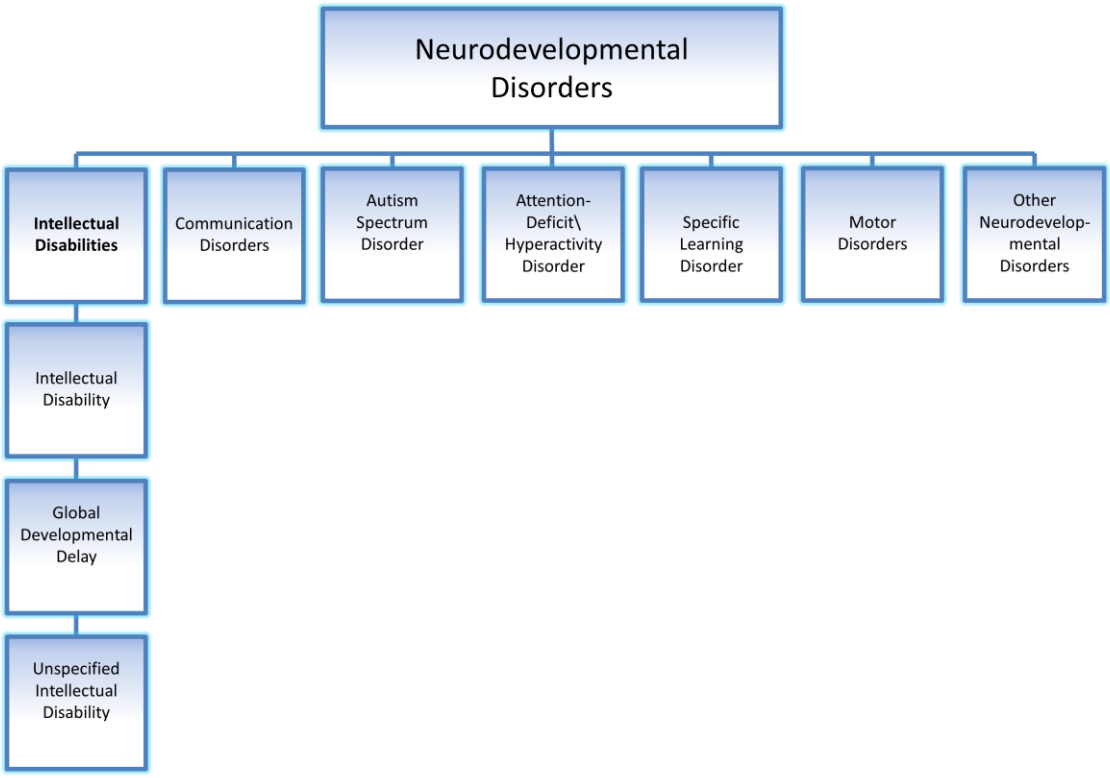


Figure 4: DSM-V classification of neurodevelopmental disorders.

1.7 The Amish

The Amish are an Anabaptist Christian church that traces its roots to the Protestant Reformation in sixteenth-century Europe. The church arose in 1693 after a schism within a group of Swiss Anabaptists in Alsace, led by the Bishop Jakob Amman. The term Anabaptist means “rebaptisers” and refers to the rejection of infant baptism, a practice that linked the church with citizenship, in favour of adult baptism where individuals could make their own decision about church membership. Amman wished to make a number of reforms to church practices; most controversially he wanted closer links between church discipline and social practices. If a member of the church was excommunicated because of an unrepented sin, Amman advocated that other members of the church should avoid, or shun the individual, in ways such as refusing to share a meal with them. This practice of shunning or “Meidung”, was not meant as a punishment, but was a lesson, a way of showing the person the seriousness of the offence to encourage them to confess and seek forgiveness. Taking his name, Amman’s followers were first called the “Amman-ish” group, and later became known as the Amish. In 1681 the Quaker William Penn founded the state of Pennsylvania on the principles of toleration and peace. The first wave of Amish migration to the state was between 1736 and 1770 and comprised around 500 individuals, largely seeking to avoid religious persecution. A second wave of Amish immigrants left Europe for the “New World” between 1815-1860. The number of migrants is estimated to have been around 3000. Compulsory military service was becoming commonplace in Europe and this combined with the economic opportunities in North America was a strong driving force to leave and seek a new life. The majority of these 3000 newcomers did not remain in Pennsylvania. Land prices were cheaper in the west and so they travelled to the states of Ohio, Indiana, Iowa and Illinois and Ontario in Canada. It is important to remember that these original settlers were already part of related family groups, so the overall gene pool that founded the modern day Amish was particularly modest.

The first three Amish settlements that were established were in eastern Pennsylvania, Indiana and Ohio. In 2014, 64% of the total Amish population still live in those three states. The Amish population doubles about every 18 to 20

years and grew by an estimated 46,000 from 235,000 to 282,000 between 2008 and 2013 (http://www2.etown.edu/).

Today, Amish people accept basic Christian beliefs but also have some of their own emphases and special interpretations that have emerged throughout their history. Professor Donald Kraybill, based at the Young Center for Anabaptist and Pietist Studies at Elizabethtown College, Pennsylvania is the world's foremost expert on the Amish. Kraybill defines the modern day Amish as: "any group that affirms the basic tenants of Amish belief, including adult baptism, non-violence and separation from the outside world", "uses horse-drawn transportation" (Figure 5), "speaks a German-derived dialect" and "considers itself Amish".

Until 1960, the majority of the Amish were farmers, now small business ownership and factory work has become commonplace. The number of different Amish affiliations has grown over the years to over 40, each comprising groups of church districts united by social and religious practices. The origin of this religious diversity is sometimes disagreement about religious or cultural practices and sometimes more to do with geographical origins. Affiliations range between low to high order dependent on the level of separation from modern society, with the most traditional termed "low order" and those most assimilated "high order".

The lowest orders would still milk their cows by hand and have outhouse latrines. The highest orders would use automatic milking machines, mobile phones and through third parties, use the internet to send emails and advertise their business. Aside from those noted by Professor Kraybill there are a number of additional practices that all the Amish groups have in common: education only up to eighth grade, rural living, selective use of technology, lay ministers, small local congregations with services held in the homes of members and no formal religious buildings. Although the style of dress differs between different affiliations, in all cases this is church-regulated, traditional and plain, hence the term "plain-people" which is frequently used to refer to the Amish (Hurst and McConnell, 2010, Kraybill et al., 2013).

Figure 5: Amish horse and buggy transportation

1.7.1 History of the Windows of Hope project

Genetic studies of the Amish date back to 1962 when the world-renowned geneticist Victor McKusick first read an article written by a family doctor David Krusen with an extensive practice in Lancaster Pennsylvania. Krusen mentioned that achondroplasia was extremely common amongst the Amish. Later that year McKusick was asked to review a manuscript written by John Hostetler entitled “Amish Society”. It struck Mckusick then that the Amish community possessed many characteristics that made them an ideal group in which to study genetic traits, particularly autosomal recessive conditions. (Francomano et al., 2003).

MIM (Mendelian Inheritance in Man) and it’s now online version OMIM date back to 1964 when Victor McKusick was working at Johns Hopkins University. The origins of MIM came from three sources, the first, Mckusick and colleagues annual reviews of Genetics between 1958-1963, the second a catalogue of X-linked traits compiled by Mckusick in 1962, and thirdly a catalogue of autosomal recessive phenotypes prepared by Mckusick in 1963 to use as resource for identifying and studying recessive disorders amongst the Amish (McKusick, 2007). Mckusick travelled to the Pennsylvanian Amish community with Krusen initially to study the different forms of dwarfism. Their studies found that in fact the majority did not have achondroplasia, and instead led to the description of the two predominate forms of recessive dwarfism amongst the Amish: Ellis-van Creveld syndrome, which had previously been reported and a newly recognised condition Cartilage Hair Hypoplasia (Francomano et al., 2003).

In 1962, a young doctor approached him; Dr Harold Cross who wanted to study for a PhD. Dr Cross was raised amongst the Indiana Amish community and speaks the most common native Amish dialect, “Pennsylvania Dutch”. Together they studied inherited neurological conditions in the Ohio Amish. These studies led to publications describing a number of such disorders including the Troyer syndrome and the Mast syndrome (Cross and McKusick, 1967), two novel forms of complex hereditary spastic paraplegia. Dr Cross, now Professor, went on to study ophthalmology and is now based at the University of Arizona.

Over recent years researchers have begun to disentangle the genetic basis of the inherited disorders that occur within the Anabaptist communities. In 2000 Professor Andrew Crosby, a geneticist with a particular interest in neurological conditions, approached Dr Cross and suggested that given the advances in genetic technology since the original studies, identification of the genes responsible for the conditions studied by Mckusick and Cross would be possible. Crosby and Cross, initially identified the genetic basis of the Troyer (*SPG20* gene), (Patel et al., 2002) and Mast (*ACP33* gene) (Simpson et al., 2003a) syndromes in 2002 and 2003 respectively. Recognising the potential impact of these studies in not only scientific terms but more importantly for the Amish community, Cross and Crosby established the Windows of Hope project (WOH, see 'www.WOHproject.org'), the largest survey of inherited conditions amongst the Amish communities which together with the non-profit Amish-led Windows of Hope Genetic Information Centre (WHGIC), have since made significant headway in determining the molecular causes and clinical manifestations of inherited diseases amongst the Ohio Amish. WOH has a regional centre based in Holmes County, Ohio, the second largest Amish settlement in the USA (population 32,630, 251 church districts (http://www2.etown.edu/). The project aims to translate molecular genetic studies into direct community benefits through education, diagnostic testing, development of therapies and support of local clinicians, leading to improved clinical outcomes.

As with any community arising from a limited number of founder individuals the Amish, together with other genealogically-related Anabaptist communities, have a unique genetic distinctiveness and heritage. A limited number of ancestors founded a rapidly expanding endogamous population (the average family size is 7-9 children), in whom particular causal genetic variants have become enriched, unfortunately leading to a high incidence of particular genetic diseases in that population. The Amish are relatively immobile given the religious constraints on transportation; they also keep extensive genealogical records, which enable the construction of comprehensive pedigrees. These factors taken together greatly facilitate the discovery of genes responsible for inherited disease, which might otherwise have been impossible in studies of other populations due to the genetic and environmental complexities of a condition.

There is a high incidence of undiagnosed childhood developmental disorders amongst the Holmes County Amish, which place a significant social and financial burden on this community who for the most part are ineligible for the Medicaid program for religious regions. Although other community schemes such as the “Ohio Crippled Children Fund”, are available to the majority of the local Amish to help with medical expenses, the rising costs of medical care means that many Amish families are not able to be supported and go without important clinical investigations and treatments. Over the last 14 years, the WOH research team has identified the genes underlying 16 inherited diseases present amongst the Amish, including 8, which cause childhood developmental disorders, 3 of which are the focus of this thesis.

Highlighting the global relevance and importance of studies of inherited conditions in such genetic isolates, virtually all of disease genes initially identified in the Amish have subsequently been shown to cause similar diseases occurring worldwide.

1.7.2 History of the Clinic for Special Children

In 1988 Dr Holmes Morton changed the face of medical care for the Amish community. Morton, a paediatrician with a special interest in metabolic medicine, saw his first Amish patient, a child who had been given a diagnosis of cerebral palsy that he subsequently diagnosed with glutaric aciduria type 1. Morton learned that there were many other Amish children with glutaric aciduria type 1 who had been diagnosed with Amish cerebral palsy but who actually had glutaric aciduria type 1 as well as other metabolic conditions which when treated appropriately could prevent devastating long term neurological sequelae. Many of the families of these children were unaware of the cause of their child’s medical problems and unable to seek the medical attention needed because of expense or the travel distances involved. Morton and his wife Carolyn decided to open a clinic in the heart of the Pennsylvania Amish community. The clinic was built following the Amish “barn raising” tradition and still remains on the original land donated by an Amish farmer whose children Morton had treated. (Morton et al., 2003).

Morton used his genetic knowledge of the community to diagnose and develop treatments for the inherited conditions found within the community. He and others developed a comprehensive in-house genetic and metabolic diagnostic service collocated with the not for profit clinical facilities. Many of the metabolic treatment strategies developed by Morton and co-workers have been taken up by clinicians treating those same disorders in individuals from outside of the Amish community. The enrichment of certain conditions within the community actually facilitates the development of therapies which would be difficult to do for the small numbers of children affected by rare disorders in the general population. This model of care has now been extended to other Amish communities including Indiana and Ohio.

1.7.3 Disorders of brain growth and development identified amongst the Amish

Amongst the Amish, chromosomal aberrations and autosomal dominant conditions such as fragile X syndrome, Angelman syndrome and Rett syndrome that can result in developmental delay, are seen at similar frequency as outside of this community. There are however a subset of rare autosomal recessive neurodevelopmental disorders which occur at increased frequency within this population as a result of the founder effect. Currently recognised autosomal recessive causes of inherited disorders of brain growth and development recognised amongst the North American Amish communities are summarised in Table 1 and Table 2. The tables have been split into metabolic and syndromic causes.

Table 1: Metabolic causes of disruption to normal brain growth and development identified amongst the North American Amish

Reference	Metabolic Disorders	Gene	Clinical Features
1	Phenylketonuria	***PAH***	Microcephaly, seizures, cognitive and behavioural abnormalities, light skin and hair
1	GM3 synthase deficiency	***ST3GAL5***	Microcephaly, infantile onset of seizures, profound psychomotor retardation
1	Gutaric aciduria, type 1	***GCDH***	Macrocephaly, basal ganglia degeneration, dystonia, retinal haemorrhages
2	Homocystinuria	***MTHFR***	Intellectual disability, dislocated lenses
1	Galactosemia	***GALT***	Cataract, liver damage, bleeding diathesis, intellectual disability
1	Biotinidase deficiency	***BTD***	Hypotonia, developmental regression, seizures, hearing loss, optic atrophy
1	3-methylcrotonylglycinuria	***MCCC2***	Metabolic acidosis, hypotonia, seizures, variable intellectual disability, some asymptomatic
1	Crigler-Najjar syndrome	***UGT1A1***	Dystonia, developmental delay, hypotonia, irritability and lack of muscle control
1	Maple syrup urine disease	***BCKDHA***	Dystonia, seizures, apnoea, global developmental delay, growth deficiency
3	GM2 synthase deficiency	***B4GALNT1***	Complex spastic paraplegia, autistic features, intellectual disability
1	Propionic acidemia	***PCCB***	Vomiting, seizures, hypotonia, developmental delay, dilated cardiomyopathy
1	Tyrosinemia, type 3	***HPD***	Intellectual impairment, Ataxia, seizures, confusion, hypotonia
1	Salla disease	***SLC17A5***	Ataxia, hypotonicity, developmental delay/intellectual disability, growth deficiency

(Puffenberger et al., 2012), (Strauss and Puffenberger, 2009); 2, (Strauss et al., 2007); 3, (Harlalka et al., 2013b)

Table 2: Syndromic causes of abnormal brain growth and development identified amongst the North American Amish.

Reference	Syndromic Disorders	Gene	Clinical Features
4	Ataxia telangiectasia	***ATM***	Ataxia, ocular and cutaneous telangiectasia, oculomotor apraxia
5	Neonatal rigidity and seizure syndrome	***BRAT1***	Microcephaly, intractable epilepsy
5	Symptomatic epilepsy, skull dysplasia	***SNIP1***	Seizures, skull dysplasia, microcephaly
6	MTPAP	***MTPAP***	Spastic ataxia, intellectual disability
7	Aicardi Goutieres	***SAMHD1***	Microcephaly, large vessel vasculopathy, chilblain lesions, glaucoma, spasticity, arthropathy, intellectual disability
5	Non syndromic mental retardation	***CRADD***	Mild-moderate intellectual disability
8	TMCO1 defect	***TMCO1***	Skeletal anomalies, intellectual disability characteristic features
5	Microcephaly with chorioretinopathy	***TUBGCP6***	Microcephaly with pachygyria, developmental delay, chorioretinopathy
1	Amish microcephaly	***SLC25A19***	Profound microcephaly, brain malformations, progressive encephalopathy, profound psychomotor retardation
1	Bardet-Biedl syndrome	***BBS1***	Rod-cone dystrophy, obesity, postaxial polydactyly, cognitive impairment, male hypogonadotrophic hypogonadism, female genitourinary malformations, and renal abnormalities
9	Cohen Syndrome	***COH1***	Microcephaly, retinal dystrophy, intellectual disability, hypotonia, characteristic facial features
1	Cortical dysplasia and focal epilepsy	***CNTNAP2***	Macrocephaly, seizures, intellectual disability, behavioural abnormalities
1	Troyer syndrome	***SPG20***	Spastic paraplegia, learning difficulties, short stature
1	Cockayne syndrome	***ERCC6***	Microcephaly, short stature, spasticity, premature aging, cataracts
1	Pelizaeus-Merzbacher-like syndrome	***GJA12***	Leukodystrophy, developmental delay, congenital nystagmus, hypertonia
1	Pretzel syndrome	***STRADA***	Polyhydramnios, megalencephaly, and symptomatic epilepsy
10	MOPD type 1	***RNU4atac***	Primordial dwarfism, contractures, developmental stagnation
11	Brittle hair syndrome (trichothiodystrophy)	***C7ORF11***	Brittle hair, dry skin, intellectual disability, autistic features
12	HERC2 deficiency (Blue eye delay syndrome, Angelman-like)	***HERC2***	Intellectual disability, autistic features, broad based gait, prominent big toe, absence of the posterior part of the corpus callosum

4,(McKusick and Cross, 1966); 5,(Puffenberger et al., 2012); 6,(Crosby et al., 2010); 7,(Xin et al., 2011); 8,(Xin et al., 2010); 9,(Falk et al., 2004); 10,(Edery et al., 2011); 11,(Nakabayashi et al., 2005) 12,(Harlalka et al., 2013a)

1.8 Autozygosity mapping

A significant contribution to our understanding of the processes involved in human brain development have been made by the study of single gene disorders which are rare in the general population, but which occur with increased frequency in certain consanguineous populations or populations in which there is a high frequency of inter-community marriage (Amselem et al., 1989, Ruiz-Perez et al., 2000). Genome scans to identify autozygous regions specific to the affected individuals in such families provides a powerful means to demarcate the chromosomal location of the disease locus. This concept of 'autozygosity mapping' in endogamous communities and consanguineous families is based on the principle that affected individuals will be identical by descent for the disease causing variant and surrounding haplotype (Lander and Botstein, 1987), Figure 6.

When the ancestry of a community comprises only a limited number of founder individuals, the frequency of autosomal recessive alleles that were present in the founders may increase within that population such that if any two individuals from that community are selected at random, the chance that both will be carriers will be high. This is known as the founder effect (Alkuraya, 2013). Aside from the Amish, Finland, is a good example of where a relatively small number of founders 100 generations back formed a current modern day population of 5.4 million. Enrichment for founder mutations has caused an increased frequency in the population of rare diseases such as the so called Finnish nephropathy (*NPHS1)* and cartilage hair hypoplasia (*RMRP*) (de la Chapelle, 1993). In consanguineous unions where the relationship between two individuals is more recent, such as first or second cousins, much larger and numerous blocks of autozygosity are likely to be present. Historically autozygosity mapping was dependent on the analysis of polymorphic microsatellite markers. Although these markers are often highly informative due to their highly polymorphic nature, performing a genome-wide analysis using microsatellite markers was costly and time consuming and may miss smaller autozygous sections due to wide marker spacing. The advent of single nucleotide polymorphism (SNP) genotyping panels and microarray technology has revolutionised autozygosity mapping, greatly increasing throughput.

Although the heterozygosity score for a SNP is typically much lower than that of a microsatellite marker, SNPs occur at much higher density throughout the genome, providing significantly increased power as markers. Although the original SNP mapping microarray platforms were comprised of 10,000 SNP's, densities of 1,000,000 SNP microarray platforms are now commonly used. These more dense mapping platforms are preferred when analysing more distantly related individuals (such as in the Amish), due to the smaller sections of autozygosity cosegregating with the disease gene typically present in such cases (Alkuraya, 2010). Despite the power of autozygosity mapping to identify the locus for autosomal recessive diseases the process of identification of the disease gene by Sanger sequencing was still very laborious, particularly if the region was gene dense. More recently autozygosity mapping has been coupled with massively parallel sequencing techniques to more rapidly identify potential disease causing sequence alterations within the disease locus, which may require prioritisation for investigation. (Alkuraya, 2013, Carr et al., 2013).

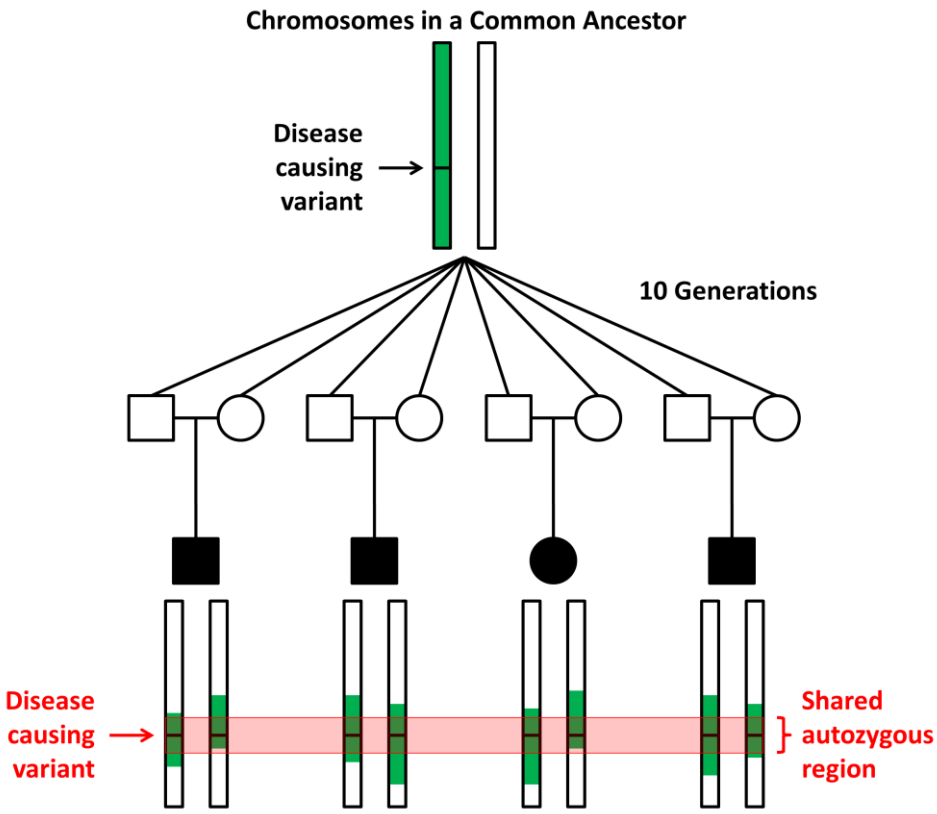


Figure 6: Inheritance of autozygous regions

1.9 Aims

The overarching aim of my studies was to clinically characterise and identify the underlying molecular cause of three new inherited disorders of brain growth and development identified amongst the Amish. To this end I pursued the following aims:

- To perform detailed phenotyping of affected individuals with each condition with particular reference to brain growth and neurodevelopment.
- To undertake in-depth genetic studies, including autozygosity mapping, linkage studies and traditional and next generation sequencing technologies, to identify the underlying molecular cause of each disorder.
- To functionally characterise the putative disease genes and pathogenic variants identified.
- To translate any findings into direct benefits to patients with improved diagnostic information, genetic counselling and eventually therapies.

2

CHAPTER TWO

MATERIALS AND METHODS

2 MATERIALS AND METHODS

2.1 Buffers, reagents and stock solutions

- **Agarose loading buffer** – 40% ficoll, 0.2% xylene cyanol, 0.2% bromophenol blue
- **Ampicillin (Melford)** – 100mg/ml Amp in ddH_2O sterilized using a 0.22µM filter unit.
- **Kanamycin (Melford)** – 50mg/ml Kan in ddH_2O sterilized using a 0.22µM filter unit
- **LB Agar** – For 1 litre: 15g Agar, 25g LB broth base, ddH_2O to final volume, pH adjusted with 5 N NaOH, then autoclaved
- **Luria Bertani (LB) broth (For medium)** – For 1 litre: 25g of LB broth base (10g tryptone, 5g NaCl, 5g yeast extract), ddH_2O to final volume, pH adjusted with 5 N NaOH, then autoclaved.
- **PBST –** PBS with 0.1% TWEEN20
- **Phosphate buffered saline (PBS)** – Dulbecco's PBS with calcium and magnesium, bought from PAA
- **Ponceau stain** – 1g Ponceau S, 15g trichloroacetic acid, 15g sulphosalicylic acid and ddH_2O
- **Protease inhibitors (Roche)** – Complete EDTA-free protease inhibitor cocktail tablets
- **SOC medium** – For 100ml, supplement LB broth with the following: KCl 2.5mM, $MgCl_2$ 10mM, $MgSO_4$, 10mM, Glucose 20mM, sterilised using a 0.2µM filter unit.
- **2x NuPAGE sample loading buffer** – For 10ml: 5ml NuPAGE LDS sample buffer (4x), 2ml NuPAGE reducing agent (10x), 3ml ddH_2O

2.2 Subjects and samples

2.2.1 Recruitment to the Amish Windows of Hope (WOH) project

All the WOH project studies were reviewed and approved (reference 10-0050-01) by the Institutional Review Board of the Office for Responsible Conduct of Research, University of Arizona (Tuscon, Arizona, USA) and the University of Exeter Research Ethics Committee (reference 14/04/048), and were conducted in accordance with the principles of the Declaration of Helsinki. Recruitment to the WOH project requires submission of the appropriate signed consent, clinical details and a blood or buccal sample.

2.2.2 Phenotyping of affected individuals

Height and weight were measured barefoot in light clothing. The Leicester Height Measure MKII, recommended by the child growth foundation was used to measure height in the family homes. Occipitofrontal circumference (OFC) was measured following standard procedures.

Height, weight and OFC standard deviations and Z-scores were calculated using a Microsoft Excel add-in to access growth references based on the LMS method (Pan H, Cole TJ. LMS growth, a Microsoft Excel add-in to access growth references based on the LMS method. Version 2.77, http://www.healthforallchildren.co.uk/;2012) using a reference European population (Cole et al., 1998).

A full medical and developmental history and systemic examination for the purposes of the research study was obtained for all individuals recruited and photographic records of pertinent examination findings were taken with informed consent. Additional clinical records (including results of any investigations) and educational reports were requested using the appropriate information release form. Where possible Intelligence Quotient (IQ) was requested using the Wechsler Adult Intelligence Scale. Whenever possible original neuroimaging was requested for review locally by Dr Phillip Rich, Consultant Neuroradiologist, St George's Hospital NHS Trust. Audiograms and

dermatology were reviewed by Dr Wendy Albuquerque and Dr Kristiana Gordon respectively, St George's Healthcare NHS Trust.

2.3 Data management

On collection of a sample, each family is awarded a pedigree ID, each individual an individual ID, and each sample a sample ID. Clinical and molecular information is recorded in a password protected database. Family pedigrees are constructed using the online Swiss Anabaptist Geneological Association database (www.saga-omii.org).and stored using Progeny (800 Palm Trail, Suite 200, Delray Beach, FL 33483 USA (http://www.progenygenetics.com/).

2.4 Molecular methods

2.4.1 DNA extraction

Lymphocyte DNA was extracted from whole blood. 10-30ng of DNA was used per reaction. The protocol for DNA extraction and purification using the Promega ReliaPrepTM Blood gDNA Miniprep System is summarised below.

The blood samples were stored at -20^{o}C and were thawed completely and thoroughly mixed prior to DNA extraction. A volume of 200µl of whole blood was used. Filter tip pipette tips were used at all stages during the procedure to prevent contamination of samples and equipment.

20µl of Proteinase K was dispensed into a 1.5ml microcentrifuge tube. 200µl of blood was mixed with the Proteinase K by repeat pipetting. 200µl of Cell Lysis Buffer was added to the microcentrifuge tube. The tube contents were vortexed for 10-20secs to ensure thorough mixing and incubated at 56^{o}C for 10mins. After incubation 250µl of Binding Buffer A was added and the microcentrifuge tube contents vortexed for 10-20secs. The appearance of the lysate was checked to ensure that it was dark green in colour. The tube was briefly centrifuged to remove any residue from the tube lid. Using a pipette, the lysate was transferred to a Reliaprep binding column that was placed into a collection tube that was then placed in a microcentrifuge and spun at maximum speed (13,000rpm) for 1min. If any lysate was still visible on top of the column

membrane the tube was spun for a further 1min. The column was placed in a clean collection tube and the flow through was discarded as hazardous waste in a sharps bin. 500µl of Column Wash Solution was added to the column which was then spun for 3mins at maximum speed, if any solution was still visible after this then the column was spun for a further 1min, the flow through was discarded as hazardous waste. This step was repeated to complete a total of 3 washes. The column was then placed in a clean 1.5ml microcentrifuge tube. 200µl of nuclease-free water was added to the column which was then spun for 1 minute at maximum speed to elute the DNA. After elution the DNA concentration and purity was measured using the nanodrop. To achieve the desired DNA concentration of 10-30ng/µl the DNA was diluted with molecular grade water.

2.4.2 RNA extraction

Lymphocyte RNA was extracted from whole blood collected in PAXgene collection tubes. The protocol for RNA extraction and purification using the Qiagen PAXgene blood RNA kit handbook is summarised below. Prior to commencing RNA extraction the lab bench top, tube racks and pipettes were cleaned with RNasezap (Ambion). Filter tip pipette tips were used at all stages during the procedure to prevent contamination of samples and equipment.

The Paxgene RNA blood tube containing the blood sample was centrifuged at 3000g for 10mins in a swing-out centrifuge. The supernatant was carefully removed by pipetting without disturbing the pellet. 4ml of RNase free water was added to the pellet and the blood tube was closed with a new BD Hemogard closure. The tube was vortexed until the pellet was entirely resuspended and then centrifuged at 3000g for 10mins. The supernatant was carefully removed by pipetting. 350µl of buffer BR1 was added to the pellet and then vortexed until the pellet was dissolved. The sample was transferred to a 1.5ml microcentrifuge using a pipette. 300µl of buffer BR2 and 40µl of proteinase K were added to the sample which was the vortexed for 5secs to achieve thorough mixing. The sample was incubated at 55^{o}C in a shaker-incubator set at 1400rpm for 10 mins. The sample was transferred by pipetting to a PAXgene shredder spin column placed in a 2ml processing tube and centrifuged at 14,000xg for 3mins.

The flow-through was then transferred to a fresh 1.5ml microcentrifuge tube using a pipette and without disturbing the pellet. 350µl of 100% purity grade ethanol was added to the sample and mixed by vortexing for 10secs. The tube was centrifuged briefly (1-2secs) to remove any fluid from the tube lid.

700µl of sample was pipetted into a PAXgene RNA spin column which was placed in a 2ml processing tube and centrifuged for 1min at 14,000xg. The spin column was placed in a fresh processing tube and the flow-through discarded as hazardous waste. This step was repeated with the remainder of the sample. 350µl of wash buffer 1 (BR3) was subsequently added to the PAXgene spin column which was centrifuged for 1 minute at 14,000xg. The spin column was then placed in a new 2ml processing tube and the old tube and flow-through were discarded.

10µl of DNase I stock solution was added to 70µl of DNA digestion buffer In a 1.5ml microcentrifuge tube (supplied with the kit). The solutions were gently mixed by flicking the tube to prevent denaturation of the enzyme. The tube was briefly centrifuged for 1-2secs to remove any liquid from the lid and the side of the tube. The 80µl solution was then carefully pipetted onto the spin column membrane and incubated at room temperature for 15mins. 350µl of wash buffer 1 (BR3) was added to the spin column and centrifuged at 14,000xg for 1 minute and then transferred to a new processing tube, the old tube and flow through were discarded. 500µl of wash buffer 2 (BR4) was then added to the spin column and centrifuged at 14,000xg for 3mins. The spin column was placed in a fresh tube and centrifuged for a further 1 min at 14,000xg. The processing tube was discarded and the spin column was placed in a 1.5ml microcentrifuge tube. 40µl of elution buffer (BR5) was pipetted onto the spin column membrane and centrifuged at 14,000xg for 1min to elute the RNA. This step was repeated using the same microcentrifuge tube and a further 40µl of elution buffer. The tube and contents were then incubated at 65^{o}C for 5mins using the desk-top shaker-incubator, in order to denature the RNA. After the incubation period the eluted RNA was placed on ice and stored in the -80^{o}C freezer until required.

2.4.3 Single nucleotide polymorphism (SNP) genotyping

SNP genotyping was carried out using Illumina CytoSNP-12v2.1 330K arrays following the Infinium® HD Assay Ultra manual protocol. Each chip holds 12 DNA samples. DNA was extracted from venous blood or buccal samples using the methodology described above.

The assay requires 200ng of DNA per sample at a concentration of 50ng/µl and is carried out over a three day period, the methodology is summarised below:

1. Day 1, the DNA samples were denatured using a buffer containing 0.1N NaOH and neutralised in the pre-PCR room prior to an overnight isothermal whole genome amplification step.

2. Day 2, the DNA was fragmented using Illumina fragmentation buffer (FBS), and then precipitated using isopropanol. The DNA was then collected by centrifugation at 4^{o}C before resuspension. The DNA was then denatured at 95^{o}C for 20mins. The samples were applied to a BeadChip and separated by an IntelliHyb® seal. The BeadChips were then placed in a humidified Hyb Chamber and subsequently incubated for 16-24hrs at 48^{o}C in the Illumina Hybridisation Oven. The amplified and fragmented DNA samples anneal to locus-specific 50-mers

3. Day 3, unhybridised and non-specifically hybridised DNA was washed away using PB1 a buffer containing an aliphatic amide. The BeadChip was then prepared for single base extension of the oligonucleotides and incorporation of detectable labels using the captured DNA on the BeadChip as a template. The BeadChips were then stained using the Illumina XStain HD BeadChip process before imaging on an Illumina BeadArray reader.

The BeadArray reader uses a laser to excite the fluorophores of the single-base extension products on the beads and the scanner records images of the light emitted allowing the genotype to be determined.

The Illumina GenomeStudio Integrated Informatics Platform was used to extract, visualise and analyse the genotyping data. The data was then analysed

and imported into Microsoft Excel for analysis using Excel macro or IBD finder software (http://dna.leeds.ac.uk/ibdfinder/), looking for notable regions of homozygosity (typically >1Mb) and to compare genotyping data across samples and patient cohorts.

Assistance and support for SNP chip genotyping and analysis was provided by Dr Barry Chioza, University of Exeter and Dr Jayne Dennis, St George's University of London.

2.4.4 Primer design

The Ensembl Genome Browser (February 2009 assembly) was used to obtain the gene sequences (website http://www.ensembl.org/index.html). Primers used for PCR amplification were designed either using Primer3 software version 0.4.0 (website http://frodo.wi.mit.edu/primer3/) or the ExonPrimer link from the UCSC genome browser (http://genome.ucsc.edu/).

Primers were designed following the criteria below:

1. Primer sizes were 18-27 bases in length (the specificity and annealing temperature are dependent on the length of the primer).

2. Melting temperatures were kept as similar as possible for both flanking primers to maintain PCR efficiency (ideally between 57 – 63^{o}C).

3. Where possible the guanine-cytosine (GC) base content was kept between 40-60%

4. Primers with inter or intra-primer efficiency extending for more than 3 bases were avoided, to prevent the formation of secondary structures and primer dimers.

5. The primer sequences selected were specific and a 100% match to the region to be amplified. *In silico* PCR analysis and BLAST searches were performed using the UCSC Genome bioinformatics website.

2.4.5 Microsatellite marker design

Microsatellite markers were chosen using the UCSC genome browser (February 2009 assembly). The region of interest was entered into the programme and "STS markers" and simple repeats and microsatellites" viewed. STS markers that are known to be polymorphic, with published primers were favoured.

2.4.6 Polymerase chain reaction (PCR)

PCR is an invitro method of cloning DNA that permits exponential amplification of specific lengths of DNA from only minute amounts of starting material. A touchdown PCR protocol was used. A “touchdown” PCR protocol was used to improve the specificity of the PCR by helping to avoid amplification of non-specific products by incrementally lowering the initial annealing temperature (by 2^{o}C) every 2 cycles until the required melting temperature (Tm) is reached (Table 3).

Table 3: Touchdown PCR program

Number of cycles	Temperature (^{o}C)	Time (Seconds)
1	98	120
2	95 Anneal plus 4 72	30 45 45
2	95 Anneal plus 2 72	30 45 45
35	95 Anneal 72	30 45 45
1	72(elongation)	300

A standard protocol for DNA amplification was followed, using a 20µl final volume mix (Table 4).

Table 4: Components of a 20µl PCR reaction

Component	Volume (µl)
DNA	1
Forward Primer (Sigma)	0.6
Left Primer (Sigma)	0.6
dNTPs (10mM)	0.6
10xBuffer (15mM MgCl2, Thermoscientific)	2.0
Molecular grade H2O	15.1
Red hot Taq polymerase (5U/µl)	0.1

Primer pairs were initially trialled on control DNA samples that were known to amplify well in previous reactions. If the PCR reaction produced weak or no product then the contents of the mix and amplification conditions were altered (see optimisation of amplification conditions below). In order to ensure that the desired DNA template was being amplified and not a contaminant, a water control was included in each reaction. An Eppendorf 96 well mastercycler, thermal cycler machine was used for the PCR reactions.

2.4.7 Optimisation of amplification conditions

In order to determine the optimal annealing temperature for the primer pair a temperature gradient of 55-67^{o}C across the PCR block was carried out. Where the GC content was above 60% a second gradient with 10% diMethyl sulfoxide (DMSO, Fisher Scientific) was performed. If amplification was still inadequate magnesium titration (1mM-3mM) across the temperature gradient was performed as well as increased primer concentration as required.

2.4.8 Agarose gel electrophoresis

To determine the adequacy of amplification of the DNA, the resulting products underwent agarose gel electrophoresis. For resolving smaller DNA fragments

such as most PCR products (typically,500bp), a 2% agarose gel was made by adding 2.0g of agarose to 100ml of 1x tri-acetate EDTA (TAE) buffer. The solution was warmed in a microwave until the agarose was dissolved. 2µl of of a 10mg/ml solution of ethidium bromide (Fisher Scientific) was added to the gel after cooling for 10mins. Ethidium bromide is a DNA intercalating agent that fluoresces brightly when exposed to ultraviolet light. The gel was then poured into a casting tray with combs in place to create wells. Once set the gel was placed into an electrophoresis tank filled with sufficient TAE buffer to cover the surface of the gel. A DNA ladder (1Kb Plus from Invitrogen) was used to assess the size of the PCR product. The ladder was made up with agarose gel loading buffer (final concentration 50ng/µl) and 2µl was loaded into one of the wells. 5µl of each PCR product was mixed with 5µl of loading dye and loaded into each subsequent well, then electrophoresed through the gel for 40mins at 100V. The gel was subsequently visualised on a UV transilluminator and photographs were taken with an orange filtered camera. Larger DNA fragments (>500bp) were resolved using lower concentration agarose gels (0.8%-1.0%) prepared similarly.

2.4.9 Purification of PCR products

Prior to sequencing it is necessary to remove unincorporated primers and dNTP's. This was done using ExoSAP-IT a product that contains exonuclease-1 (EXO) and shrimp alkaline phosphatase (SAP). Exonuclease-1 degrades single stranded DNA in a 3 prime to 5 prime direction, releasing deoxyribonucleoside 5 prime-monophosphates in a stepwise manner. This releases dNTPs from the unincorporated primers. shrimp alkaline phosphatase (SAP) catalyses the release of 5 prime - and 3 prime-phosphate groups from extra nucleotides. 2µl of ExoSAP-IT were added to 5µl of each PCR product. The samples were then incubated at 37^{o}C for 15mins the optimal temperature for the enzyme activity and subsequently at 85^{o}C for 15mins to inactivate it.

2.4.10 Sequencing reaction

Purified PCR products were sequenced using BigDye Terminator Cycle Sequencing Kit v3.1 (ABI, Applied Biosystems) with the forward and reverse primers in separate reactions. The kit contains Taq DNA Polymerase, dNTPs,

ddNTPs-dye terminators, $MgCl_2$ buffer. In dye-terminator sequencing, each of the four dideoxynucleotide chain terminators is labelled with fluorescent dyes, each of which emit light at different wavelengths. The protocol used for the sequencing reaction is detailed in Table 5 below.

Table 5: Components of a 10µl sequencing reaction

Component	Volume (µl)
Big Dye	0.5
Big Dye Buffer	1.75
Molecular grade H_2O	4.25
Forward or Reverse Primer(10mM)	0.5
Cleaned PCR product	3.0

Sequencing Reaction Program:

1. 96^{o}C for 30secs (denaturation)

2. 50^{o}C for 15secs (annealing)

3. 60^{o}C for 4mins (elongation)

4. Steps 1-3 are repeated to complete 25 cycles.

2.4.11 Sequencing reaction purification

The products were purified prior to running on an automated DNA sequencer. The BigDye XTerminator purification kit protocol was followed (ABI, Applied Biosystems), briefly 45µl of SAM solution and 10µl of XTerminator solution were added to each sample in a 96 well plate. The contents of the plate were mixed thoroughly using a desk top shaker set at full speed for 30mins then centrifuged for 2mins at 1000xg. The samples were run on a sixteen capillary 3130XLA Applied Biosystems Automated Sequencer. A 36cm array POP7 polymer

programme setting was used on the sequencer. The genomic sequence was checked using Chromas light (Technelysium Pty Ltd), Gene tool (Biotool Incorporated) and Mutation Surveyor (Soft Genetics) software.

2.4.12 Restriction digest

A restriction fragment length polymorphism (RFLP) is a DNA polymorphism that creates or abolishes a recognition site for a particular restriction enzyme. Most restriction enzymes recognise a palindromic DNA sequence. RFLP analysis is a technique by which a variation in a DNA sequence can be detected by using restriction enzymes to fragment DNA. The resulting fragments are then sized by electrophoreses on agarose gels. This technique was used to assess if any previously unrecognised variant detected during sequencing was present in a control population. NEBcutter V2.0 website (http://tools.neb.com/NEBcutter2/) was used to select the restriction enzyme. The forward and reverse primer sequences with their intermediate sequence (wild-type and variant) were entered into the NEBcutter programme. Enzymes that cut the sequence containing the variant differently to the wild-type (WT) sequence were used. Prior to selection of an enzyme it was checked that the expected product sizes produced by the enzyme were sufficiently different between the WT sequence and the sequence containing the variant so that they would be clearly distinguishable when run on an agarose gel. If the variant had no effect on a recognised restriction site or the presence of pseudogenes complicated interpretation of results, then the controls were sequenced to look for the variant.

Prior to performing a restriction digest, a 20µl PCR reaction was set-up for each sample to amplify the sequence where the variant was detected using the method previously described. 5µl of each PCR product was subsequently run on an agarose gel to check for successful amplification. A restriction digest was set-up using 5µl of PCR product and conditions determined by the specific enzyme to give a final volume of 30µl. Samples were incubated at the required temperature. In addition to the control DNA samples, a negative control (water replacing DNA) and a known homozygote and heterozygote for the variant were included. The latter provides control bands to confirm digestion and to aid

viewing fragments on agarose gels corresponding to both the WT and variant sequences. Following the incubation step samples were electrophoresed on a 3% agarose gel for 60mins at 100V. The bands produced were analysed to identify if the specific variant under investigation was also present in the control DNA samples.

2.4.13 Denaturing polyacrylamide gel electrophoresis (PAGE)

Microsatellites (also known as short tandem repeat polymorphisms, STRPs) are repeats of 1-9 nucleotide sequences that are interspersed throughout the human genome. They serve as highly polymorphic markers which are typically significantly more informative in inheritance studies than single nucleotide polymorphisms (SNPs) due to the often large number of alleles in the population. DNA polymerase stuttering during replication can alter the length of microsatellites by one or two repeats and is also thought to be the primary mechanism for generating microsatellite polymorphism during meiosis (Strachan and Read, 2011). Microsatellite marker design has been described in section 2.4.5. PCR was carried out using the primers designed for each marker and DNA from affected individuals, parents and unaffected siblings (where available). To determine if the amplification was successful the PCR products were initially run on a 2% agarose gel as previously described. Successfully amplified products from each individual were then size fractionated by electrophoresis on a denaturing polyacrylamide gel. The size of an individual's microsatellite DNA can be assessed to produce a genotype which can be compared to that of other family members.

The polyacrylamide gel was made by mixing 50ml of 8% PAGE solution (contains acrylamide and BIS, Amresco products), 80µl TEMED (N'N'N'N'-tetramethylenediamine) and 80µl of 25% APS (ammonium persulphate). The solution was mixed for 20secs using a magnetic stirrer and then poured in to the gel casting apparatus. Combs were inserted into the gel to create wells. The gels were then left to set for 40mins. The electrophoresis tank was assembled, each tank holds two gels. The tank was then filled with 1 litre of TBE (warmed in a microwave).

The PCR products were mixed with denaturing PAGE loading buffer (formamide, bromophenolblue and xylene cynol). The relative volumes of PCR product and loading buffer were determined depending on the strength of the PCR band on the agarose gel after UV transillumination, typically 2µl of PCR product: 2µl loading dye. The solution was then denatured by heating to 95^{o}C for 4mins in a thermocycler and then stored on ice. Prior to loading the products each well is washed out with TBE buffer using a syringe and needle in order to remove the urea from the wells. 1.1µl of denatured product was loaded into subsequent wells. The gel was then run at 100mAmp, the length of running time is dependent on the expected size of the microsatellite PCR product (approximately 50mins per 100 base pairs).

Once the electrophoresis was completed the gel was washed and silver stained. The gel was placed in the following solutions sequentially on a desk-top shaker. Between each step the gel was washed thoroughly in distilled water.

1. Solution 1 - 268.5ml dH20, 30ml ethanol and 1.5ml acetic acid (5-10mins).

2. Solution 2 - 300ml ddH20 and 0.3g silver nitrate (15mins).

3. Solution 3 - 300ml ddH2O, 4.5g sodium hydroxide and 450µl formaldehyde (until visible bands appear, approximately 10-15mins).

4. Solution 4 - 300ml ddH2O and 2.25g sodium carbonate (5mins, to prevent further development of the bands).

The gel was then wrapped in Saran wrap, labelled and the bands read.

2.4.14 Reverse transcription PCR (RT PCR)

Reverse transcriptase PCR (RT PCR) is a method for indirectly amplifying RNA by initially making a complementary DNA (cDNA) copy. The amplified cDNA can subsequently be sequenced and the RNA sequence inferred.

Prior to starting the procedure the bench top, pipettes and racks were all cleaned with 70% ethanol and subsequently RNaseZap (Ambion) to remove any traces of RNase. Filter tip pipette tips were used for all RNA work. RNA was extracted from whole blood as previously described and stored on ice until required. The Titanium one-step RT-PCR kit and protocol (Clontech) were used. RT-PCR primers were designed to target exonic sequences that are separated by at least one intron. Primers were designed using Primer 3, programme settings were changed so that the primers would meet the following specifications:

1. Primer melting temperature of approximately 70^{o}C.

2. Primer length at least 22 nucleotides (25-30 preferred).

3. G-C content of 45-60%.

Primers were diluted and a mix made with a final concentration of 45μM for each primer. A master mix was prepared as below:

1ng - 1μg of patient RNA was added to each subsequent reaction and the total volume was made up to 50μl with RNase free water. In addition to a negative control, where water replaced RNA to control for contamination, a further negative control was included. A concern in RT-PCR is that genomic DNA is being amplified as opposed to RNA, to control for this the titanium taq was replaced with the standard taq used for PCR and the RNA was replaced with water. A positive control was also included, primers specific for mouse βactin were used to amplify βactin from mouse liver total RNA (supplied with the one-step kit).

See Table 6 for details of reaction set up.

Table 6: Components of a 50µl PCR reaction

Component	Volume (µl)
Primer mix 45mM	1.0
50x dNTP mix	1.0
Recombinant RNase inhibitor (400u/ µl)	0.5
Thermostabilizing reagent	25.0
GC melt	10.0
Oligo (dT) primer	1.0
50x Titanium Taq RT enzyme mix	1.0

The reactions were placed in a thermal cycler and the programme in Table 7 below followed:

Table 7: RT- PCR program

Number of cycles	Temperature (^{o}C)	Time
1	50	1hr
1	94	5mins
	94	30secs
40	65	30secs
	68	1min
1	68	2mins

To determine successful cDNA amplification and the size of the products they were electrophoresed on a 2.5% agarose gel for 90mins with the appropriate DNA ladder.

2.4.15 Growth and maintenance of bacteria

Transformed bacteria were propagated in liquid Luria Bertani (LB) broth or on solid media, supplemented with antibiotic and/or other additives where appropriate, at 37°C. Frozen stocks were made from liquid cultures by mixing 0.5ml overnight culture with 0.5ml of 100% glycerol (sterile) followed by vortexing briefly and long-term storage at -80°C.

2.4.16 Heat shock transformation of chemically competent bacteria

In all cases the chemically competent bacteria were initially thawed on ice.

With intact plasmid: 50ng DNA was gently mixed with 20µL chemically competent DH5α and incubated for 30min/ice. Cells were then heat pulsed for 45 sec in a 42 ^{o}C water bath before being immediately transferred back to ice for 2mins to recover. 80µl of SOC medium was then added and the cells incubate at 37 ^{o}C for1hr. Cells were then plated onto a pre-prepared LB agar plate containing antibiotic and incubated overnight at 37°C.

XL-10-Gold® ultracompetent cells with intact plasmid: For each reaction 45µl of ultracompetent cells were aliquoted into a 15ml falcon kept on ice. 2µl of XL-10-Gold® β-mercaptoethanol was then added to each falcoln tube and mixed with cells by swirling the tube before incubation on ice for 2mins. 2µl of Dpn treated DNA from each reaction was then gently mixed with the cells and incubated for 30min on ice. The tubes were then heat pulsed for 30secs in a 42^{o}C water bath before being immediately transferred back to ice for 2mins to recover. 0.5ml of preheated (42^{o}C) SOC medium (used as a substitute for NZY^{+} broth) was added to each tube and tubes incubated 1 hr/37°C in a shaker incubator (250 rpm). Cells (250 µl) were then plated onto prewarmed (37^{o}C) LB agar plate containing antibiotic and 80 µg/ml X-gal and 20mM IPTG and incubated overnight at 37°C.

With ligation mix: 2µl of the ligation reaction was added to 50µl of the DH5α competent cells and gently mixed by flicking the tube before being placed on ice for 5mins. The cells were subsequently placed in a water bath set at 42^{o}C and

heat shocked for 45secs then immediately transferred back to the ice for 2mins in order to recover.

1ml of pre-prepared LB broth was then added to the transformed cells which were incubated horizontally for 1hr at 37^{o}C in a shaker incubator set at 150rpm.The mixture was then briefly centrifuged for 5secs. The cells formed a pellet at the bottom of the tube and 950µl of the broth was removed with a pipette taking care not to disturb the pellet. The cells were resuspended by gently pipetting up and down. Using a sterile spreader the remaining 100µl of transformation culture was spread as evenly as possible across a pre-prepared LB/ampicillin/agar plate. The plate was then incubated overnight at 37^{o}C.

2.4.17 Cloning of PCR product

In order to clearly visualise and interpret the effect of compound heterozygous changes, both individual DNA strands of the PCR product were cloned for DNA sequencing using the pGEM-T Easy Vector (Promega) and pGEM-T Easy Vector System II kit. The pGEM-T easy vector is a linearized vector with a 3 prime terminal thymidine at both ends. This is compatible with the Thermoprime taq polymerase used for the PCR which adds an “A” nucleotide overhang to the PCR product. The ligation buffer and T4 DNA ligase were thawed on ice and the ligation buffer vortexed for 10secs to ensure it was thoroughly mixed.

The ligation reaction mix was prepared as in Table 8 below in a 0.5ml tube. The solution was then incubated at room temperature for 1hr.

Table 8: Components of a 5µl ligation reaction mix

Component	Volume (µl)
2x Rapid ligation buffer	2.5
pGEM-T easy vector 50ng/µl)	0.5
T4 ligase (3 Weiss units/µl)	0.5
Fresh (less than 1 day old) PCR product	1.5

50µl of chemically competent DH5α cells were thawed on ice and mixed by gently flicking the tube. The transformation reaction was carried out as described in section 2.4.16, in a 1.5ml microcentrifuge tube.

16 wells of a 96 well plate were filled with 50µl of molecular grade water. A sterile 10µl pipette tip was used to remove individual colonies gently from the agar plate and each colony was placed into subsequent wells containing water (the pipette tip was agitated in the well to remove the cells). 1µl was taken from each well and used as template. PCR was carried out as previously described with appropriate forward and reverse primers. The products were then run on a 2% agarose gel to determine successful amplification.

2.4.18 Site-directed mutagenesis (SDM) to generate mutant constructs

SDM was carried out using the Agilent QuikChange Lightning site-directed mutagenesis kit, following the manufacturer's protocol which is summarised below.

A 50µl reaction was prepared comprising, 5µl 10X reaction buffer, xµl dsDNA template (corresponding to between 10 - 100ng), yµl oligonucleotide primer 1 (corresponding to 125ng) , yµl oligonucleotide primer 2 (corresponding to 125ng) , 1µl dNTP mix ,1.5µl QuikSolution™ reagent, 1µl of Quikchange Lightening Enzyme (derivative of PfuUltra™ High Fidelity DNA polymerase) and ddH_20. Reaction conditions: 95°C/2min, [95°C/20sec, 60°C/10sec, 68°C/30sec/kb of plasmid length] x18 cycles, 68°C/5min.

The amplification products were subsequently mixed with 2µl per reaction of DpnI enzyme provided in the kit. The reactions were then immediately incubated in a water bath at 37^{o}C for 5mins, in order to digest the parental supercoiled dsDNA. XL10-Gold® ultracompetent cells were then used for transformation reactions (section 2.4.16).

To assess the efficiency of mutant plasmid generation, a further reaction was set-up in parallel using the 4.5kb pWhitescript™ control plasmid (which contains a stop codon at the position where a glutamine codon would usually appear in

the β-galactosidase gene), and oligonucleotide control primers 1 and 2 contained within the kit. XL10-Gold ultracompetent cells transformed with this plasmid appear white on LB-ampicillin agar plates containing IPTG and X-gal due to a lack of β-galactosidase activity. The oligonucleotide control primers create a point mutation that reverts the sequence back to the wild type β-galactosidase gene sequence. Successful mutagenesis is indicated by the presence of blue colonies as opposed to white, following transformation. Expected results of the control reaction indicating successful mutagenesis and transformation were considered to be >100 colonies, >85% of which are blue.

2.4.19 Plasmid preparation and DNA sequencing

Single colonies were picked from the plate using a sterile inoculating loop and grown in 2-5ml LB containing the appropriate selective antibiotic for 8hrs at 37^{o}C in a shaker incubator (300rpm). Plasmid mini-preps were then carried out using the Qiagen plasmid Mini kit to extract DNA from the *E.coli* transformants, following the Qiagen plasmid purification handbook. The protocol is based on an alkaline lysis procedure, followed by binding of plasmid DNA to an anion-exchange resin under low-salt and pH conditions. Impurities such as RNA and proteins are removed by medium salt washes. Elution of the plasmid DNA is by a high-salt buffer. DNA is subsequently precipitated and desalted with isopropanol.

The plasmid DNA was subsequently sequence verified by dideoxy sequencing. Once the plasmid sequence was confirmed larger-scale preparations of the plasmid were performed.100µl of the 2-5ml starter culture was diluted 1 in 1000 into selective media in a conical flask and grown overnight at 37^{o}C in a shaker incubator (300rpm) followed by Midi-preps using the Qiagen plasmid Midi kit.

2.5 Basic protein methods

2.5.1 Denaturing SDS-PAGE using the Invitrogen NuPAGE gel system

Denaturing SDS gels were used to separate proteins for western blot analysis. Invitrogen Bis Tris Noves NuPAGE 4-12% gradient gels were used. Protein gels were run using the Invitrogen system, the gel running apparatus was

assembled and the tank filled with fresh 1x NuPAGE MES running buffer, 435µl NuPAGE reducing agent was added to the buffer chamber just prior to electrophoresis. Samples were prepared for electrophoresis by mixing with NuPAGE loading buffer to a final concentration of 1x and boiling for 3mins at 95^{o}C prior to loading. 5µl of prestained broad range protein size markers (NEB) were loaded alongside the protein samples. Gels were routinely electrophoresed at 200V for 50mins at room temperature.

2.5.2 Semi-dry transfer

The proteins separated by SDS-PAGE were transferred directly onto nitrocellulose membrane (Whatman-Protran) using the BioRad Transblot semi-dry transfer cell. The membrane and two sheets of extra thick blotting paper (BioRad) were cut according to the size of the gel, and soaked in 1x NuPAGE transfer buffer for 10mins prior to transfer. The gel was placed on top of the membrane, which was then sandwiched between the two sheets of blotting paper. A test tube was gently rolled over the gel sandwich to get rid of any trapped air bubbles. The proteins were transferred at 13V for 35mins. The membrane was soaked in Ponceau stain for 2-3mins and rinsed with ddH_2O to visualise transferred proteins on the membrane. The stain was removed by washing twice in PBST.

2.5.3 Western blotting

Following the transfer of proteins, the membrane was blocked in 5% skimmed milk powder (w/v) in PBST. This was carried out at room temperature, with gentle agitation. The membrane was then incubated in the appropriate primary antibody (Table 9) for between 1 hour at room temperature and overnight at 4^{o}C and washed 3x in PBST for 5mins/wash with gentle agitation. The membrane was then incubated in the respective goat-anti-mouse-IgG-HRP or goat-anti-rabbit-IgG-HRP secondary antibody (Dako), used at a concentration of 1 in 5000 in 1% skimmed milk in PBST for 1-2hrs at room temperature with gentle agitation. The transferred proteins were then detected using ECL reagents (Perkin-Elmer) according to the manufacturer's protocol and digitally visualised and photographed on a Biorad imager.

Table 9: Primary antibodies and conditions used in western blot

Antibody name	Antibody type	Primary
Anti-PCNA: PC10 (Abcam)	Mouse monoclonal	1 in 4000, 1% milk
Anti-PCNA (Abcam)	Rabbit polyclonal	1 in 1000, 1% milk
Anti-Actin:Ac40 (A4700) (Sigma)	Mouse monoclonal	1 in 1000, 1% milk
Anti-Fen1: 4E7 (GeneTex)	Mouse monoclonal	1 in 1000, 1% milk
Anti-Fen1: EPR4459(2) (GeneTex)	Rabbit monoclonal	1 in 1000, 1% milk
Anti-Lig1: 10H5 (Thermo Scientific)	Mouse monoclonal	1 in 1000, 1% milk
Anti-Lig1 (Thermo Scientific)	Rabbit polyclonal	1 in 1000, 1% milk
Anti-Lig1: 10G12 (Abcam)	Mouse monoclonal	1 in 1000, 1% milk
Anti-XPA: 5F12 (Abcam)	Mouse monoclonal	1 in 1000, 1% milk
Anti-XPG (Abcam)	Rabbit polyclonal	1 in 1000, 1% milk
Anti-XPG: 8H7 (Abcam)	Mouse monoclonal	1 in 1000, 1% milk

2.5.4 Coomassie staining

The gel was washed in ddH_2O then stained with Coomassie SimplyBlue™ SafeStain (Invitrogen) overnight with gentle rocking, followed by repeated washes in ddH_2O until the excess stain was removed and the gel could be analysed.

2.6 Cell culture techniques

2.6.1 Cell Culture

Human primary fibroblast cells were maintained in 5% CO_2, at 37^{o}C in DMEM supplemented with 20% Foetal Calf Serum (FCS), 2mM Glutamine (Invitrogen) and 1% penicillin and streptomycin.

Human lymphoblastoid cells lines were Epstein Barr virus (EBV) transformed by Public Health England (formerly Health Protection Agency) and grown in RPMI 1640 (Sigma Aldrich) supplemented with 20% FCS, 2mM Glutamine (Invitrogen) and 1% penicillin and streptomycin and maintained in 5% CO_2, 37^{o}C at a density of between $3x10^5$ - $2x10^6$ cells/ml.

Human MRC5 lung fibroblast cells were grown at 37^{o}C, 5% CO_2 in MEM (Sigma Aldrich) supplemented with 10% FCS, 2mM Glutamine (Invitrogen) and 1% penicillin and streptomycin.

2.6.2 Freezing cell line stocks

Fibroblast or lymphoblastoid cells in exponential growth phase were harvested and resuspended in freezing medium before immediate transfer into CryoTubes, followed by overnight storage in a Mr. Frosty and long term storage in liquid nitrogen.

2.6.3 Lymphoblastoid UV survival assay

UV sensitivity of lymphoblastoid cells was determined by irradiating cells in PBS (1x106 viable cells in 1ml) with UVC (254nm, 1Jm-2sec-1), adding 4ml growth medium and counting viable cells using trypan blue exclusion after a 72hr recovery period.

2.6.4 Fluorescence activated cell sorting (FACS)

The BD LSRII flow cytometer fitted with Blue (488nm), Violet (405nm), Red (633nm) and UV lasers (35-60nm) at Barts and The London School of Medicine and Dentistry Flow Cytometry Core Facility or the CyAn ADP Analyser

(Beckman Coulter Inc) in the Flow Cytometry Core Facility of the Jenner Institute Oxford, was used to analyse and characterise primary fibroblast cells actively synthesising DNA (bromodeoxyuridine (BrdU) incorporation) in terms of their cell cycle position defined by a total DNA stain (7-amino-actinomycin D (7-AAD) or propidium iodide (PI) intensities.

Asynchronously growing primary fibroblast cells passage 10 (P10) or below were plated at a density of 4x 10^5 for 24hrs in 10ml of complete media in 10cm dishes. The cells were then pulse labelled for 30mins with BrdU at a final concentration of 10μM. Cells were subsequently trypsinised, fixed and stained using the BD Pharmingen™ FITC BrdU Flow Kit (catalogue number 559619) The fixative in the kit is a mixture of paraformaldehyde and the detergent saponin which permeabilises the cells. DNase treatment of the cells exposes incorporated BrdU before staining with an FITC-conjugated anti-BrdU antibody and 7-AAD. The protocol was followed according to the manufactures instructions. In addition to the labelled cells, an aliquot of unlabelled cells was stained as a negative control. Samples were resuspended in BD Pharmingen™ Stain Buffer (FBS) and kept at 4°C prior to FACS analysis.

Alternatively the cells were fixed by the addition of methanol to 70%. After storage at -20 °C cells were treated with 0.2 mg/ml pepsin in 2M HCl for 20mins, then incubated in 0.5% BSA in PBS with 0.5% tween 20 and a 1/50 dilution of anti-BrdU antibody (Becton Dickinson) for 1hr. After washing in 0.5% BSA, incubation with 1/100 anti-mouse secondary antibody coupled to Alexa Fluor 488 (Molecular Probes) was for 30mins, followed by washing and resuspension in PBS with 0.5 mg/ml RNaseA for 15mins. Propidium iodide was added (10 μg/ml) just before analysis.

The fixed samples were then analysed on the flow cytometer (20,000 events collected) to detect 7-AAD/PI and BrdU with the width parameter used to exclude doublets of cells. Histogram analysis of the signal allowed the determination of the percentage of cells in Sub G1, G1, S phase and G2M phases. Two-colour flow cytometric analysis permitted the enumeration and characterisation of cells that were actively synthesising DNA in terms of their cell cycle position.

Data was analysed by Summit software (Dako) in Oxford and BD FACSDiva software version 6.1.3 in London.

FACS analysis was performed with Mr Gary Warnes (Flow Cytometry Core Facility, Blizard Institute, Barts and The London School of Medicine and Dentistry) and by Mr Drew Worth (Flow Cytometry Core Facility, Jenner Institute Oxford).

2.6.5 Generation of mammalian cells stably expressing normal levels of His-PCNA

To generate silence mutated His-PCNA and His-PCNA p.Ser228Ile expressing cell lines, MRC5 cells were transfected with DNA encoding wild-type PCNA or PCNA-Ser228Ile tagged with 6 histidine residues in pcDNA3.1 (G418 resistant) following the X-tremeGene 9 transfection protocol (Roche). The cDNA constructs carried in addition the following silent mutations to render them refractory to targeting by the small interfering (si)RNA.

Wild type: GC CGA GAT CTC AGC CAT AT

Mutant: GC **A**G**G** GA**C** **T**T**A** **TCT** CAT AT

24hrs prior to transfection, $5x10^5$ MRC5 cells were plated in each well of a 6 well plate in 5ml of complete medium. 100µl of serum free medium containing 1µg of DNA and 3µl of X-tremeGene 9 DNA transfection reagent was incubated for 15mins at room temperature and then added dropwise to the cells. A negative control with no plasmid was also included. Cells were trypsinised after 40hrs and each well was resuspended in 5ml of medium containing 1200µg/ml G418 (Geneticin PAA). Cells were then moved into 10cm dishes with 10ml of G418 containing medium, 2ml of cells were added to 2 dishes, 200µl to 2 dishes and 20µl to the last 2 dishes. The dishes were then agitated to spread the cells evenly and allowed to grow for 10 days at 37^oC, with media change as appropriate. The negative control was checked to ensure that it did not contain any colonies. Well separated colonies were then selected using cloning cylinders (Bel-Art) and sterile silicone grease. Each colony was trypsinised and expanded into a single well of a 6 well plate. When the well reached confluence,

cells were trypsinised and half were expanded into a T25 flask, the remainder were analysed by western blot.

2.6.6 Transfection of MRC5 cells with siRNA

MRC5 cells stably expressing silence mutated His-PCNA at normal levels were transfected with 10nM PCNA siRNA [GCCGAGAUCUCAGCCAUAUTT(Eurofins MWG Operon)] or the non-specific siRNA pool (Eurofins MWG Operon), using the HiPerFect transfection reagent (Qiagen).

The day before transfection, $1x10^4$ cells were seeded per well of a 24 well plate in 500µl of appropriate complete medium without G418. The cells were incubated for 24hrs at 37^oC, 5% CO_2. On the day of transfection 75ng PCNA or non-specific control siRNA was diluted in 100µl MEM culture medium (without serum or antibiotics), 3µl of HiPerFect reagent was then added to the diluted siRNA and mixed by vortexing. The samples were incubated for 10mins at room temperature to allow transfection complexes to form. The complexes were then added dropwise on to the cells, plates were gently swirled to ensure uniform distribution of the transfection complexes. For each well transfected with PCNA siRNA two control wells were included, one with 100µl of culture medium alone and one with 100µl of medium and 3µl of HiPerFect reagent. Cells were then incubated for 72hrs, harvested and analysed by western blot.

2.7 Protein Interaction analyses

2.7.1 Purification of 6xHIS-tagged proteins from E.coli under denaturing conditions

First pET16b-Lig1, pET16b-Lig1 PIP*, and pET16b-empty vector were transformed into BL21 competent *E.coli* on ampicillin containing plates. A single colony was picked into typically 6ml of ampicillin containing LB and grown for 8hrs in the shaker incubator. This culture was then inoculated into a 100ml culture and left overnight until the optical density (OD) was 0.6-0.9. A 1ml sample (un-induced control) was taken, centrifuged at 3000xg for 2mins and stored at -20^oC.

IPTG 0.5mM was added to each culture to induce protein expression and the temperature in the shaker incubator reduced to 24^{o}C for the next 3hrs. A further 1ml sample was taken (induced) and the rest of the cells were harvested by centrifuging at 3000g for 2mins. The bacterial cell pellets were stored at -20^{o}C.

The protocol below was modified from the Qiagen QIAexpressionist.

Buffer composition

100mM NaH_2PO_4

10mMTris-Cl

8M urea

(B) pH adjusted to 8.0 using NaOH,

(C) pH adjusted to 6.3 using HCL,

(D) pH adjusted to 5.9 using HCL,

(E) pH adjusted to 4.5 using NaOH.

Bacterial cell pellets were thawed on ice for 15mins and then resuspended in buffer (B) at 5ml/g wet weight. The cells were then lysed at room temperature on a roller mixer, taking care to avoid foaming. Lysis was considered to be complete when the solution became translucent (about 60mins). The Lysate was then centrifuged at 10,000xg for 30mins at room temperature (to avoid precipitation of the urea) to pellet the cellular debris. 5µl of 2x NuPAGE loading buffer was added to 5µl of supernatant (cleared lysate).

The appropriate volume of Ni-NTA resin was washed in buffer (B) (centrifuged at 300 x g for 3mins). 1ml of 50% Ni-NTA slurry was added to the total lysate and mixed at room temperature on a rotary mixer at 200rpm for 60mins. The Lysate-resin mixture was added to an empty polypropylene column. The flow-through was kept for SDS-PAGE analysis. The column was washed twice with 4ml of buffer (C). The pooled wash fractions were kept for SDS-PAGE analysis. The recombinant proteins were eluted four times with 0.5ml buffer (D), followed by four times with buffer (E). Each fraction was analysed by SDS-PAGE on a 4 - 12% Bis Tris Noves NuPAGE gel which was subsequently stained with Coomassie.

2.7.2 Surface Plasmon Resonance (SPR)

HBS-EP running buffer composition:

- 10mM Hepes pH7.4
- 150mM NaCl
- 3mM EDTA
- 0.005% (v/v) Tween 20

TBS running buffer composition:

- 50mM Tris pH 8.5
- 150mM NaCl
- 1mM EDTA
- 0.005% Tween 20

Gomes running buffer composition:

- 30mM HEPES–NaOH pH7.5
- 1mM EDTA
- 0.5% inositol
- 0.01% NP-40
- 8mM magnesium acetate
- 100mM NaCl.

The SPR phenomenon occurs when polarised light, under conditions of total internal reflection, strikes an electrically conducting gold layer at the interface between media of different refractive index: the glass of a sensor surface (high refractive index) and a buffer (low refractive index).

Surface plasmon resonance (SPR) experiments were performed using a Biacore T200 (Center for Cellular and Molecular Physiology, Wellcome Trust Human Genetics Unit, Oxford) to monitor the interactions between PCNA, Lig1 and Fen1, following the manufacturer's instructions and best practice guidelines recommended by (Myszka, 1999).

Surface/pH scouting was performed by running recombinant wild type PCNA protein diluted 1 in 10,000 (stock concentration 0.25mg/ml) in 10mM sodium

actetate pH 4.0 (i.e. below the isoelectric point of PCNA, pH 4.57) over the chip. An increase in SPR signal (expressed in response units, RU) is looked for, indicating adherence to the chip, running buffer is then run over the chip surface and the SPR signal checked to ensure that it returns to baseline indicating that the adherence is reversible and experimental coupling conditions.

About 500 response units (RU) (~2 fmol) of either recombinant wild-type PCNA, PCNA IDCL* (Leu126Ala, Ile128Ala, mutations disrupt the PCNA Inter-domain connector loop) and PCNA Ser228Ile were immobilised on separate running channels on the surface of a BIAcore Series S Sensor CM5 or CM4 chip by the carbodiimide coupling method as per the manufacturer's instructions. Channel 1 was left empty. The running buffers used in the analysis were either HBS-EP, TBS or Gomes buffer, The pH and salt were also varied, pH was increased to 8.8 and salt concentration to 300mM within these buffers with the aim of optimising experimental conditions. The PCNA–Fen1 and PCNA-Lig1 interactions were measured by injecting 55µl/2mins of increasing concentrations of recombinant Fen1 PIP* (Leu340Ala, Phe343Ala, Phe344Ala mutations disrupt the PIP box), Lig1 PIP* (Ile5Ala, Phe8Ala, Phe9Ala mutations disrupt the PIP box), WT Fen1 or Lig1 over the immobilised WT or mutant PCNAs at a flow rate of 30µl/min. To determine the relative affinities of Fen1 or Lig1 for PCNA, steady-state binding levels (RUeq) were plotted as a function of Fen1 or Lig1 concentration.

2.7.3 Investigation of the salt resistance of the PCNA-Fen1 interaction

Interaction buffer (stored at 4^{o}C):

- 10% Glycerol
- 25mM NaCl
- 0.01% NP-40
- 25mM Tris pH7.5
- Protease inhibitors (Roche)
- Volume adjusted with ddH_2O

HeLa nuclear extract and recombinant WT PCNA and PCNA p.Ser228Ile both stored in Gomes buffer were thawed gently on ice. HeLa nuclear extract was

then centrifuged at 1500xg for 15mins at 4^{o}C to remove precipitate. Supernatant was then transferred to a new 1.5ml Eppendorf tube. 180µl of HeLa nuclear extract was added to each of four prechilled 1.5ml Eppendorf tubes. Equal concentrations of each PCNA protein (10µg) in IP buffer (total volume 30µl) were then combined with the HeLa nuclear extract in separate tubes and vortexed gently. The 3rd tube contained HeLa nuclear extract and 30µl of IP buffer only as a negative control. Samples were incubated on ice for 1 hour, prior to centrifuging associated extracts at 21,000xg for 10mins at 4^{o}C. 10µl of supernatant was removed and mixed with x2 NuPAGE loading buffer as start material. 40µl of 50% slurry Ni-NTA beads, prewashed in interaction buffer, were added to each sample, prior to incubation on a rotary mixer at 4^{o}C for 1hr. Post-incubation samples were spun at 3,000rpm for 3mins, and tubes left on ice for 10mins to allow the beads to settle at the base of the tube. 10µl of flow-through was removed and combined with x2 LB, the remaining supernatant was removed carefully and discarded. Beads were then carefully washed 2x in 100µl of interaction buffer followed by elutions at 100mM, 500mM and 1M NaCl. Beads were dried with 3MM blotting paper strips, resuspended and boiled in 20µl of interaction buffer and 50µl 2x NuPAGE loading buffer. Proteins were separated by SDS-PAGE on a 4-12% NuPAGE gradient gel (200V for 50mins).

2.7.4 Cell extract preparations

Benzonase lysis buffer (made on ice):

- 40mM NaCl
- 50mM Hepes pH 8.5
- 2mM MgCl
- 0.1% NP-40
- 10% Glycerol
- Protease inhibitors (Roche)
- Volume adjusted with ddH_2O

1µl/ml Benzonase nuclease (Merck) was added immediately before use.

Human lymphoblastoid cells were pelleted, washed in cold PBS and resuspended in benzonase lysis buffer (1ml per 10^8 cells) in chilled eppendorf

tubes on ice for 20mins.The NaCl concentration was adjusted to 150mM. The lysate was incubated on ice for 15mins and subsequently centrifuged at 4^{o}C at 20,000xg for 10mins. The supernatant was then transferred to a clean tube and buffer exchanged into interaction buffer using 5ml Zeba desalt spin columns (Thermo Scientific) according to the manufacturer instructions. Briefly, columns were centrifuged at 1,000xg at 4^{o}C for 2mins to remove storage solution. The columns were prepared by adding 2.5ml of interaction buffer slowly to the resin and centrifuging at 1,000xg at 4^{o}C for 2mins, then discarding the flow-through. This step was repeated three times. The supernatant was then slowly applied to the centre of the compacted resin bed. A stacker of 100µl of interaction buffer was added to samples <750µl.

2.7.5 Bradford Assay

Bradford assays were performed to determine protein concentration of the samples. 2µl of the cell extract was added to 800µl H_2O and 200 µl of Bradford reagent (Sigma). The samples were mixed in an Eppendorf tube and transferred to plastic cuvettes, and the absorbance at 595nm determined using a spectrophotometer. The protein concentrations were then estimated in comparison to a standard curve plotted from the absorbance of known concentrations of BSA solutions.

2.7.6 Co-immunoprecipitation assays - Interaction between endogenous PCNA and endogenous Fen1, Lig1, XPG, XPA

Lymphoblastoid cell extracts were prepared as described above. Protein concentration was determined by Bradford assay and the concentration adjusted with interaction buffer. A sample was taken as input material.

50µl of a 50:50 mix of protein A and G magnetic Dynabeads (Diagenode) blocked with 1% BSA for 1hr were used per sample. The beads were washed 2x with interaction buffer and resuspended in the same buffer prior to use.

1µg of antibody (Table 10) was used per cell extract sample (200-250µl containing 2-mg/ml protein), vortexed briefly and left incubating on ice for 1hr to allow binding, before centrifuging at 4^{o}C at maximum speed for 15mins. A

control IgG was included for each extract. Each sample was then transferred to a new cold tube. (10µl of each sample was then removed as pre-cleared load material). 50µl of beads were added to each sample and allowed to bind at 4^{o}C for 1 hour on the rotary tube mixer. The tubes were then placed on a chilled magnet and the supernatant removed (20µl of flow-through taken for each sample). The beads were then washed x4 in 500 µl of interaction buffer, using the magnet and on the last wash the beads were moved to a new tube. The supernatant was discarded and the beads resuspended in 30 µl of 1x NuPAGE loading buffer and boiled for 3mins.

0.5-5% of input, pre-cleared load and flow-through material and 30-60% of bead material were loaded on a 4-12% gel and proteins were separated by SDS-PAGE (200 volts for 40mins). Western blot analyses were performed as described in section 2.5.3.

Table 10: Antibodies used in co-immunoprecipitation assays

Antibody name	Antibody type
Anti-Fen1: 4E7 (GeneTex)	Mouse monoclonal
Anti-Fen1: EPR4459(2) (GeneTex)	Rabbit monoclonal
Anti-Lig1: 10H5 (Thermo Scientific)	Mouse monoclonal
Anti-Lig1 (Thermo Scientific)	Rabbit polyclonal
Anti-Lig1: 10G12 (Abcam)	Mouse monoclonal
Anti-XPA: 5F12 (Abcam)	Mouse monoclonal
Anti-XPG (Abcam)	Rabbit polyclonal
Anti-XPG: 8H7 (Abcam)	Mouse monoclonal

2.7.7 Interaction between glutathione S transferase (GST) tagged PCNA partner proteins and PCNA

All of the expression vectors used in this assay were cloned by Dr Catherine Green using either standard cloning procedures or the Invitrogen Gateway system.

Expression vectors used in this assay (pGEX vectors are all ampicillin resistant. All of the pET30a vectors are Kanamycin resistant):

- pGEX4T – 1 empty vector
- pGEX Fen1 PIP box (amino acids 328-355)
- pGEX XPG PIP box (amino acids 981 -1009)
- pGEX Lig1 PIP box (amino acids 1-21 with an 8 amino acid glycine-serine space to maintain equal distance between the GST and PIP box in all cases)
- pET30a – empty vector
- pET30a PCNA WT
- pET30a PCNA IDCL* (Leu126Ala, Ile128Ala, mutations disrupt the PCNA Inter-domain connector loop)
- pET30a PCNA Ser228Ile

Binding buffer composition:

- 50mM Tris
- 100mM KH2PO4/K2HPO4 pH7
- 2mM EDTA
- 150mM NaCl

Lysates from E-Coli expressing one of the PCNA constructs or pET30a empty vector as a negative control (protein concentration was adjusted with PBS) were mixed with 250µl of binding buffer in an 1.5ml eppendorf and added pairwise to lysates from cells expressing either GST (empty vector, negative control). or fusion proteins comprising GST and the PIP boxes of Fen1, Lig1 or XPG. 25µl was removed from each as input and combined with 25µl of 2x NuPAGE loading buffer. Glutathione beads (GE) were washed with binding buffer. 50µl of bead slurry was used per reaction. The beads were added to the combined cell

lysates and allowed to bind at 4^{o}C for 3hrs on a rotating shaker. The beads were then washed twice with binding buffer and subsequently with 50% binding buffer and 50% PBS (to remove the potassium and thus prevent SDS precipitation). Beads were then dried with 3MM blotting paper strips. Washed beads were resuspended in 37.5µl of 1.5 x NuPAGE loading buffer. 5µl of input and 10µl of bead material were loaded on a 4-12% gel and proteins were separated by SDS-PAGE. Gels were then washed 2 x in ddH_2O for 5mins/wash and and stained with Coomassie as described in section 2.5.4.

2.7.8 Interaction between S-tagged PCNA and endogenous Lig1

S (Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Glu-Arg-Gln-His-Met-Asp-Ser) – fusion protein association studies were performed using lysates from *E.coli* harbouring pET30a empty vector as a negative control or PCNA (WT or p.Ser228Ile) expressed from pET30a (protein concentration was adjusted with PBS)

A total volume of 250µl of S-agarose beads (Novagen) were washed in PBS and blocked with 3% BSA for 30mins before being washed and resuspended in PBS. 50µl HeLa nuclear extract (CilBiotech) resuspended in 200µl of binding buffer (as used in section 2.7.7) was added to each lysate, 5% was then removed as input. 50µl of S- agarose beads (previously washed and blocked) were added to each reaction and allowed to bind for 3 hrs at 4^{o}C on a rotating shaker. The beads were then washed again 3x with PBS and dried with 3MM blotting paper strips. Beads were then boiled in 37.5µl of 1.5 x NuPAGE loading buffer. Proteins were separated by SDS-PAGE on a 4-12% gel and transferred onto a nitro-cellulose membrane. Western blot analyses were performed as described in section 2.5.3 and membranes probed with rabbit polyclonal anti-Ligase antibody (PA5-27820), mouse monoclonal anti-PCNA antibody (PC10), mouse monoclonal anti-Actin antibody (Ac40) and the appropriate secondary antibody.

3

CHAPTER THREE

HYPOMORPHIC PCNA MUTATION UNDERLIES A NOVEL HUMAN DNA REPAIR DISORDER

3 HYPOMORPHIC PCNA MUTATION UNDERLIES A NOVEL HUMAN DNA REPAIR DISORDER

3.1 Summary

A number of human disorders, including Cockayne syndrome, UV-sensitive syndrome, xeroderma pigmentosum and trichothiodystrophy, result from the mutation of genes encoding molecules important for nucleotide excision repair. We describe a novel DNA repair disorder identified amongst the Ohio Amish community. The cardinal clinical features of this disorder include postnatal growth retardation, hearing loss, premature aging, telangiectasia, neurodegeneration, photophobia, photosensitivity and predisposition to sun induced malignancy. Assuming autosomal recessive inheritance and that a founder mutation was responsible for the condition, we used a combination of autozygosity mapping and linkage analysis to identify the underlying molecular cause. Our genetic investigation identified a homozygous missense (p.Ser228Ile) sequence alteration of the proliferating cell nuclear antigen (PCNA) causing the disease phenotype. PCNA is a highly conserved sliding clamp protein essential for DNA replication and repair. Due to this fundamental role, mutations in PCNA that profoundly impair protein function would be incompatible with life. Interestingly, while the p.Ser228Ile alteration appears to have no effect on protein levels or DNA replication, patient cells exhibit significant abnormalities in response to UV irradiation displaying substantial reductions in both UV survival and RNA synthesis recovery. Importantly the defective transcriptional responses to UV light are completely rescued by wild-type PCNA molecule, demonstrating that the mutation is causative. Furthermore we show that the p.Ser228Ile change profoundly alters the capacity of PCNA to interact with a specific subset of partner proteins. These proteins include XPG and the DNA metabolism enzymes Flap endonuclease 1 (Fen1) and DNA Ligase 1 (Lig1), molecules fundamental to genomic integrity, thus providing an explanation for the separation of function effect. Taken together our findings detail the first mutation of PCNA in humans, associated with a unique neurodegenerative disease displaying clinical and molecular features common to other DNA repair disorders, which we show to be attributable to a hypomorphic amino acid alteration. Further investigation of the altered

biological processes underlying this syndrome should provide valuable insight into the neurodegenerative disease mechanisms involved in DNA damage tolerance and repair disorders.

3.2 Introduction

Maintenance of genomic integrity is fundamentally important for normal cell division and basic cellular biological processes. The molecular characterization of human DNA damage sensitivity disorders such as ataxia telangiectasia, xeroderma pigmentosum (XP) and Cockayne syndrome (CS) has provided unique insight into the cellular mechanisms required for DNA damage tolerance and repair, and their importance for human health. The key features of such syndromes include predisposition to premature aging, malignancy, neurodegeneration, immunodeficiency, photosensitivity and growth insufficiency (DiGiovanna and Kraemer, 2012, Kamileri et al., 2012, Nance and Berry, 1992, Perlman et al., 2003).

Cockayne syndrome is caused by mutations in the *ERCC6/CSB* or *ERCC8/CSA* genes. Patients with this condition present with postnatal growth retardation, acquired microcephaly, immature sexual development, progeria, cachexia, frequent dental caries, progressive sensorineural hearing loss and learning difficulties. Ocular abnormalities are progressive and include pigmentary retinopathy, cataracts and optic atrophy. Neurodegeneration with predominantly white matter involvement, frequently manifests as spasticity and ataxia, contractures develop with age. Characteristic neuroimaging findings include: hypomyelination, supratentorial white matter loss, cerebellar atrophy/hypoplasia and bilateral putaminal calcifications (Koob et al., 2010). Affected individuals exhibit marked photosensitivity but lack the predisposition to cutaneous malignancies that is seen in xeroderma pigmentosum. The classic gestalt is described as “cachectic dwarfism” with deep set eyes, prominent ears and a pinched appearance of the nose, due to loss of subcutaneous tissue (Kraemer et al., 2007). Clinical presentation and diagnostic criteria are described in (Nance and Berry, 1992)

Xeroderma pigmentosum is characterised by abnormal pigmentation changes, extreme photosensitivity and photophobia in some but not all individuals, and an estimated 10,000 fold increased risk of non-melanocytic skin cancer and 2,000 fold risk of melanoma. Malignancies of the eye and oral cavity are also increased (Bradford et al., 2011, Lehmann et al., 2011). Ocular findings are

limited to sun-exposed structures and include chronic UV induced conjunctivitis, keratitis, corneal opacification and vascularisation and inflammatory lesions. Neurodegeneration is described in 20-30% of xeroderma pigmentosum patients. Findings include progressive intellectual deterioration, sensorineural hearing loss, peripheral neuropathy, areflexia, ataxia and speech abnormalities. The underlying pathological cause appears to be neuronal degeneration reflected by the cortical and cerebellar atrophy and sparing of the white matter seen on neuroimaging (Bradford et al., 2011, Kraemer et al., 2007, Lehmann et al., 2011).

The process of nucleotide excision repair (NER, summarised in Figure 7) comprises two DNA repair pathways that are crucial for the removal of bulky lesions in DNA including photoproducts resulting from exposure to ultraviolet (UV) light. Transcription-coupled nucleotide excision repair (TC-NER) rapidly undertakes preferential repair of DNA lesions on actively transcribed DNA strands and is deficient in Cockayne syndrome and UV-sensitive syndrome (Cleaver, 2012). Global genome repair (GG-NER), removes photoproducts more slowly from across the genome and is deficient in xeroderma pigmentosum and trichothiodystrophy (Cleaver et al., 2009)

There are 8 nucleotide excision repair-deficient complementation groups of xeroderma pigmentosum and trichothiodystrophy (TTD) (XP-A to G and TTD-A). XP-C and XP-E are deficient only in GG-NER whereas the other groups are deficient in both TC-NER and GG-NER. There is a further xeroderma pigmentosum variant group that is not defective in nucleotide excision repair but has a defect in DNA polymerase η that is required for replication of UV-damaged templates. A small number of patients have features of both xeroderma pigmentosum and Cockayne syndrome (xeroderma pigmentosum /Cockayne syndrome complex) and were found to harbour mutations in XPB, D, F or G. These patients have the dermatological and ocular findings characteristic of xeroderma pigmentosum and the neurological and developmental problems found in Cockayne syndrome (Kraemer et al., 2007, Lehmann et al., 2011). Genotype-phenotype correlations are observed, with XPD the best example where specific mutations are respectively associated

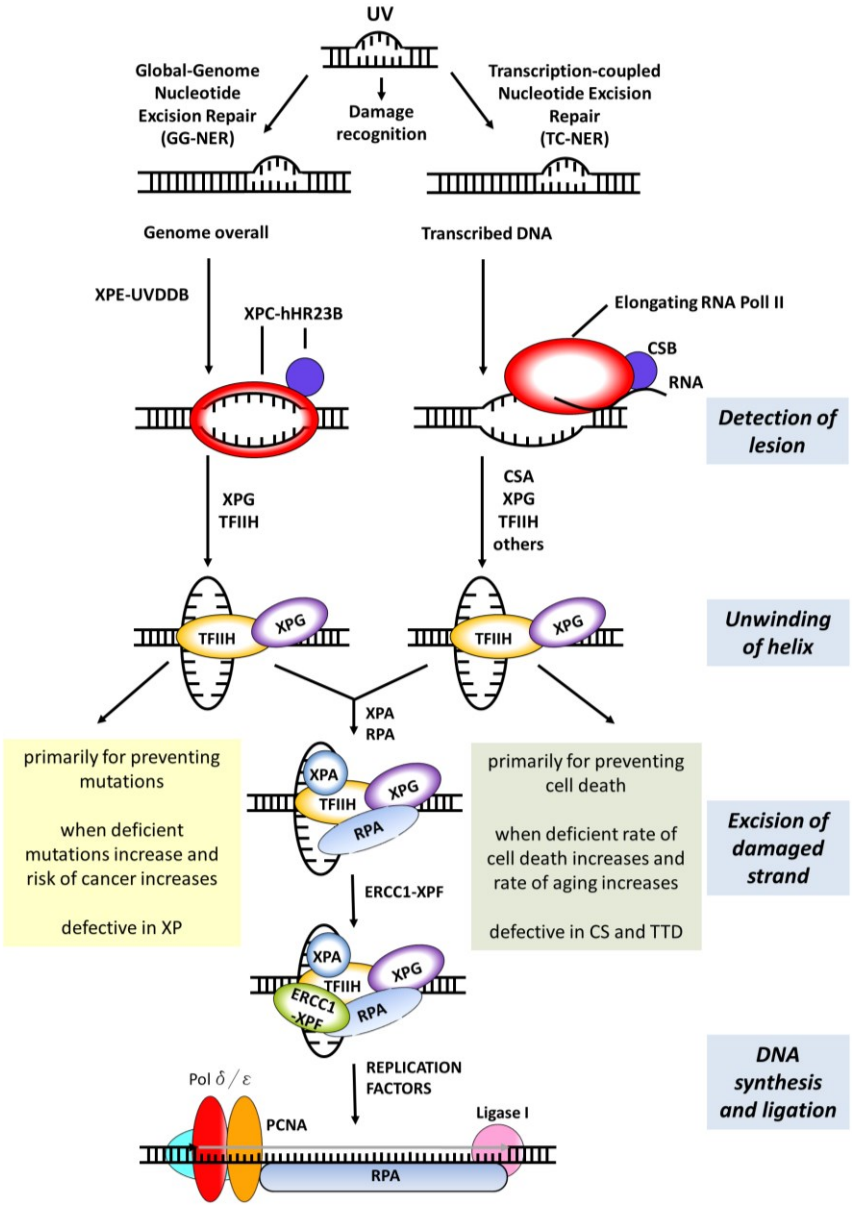


Figure 7: The process of nucleotide excision repair (NER).

NER comprises two pathways, global-genome NER (GG-NER) and Transcription-Coupled NER (TC-NER). In GG-NER, the XPC/hHR23 protein complex detects the DNA helix-distorting lesion. The XPE/UV–DDB complex detects some less significant lesions not recognised by XPC/hHR23. TC-NER is initiated by the CSB protein, when RNA polymerase II progress is stalled by a distorting lesion in the template strand, and is later followed by the *ERCC8/CSA* gene product. In both GG-NER and TC-NER, TFIIH.recruitment follows lesion detection. The XPB and XPD helicases from the TFIIH complex unwind the DNA helix around the lesion. The complex is stabilised by XPG and XPA, which verify the lesion, and RPA which binds the opposite intact single-stranded DNA. The endonucleases XPG and ERCC1/XPF excise 3' and 5' of the lesion, respectively. This results in the excision of about 30 nucleotides encompassing the damaged DNA. PCNA plays a crucial role in the repair synthesis of the remaining single-strand gap. The nick is then ligated by DNA ligase I/III. This figure was adapted from (Diderich et al., 2011, Schumacher, 2009).

with xeroderma pigmentosum, xeroderma pigmentosum/Cockayne syndrome complex and trichothiodystrophy patients (Lehmann, 2001)

After DNA damage is detected, both transcription-coupled nucleotide excision repair and global genome- nucleotide excision repair processes involve the excision of about 30 nucleotides encompassing the damaged DNA, this process requiring the xeroderma pigmentosum proteins to recognise (XPC, E), open up (XPB,D) and verify (XPA) the damaged site, and cleave the damaged strand on both the 5' (XPF) and 3' (XPG) sides. This is followed by repair synthesis of new DNA to replace the damaged section. Repair synthesis is absolutely dependent on the DNA polymerase accessory protein PCNA (Shivji et al., 1992). This protein's major cellular role is to recruit and retain the replicative DNA polymerases at the sites of DNA synthesis during DNA replication. It forms a homotrimeric ring encircling and freely sliding along the DNA helix. PCNA interacts with a large number of accessory proteins and acts as a protein recruitment platform to coordinate the multiple enzymatic activities that are required for DNA replication and repair (Moldovan et al., 2007).

Due to the large size of its genome, eukaryotic DNA replication is initiated at numerous sites called replication origins. Licensing is a crucial step in the initiation of DNA replication, because it determines which origins of replication may be used (Blow and Dutta, 2005, Diffley, 2004). After licensing, the origin is activated in the S-phase of the cell cycle, to generate a replication fork. PCNA is present at the core of the replication fork and is required to tether and stimulate the processivity of the replicative polymerases on the DNA. Due to the antiparallel nature of DNA, replications of the two parental strands are by different mechanisms, named the leading and lagging strand. First the DNA is unwound by the helicase, the parental strand is replicated in the 5'-3' direction in a continuous manner (same direction as the helicase) and is known as the leading strand. Polymerases cannot synthesis DNA in the 3'-5' direction, thus the second strand which would need to be replicated in that direction is synthesised discontinuously in the form of short fragments in the 5'-3' direction in order to overcome this problem. These fragments are called Okazaki fragments and the replicated strand is called the lagging strand (Tsurimoto and Stillman, 1991). Fen1 and Lig1 play an important part in Okazaki fragment

maturation and carry out their functions by interacting with PCNA. During lagging strand synthesis when the replicative polymerase reaches the end of a previous Okazaki fragment it displaces it generating a flap structure, this is then cleaved by Fen1 and the resulting nick sealed by Lig1 (Moldovan et al., 2007). An overview of the processes occurring at the replication fork is shown in Figure 8.

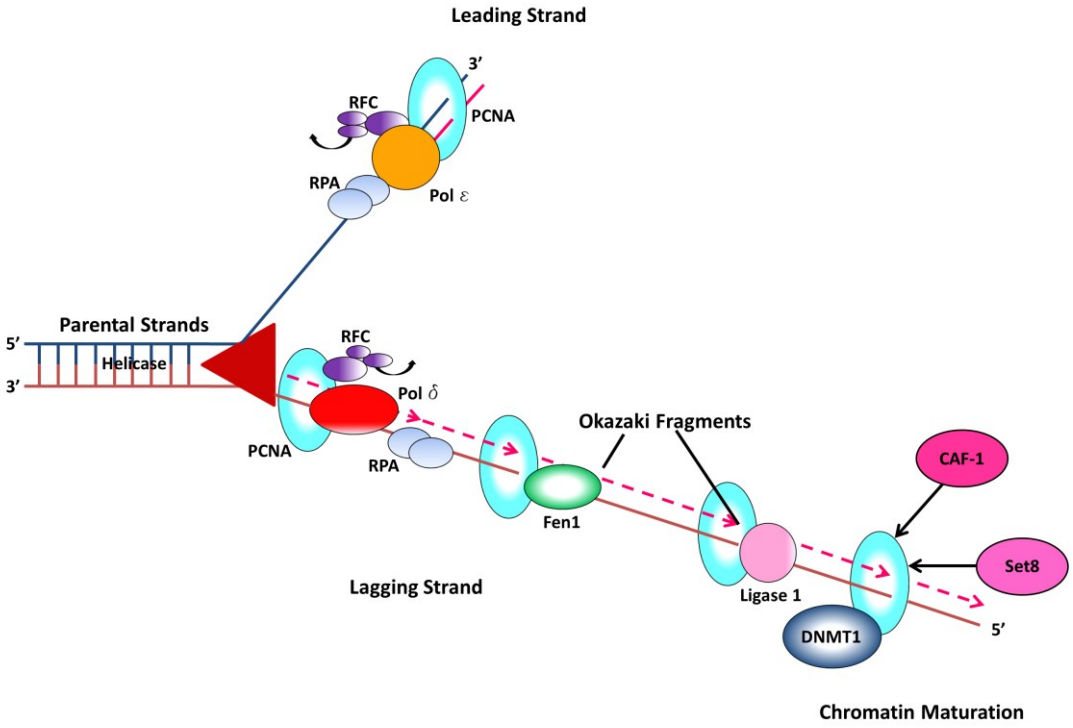


Figure 8:Simplified overview of events occurring at the replication fork

Once activated, the helicase moves along the replication fork unwinding the DNA. The clamp loader RFC, loads PCNA onto the primer template junction where it binds the replicative polymerases. Pol ε continuously synthesises DNA as the leading strand, and Pol δ synthesises DNA discontinuously as Okazaki fragments (lagging strand). The replication fork associated proteins Fen1 and Lig1 play an important role in the fragment maturation process during DNA replication. Additional proteins involved in chromatin maturation perform their functions during progression of the fork.

3.3 Methods

3.3.1 Exome sequencing

Whole exome sequencing was performed by the Otogenetics Corporation using the SureSelect Human All Exon V4 (Agilent Technologies) exome enrichment kit on an Illumina HiSeq2000. The exome sequencing produced 31,783,299 mapped reads, corresponding to 93% of targeted sequences covered sufficiently for variant calling (>10X coverage, mean depth 45X).

The fastq files were subsequently obtained from Otogenetics, BWA was used to align reads to the reference genome. SAM format was converted to binary SAM format (BAM). Duplicate reads were excluded. Variants were called using GATK UnifedGenotyper and subsequently quality filtered based on parameters including: mapping quality, strand bias, the average position of a variant in a read and SNP quality. The VCF file was annotated using the Alamut Software Suite (Interactive Biosoftware). Support for analysis of exome sequencing data was provided by Dr Barry Chioza and Dr Michael Weedon, based at the University of Exeter.

3.3.2 Functional studies

Functional studies were carried out in collaboration with Dr Catherine Green based at Cambridge University and subsequently at the Wellcome Trust Human Genetics Unit in Oxford. The aim of these studies was to ascertain whether the PCNA mutation disrupted the function of PCNA protein. These studies were carried out working closely with Dr Catherine Green and Dr Helen Chambers, a post-doctoral research fellow from the Green laboratory. Where specific experiments were performed solely by Dr Green or Dr Chambers this is mentioned in the figure legends.

Assays assessing fibroblast responses to UV were carried out in collaboration with Professor Alan Lehmann, (Genome Damage and Stability Centre, University of Sussex and Professor Tomoo Ogi (Nagasaki University Research Centre for Genomic Instability and Carcinogenesis).

Exclusion of Ataxia Telangiectasia in the affected individuals was carried out in collaboration with Professor Alexander Malcolm Taylor based at the School of Cancer Sciences, University of Birmingham. Chromosome breakage studies were performed by GSTS Pathology, and karyotyping was performed at South West Thames Genetics Laboratory.

3.4 Results

3.4.1 Phenotype of Amish individuals with a previously unrecognised DNA repair disorder

Four affected individuals aged between 11-31 years affected by a novel syndrome, were identified from a single extended pedigree comprising two kindreds within the Ohio Amish community (Figure 9). The principal features include neurodegeneration, postnatal growth retardation, prelingual sensorineural hearing loss, premature aging, ocular and cutaneous telangiectasia, learning difficulties, photophobia and photosensitivity with evidence of a predisposition to sun-induced malignancy. Phenotypic similarities with xeroderma pigmentosum, Cockayne syndrome and ataxia telangiectasia were noted (Table 11). All four affected individuals displayed short stature ranging from -3.8 to -5.2 SDS, with an absence of pubertal growth spurt in the two eldest individuals. Neurodegeneration was a consistent feature, characterised by progressive gait instability, muscle weakness, foot deformity, difficulties with speech and swallowing, learning difficulties and cognitive decline with advancing age. Prelingual onset of moderate to profound (worse at high frequency) sensorineural hearing loss was universal, a feature not commonly associated with the recognised DNA repair defects, and there was clear evidence of progression in VI:9. Muscle weakness was progressive (VI:6 has been wheelchair bound since the age of 16 years with contractures in all 4 limbs). Speech was unclear from onset and deteriorated with age (VI:6 currently has no speech and communicates with basic gestures). Progressive difficulty with swallowing solids/drooling was present in 3/4 affected subjects. Neuroimaging was only available in one case (VI:7) Magnetic resonance imaging of the brain aged eight years demonstrated cerebellar atrophy involving the cerebellar vermis and hemispheres. Inner ear abnormalities included bilateral mild labyrinthine dysplasia with mildly dilated vestibules.

All affected individuals exhibited photophobia and photosensitivity, with evidence of premalignant changes in VI:11 (basal cell carcinoma *in situ*). All have several cutaneous telangiectasias and 3/4 have conjunctival telangiectasia. Telangiectasia were not present from birth, but became apparent during early childhood. Images of affected individuals are shown in Figure 10.

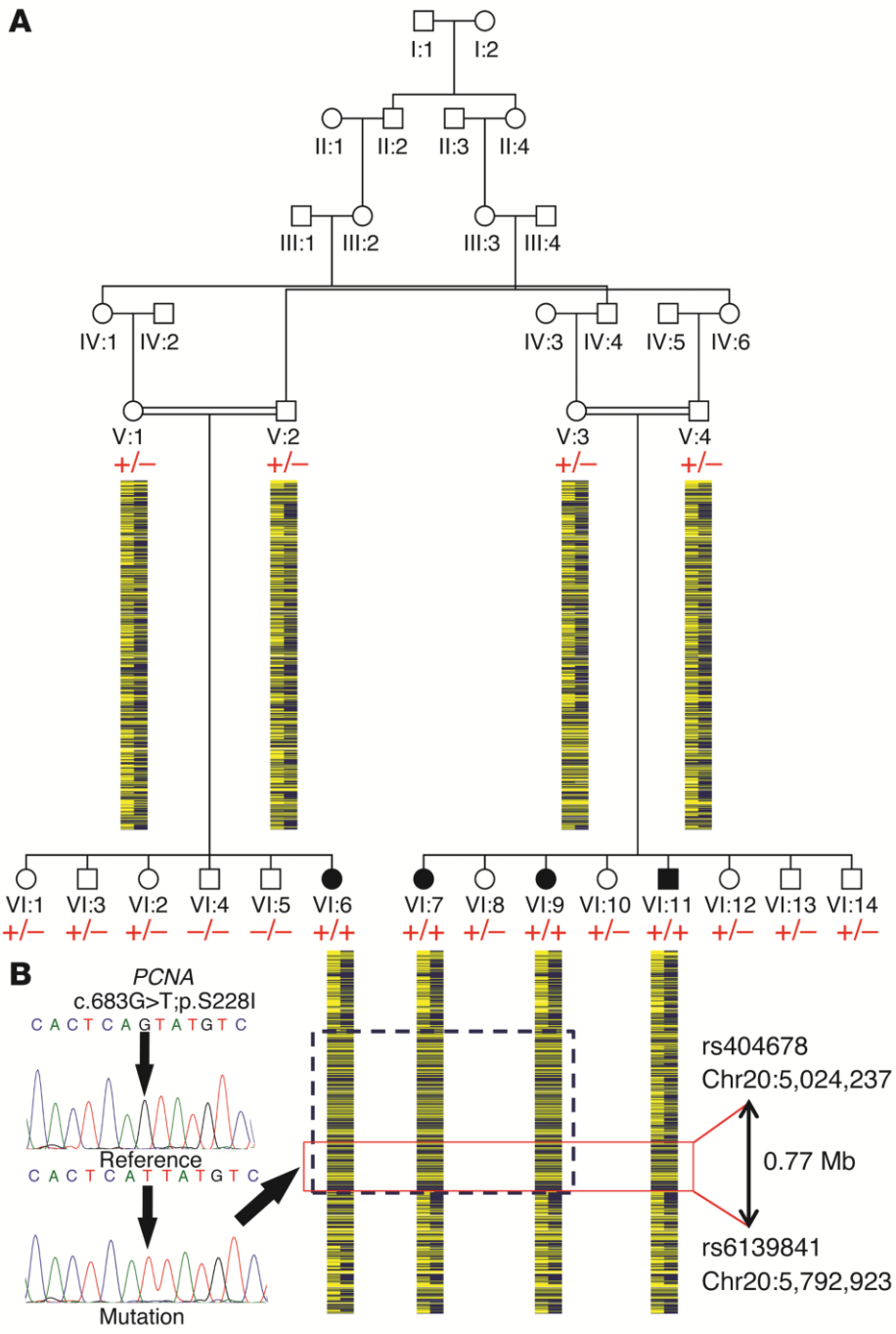


Figure 9: Family pedigree showing PCNA c.683G>T genotype data.

(**A**) simplified pedigree of the extended Amish family investigated, with pictorial representation of genotypes across ~6Mb of chromosome 20 encompassing the disease locus (2.7Mb autozygous section in affected females indicated by blue hashed box, and the common 0.78Mb region indicated by red box). All affected individuals were subsequently shown to be homozygous for the PCNA (NM_002592.2 c.683G>T) variant (indicated). Parental samples were heterozygous and unaffected siblings were either wild type or heterozygous carriers. The genotype is shown in red under individuals in generations V and VI, with (+) denoting mutant and (-) wild type. (**B**) Electropherograms showing the DNA sequence at the position of the (NM_002592.2 c.683G>T) PCNA variant in a wild-type control, and homozygous affected individual.

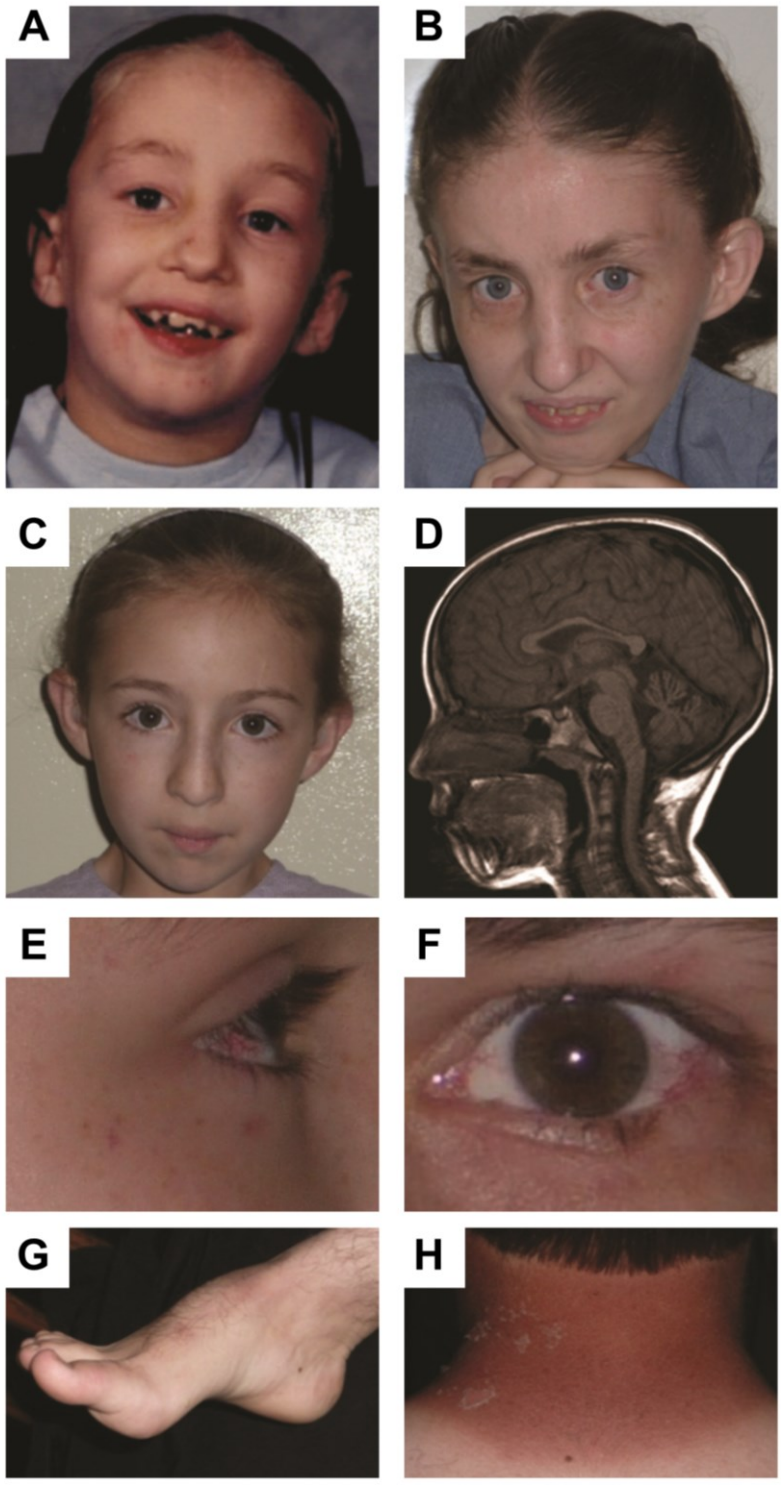


Figure 10: Images of affected individuals homozygous for PCNA c.683G>T

(**A-H**) The clinical features of individuals homozygous for PCNA c.683G>T. Individual VI:6 aged 8 years (**A**) and 31 years (**B**) showing signs of premature aging. (**C**) Individual VI:7 aged 11 years with bilateral hearing aids insitu. (**D**) Midline sagittal T1-weighted brain scan (individual VI:9 aged 8 years). There is atrophy of the cerebellar vermis resulting in enlargement of vermian sulci and mild widening of the fourth ventricle. The brainstem is normal in appearance and no supratentorial abnormality is shown. (**E** and **F**) Ocular and cutaneous telangiectasia (Individuals VI:7 and VI:11). (**G**) Pes Cavus (Individual VI:11). (**H**) Photosensitivity after minimal sun exposure (individual VI:11).

Table 11: A comparison of the clinical findings of individuals homozygous for PCNAc.683G>T and features of ataxia telangiectasia, Cockayne syndrome and the neurological form of xeroderma pigmentosum.

	VI:7	VI:9	VI:11	VI:6	ATAXIA TELANGIECTASIA	COCKAYNE SYNDROME	XERODERMA PIGMENTOSUM (neurological form)[a]
GENDER	F	F	M	F			
AGE (YRS)	11.5	14.3	26	31			
GROWTH							
Short stature	- 3.8 SDS	- 5.2 SDS	- 4.3 SDS	significant*	+	++	+
Head circumference	- 3.2 SDS	- 2.8 SDS	0.1 SDS	- 2.8 SDS	-	++	+
EYES							
Ocular telangiectasia	✓	✓	✓	-	++	-	-
Photophobia	✓	✓	✓	✓	-	+	++
Other eye abnormalities	-	-	-	UV induced conjunctivitis	Oculomotor apraxia	Cataracts, Pigmentary retinopathy, Optic atrophy	(limited to sun exposed structures) UV induced conjunctivitis, Inflammatory lesions
SKIN							
Cutaneous telangiectasia	✓	✓	✓	✓	++	-	-
Photosensitivity	✓	✓	✓	✓	-/+	++	++
Predisposition to sunlight induced skin cancer	N/K	N/K	+	N/K	-/+	-	++
NEUROLOGY							
Developmental delay	✓	✓	✓	✓	-/+	++	++
Ataxia/gait instability	✓	✓	✓	✓	++	++	+
Neurodegeneration	✓	✓	✓	✓	++	++	++
Cerebellar atrophy	N/K	✓	N/K	N/K	++	++ Also hypomyelination, Putaminal calcifications	++
Hearing loss	✓ Prelingual	✓ Prelingual	✓ Prelingual	✓[§] Prelingual	-	++ Progressive, postlingual	++ Progressive, postlingual
ADDITIONAL FEATURES							
Premature aging	N/K	N/K	-	✓	+	++	++
Other physical findings	-	-	Absent pubertal growth spurt	Diaphragmatic hernia, Absent pubertal growth spurt	Increased risk of leukemia and lymphoma	Incomplete puberty, Dental caries	

Abbreviations: *, not possible to obtain accurate measurement; N/K, not known; (✓), indicates presence of a feature in an affected subject; (++), indicates traits that are hallmarks of the disease; (+), indicates traits that are sometimes associated; (-), indicates traits that are absent or not associated, [§]; aided, but not cognitive impairment prevented accurate assessment of severity, [a]; about 25% of XP patients (from groups A, D, F and G) have neurological abnormalities (the information in the table applies to the neurological form of the disorder).

3.4.2 Identification of a potentially pathogenic variant in PCNA

3.4.2.1 Autozygosity mapping reveals a single notable autozygous region on chromosome 20p13 in affected individuals.

Assuming that a founder mutation was responsible for the condition, we used a combination of autozygosity mapping and linkage analysis in order to study this novel syndrome and identify the underlying molecular cause.

A genome-wide SNP microarray scan of DNA from affected individuals, unaffected siblings and parents was undertaken using Illumina Human CytoSNP-12v2.1 330K arrays. Inspection of the resultant genotypes identified a single notable region of homozygosity of ~2.72Mb on chromosome 20p13, shared solely by the three affected females. In the affected male, a de novo telomeric recombination event reduced this region to ~0.77Mb, delimited by markers rs404678 and rs6139841, NC_000020.10:g.5024237-5792923 (Figure 9), likely to correspond to the disease locus and containing 13 genes, 3 of which are pseudogenes, 3 hypothetical proteins and 1 antisense RNA. Microsatellite marker analysis confirmed autozygosity across this region. Multipoint linkage analysis was performed with MERLIN (Abecasis and Wigginton, 2005) under a model of autosomal-recessive inheritance with full penetrance, assuming a disease allele frequency of 0.0001 which gave a LODmax of 6.4 for the putative disease locus.

All six protein coding genes in the region were sequenced: *PCNA, CDS2, PROKR2, GPCPD1, C20orf30 and C20orf196* (for primer sequences see appendix), revealing only a single potentially pathogenic sequence variant (NM_002592.2 c.683G>T) in exon 6 of PCNA. This variant, which cosegregated with the disease phenotype and was verified in the RNA transcript obtained from whole blood of an affected individual (VI:11), is predicted to result in a substitution of a stringently conserved serine at position 228 for isoleucine (p.Ser228Ile) (Figure 11).

The variant was not detected in 360 control chromosomes of European ancestry and only two heterozygous carriers were detected in the analysis of 310 Ohio Amish control chromosomes, which is not unexpected in this

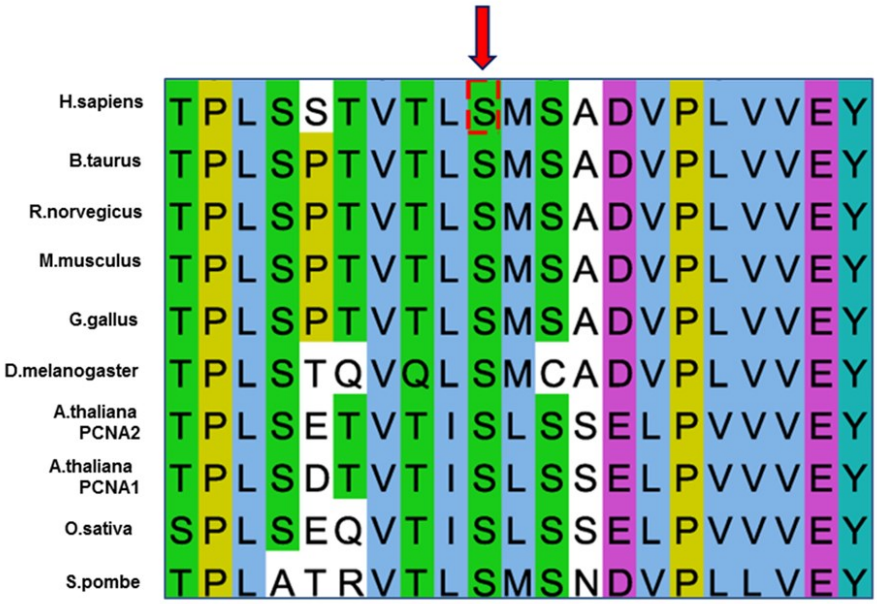


Figure 11: Conservation of Ser228.

A ClustalW2 alignment of amino acid sequences of PCNA in Homo sapiens and 9 other species, illustrating stringent conservation of the Ser228 residue, indicated with the red arrow (created using Jalview software).

endogamous community. Only a single heterozygous carrier of European origin was identified in online sequencing project databases (1000 genomes project and the Exome Variant Server, ESP6500SI release). This single mutation carrier in 7775 non-Amish individuals genotyped corresponds to a homozygous PCNA c.683T allele incidence in the region of ~1:241 million.

3.4.2.2 Other possible genetic causes of the phenotype are comprehensively excluded using a combination of genetic and functional studies.

Given that the phenotype of the affected individuals studied here has significant overlap with other DNA repair disorders, DNA from a single affected individual was subjected to whole exome sequence analysis to exclude other possible genetic causes.

As expected this confirmed the presence of the PCNA variant but identified no other known or possible pathogenic variants elsewhere after excluding variants not compatible with recessive inheritance or cosegregating with the disease phenotype. The most significant phenotypic overlap was with ataxia telangiectasia and as such, further conclusive support for the exclusion of this diagnosis as well as "ataxia telangiectasia-like" disorders was sought. The typical γ-ray-induced chromosome breakage associated with ataxia telangiectasia was investigated in two of the affected individuals. Analysis of a peripheral blood sample from VI:11 and a primary fibroblast sample from (VI:7) revealed normal chromosomes (untreated or γ-ray-induced). Alpha fetoprotein levels are typically raised in ataxia telangiectasia and immunoglobulin patterns abnormal with often a specific reduction in IgA levels (Perlman et al., 2003) both were found to be within normal limits for age in samples obtained from VI:7 and 9. Epstein-Barr virus (EBV) transformed lymphoblastoid cell lines were established from affected subjects VI:6 and VI:11. ATM and associated protein and kinase activity levels were measured in the two lymphoblastoid cells and found to be normal (Appendix). Taken together, these data provide robust and multi-layer exclusion of a diagnosis of ataxia telangiectasia. An increased number of sister chromatid exchanges after incubation with BrdU is observed in patients with Bloom syndrome who have a deficiency of the BLM RecQ DNA helicase a protein involved in several aspects of DNA repair and recombination.

No increase in sister chromatid exchanges was observed in peripheral blood lymphocyte's derived from individual VI:13, implying no gross disturbance of this pathway.

3.4.3 Functional studies of PCNA p.Ser228Ile

3.4.3.1 Crystal structure analysis and in-silico modelling of the PCNA p.Ser228Ile variant

Insilico analysis programs Polyphen-2 (HumVar), PROVEAN, SIFT and Mutation Taster were used to predict the possible impact of the p.Ser228Ile amino acid substitution on the structure and function of PCNA (Table 12). All but SIFT predicted that the variant was likely to be damaging.

Table 12: Insilico prediction of the possible impact of the p.Ser228Ile substitution on the structure and function of the human PCNA protein.

Insilico prediction tool	Predicted pathogenicity of PCNA p.Ser228Ile variant
Polyphen-2 (HumVar)	Possibly Damaging
PROVEAN	Deleterious
SIFT	Tolerated
Mutation Taster	Disease causing

PCNA is a member of the family of DNA sliding clamps and is an essential DNA replication accessory protein, highly conserved throughout evolution. Mammalian PCNA was initially identified as a cell-cycle dependent antigen, hence the name (Miyachi et al., 1978). Eukaryotic PCNA is a homotrimer comprised of three identical monomers containing two domains that can bind to each other in a head-to tail fashion to form a ring-shaped molecule that is loaded around the DNA (Krishna et al., 1994) (Figure 12). As described above, PCNA plays a central role at the replication fork, recruiting and retaining many of the enzymes required for DNA replication (Moldovan et al., 2007). Due to its fundamentally important cellular role, mutations that significantly affect PCNA

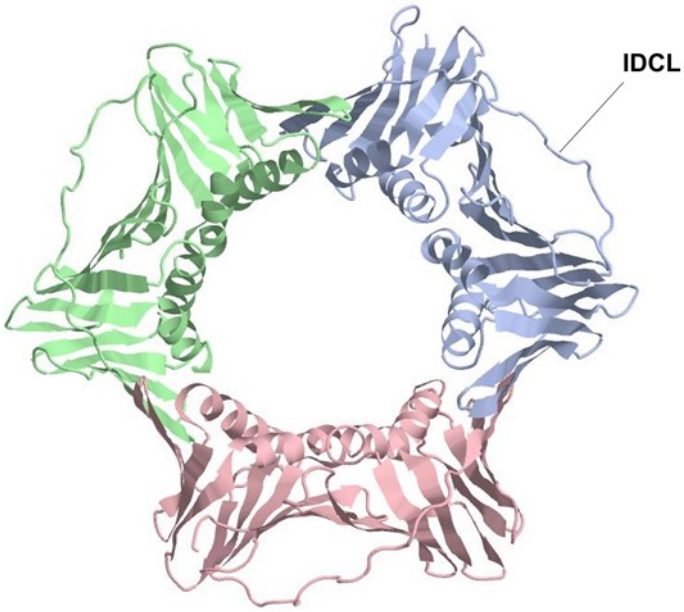


Figure 12: Crystal structure of human PCNA (PDP ID:1VYM, Kontopidis G et al. 2005).

PCNA forms a homotrimeric ring structrure. The three monomers (coloured green, pink and blue) are identical to each other and are arranged in a head-to-tail fashion. Each of the monomers are made up of two similar globular domains connected by an interdomain connector loop (IDCL). This image was rendered using FirstGlance in Jmol.

protein function would be predicted to result in embryonic lethality. However the survival of $PCNA^{Lys164Arg/Lys164Arg}$ knock-in mice, generated to determine the role of $PCNA^{Lys164}$ modifications in somatic hypermutation, demonstrates that certain hypomorphic mutations of PCNA can be tolerated in mammals (Langerak et al., 2007). No inherited human disorder arising from PCNA mutation has been described previously.

3.4.3.2 The PCNA p.Ser228Ile mutation does not have any significant effect on PCNA protein stability

EBV-transformed lymphoblastoid cells (L-) were established from 2 heterozygous carriers (L8 and L10), and 2 Amish WT controls (L7 and L9).

The Ser228 residue affected by the c.683G>T sequence alteration is distant from the regions of PCNA that mediate the formation of the trimeric ring structure (Gulbis et al., 1996), thus we hypothesised that protein stability would be unaffected by the Ser228Ile mutation. To investigate this further, extracts were prepared from lymphoblastoid cells derived from affected, heterozygous and WT individuals. Equal numbers of cells from each different cell line were lysed. Western blot analysis of PCNA detected using mouse monoclonal anti-PCNA antibody (PC10) did not reveal any difference between the levels of endogenous PCNA in the different cell lines. (Figure 13).

3.4.3.3 Cells from individuals homozygous for PCNA p.Ser228Ile show increased sensitivity to UV irradiation

Primary fibroblasts (P-) were established from 3 affected individuals, patients VI:11, VI:9, and VI:7 (P1, P2, and P3, respectively), 2 heterozygous carriers (Het1 and Het2), and wild type (WT) controls.

The reported photosensitivity in p.Ser228Ile homozygotes prompted us to examine the effect of the p.Ser228Ile substitution on cellular sensitivity to UV. Primary fibroblast line P3 cellular sensitivity to UV was compared to that of a WT cell line. Cell survival was measured by colony forming ability after exposure to doses of UV between 0-12 J/m^2. (Figure 14A). P3 cells showed increased sensitivity to UV, most noticeable at higher levels of exposure.

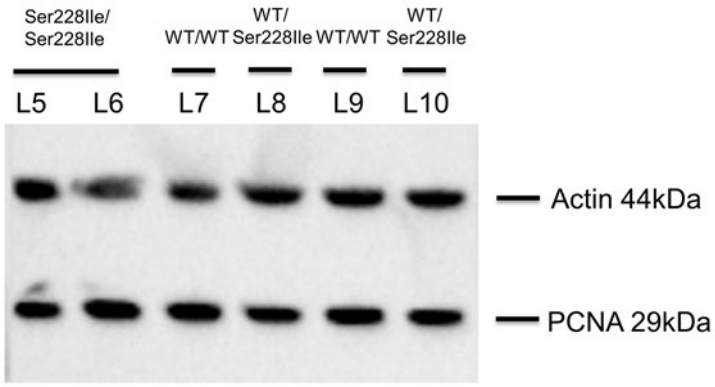


Figure 13: PCNA protein levels are not affected by the Ser228Ile alteration.

Extracts from lymphoblastoid cell lines derived from affected individuals, PCNA p.Ser228Ile heterozygous family members, and unrelated WT Amish controls were analysed by Western blot for PCNA. β-actin was included as a loading control. Levels of PCNA in affected individuals were indistinguishable from those of heterozygotes or WT controls.

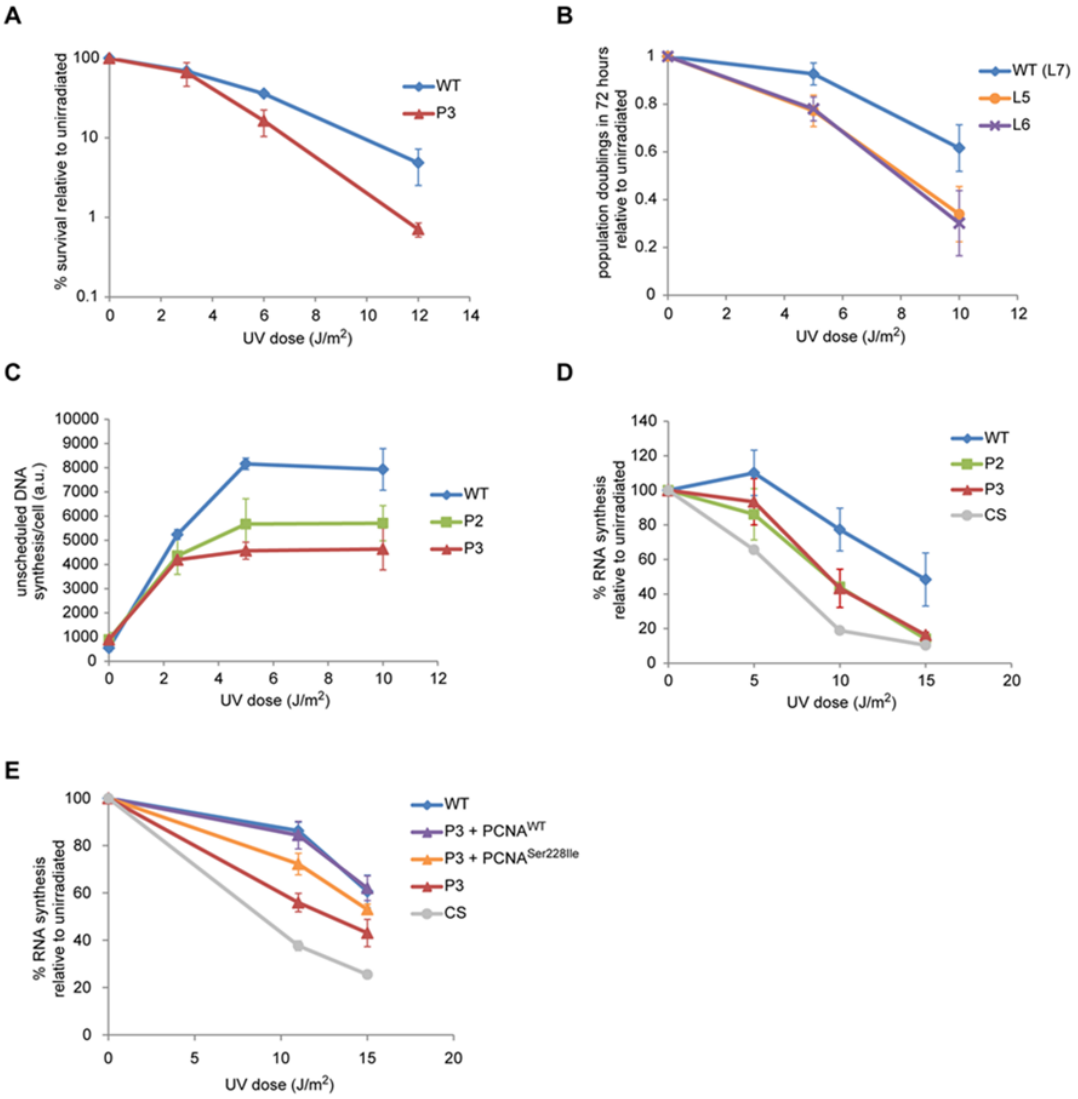


Figure 14: Abnormal cellular responses to UV.

(A) Primary fibroblasts from an affected individual are sensitive to UV. Cell survival was measured by colony-forming ability following exposure to the indicated doses of UVC. **(B)** Lymphoblastoid cells from two different affected individuals are sensitive to UV. Viable cells were counted 72hrs after exposure to UVC using trypan blue exclusion. **(C)** Cells from affected individuals have reduced Unscheduled DNA synthesis (UDS). UDS activity was measured by incorporation of radiolabelled thymidine immediately after UV-irradiation of non-dividing cells with the indicated doses of UVC. **(D)** RNA synthesis measured by the incorporation of radiolabelled uridine 24hrs after UVC-irradiation of non-dividing cells with indicated doses. The reference line in grey shows a typical response of cells from a Cockayne syndrome (CS) patient. **(E)** Defective RNA synthesis recovery (RRS) is complemented by lentiviral transduction of primary patient fibroblasts with wild-type PCNA. RRS was measured using EdU incorporation and automated high content microscopy in cells expressing high levels of ectopic PCNA. Values plotted are mean and range of 2 experiments (**A**) or mean +/- SEM of 3 (**B-D**) or 4 (**E**) experiments. Data shown in A, C and D, courtesy of Professor Alan Lehmann, Mr Jon Wing and Ms Heather Fawcett, University of Sussex. Data in E was provided by Professor Tomoo Ogi, Nagasaki University.

Primary fibroblast cell lines have only limited viability in culture, in addition to this we had difficulties growing fibroblasts from affected individuals in the colonies required for the UV survival assay. In order to maintain these precious resources for assays where they are the only suitable cell type easily obtainable, we sought to optimise an assay for UV survival in lymphoblastoid cell lines. Initial attempts involved an assay using a WT lymphoblastoid cell line in 96 well plates. Ideally a dilution where half of the wells in each dish show growth at 7 days (p0 method) would be identified in order to ascertain an accurate plating efficiency. Lymphoblastoid cells grow in clumps when observed in culture and as such would only grow at very high cell densities of at least 500 cells per well. Next cells were cultured in U bottomed 96 well plates, in which they grew at much lower densities of 10 cells per well.

The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide tetrazolium (MTT) reduction assay is commonly used to assess cell viability. Live cells are capable of reducing MTT 3-diphenyltetrazolium bromide to its insoluble formazan, which has a purple color, when cells die they lose this capability (Riss et al., 2004). This assay is quite straightforward using fibroblasts. Given their relatively large size, it is possible to count the number of cells that are viable (contain purple formazan crystals) by eye using a light microscope. Due to the small size of the lymphoblastoid cells and their propensity to grow in clumps, cells could not be counted in this way. The MTT assay can be performed by lysing the cells and measuring changes in absorbance using a plate reading spectrophotometer. Unfortunately the spectrophotometer was not designed for U bottomed 96 well plates and we were concerned about the accuracy of the assay if we were to transfer the cells to a flat bottomed plate.

Given that quite significant differences in UV sensitivity between cells derived from affected individuals and WT cell lines was expected, we next proceeded to use a less sensitive assay. Equal number of L5, L6 and WT (L7) lymphoblastoid cells were irradiated in PBS and then cultured as normal for 72 hrs, the number of viable cells for each cell line was determined by tryphan blue exclusion after the recovery period. Similar to the results seen when comparing UV survival in fibroblasts from affected individuals to those of WT controls, a significantly increased sensitivity to UV irradiation in cells homozygous for PCNA

p.Ser228Ile was observed, suggesting that the mutation specifically impairs the function of PCNA in DNA repair (Figure 14B).

3.4.3.4 The PCNA p.Ser228Ile mutation causes a defect in nucleotide excision repair

Repair synthesis and downstream events of nucleotide excision repair are absolutely dependent on PCNA (Shivji et al., 1992). To investigate this aspect further, we evaluated global genome-nucleotide excision repair which contributes to ~90% of nucleotide excision repair activity, and is, termed unscheduled DNA synthesis (UDS). The unscheduled DNA synthesis assay measures the repair synthesis step of nucleotide excision repair and is standardly used for detecting defects in response to UV irradiation in xeroderma pigmentosum patients. Replicative DNA synthesis is abolished by using non-dividing cells and the DNA replication inhibitor hydroxyurea. Cells were serum starved then incubated for 1 hr with hydroxyurea, UV irradiated and incubated for a further 3hrs with radiolabelled thymidine and hydroxyurea. Radiolabelled thymidine incorporation into repair patches was measured by liquid scintillation counting (Lehmann and Stevens, 1980). In primary fibroblast cells from affected individuals unscheduled DNA synthesis was reproducibly reduced to approximately 50-60% of normal values (Figure 14C).

RNA synthesis recovery levels following UV radiation (RRS) are used as a marker of transcription-coupled repair (Mayne and Lehmann, 1982). Cells are UV-irradiated and 24hrs later RNA synthesis is measured by incorporation of uridine. This assay is standardly used as a measure of transcription-coupled repair which is deficient in Cockayne syndrome. The assay is done in serum starved non-proliferating cells, so is not associated with defects in DNA replication. RNA synthesis recovery was significantly decreased, to approximately 30-50% of normal in cells from affected individuals (Figure 14D, Appendix Figure). The deficiency approaches that of Cockayne syndrome cells, known to be defective in transcription-coupled nucleotide excision repair.

3.4.3.5 The deficient UV responses are due to the PCNA p.Ser228Ile sequence alteration

In order to demonstrate that the observed cellular defects in nucleotide excision repair were due to the PCNA p.Ser228Ile sequence alteration, we sought to recapitulate the defect in MRC5 human fibroblasts by generating cell lines stably expressing His-tagged PCNA, either wild-type or mutated at Serine-228, at levels similar to those of endogenous PCNA and then knocking down the endogenous protein using small interfering RNA (siRNA). Given that responses to UV irradiation in cells heterozygous for PCNA p.Ser228Ile were indistinguishable from WT controls, it was difficult to predict to what extent endogenous PCNA would need to be knock down in order to recapitulate the defect seen in cells from affected individuals. Sustained knockdown of PCNA using small interfering siRNA approaches is complicated by the abundance and stability of the protein.

Niimi et al. had previously optimised such an assay to demonstrate substantial UV sensitivity in cells expressing PCNA p.Lys164Arg. Depletion such that >80% of the PCNA expressed in the cells was from the transfected cDNA was achieved (Niimi et al., 2008).

The His-tagged PCNA construct used in the PCNA p.Lys164Arg work, contained silent mutations to make it refractory to targeting by siRNA directed against endogenous PCNA. PCNA p.Ser228Ile mutant construct was generated from this siRNA resistant construct by site directed mutagenesis (primer sequences in Appendix). WT and Ser228Ile expression vectors were transfected into human MRC5 fibroblast cells. Cells were then grown in the presence of the selection antibiotic G418 at a concentration of 900µg/ml. Individual G418 resistant colonies were picked and expanded. When analysed by western blotting, none of the clones were expressing detectable levels of either WT or Ser228Ile His-tagged PCNA. Advice was sought from Dr Atsuko Niimi who had carried out the PCNA p.Lys164Arg work. Dr Niimi had had similar difficulties obtaining sufficient levels of stably expressed His-tagged PCNA, and had found that using a higher concentration of G418 (1200µg/ml) in the cloning step was critical. Increasing the concentration of G418 as advised, we were able to produce MRC5 clones expressing siRNA-resistant His-tagged

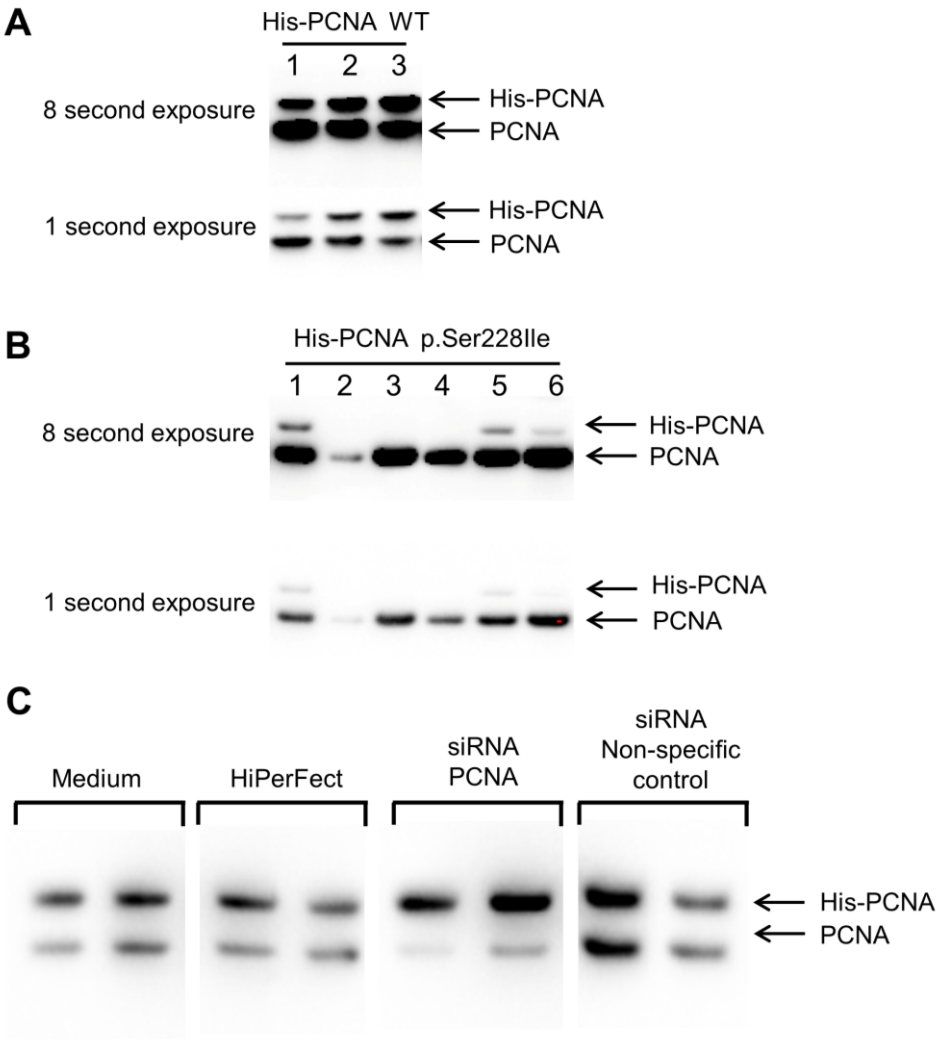


Figure 15: MRC5 cell clones stably expressing His-PCNA

MRC5 cell clones expressing (**A**) His-tagged siRNA resistant WT PCNA or (**B**) His-tagged siRNA resistant PCNA p.Ser228Ile were analysed by Western blot for PCNA. Similar expression levels of endogenous and His-tagged WT PCNA were seen. PCNA p.Ser228Ile was expressed, but only at much lower levels than the endogenous WT protein. Membrane exposures after 1sec and 8secs are shown. (**C**) Shows partial siRNA depletion of endogenous PCNA in MRC5 cells stably expressing His tagged siRNA resistant WT PCNA, levels were unaffected in cells either mock transfected or treated with non-specific control siRNA.

WT PCNA at levels similar to the levels of endogenous PCNA. However only very low levels of siRNA-resistant His-tagged Ser228Ile PCNA expression were seen after western blot analysis of large numbers of clones (Figure 15). Previously we had only sequenced verified successful Ser228Ile mutagenesis and the presence of the sequence alterations rendering the PCNA siRNA-resistant. Consequently, a mutation elsewhere in the construct perhaps in the promotor region could be affecting His-tagged PCNA p.Ser228Ile expression. Sequencing primers were therefore designed and the integrity of the entire construct sequence was confirmed eliminating this as an explanation. However it remains a possibility, that although there were no detectable differences in trimer formation or stability between PCNA p.Ser228Ile and WT PCNA, there may be subtle variations, which might account for the poor expression levels of His-tagged PCNA p.Ser228Ile. The final amount (steady state) of PCNA p.Ser228Ile that is found in the cells of affected individuals appears to be approximately the same as the WT PCNA protein in WT individuals. However, the relative stability of the two PCNA proteins in heterozygotes is unknown. It remains a possibility that the p.Ser228Ile protein has a faster rate of synthesis and degradation than the WT equivalent. Thus, if it is absolute levels of PCNA protein that are monitored by the cell, the p.Ser228Ile protein might be at an expression disadvantage.

In parallel to this attempts were made to optimise siRNA depletion of endogenous PCNA in an MRC5 cell clone stably expressing siRNA-resistant His-tagged WT PCNA. Cells were either mock-transfected, treated with 10nM non-specific control siRNA or 10nM PCNA-specific siRNA. The PCNA-specific siRNA did deplete the endogenous PCNA but not at levels that would be sufficient enough to enable us to assess whether UV-sensitivity is increased as a result of the Ser228Ile alteration. It would be possible to treat the cells with higher doses of the siRNA. However given that we were unable to generate clones expressing siRNA-resistant His-tagged PCNA p.Ser228Ile at levels approaching endogenous PCNA levels, we were unlikely to ever achieve experimental conditions suitable to assess whether the PCNA p.Ser228Ile was responsible for the abnormal responses to UV irradiation observed in cells from affected individuals.

In order to demonstrate that the PCNA p.Ser228Ile was responsible for the increased UV sensitivity observed, we next attempted to complement the defective recovery of RNA synthesis by lentiviral transduction of WT PCNA into primary cell lines from affected individuals. WT PCNA fully rescued the recovery of RNA synthesis deficiency in cells from affected individuals (Figure 14E), whereas such expression had only a minor effect on recovery of RNA synthesis in normal fibroblasts (Appendix Figure). This provides important conclusive verification that the deficient UV responses are caused by the p.Ser228Ile sequence alteration.

3.4.3.6 The PCNA p.Ser228Ile alteration has no effect on DNA replication

The most well recognised function of the DNA sliding clamp PCNA is the fundamental role that it plays in DNA replication, forming a homotrimeric ring that encircles DNA and can slide freely along it acting as a processivity factor for DNA polymerases. Given this and the generally poor growth exhibited by the affected individuals, parameters of DNA replication were next investigated.

5-Bromo-2´-Deoxyuridine (BrdU) is an analogue of the DNA precursor thymidine and as such is incorporated into newly synthesized DNA by cells entering and progressing through the S-phase of the cell cycle. When coupled with a DNA stain such as propidium iodide (PI) or &-aminoactinomycin D (7-AAD), BrdU staining permits enumeration and characterisation of cells that are actively synthesising DNA in terms of their cell cycle position (G0/1, S, G2/M phase). In order to assess the influence of the p.Ser228Ile alteration on DNA replication we incubated asynchronously growing primary fibroblast cells (at < passage 10) with BrdU for 30mins and analysed the incorporation (stained with anti-BrdU fluorescent antibodies) and DNA content by Fluorescence Activated Cell Sorting (FACS) (Figure 16). No apparent differences in cell cycle profile or progression through S-phase were observed when comparing the fibroblasts from affected individuals, with related heterozygote and unrelated WT controls.

During S phase, PCNA is recruited to so called “DNA replication factories”, which are the sites of nucleotide incorporation and are visible as foci in the nuclei of replicating cells. Immunofluorescence analysis of endogenous PCNA

in fibroblasts from affected individuals revealed no abnormalities in these foci (Figure 17). In addition, no difference in replication fork rate was detected during unperturbed S phase using DNA fibre analysis (Figure 18). Taken together, these data suggest that the Ser228Ile variant does not dramatically interfere with the major replicative function of PCNA and that bulk DNA replication is not grossly perturbed.

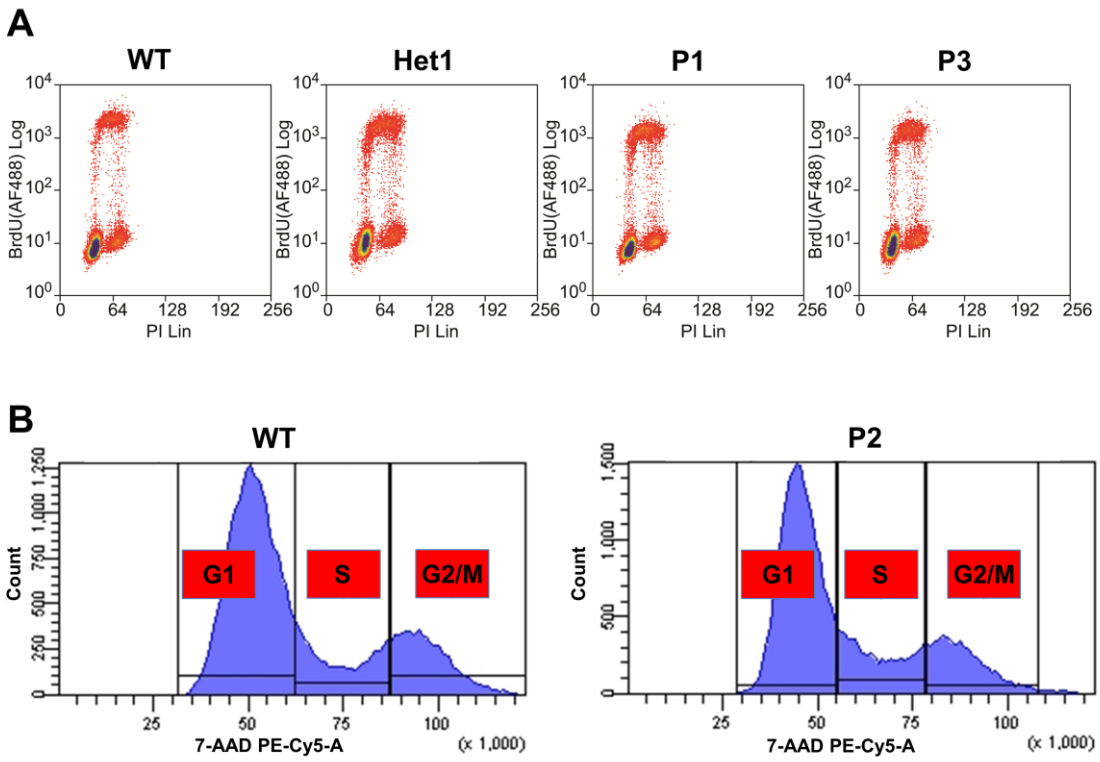


Figure 16: Cells homozygous for PCNA p.Ser228Ile have normal replication parameters.

(A) FACS analysis of S phase in primary cells. Cells at passage number 10 or less were labelled with BrdU for 30mins and BrdU incorporation and DNA content analysed by FACS. No significant differences were observed between cells from affected individuals (P1, P3) and related (Het1) or unrelated (WT) control cells. Representative scatter plots are shown, each line was independently analysed at least twice. **(B)** FACS analysis of DNA content in asynchronously growing primary cells. Cellular DNA was stained with 7-AAD; accordingly, DNA content is represented by 7-AAD intensity. A G1 peak, representing cells with 2N DNA content, and a G2/M peak, representing cells that have undergone replication and therefore possess a 4N DNA content, are labelled. The area between these two peaks is representative of cells in S phase. Typically 60% of cells are seen to be in G1 and 20% in each of S and G2/M phases. No significant differences were observed between the cell cycle profiles of affected individuals and WT controls, each line was independently analysed at least twice.

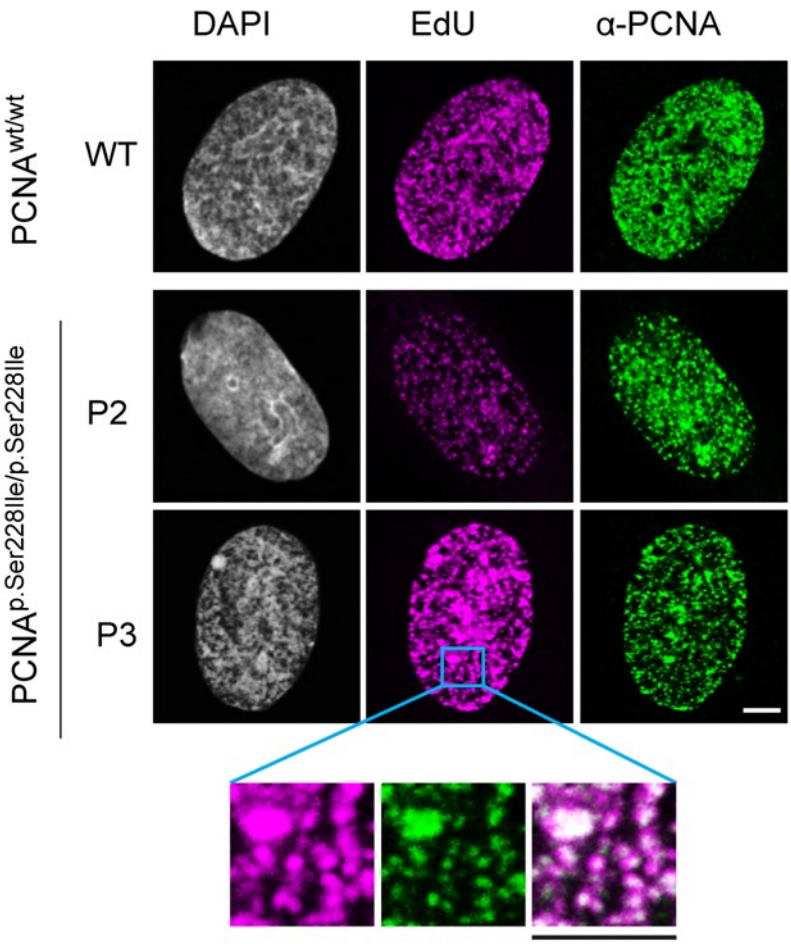


Figure 17: Cells homozygous for PCNA p.Ser228Ile have normal replication factories

Immunofluorescence of primary control fibroblast cells (WT) and fibroblast cells from affected individuals homozygous for PCNA p.Ser228Ile (P2 and P3). Soluble proteins were removed by triton extraction prior to fixation, and cells were immunostained for PCNA (right hand panel). The centre panel shows sites of active DNA replication, visible as foci of incorporation of EdU (5-ethynyl-2´-deoxyuridine), a nucleoside analogue, into the DNA. Enlargement shows colocalisation between these signals and reveals no notable differences between the individuals investigated. DNA throughout the nucleus was counterstained with DAPI (left hand panel). Data courtesy of Dr Helen Chambers, University of Cambridge.

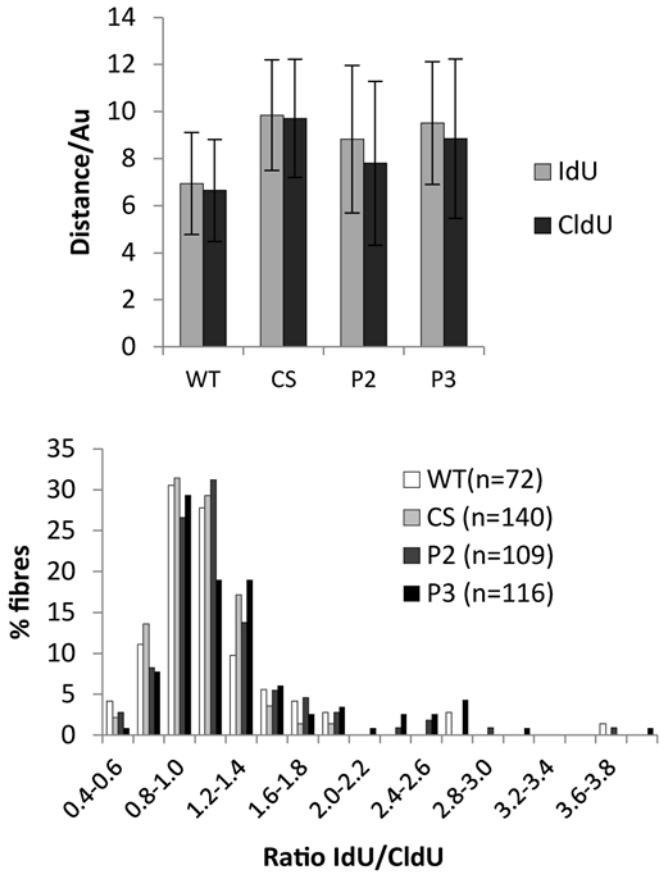


Figure 18: Cells homozygous for PCNA p.Ser228Ile have normal replication fork rates

Analysis of replication fork progression. Primary control cells and Cockayne syndrome cells (WT and CS) and cells homozygous for PCNA p.Ser228Ile (P2 and P3) were incubated sequentially in medium supplemented with IdU (5-iodo-2'-deoxyuridine) and CldU (5-chloro-2'-deoxyuridine), and the length of the DNA fibre replicated during each pulse measured. Comparable lengths of DNA were replicated during each pulse (top panel, compare light grey [IdU] and dark grey [CldU] bars for each cell line). The frequency distribution of the ratio of IdU/CldU within each fibre was similar between cell lines (bottom panel). The error bars indicate SEM of two experiments. Data courtesy of Dr Catherine Green, University of Oxford.

3.4.3.7 Specific PCNA interactions are perturbed by the Ser228Ile mutation

Serine 228 lies near the external face of PCNA and the mutation is therefore unlikely to affect its interaction with DNA (Figure 19). Interestingly however, the altered residue is close to the inter-domain connecting loop (IDCL) region which mediates the interaction of PCNA with many of its protein partners. PCNA partner proteins, including DNA polymerase delta (PolD), DNA Ligase 1 (Lig1) and Flap endonuclease 1 (Fen1), often contain a PCNA interacting protein (PIP) box, a sequence motif (Gln xx [Leu or Ile or Met] xx [Phe or Try][Phe or Try]) which mediates the inter-domain connecting loop interaction (Figure 20) (Warbrick, 1998). The crucial amino acids on PCNA that directly contact the PIP box residues are Met119 to Pro129, but other regions of PCNA either side of the inter-domain connecting loop are also implicated in binding events. In published crystal structures serine 228 does not directly contact the PIP box-containing peptide (Gulbis et al., 1996, Bruning and Shamoo, 2004), although it's close proximity to this crucial interaction region may generate subtle positional alterations within this crucial domain, with functional consequences.

To determine whether specific protein interactions are perturbed by the p.Ser228Ile alteration, a SILAC (stable isotope labelling by amino acids in culture) affinity purification assay was performed. Extracts were made from HeLa cells grown in either normal medium, or SILAC medium containing "heavy" amino acids: [$^{13}C_6/^{15}N_2$] lysine and [$^{13}C_6/^{15}N_4$] arginine. These extracts were individually exposed to affinity columns comprising recombinant PCNA, either WT or PCNA p.Ser228Ile. Proteins that bound to each column were eluted, the eluates combined, and the relative binding to each column was determined by mass spectrometry (for more detailed methodology please see Appendix). This identified clear perturbations in the interaction of PCNA p.Ser228Ile with a limited number of interacting partners, most strikingly with Fen1 and Lig1 (Figure 21A, Appendix Table). This experiment revealed relatively increased amounts of Fen1 and Lig1 in the 500mM eluates from the

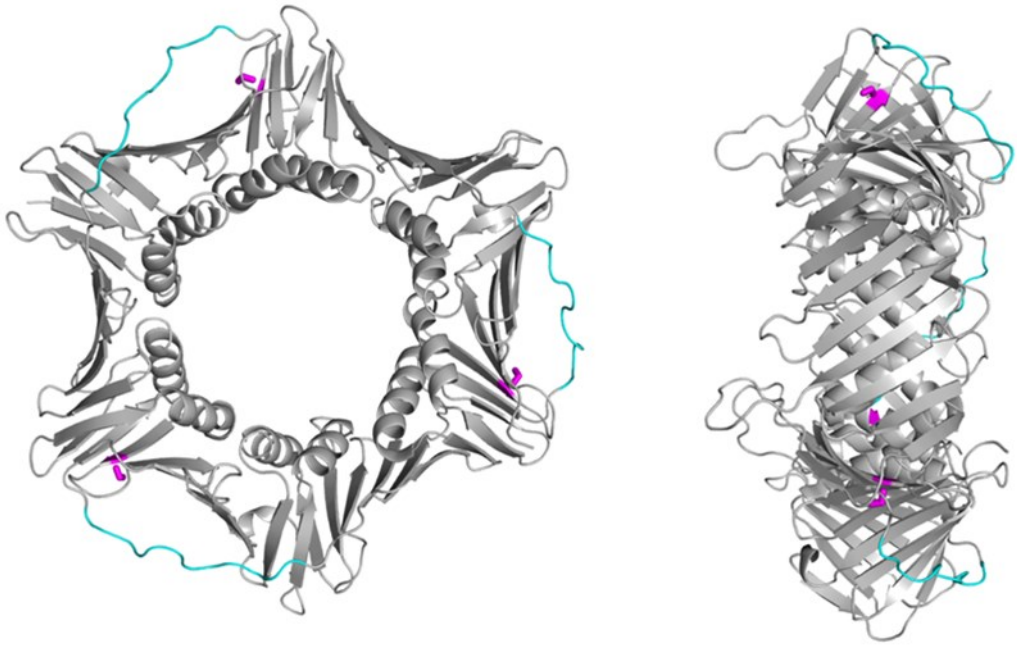


Figure 19: The Ser228 residue of PCNA lies close to the Interdomain connector loop.

The PCNA homotrimer from the front (left panel) and side (right panel), with Ser228 highlighted in magenta and the interdomain connector loop highlighted in cyan. Generated using POLYVIEW-3D (Porollo and Meller, 2007), based on the 1YM crystal structure, rendered using PyMol.

Activities	Proteins
DNA polymerases	Polδ, Polε, Polη, Polι, Polκ, Polζ, Polλ, Polβ, Rev1
Clamp loader	Rfc1, Rfc3, Rfc4
Flap-endonuclease	FEN-1
DNA ligase	DNA Ligase 1
Topoisomerase	Topo IIα
Replication licensing factor	Cdt1
E3 ubiquitin ligases	Rad18, Rad5
E2 SUMO-conjugating enzyme	Ubc9
Helicases, ATPases	Srs2, Rrm3, Mgs1, WRN, RECQ5
Mismatch repair enzymes	Msh3, Msh6, Mlh1, EXO1
Base excision repair enzymes	UNG2, MPG, NTH1, hMYH, APE1, APE2, XRCC1
Nucleotide excision repair enzyme	XPG
Poly (ADP-ribose) polymerase	PARP-1
Histone chaperone	CAF-1
Chromatin remodeling factor	WSTF
Histone acetyltransferase	p300
Histone deacetyltransferase	HDAC1
DNA methyltransferase	DNMT1
Sister-chromatid cohesion factors	Eco1, Chl1, Ctf18
Protein kinases	CDK2, EGF Receptor
Cell-cycle regulators	p21, p57, Cyclin D1
Apoptotic factors	Gadd45, ING1b, p53

Figure 20: PCNA-binding proteins.

PCNA interacts directly with a host of proteins involved in many different cellular processes. Listed are a selection of key PCNA-dependent activities and the corresponding PCNA-interacting proteins. Proteins in orange are known to contain PIP-box sequences, which bind to a groove in PCNA (orange) buried underneath the interdomain connector loop. Taken with permission from (Moldovan et al., 2007)

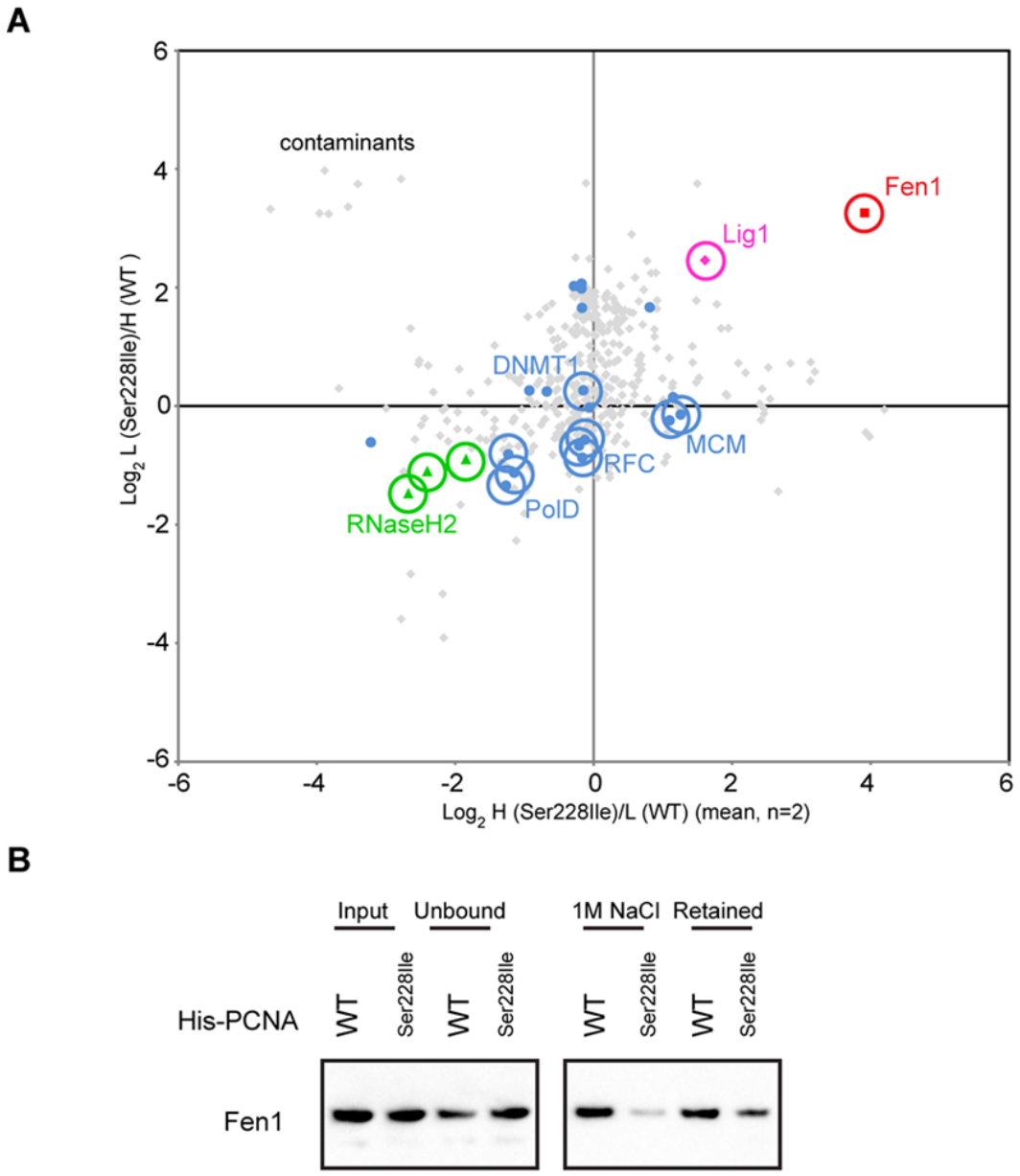


Figure 21: Perturbed PCNA interactions resulting from the p.Ser228Ile mutation.

(A) The PCNA Graphical representation of SILAC-based comparative PCNA-interaction analyses. Lysine [$^{13}C_6/^{15}N_2$] and arginine [$^{13}C_6/^{15}N_4$] labelled (H), or unlabelled (L), cell extracts were purified over either wild type (WT) PCNA or p.Ser228Ile PCNA affinity columns and eluted proteins were analysed by mass spectrometry. Each point represents a protein present in 3/3 experiments. X and Y axis positions are the H:L ratios of each protein in experiments performed in the forward or reverse directions, respectively. Highlighted in blue are known PCNA interacting proteins (Appendix Table). Proteins of particular interest are labelled by name. Data shown in A is courtesy of Dr Helen Chambers. **(B)** A pull down experiment in which recombinant His-PCNA (WT or p.Ser228Ile) was added to HeLa cell extracts and proteins bound to nickel-NTA. The resin was washed with up to 500mM NaCl and then proteins eluted at 1M NaCl or retained on the beads were analysed by western blot for Fen1.

PCNA p.Ser228Ile column, when compared to the WT column, suggesting that the mutant protein bound more tightly to these partner proteins.

To investigate this unexpected finding further, the salt resistance of the PCNA-Fen1 interaction was analysed. HeLa nuclear extract was incubated with His-PCNA WT or p.Ser228Ile) beads, which were then washed with increasing NaCl up to 500mM. The material eluted at 1M NaCl and that which was retained on the beads was subsequently analysed by western blot for Fen1 (Figure 21B). This clearly demonstrates that the interaction between Fen1 and WT PCNA is much more resistant to high salt concentrations that the interaction between Fen1 and PCNA p.Ser228Ile. This leads us to conclude that in the SILAC experiments, Fen1 remained bound to the WT-PCNA column during the elution with 500mM NaCl, resulting in a dramatic reduction of this protein in the WT eluate. This conclusion, that the p.Ser228Ile change impairs the ability of PCNA to bind Fen1, was further validated by subsequent immunoprecipitation and GST pull down experiments (described below).

3.4.3.8 Optimisation of an assay to assess the PCNA – Fen1 interaction

In order to investigate the perturbed interaction between PCNA and Fen1 further, attempts were made to examine the kinetics of this interaction using the phenomenon of surface plasmon resonance (SPR) via the Biacore system. Complete profiles of recognition, binding and dissociation can be generated in real time using this methodology. This allows data such as specificity, affinity, kinetic behaviour and sample concentration to be determined (Biacore, 1999)

Recombinant His-tagged WT PCNA, PCNA p.Ser228Ile and PCNA IDCL* (PCNA interdomain connecting loop mutant, sequence alterations detailed in methods) were coupled to channels 2-4 of a CM5 chip. Recombinant N-terminus His-tagged Fen1 (stock concentration 1.5mg/ml) was diluted 1/100 in HBS and run over the coupled PCNA, the binding in channel 1 was the same as the binding in the other channels. This was likely due to a problem with Fen1 aggregation, given that precipitation of purified Fen1 protein solutions had been observed even at low concentrations, and as such a buffer that does not destabilise Fen1 was sought. The analysis was repeated diluting the Fen1 in PBS, with no improvement in the results. The next experiment used

recombinant Fen1 harbouring a sequence alterations that result in disruption of the Fen1 PIP box, completely abrogating it's interaction with the interdomain connector loop of PCNA (Fen1 PIP*, sequence alterations detailed in methods). With this analysis, no binding of the Fen1 PIP* would be expected in any of the channels, however similar results were seen as with WT Fen1.

The isolelectric point (pI) of Fen1 is unusually high at 8.7, which means that the protein is positively charged at a pH of <8.7. It was therefore highly likely that Fen1 was adhering non-specifically to the highly carboxymethyl dextran matrix on the chip as has been reported for other basic proteins (Myszka, 1999) However a further trial using TBS buffer pH8.5 did not result in any improvement.

The only published SPR analysis of the PCNA – Fen1 interaction was carried out by Gomes and Burgers using recombinant yeast proteins (Gomes and Burgers, 2000). This analysis was carried out using recombinant yeast proteins as opposed to human. Interestingly, the authors had modified the standard SPR buffer by adding Inositol and $MgCl_2$. The Fen1 nuclease requires a divalent metal in its active site, absence of magnesium might destabilise protein folding. Inositol increases the hydrophobicity of the solution, so might stabilise proteins like Fen1 which have solvent exposed hydrophobic patches. Fen-1 was buffer exchanged into Gomes buffer, significant precipitation of Fen1 still occurred. Using the supernatant from this, the analysis was reattempted, however only non-specific binding of Fen1 to the chip was observed.

The other modification made by Gomes and Burgers was to use a Pioneer chip B1 as opposed to the standard CM5 chip. The B1 chip has a lower charge density, and has been shown previously to profoundly reduce non-specific binding. The B1 chip is no longer manufactured, the Biacore representative suggested that the CM4 chip was similar to and had replaced the B1 chip. Using Gomes buffer and the CM4 chip, reduced non-specific binding in channel 1 was observed, but this was still too high to permit any meaningful analysis. Despite additional attempts to optimise conditions including varying the salt concentration up to 300mM NaCl and digestion of any DNA that might be contaminating the Fen1, we were unable to produce conditions suitable for

analysis of the PCNA-Fen1 interaction. At this stage, it was apparent that a result with Fen1 in flow was not going to be obtainable, leading to attempts to couple Fen1 to channels 2-4 on the chip. Fen1 would not dissociate from the chip surface even if it was not covalently coupled. This strongly suggests that the recombinant Fen1 protein was interacting in a non-physiological manner with the chip surface, probably unfolding and permanently precipitating on it, both when used as an anylate or ligand. These results rendered the experiment non-viable suggesting that the recombinant Fen1 being used was not amenable to invitro protein-protein interaction studies.

Gomes and Burgers also used N-terminal tagged Fen1, however there are differences between the aminoacid sequences of yeast and human PCNA and perhaps C-terminal tagging of human Fen1 would produce better results. The N-terminus methionine of human Fen1 is likely to be internal and therefore adding a tag to this end of the protein may well destabilise the tertiary structure by exposing the hydrophobic centre of the protein leading it to unfold.

The interaction of PCNA p.Ser228Ile with Fen1 was next assessed in a more physiological setting by immunoprecipitation from lymphoblastoid cell extracts. WT line L7 and homozygous PCNA p.Ser228Ile line L5 were initially used in attempts to optimise the assay. Whole cell extracts from each cell line were immunoprecipitated under standard conditions (NaCl final concentration 40mM) using anti-PCNA rabbit polyclonal antibody conjugated to Protein G Dynabeads. The resulting immunoprecipitates were separated by SDS-PAGE and probed with anti-PCNA mouse monoclonal antibody (PC10) and anti-Fen1 rabbit polyclonal antibody (ab17993). PCNA was clearly visualised but no Fen1 was co-immunoprecipitated. Given that co-precipitation of Fen1 with PCNA had been unsuccessful, an antibody that would immunoprecipitate Fen1 was sought. Anti-Fen1 mouse monoclonal antibody (4E7), which had previously been shown to immunoprecipitate Fen1 (Guo et al., 2010) was obtained along with and anti-Fen1 monoclonal rabbit antibody (EPR4459(2)), which was raised to a peptide taken from within the general region of amino acids 100-150 of the human Fen1 protein, away from the PCNA binding region (between amino acids 252-371). Attempts made to precipitate Fen1 from the WT line L7 using either anti-Fen1 antibody gave further unsatisfactory results.

The PCNA protein interactions in the SILAC assay had been analysed in the presence of a lower salt concentration of 25mM NaCl. Given this and that the interaction of PCNA with another PIP box containing partner protein, CAF1, had been shown to be very salt sensitive with the optimal NaCl concentration being 25mM (Hoek et al., 2011), whole cell extracts from lines L5 and L7 were buffer exchanged into an interaction buffer at that lower salt concentration, (see materials and methods section 2.7.4 for a detailed description). Anti-Fen1 mouse monoclonal antibody (4E7), and anti-Fen1 monoclonal rabbit antibody (EPR4459(2)) were each conjugated to a 50:50 mix of protein A and G Dynabeads. After separation of immunoprecipitates by SDS-PAGE and probing with anti-Fen1 and anti-PCNA antibodies, Fen1 was seen not to have precipitated with the mouse antibody. However, an interaction between Fen1 and PCNA was clearly demonstrated in the WT line L7 when the rabbit antibody was used to precipitate Fen1, whereas the interaction was barely detectable in cell extract L5 which is homozygous for the Ser228Ile PCNA alteration. An IgG control was run in parallel which revealed that the Fen1 co-precipitates with PCNA and is not just adhering non-specifically to the Dynabeads (Figure 22). Precipitation of PCNA using the PCNA polyclonal antibody was also attempted to see if Fen1 would co-precipitate. Although PCNA was clearly seen to have precipitated using extracts from both WT and homozygous p.Ser228Ile lines, Fen1 did not co-precipitate (Figure 23).

3.4.3.9 The PCNA p.Ser228Ile alteration perturbs the PCNA – Fen1 interaction

Having optimised the conditions for co-immunoprecipitation of PCNA with Fen1, The analysis was repeated including whole lymphoblastoid cell extracts from two WT lymphoblastoid cell lines (L7 and L9), two heterozygous lines (L8 and L10) and two lines homozygous for PCNA p.Ser228Ile (L5 and L6). Fen1 protein was precipitated using the anti-Fen1 rabbit antibody (EPR4459(2)) and protein A/G Dynabeads. Proteins in the extracts and precipitates were analysed by western blot with anti-PCNA (PC10) or anti-Fen1 (4E7) antibodies. While an interaction between Fen1 and PCNA was clearly demonstrated in controls, an intermediate level of interaction was detected in heterozygote cell lines and the interaction was barely detectable in cells homozygous for the PCNA p.Ser228Ile alteration (Figure 24). These data provide unambiguous confirmation that

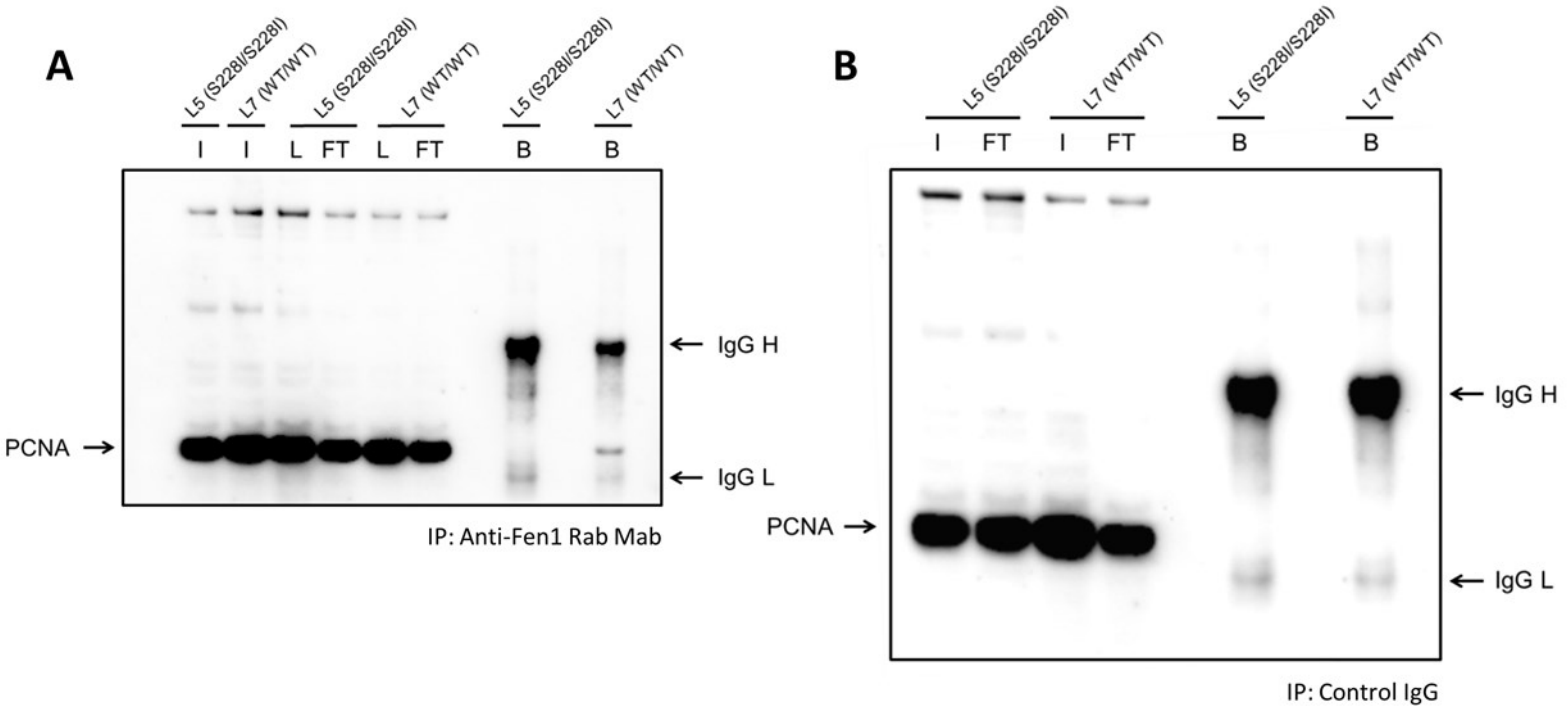


Figure 22: Anti-Fen1rabbit monoclonal antibody (EPR4459(2)) is suitable for use in co-immuoprecipitation analyses of the PCNA-Fen1 interaction.

Anti- Fen1 immunoprecipitates, from extracts made from lymphoblastoid lines derived from affected individuals or family members, were analysed by western blot for the presence of PCNA and Fen1. Endogenous PCNA was co-precipitated with Fen1 in cells from WT individuals, but was almost undetectable in the precipitation from cell extracts derived from affected individuals. No PCNA is seen in the IgG control.. I =input, L=load, FT=flow-through, B=beads.

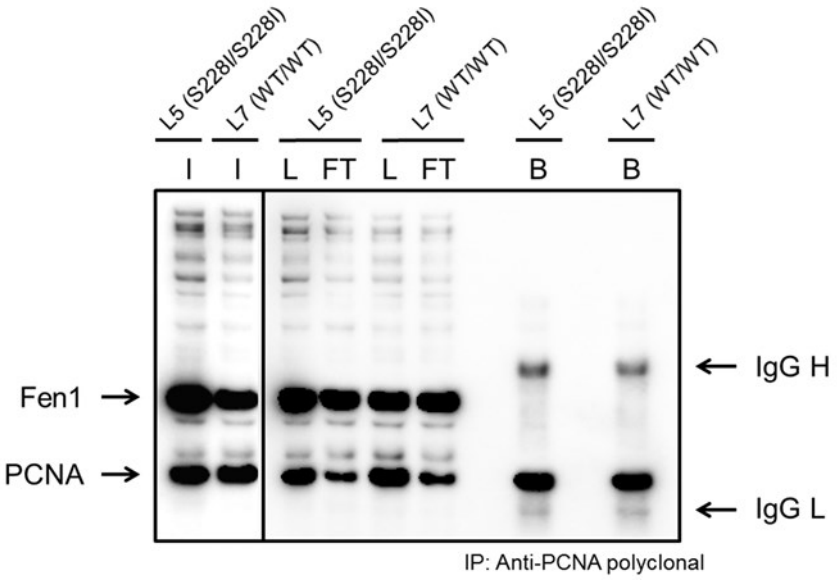


Figure 23: PCNA rabbit polyclonal antibody is not suitable for use in co-immuoprecipitation analyses of the PCNA-Fen1 interaction

Anti-PCNA immunoprecipitates, from extracts made from lymphoblastoid lines derived from affected individuals or WT individuals, were analysed by western blot for the presence of PCNA and Fen1. Endogenous Fen1 was not co-precipitated with PCNA in cells from either affected or WT individuals. I =input, L=load, FT=flow-through, B=beads.

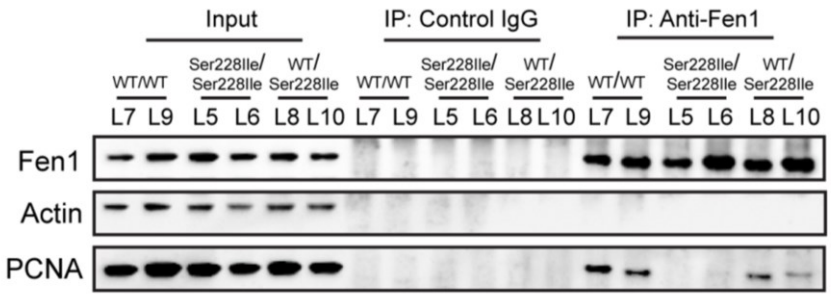


Figure 24:The PCNA-Fen1 interaction is perturbed by the p.Ser228Ile mutation

Anti-Fen1 immunoprecipitates, from extracts made from lymphoblastoid lines derived from affected individuals or family members, were analysed by western blot for the presence of PCNA, Fen1 and actin. PCNA was almost undetectable in the precipitation from cell extracts derived from affected individuals.

interaction of Fen1 with PCNA is dramatically reduced by the p.Ser228Ile mutation and we postulated that interactions with other PCNA partner proteins would be similarly affected, providing a plausible explanation for the reduced nucleotide excision repair activity

3.4.3.10 Optimisation of an assay to assess the PCNA – Lig1 interaction.

We first attempted to assess the PCNA-Lig1 interaction using SPR. 6xHis-tagged human Lig1 protein, Lig1 PIP* (Lig1 mutant that results in disruption of the Lig1 PIP box, completely abrogating it's interaction with the interdomain connector loop of PCNA) and empty vector were expressed in *E.coli.* After transformation, overnight culture and induction of protein expression with Isopropyl β-D-1-thiogalactopyranoside (IPTG), the bacterial cells were harvested and lysates prepared under native conditions before protein purification using Ni-NTA affinity chromatography. Analysis by SDS-PAGE and Coomassie blue staining revealed that the Lig1 and Lig1 PIP* proteins were induced but the His-tag was not binding to the nickel, instead all of the protein was in the flow-through.

PCNA (WT/p.Ser228Ile/IDCL*) were coupled to 3 flow cells of the CM5 chip as described above and given that the protein purification had failed, bacterial supernatant containing unpurified human Lig1 was run in flow. Although we knew that we would not obtain clean curves, we were hopeful that this analysis would provide us with some indication as to the likelihood that interpretable results would be possible if sufficient amounts of pure Lig1 could be produced. The concentration of Lig1 was unknown, thus it was not possible to generate absolute affinity values. However, binding of Lig1 to WT PCNA appeared to always be greater than to PCNA pSer228Ile. As expected given the complex nature of the mixture, the curves were not easy to interpret. However, equilibrium did appear to be reached quickly and dissociation was rapid and complete on the WT PCNA channel. The curves obtained from the PCNA p.Ser228Ile channel were of lower quality, but it was clear that equilibrium was not reached rapidly and dissociation was less rapid (Figure 25).

It was concluded that large quantities of purified Lig1 protein would be required to be able to make an assessment of kinetics of the PCNA-Lig1 interaction. One

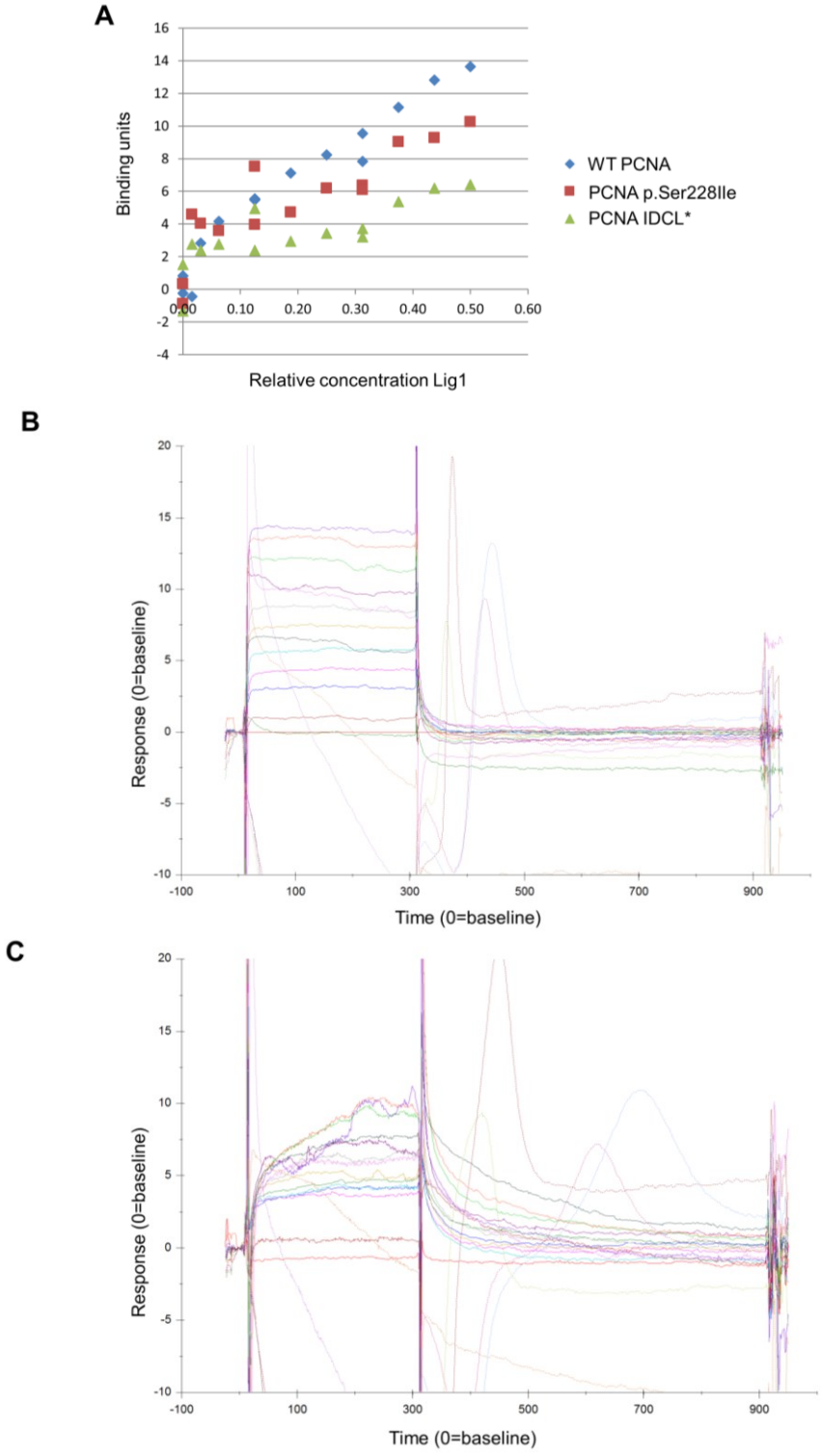


Figure 25: Surface plasmon resonance analysis of the PCNA-Lig1 interaction

Recombinant PCNA WT, p.Ser228Ile and IDCL* were coupled to a biacore chip and unpurified human Lig1 present within bacterial supernatant was run in flow. (**A**) Steady state equilibrium binding curves for Lig1 on the three different PCNA surfaces. More Lig1 binds WT than S228I at equal Lig1 concentrations, suggesting a lower K_d for Lig1 and WT PCNA. Inspection of the sensorgrams for PCNA WT (**B**) and p.Ser228Ile (**C**) surfaces show that for WT PCNA the equilibrium is reached more rapidly and dissociation is also more rapid compared to p.Ser228Ile.

possible reason to explain why we had been unable to obtain purified Lig1 using Ni-NTA affinity chromatography was because the 6xHis-tag on the protein was not available for binding. Hence, if bacterial lysate was prepared and protein purified under denaturing conditions, this should render the 6xHis tag available for binding. An important consideration was that it was not known whether once purified and renatured the Lig1 protein would behave normally. Purified Lig1 was obtained using this method, but the protein concentration was very low (Figure 26). Surface plasmon resonance analysis gave results similar to those obtained using the bacterial lysate, but there was insufficient protein to make any firm conclusions. It might be possible to produce sufficient Lig1 protein by expressing in baculovirus as opposed to *E.coli*, however the facilities were not available to try this method.

In an attempt to study the PCNA-Lig1 interaction by another method, *E.coli* lysates containing recombinant S-tagged PCNA (WT, p.Ser228Ile or IDCL*) were combined with HeLa nuclear extract. S-protein agarose beads were added and proteins were subsequently analysed by western blot with anti-PCNA, anti-Lig1 or anti-actin antibodies. PCNA bound to the beads as expected, but no Lig1 was seen to precipitate even with WT PCNA (Figure 27). Given the earlier success with co-immunoprecipitation studies, immunoprecipitation of Lig1 from L7 (WT) or L5 (Homozygous PCNA p.Ser228Ile) lymphoblastoid extracts was attempted, using either Lig1 mouse monoclonal (10H5), Lig1 mouse monoclonal (10G12) or Lig1 rabbit polyclonal antibody and the same methodology as for the Fen1-PCNA studies described above. Lig1 precipitated with the rabbit polyclonal antibody, however there was no co-immunoprecipitation of PCNA from either the WT or homozygous p.Ser228Ile extract.

3.4.3.11 The PCNA p.Ser228Ile alteration perturbs the PCNA – Lig1 interaction

Given that no antibody against Lig1 that is effective for immunoprecipitation from lymphoblastoid extracts was commercially available, an invitro approach to study the PCNA-Lig1 interaction was used. Gary et al. published an assay utilising lysates from *E.coli* expressing a fusion protein consisting of glutathione

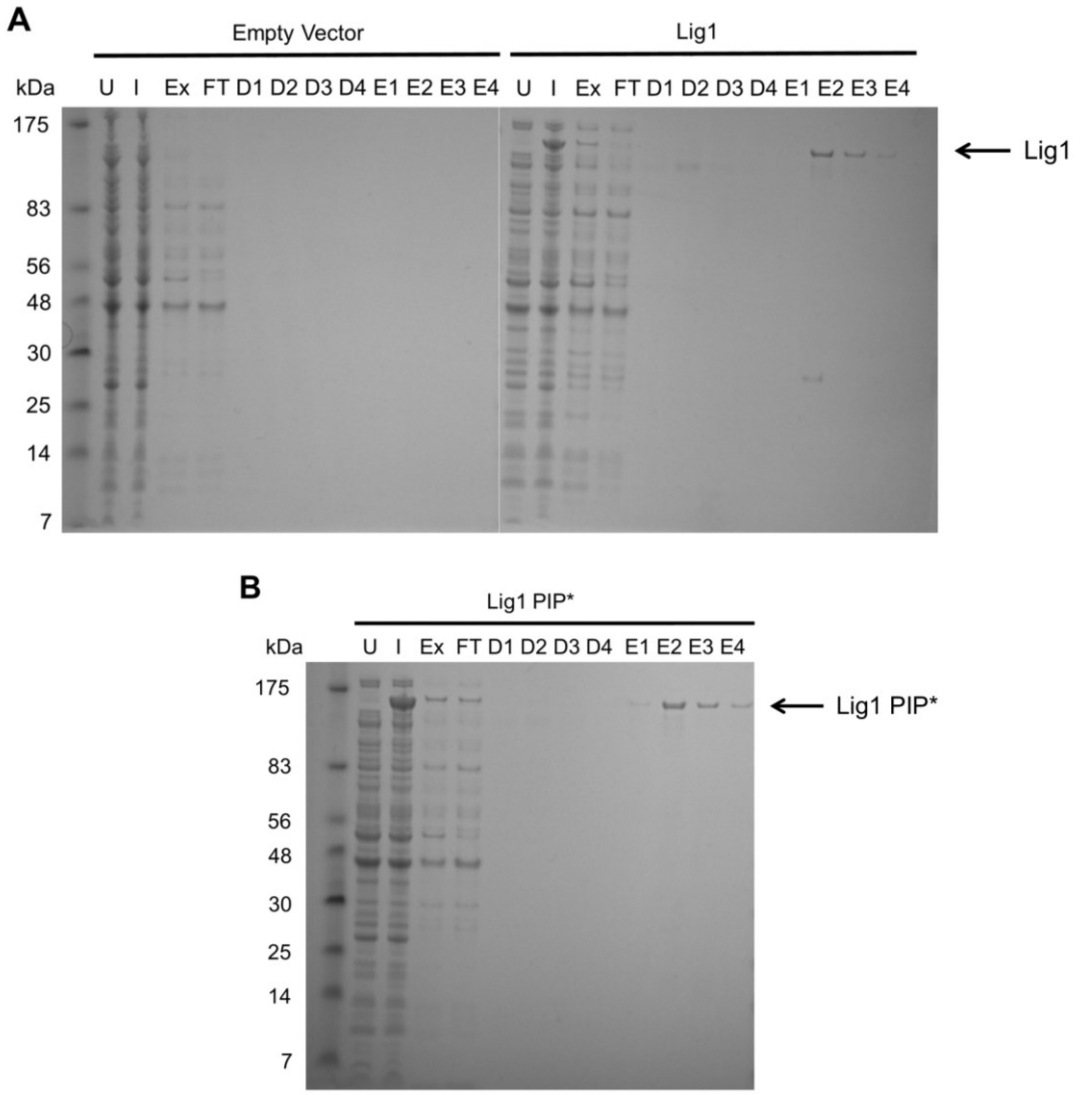


Figure 26: Recombinant human DNA Ligase 1 protein prepared under denaturing conditions

His-tagged human Lig1 protein, Lig1 PIP* and empty vector were expressed in *E.coli* before protein purification using Ni-NTA affinity chromatography. Purified recombinant human Lig1 (A) and Lig1 PIP* (B) can be seen in eluates 2-4. U = total E.coli extract (uninduced), I = total E.coli extract (induced) Ex = Urea extracts (8M) FT = flow-through (unbound from Ni-NTA column in 8M urea) D1-D4 = column washes with D buffer (pH 5.9) E1-E4 column elutions with E buffer (pH 4.5).

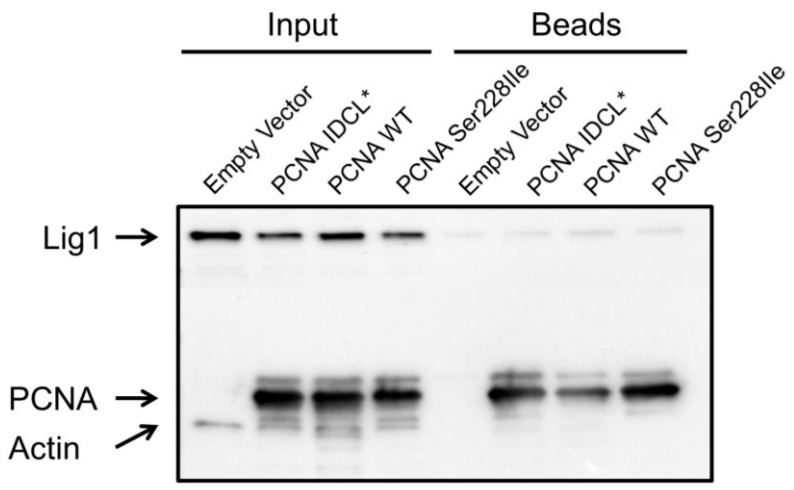


Figure 27: S-tagged PCNA does not bind to endogenous Lig1 at detectable levels

E.coli lysates containing recombinant S-tagged PCNA (WT, p.Ser228Ile or IDCL*) combined with HeLa nuclear extract. Inputs and S-agarose purified proteins were analysed by western blot for the presence of PCNA, Lig1 and actin. No detectable interaction of PCNA with Lig1 is seen under these conditions.

S-transferase (GST) and PIP box motifs of either Fen1 or XPG and then studied the interaction with PCNA (Gary et al., 1997).

Plasmids that express a fusion protein of GST and the PIP boxes of Fen1 (amino acids 328-355) or Lig1 (amino acids 1-21 with an 8 amino acid glycine-serine spacer to maintain an equal distance between the GST and the PIP box) were generated. Lysates from *E.coli* harbouring GST, or the GST-fusion proteins were prepared. The lysates were mixed with lysate containing PCNA (WT, p.Ser228Ile or IDCL*) expressed from pET30a or pET30a as a control, in a buffer containing 100mM KH_2PO_4/K_2HPO_4. Glutathione beads were added and the proteins that associated with the beads were subsequently analysed by SDS-PAGE stained with Coomassie blue. This assay recapitulates the defective interaction between PCNA p.Ser228Ile and Fen1 (Figure 28A) and most significantly shows that the Ser228Ile protein also has a reduced affinity for the PIP box of Lig1 (Figure 28B). In both cases the defective interaction was similar to that seen with PCNA IDCL*, which has been specifically mutated to destroy the interaction of PCNA with partner proteins that bind to its interdomain connector loop.

3.4.3.12 The PCNA p.Ser228Ile alteration perturbs the PCNA–XPG interaction

Given that Fen1 and Lig1 are both implicated in the nucleotide excision pathways (Araujo et al., 2000, Wu et al., 2012), it is plausible that the cellular nucleotide excision repair defects observed in cells from affected individuals result from the compromised interactions of these proteins with p.Ser288Ile PCNA. It also seemed likely that the Ser228Ile alteration would perturb the PCNA interaction with other PIP box containing partners, and that this may contribute to the cellular phenotypes observed. A clear candidate known both to be required for nucleotide excision repair and to bind PCNA through a PIP box interaction is XPG. Peptides derived from XPG were not detected in the SILAC assay, however the PCNA-XPG interaction has previously been shown to require phosphate (Gary et al., 1997), which was not present in that assay.

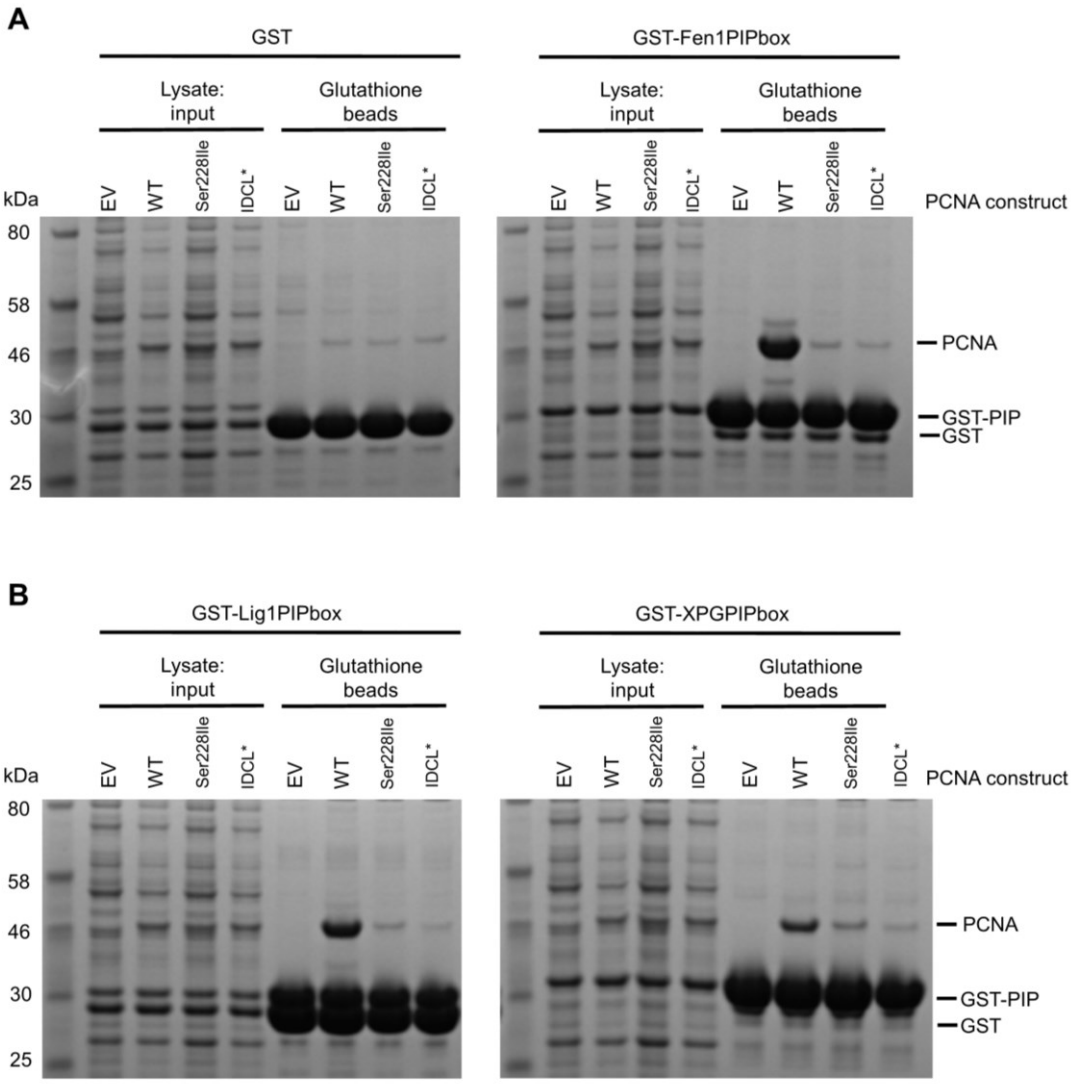


Figure 28: Perturbed PCNA interactions resulting from the p.Ser228Ile mutation

(**A**) *E. coli* lysates expressing empty vector (EV) or PCNA (WT, p.Ser228Ile or IDCL*) were mixed with lysates expressing GST or GST fused to the PIP box of Fen1. Inputs and glutathione-purified proteins were analysed by SDS-PAGE and Coomassie staining. (**B**) Similar to A, except GST fusions comprising the PIP box of Lig1 (left panel) and XPG (right panel) were used. Greatly reduced interaction of PCNA p.Ser228Ile or PCNA PIP* is seen with either the Fen1, Lig1 or XPG PIP boxes. Expression vectors were cloned by Dr Catherine Green.

Immunoprecipitation of XPG from whole cell lymphoblastoid extracts was attempted using either an XPG mouse monoclonal antibody (8H7) or an XPG rabbit polyclonal antibody known to be raised against the N-terminus of XPG away from the PIP box. PCNA did not co-immunoprecipitate in assays utilising either of these antibodies. In the absence of an XPG antibody suitable for immunoprecipitation of endogenous XPG, the GST assay was again used to assess the binding of WT PCNA and p.Ser228Ile to the PIP box of XPG (amino acids 981-1009). Crucially, although less WT PCNA was bound to the XPG PIP box when compared to its binding to the Fen1 PIP box, the p.Ser228Ile PCNA displayed greatly reduced interaction with the XPG PIP box (Figure 28B). Thus, it seems reasonable to suggest that a defective interaction with XPG is also likely to contribute to the defective nucleotide excision repair phenotype seen in cells homozygous for p.Ser228Ile PCNA.

3.5 Discussion

The genetic and cellular data described here implicate PCNA in a new distinct neurodegenerative DNA repair disorder displaying some features in common with other conditions associated with deficient DNA metabolism including Cockayne syndrome, xeroderma pigmentosum and ataxia telangiectasia. The most notable similarities between Cockayne syndrome, the neurological form of xeroderma pigmentosum and the individuals with PCNA mutation described here include growth abnormalities, premature ageing, cognitive decline, photosensitivity and photophobia. The presence of a basal cell carcinoma *in situ* in one affected individual (VI:11) likely indicates a predisposition to sun-induced malignancy in individuals homozygous for PCNA p.Ser228Ile mutation, and further clinical overlap with xeroderma pigmentosum. However the phenotype associated with p.Ser228Ile mutation of PCNA appears in general to be milder than that seen with either typical Cockayne syndrome or xeroderma pigmentosum, most noticeably the rate of premature ageing seems to be slower, with none of the ophthalmic manifestations associated with Cockayne syndrome reported and the effects of UV sensitivity less significant. Interestingly, although hearing loss is found in Cockayne syndrome and some xeroderma pigmentosum patients, the age of onset is typically later (DiGiovanna and Kraemer, 2012, Nance and Berry, 1992). Thus, prelingual sensorineural hearing loss may, in association with the other clinical features described, be a helpful diagnostic feature in patients with this condition. The ocular and cutaneous telangiectasia seen in patients with the PCNA p.Ser228Ile mutation are reminiscent of those seen in ataxia telangiectasia, as is the pattern and progression of neurological abnormalities, although as with Cockayne syndrome and xeroderma pigmentosum, PCNA p.Ser228Ile homozygotes generally appear to be less severely affected than typical ataxia telangiectasia patients. As with Cockayne syndrome and xeroderma pigmentosum, there are also striking differences between the phenotype associated with PCNA p.Ser228Ile and ataxia telangiectasia, including an absence of both the immunodeficiency and oculomotor apraxia commonly seen in ataxia telangiectasia affected individuals.

Complete disruption of PCNA in flies or yeast is lethal (Henderson et al., 1994, Waseem et al., 1992) an expected finding given its fundamental involvement in DNA replication. Until now, no pathological mutations in the PCNA gene have been reported in human populations. In other species however, some mutations of PCNA are compatible with life. For example, point mutations have been produced in the *S. cerevisiae* PCNA gene (*POL30*) that result in a wide variety of phenotypes, including DNA damage sensitivity, alterations in mutation rate and epigenetic silencing defects (Zhang et al., 2000). Similar phenotypes are seen in the various reported mutations in Drosophila PCNA (*mus209*) (Henderson et al., 1994). The only previously reported pathological mutation of PCNA in mammalian systems is a targeted missense mutation of lysine 164 to arginine ($PCNA^{p.Lys164Arg}$) in mice, generated to assess the role of ubiquitination at this residue in somatic hypermutation in the immune system (Langerak et al., 2007). Ubiquitination of PCNA at lysine 164 is essential for translesion synthesis, a DNA damage tolerance process (Kannouche et al., 2004). Mice homozygous for $PCNA^{p.Lys164Arg}$ are viable, but infertile and have an altered mutation spectrum of hypermutated immunoglobulin genes (Langerak et al., 2009, Langerak et al., 2007). Thus a subtle mutation of PCNA that impinges upon a specific function can lead to characteristic loss of function effects on DNA metabolism.

Consistent with this, our assays suggest that the PCNA p.Ser228Ile mutation described here is hypomorphic in nature. We did not detect any gross abnormalities of DNA replication associated with this mutation. Instead, we showed that the mutation has distinctive consequences for the cellular responses to UV, affecting both global genome and transcription coupled-nucleotide excision repair. The effects on recovery of RNA synthesis seen in Cockayne syndrome cell lines, and on unscheduled DNA synthesis in xeroderma pigmentosum cell lines, are typically more dramatic than those observed in homozygous PCNA p.Ser228Ile cells. This may in part explain why the phenotypic manifestations associated with these cellular abnormalities are not as profound in the individuals described here when compared to those seen in Cockayne syndrome and xeroderma pigmentosum.

Our studies define particular defective interactions of p.Ser228Ile PCNA which, together with others not yet recognised, are likely to be responsible for the specific alterations in cellular activities observed. The altered protein interaction capability of PCNA p.Ser228Ile is also likely to account for the clinical manifestations of this syndrome. Complete abrogation of the PCNA-Fen1 interaction, mediated by a homozygous mutation of the Fen1 PIP box, is lethal in mice (Zheng et al., 2007). Hence, it seems likely that while the p.Ser228Ile alteration dramatically reduces the association between PCNA and Fen1 in cell extracts, it is unlikely to completely abrogate their functional partnership. Consistent with this, both our SILAC and interaction assays using recombinant proteins show that the p.Ser228Ile protein remains capable of interacting with Fen1, although with significantly altered affinity. It is not possible to conclude from our assays whether the defective interactions with XPG, Fen1 and Lig1 are solely responsible for the observed effects on nucleotide excision repair, although it is important to note that all these proteins are reported to be involved in this repair pathway (Araujo et al., 2000, Cleaver et al., 2009, Wu et al., 2012). Thus, the perturbation of their interactions with PCNA could well be a major contributor to this abnormal cellular phenotype.

Mutations in *XPG* are rare, and can be divided into two categories: missense mutations in the nuclease domains abrogate nucleotide excision repair but result in relatively mild xeroderma pigmentosum phenotypes, while truncating mutations cause a very severe xeroderma pigmentosum/Cockayne syndrome complex disorder (Clarkson, 2003, Scharer, 2008). The *xpg*$^{-/-}$ mouse is barely viable (Harada et al., 1999). These varying phenotypes are thought to result because, in addition to its function in nucleotide excision repair, XPG is involved in other processes (Scharer, 2008). In particular, it associates with the transcription factor TFIIH, and RNA polymerase II, and has a role in transcription (Clarkson, 2003, Sarker et al., 2005, Ito et al., 2007). It is the disruption of these other functions of XPG that is thought to be responsible for the more severe features of truncation mutations. As an example, patient XPCS1RO, (94RD27), had severe early-onset Cockayne syndrome and died aged 7 months. He was homozygous for a frame-shift mutation at codon 926 of *XPG*. This mutation leaves the nuclease domains intact, but truncates the protein before the PIP-box and nuclear localisation signal (Nouspikel et al.,

1997). Thus, the C-terminal region of XPG, including the motif responsible for interacting with PCNA, is very important for XPG function. These observations lead us to suggest that perturbation of PCNA's interaction with XPG may contribute to the neurological features associated with PCNA p.Ser228Ile.

In contrast to the homozygous state, mice heterozygous for the Fen1 PIP box mutant are viable, though they show predisposition to malignancy (Zheng et al., 2011). Although specific PIP box mutations of Lig1 have not been generated in mammals, loss of function mutations in *LIG1* have been reported in a single individual and result in a complex human phenotype comprising growth restriction, immunodeficiency and photosensitivity (Barnes et al., 1992), features reminiscent of those seen in p.Ser228Ile homozygotes. It is possible that Lig1 deficiency and PCNA p.Ser228Ile homozygosity together represent a group of conditions that are the result of disruption to multiple different DNA repair pathways. In addition, while the observed cellular defects correlate with the Cockayne syndrome and xeroderma pigmentosum like features of the syndrome, the clinical overlap with ataxia telangiectasia will require further investigation.

As well as being crucial to the maintenance of genomic integrity and prevention of neoplastic changes, DNA repair is known to be fundamentally important both during the rapid proliferative phase characteristic of early neurogenesis, and in the prevention of early cell death. The affected individuals with PCNA mutation described in this study display signs of neurodegeneration, a recognised feature of Cockayne syndrome, ataxia telangiectasia as well as neurological forms of xeroderma pigmentosum, although the pathological patterns seen in each are distinct. More detailed investigation of the altered biological processes resulting from the p.Ser228Ile PCNA alteration will thus provide invaluable insight into the biological basis of this novel human disorder, as well as the neurodegenerative disease mechanism(s) involved in DNA damage tolerance and repair disorders. Further, while mutations resulting in complete loss of function of PCNA in humans may be incompatible with life, it remains an intriguing possibility that additional sequence variants in this gene, affecting other distinct aspects of PCNA function, might be viable and result in a phenotype more or indeed less severe than that associated with the p.Ser228Ile alteration described here.

4

CHAPTER FOUR

MUTATIONS IN KPTN CAUSE A NEURODEVELOPMENTAL DISORDER ASSOCIATED WITH MACROCEPHALY AND SEIZURES

4 MUTATIONS IN KPTN CAUSE A NEURODEVELOPMENTAL DISORDER ASSOCIATED WITH MACROCEPHALY AND SEIZURES

4.1 Summary

The proper development of neuronal circuits during neuromorphogenesis and neuronal-network formation is critically dependent on a coordinated and intricate series of molecular and cellular cues and responses. Although the cortical actin cytoskeleton is known to play a key role in neuromorphogenesis, relatively little is known about the specific molecules important for this process. We identified nine affected individuals from four Ohio Amish families affected by an inherited variable neurodevelopmental disorder. The most consistent features were global developmental delay/intellectual disability, macrocephaly, anxiety, and some features suggestive of a pervasive developmental disorder. A primary seizure disorder was described in three cases. Using a combination of autozygosity mapping, linkage analysis and whole-exome sequencing, we demonstrated that two founder mutations in KPTN, encoding kaptin, can result in this phenotype. Our immunofluorescence analyses in primary neuronal cell cultures showed that endogenous and GFP-tagged kaptin associates with dynamic actin cytoskeletal structures and that this association is lost upon introduction of the identified mutations. Histopathological assessment of kaptin deficient mice revealed megalencephaly. Additionally mutant mice displayed cognitive deficits mirroring the human phenotype. Taken together, our studies have identified kaptin alterations responsible for both the human and mouse disorder and define kaptin as a molecule crucial for normal human and mouse neuromorphogenesis.

4.2 Introduction

Extremes of brain growth have frequently been associated with impaired neurodevelopment and cognition. Occipitofrontal circumference (OFC) is an indirect measure of brain growth and the one most widely accepted in clinical practice with microcephaly (greater than or equal to 3 standard deviations/SDS below the mean) signifying significant brain undergrowth, and macrocephaly (greater than or equal to 3 SDS above the mean) indicative of brain overgrowth/megalencephaly in the absence of hydrocephalus and cranial thickening (Stevenson et al., 1997). The differential diagnosis relates to the underlying presence or absence of structural brain anomalies. Known pathologies associated with megalencephaly include generalised overgrowth syndromes, metabolic syndromes including storage disorders and Fragile X syndrome. The strong association between macrocephaly with neurodevelopmental disability, autism and other pervasive developmental disorders is well recognised (Kanner, 1968, Klein et al., 2013, Nevo et al., 2002). Where macrocephaly is associated with developmental disability, there is a significantly increased risk of seizures (Nevo et al., 2002).

Over recent years family studies have begun to identify genes which may be mutated to cause inherited forms of developmental disability. These studies have shed important new light into the molecular and cellular processes which orchestrate the human neuronal circuitry, and which may malfunction in neurological disorders. The establishment of the incredibly intricate human neural circuitry is critically dependent upon a complex and tightly regulated myriad of cellular processes and migrational cues. The actin cytoskeleton is known to play an important role in the formation, propagation and steering of cell motility and migration during brain development. This in turn leads to the astonishing morphological intricacy that neurons acquire during neuronal differentiation that is required for the formation of the complex functional neuronal networks underlying human higher brain functions. Mutations in genes encoding molecules important for normal function of the actin cytoskeleton have previously been implicated in inherited forms of human developmental disability and brain development, (Drevillon et al., 2013, Riviere et al., 2012, Ropers et

al., 2011), highlighting the important role of the actin cytoskeleton in neuromorphogenesis.

4.3 Methods

4.3.1 Exome sequencing

Coding regions were captured using SureSelect Target Enrichment System (50Mb) with sequencing on a HiSeq (Illumina) with 76bp paired end reads. Duplicates were marked by Picard (v1.46) and were excluded before mapping to the human genome reference sequence (GRCh37). GATK (v1.0.5777) was used to realign reads near potential insertion-deletion (indel) sites and to recalibrate base qualities, and single nucleotide variants were called using GATK and SamTools (v0.1.16) whilst indels were called using GATK and Dindel (v1.01). All variants were annotated using dbSNP (134) and the 1000 genomes pilot study (May 2011) for minor allele frequency. The variant consequences on protein structure were predicted by SNP Effect Predictor (VEP v2.1) based on Ensembl (version 63). Variants were filtered out if the read depth <4x or >1200x, if the consensus quality <20 or if the SNP quality <25 (this work was undertaken by Dr Saeed Al-Turki).

The VCF file obtained was further custom filtered using EVA (exome variant analysis suite now know as FEVA, family based exome variant analysis suite), a desktop application from the Wellcome Trust Sanger Institute which can handle VCF files and permits the application of customised filters to select a subset of variants. *Insilico* analysis of variants was performed using PROVEAN, Sift, Mutation Taster and VEP (variant effect predictor) online software

4.3.2 Functional studies

Functional studies were undertaken to understand the impact of the two variants identified in this study on kaptin function in collaboration with Dr Michael Kessels, Professor Britta Quamann and Ms Maryam Izadi based at the Research Institute of the FSU Jena, Germany. Methodology for these studies is included in the appendix.

Sequence analyses of kaptin was performed using online protein domain retrieval tools BLAST and SMART.

4.3.3 Mouse studies

The construct used to generate the *Kptn* $^{-/-}$ (*Kptn*$^{Tm1a/Tm1a}$) mice is shown in the appendix. The main insertion cassette is placed between exons 4 and 5, with a further LoxP between exons 8 and 9. This allele is known as the Tm1a allele, given that none of the endogenous allele is actually removed, only disrupted as the insertion generates a frame shift followed by a premature stop codon. Any shortened mutant mRNAs transcribed from the Tm1a allele should be targeted for destruction. Wellcome Trust Sanger Institute in house data suggests that gene expression is completed ablated by the insertion, at least in adipose and liver tissues.

Mouse studies were designed by Dr Darren Logan and Dr Chris Lelliott based at the Wellcome Trust Sanger Institute after discussion of the human phenotype with myself and Professor Crosby .The analysis was performed by Ms Maria Levitin, I observed this work.

4.4 Results

4.4.1 Identification of a potentially pathogenic mutation in *KPTN* identified by autozygosity mapping and whole exome sequence analysis

Autozygosity mapping studies were initially undertaken in a single nuclear Mennonite family. Three boys had a history of early onset of seizures refractory to treatment, global developmental delay, significant intellectual impairment and macrocephaly (Figure 29A, family 1). Only one of the brothers was alive at the time of recruitment to the study. A detailed history was taken for each of the affected boys from their mother. Individual IX:3 was examined. We obtained consent to review the existing medical records for all three affected individuals.

Unfortunately although neuroimaging had been undertaken, the films had been destroyed. Clinicians who had reviewed the siblings previously, had suspected Fragile X syndrome, but this had subsequently been excluded.

Assuming that this was an autosomal recessive condition and that a founder mutation was responsible, a genome-wide screen was undertaken in order to map the chromosomal location of the pathogenic variant. Samples from individuals IX:3 and IX:4 were genotyped using Illumina Human Cyto12 Beadchip arrays which revealed a large 8.74Mb homozygous region on chromosome 19q13.13-19q13.33 containing 394 genes. Comparison of the haplotype across that region with that of another individual with notable phenotypic similarities (Figure 29A, family 2, individual IX:6) revealed an overlapping 2.59Mb region on 19q13.33 and shared heterozygous haplotype extending beyond the homozygous segment. The homozygous region identified in families 1 and 2, considered likely to correspond to the disease locus in these families, is delimited by recombinant SNP markers rs2253022 and rs7246244 and contains 149 genes (Figure 29B). Autozygosity across this interval was corroborated by microsatellite marker analysis in all family members.

In order to identify the causative mutation, whole exome sequence analysis was undertaken. A single affected individual (Figure 29A, IX:9,) was sequenced to generate a novel variant profile. We obtained 7.6G bases of which 93% have a quality >= Q30. Approximately 8% of the reads were marked as duplicates by

Picard (v1.46) and were excluded before mapping to the human genome reference sequence (GRCh37). The VCF file obtained was then further filtered using EVA to select all the novel homozygous variants within the critical region on chromosome 19. After filtering, we were left with only a single likely deleterious variant, in exon 8 of the *KPTN* gene encoding the 436 amino acid protein kaptin (chromosomal variant: g.47983131G>T[NC_000019.9]; NM_007059.2:c.776C>A;NP_008990.2: p.S259*). The presence of the variant was confirmed by dideoxy sequencing which also confirmed its cosegregation within families 1 and 2 (Figure 29A). The KPTN c.776C>A nonsense variant destroys an EagI restriction site. Restriction digest was performed and 7 heterozygous carriers were identified in 560 control chromosomes examined, indicating an allele frequency of approximately 0.012 in this community. The variant was also listed in the Exome Variant Server (NHLBI GO Exome Sequencing Project (ESP), Seattle, WA; http://evs.gs.washington.edu/EVS/), with 1 heterozygote being reported in 8285 European American chromosomes.

4.4.2 Identification of a second potentially pathogenic mutation in *KPTN* within the same Anabaptist community

It was interesting that no notable homozygous regions were detected in affected members of families 3 and 4 despite them showing significant phenotypic overlap with the affected individuals from families 1 and 2. We next investigated these and 18 other families with individuals affected by developmental delay and features suggestive of a pervasive developmental disorder, but without macrocephaly. All five affected individuals from families 3 and 4 were found to be heterozygous for *KPTN* c.776C>A. This prompted evaluation of *KPTN* for a second mutation which may be acting in conjunction with the c.776C>A alteration in these families. Subsequent dideoxy sequence analysis of all coding regions and associated splice junctions of *KPTN* (primer sequences Appendix) in these families revealed that all affected children were indeed compound heterozygous for the c.776C>A variant, as well as an in-frame 18bp duplication (NC_000019.9:g.47983179_47983196dup;NM_007059.2:c.711_728dup)

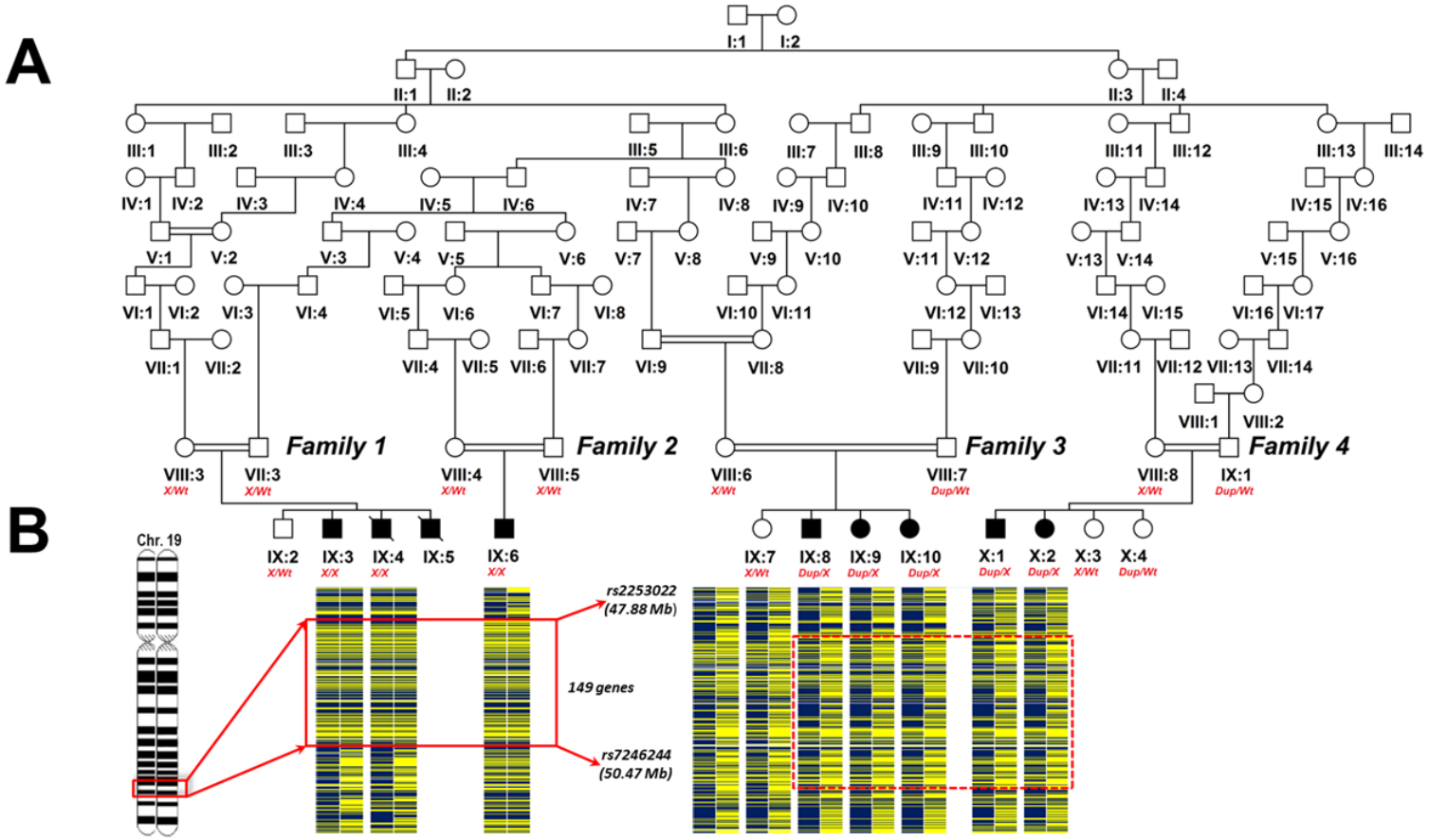


Figure 29: Family Pedigree and Gene Mapping

(**A**). Pedigree diagram showing all 4 (families 1-4) nuclear families investigated, which interlink into a single extended family. Segregation of the two mutations identified (X denotes p.S259*, and Dup denotes p.M241_Q246dup) is shown (all genotypings were validated by dideoxy sequence analysis). (**B**). Pictorial representation of the SNP genotype data encompassing the chromosome 19 homozygous (solid box) and compound heterozygous (dashed box) regions in affected individuals, with the disease locus being demarcated by SNPs rs2253022 and rs7246244 (2.59Mb; families 3 and 4). Support with SNP mapping data analysis was provided by Mr Reza Maroofian, University of Exeter.

in exon 8, predicted to result in the duplication of amino acids 241-246 (NP_008990.2:p.M241_Q246dup), the presence of this variant was confirmed in the transcript. Both the c.776C>A and c.711_728dup sequence alterations completely cosegregate with the disease phenotype as would be expected of causative compound heterozygous mutations (Figure 29A, Figure 30), a LODMAX=10.33 was calculated using Simwalk2 (Sobel and Lange, 1996). The c.711_728dup sequence duplication is not listed in genomic sequence databases and one heterozygote was detected in 560 Amish control chromosomes analysed by PAGE.

4.4.3 Phenotype of individuals homozygous or compound heterozygous for mutations in *KPTN*

In total 9 family members present in 4 nuclear families were included in the original study. All are affected by an inherited variable form of neurodevelopmental disorder. The most consistent features were global developmental delay, macrocephaly with frontal bossing, high levels of anxiety, a characteristic pattern of speech with little alteration in tone and some features suggestive of a pervasive developmental disorder. Neuroimaging was performed in 4 cases with no significant intracranial abnormalities reported. A primary seizure disorder was described in 3/9 cases. The clinical features are summarised in Table 13. Dysmorphic features were subtle and appeared to evolve with age. The facial phenotype included frontal bossing, a broad nasal tip, scaphocephaly, hooded eyelids with small downslanting palpebral fissures and a prominent chin (Figure 31).

4.4.4 KPTN is a stringently conserved but relatively uncharacterised protein

The *KPTN* gene encodes a largely uncharacterised protein kaptin. Sequence and predicted structural conservation of both the N-terminal and C-terminal half of kaptin are very high even between evolutionarily-distant mammalian species (Figure 32). Our sequence analyses of kaptin using online protein domain retrieval tools identified no protein domains, or homologous human proteins, which might provide clues to the functional basis of the neurological deficits associated with its mutation (Figure 30).

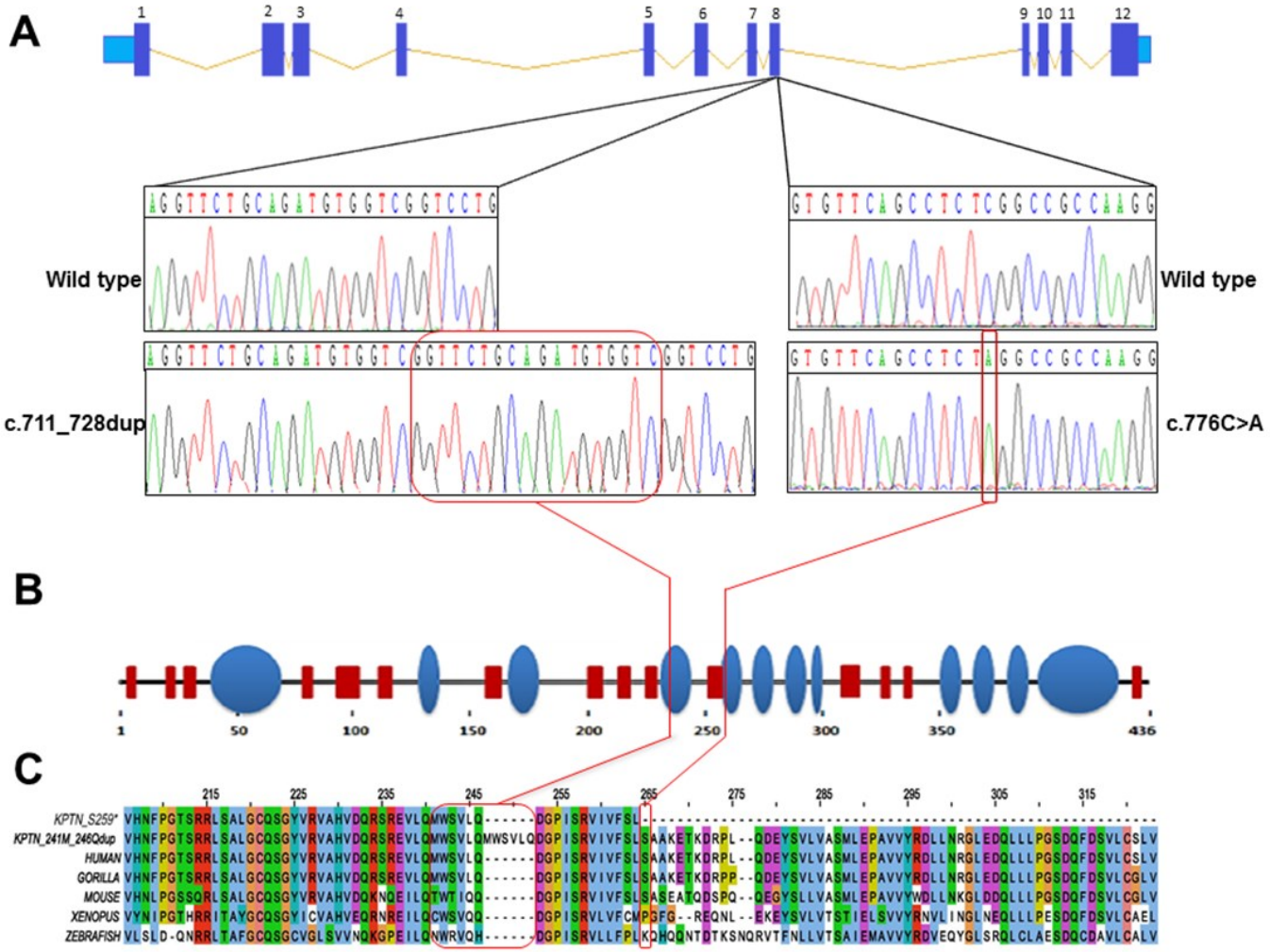


Figure 30: *In silico* analysis of the predicted effects of the *KPTN* sequence variants on the KPTN protein

(**A**) Genomic organisation of the *KPTN* and position of the two exon 8 *KPTN* mutations (c.776C>A; c.714-731dup) identified and translated outcomes (p.Ser259*; p.Met241_Gln246dup) in the corresponding polypeptide products (**B**, **C**). Red boxes denote beta-sheets and blue ellipsoids denote alpha-helices.

Table 13: A comparison of the clinical findings of individuals homozygous or compound heterozygous for KPTN mutations

	IX:3	IX:4	IX:5	IX:6	IX:8	IX:9	IX:10	X:1	X:2
Genotype	p.Ser259*/p.Ser259*	p.Ser259*/p.Ser259*	p.Ser259*/ p.Ser259*	p.Ser259*/ p.Ser259*	p.Ser259*/ p.Met241_Gln246dup	p.Ser259*/ p.Met241_Gln246dup	p.Ser259*/ p.Met241_Gln246dup	p.Ser259*/ p.Met241_Gln246dup	p.Ser259*/ p.Met241_Gln246dup
Gender	M	M	M	M	M	F	F	M	F
Age at evaluation (years)	28.2	Deceased age 29 yrs (head injury)	Deceased age 30 yrs(pneumonia	16.5	13.2	22.7	24.9	1.4	7.8
Growth parameters									
Birth weight kg (SDS)	2.95(0.7)	2.92 (0)	3.46 (1.3)	1.59 (-0.2)	3.35(-0.4)	2.89 (-1.2)	3.16 (-0.5)	3.2 (2.0)	2.75 (1.1)
Birth OFC cm (SDS)		40.6 cm at 6 wks (2.9)	35.6 cm (1.75)		51cm at 10mo (3.9)			"macrocephaly"	"macrocephaly"
Height cm (SDS)	166.7cm (-1.6)	N/K	165.1 (-1.8) at 19.1yrs	169 (-0.7)	161.6 (0.6)	156.2 (-1.3)	160 (-0.6)	NK	123.5 (-0.4)
Weight kg (SDS)	121.2 (3.4)	N/K	66.5 (-0.2) at 19.1yrs	63.1 (0.1)	51.5 (0.8)	107.9 (3.6)	82 (2.2)	NK	23.1 (-0.5)
OFC cm (SDS)	62 (3.0)		63.5(3.6) at 19.1yrs	62.5 (3.4)	61 (3.3)	63 (5.4)	60 (3.2)	52.5 (3.0)	55.4 (2.1)
Parental OFC cm (SDS)		Mother 55.5 (0.1) Father 60 (1.7)		Mother 58 (1.9) Father 59 (1.2)		Mother 58.5 (2.3)		Mother 57 (1.2) Father 59.5 (1.5)	
Development									
Walked (years)	1	1	1.33	1.9	4	3.8	2.4	>2.2	2.2
Expressive & receptive language deficit	✓	✓	✓	✓	✓	✓	✓	✓	✓
Intellectual disability	Moderate	Mild/Moderate	Severe	Moderate	Moderate	Moderate IQ 45	Moderate/Severe IQ 42	Mild	Mild
Neurology									
Childhood hypotonia	-	-	-	✓	✓	✓	✓	✓	✓
Seizures	onset age 3 mo AS and GTCS	onset age 7 yrs AS and GTCS	onset age 7 yrs GTCS	-	-	-	-	-	-
Neuroimaging	MRI - normal[a]	N/A	CT mild ventriculomegaly[a]	N/P	N/A	N/P	N/P	CT– widening of the metopic suture	CT normal
Behavioural characteristics	Repetitive speech Anxiety		Stereotypies	Stereotypies Hyperactivity Anxiety	Stereotypies Repetitive speech Anxiety	Stereotypies Repetitive speech Anxiety	Stereotypies Repetitive speech Anxiety	Phobias Anxiety	Anxiety
Physical anomalies									
Head shape	Frontal bossing Prominent chin	Frontal bossing Prominent chin	Frontal bossing Prominent chin	Frontal bossing Scaphocephaly Prominent chin Hypertelorism	Scaphocephaly Sagittal synostosis-operated age 10mo Frontal bossing	Frontal bossing	Scaphocephaly Frontal bossing	Plagiocephaly Frontal Bossing	Frontal bossing Prominent chin
Other physical findings	5th finger clinodactyly	5th finger clinodactyly	5th finger clinodactyly				Hepatosplenomegally Liver cirrhosis Recurrent pneumonia in childhood	Splenomegally Anaemia Recurrent pneumonia Fetal finger pads	Fetal finger pads

Abbreviations: F, female; M, male; OFC, occipitofrontal circumference; SDS, standard deviation scores; N/P, not performed; N/A, no longer available; N/K, not known; (□), indicates presence of a feature in an affected subject; IQ, Intelligence quotient (Wechsler Adult Intelligence Scale); AS, absence seizures; GTC, Generalised tonic clonic seizures; CT, computerised tomography; a, original imaging no longer available.

Figure 31: Clinical features of KPTN

Panels (**A-C**) show the facial features of individuals homozygous for the *KPTN* p.Ser259* mutation. Dysmorphic features are subtle and include macrocephaly, scaphocephaly, frontal bossing with a high frontal hairline, long face with a prominent chin, down-slanting and small palpebral fissures, hooded eye lids and broad nasal tip.

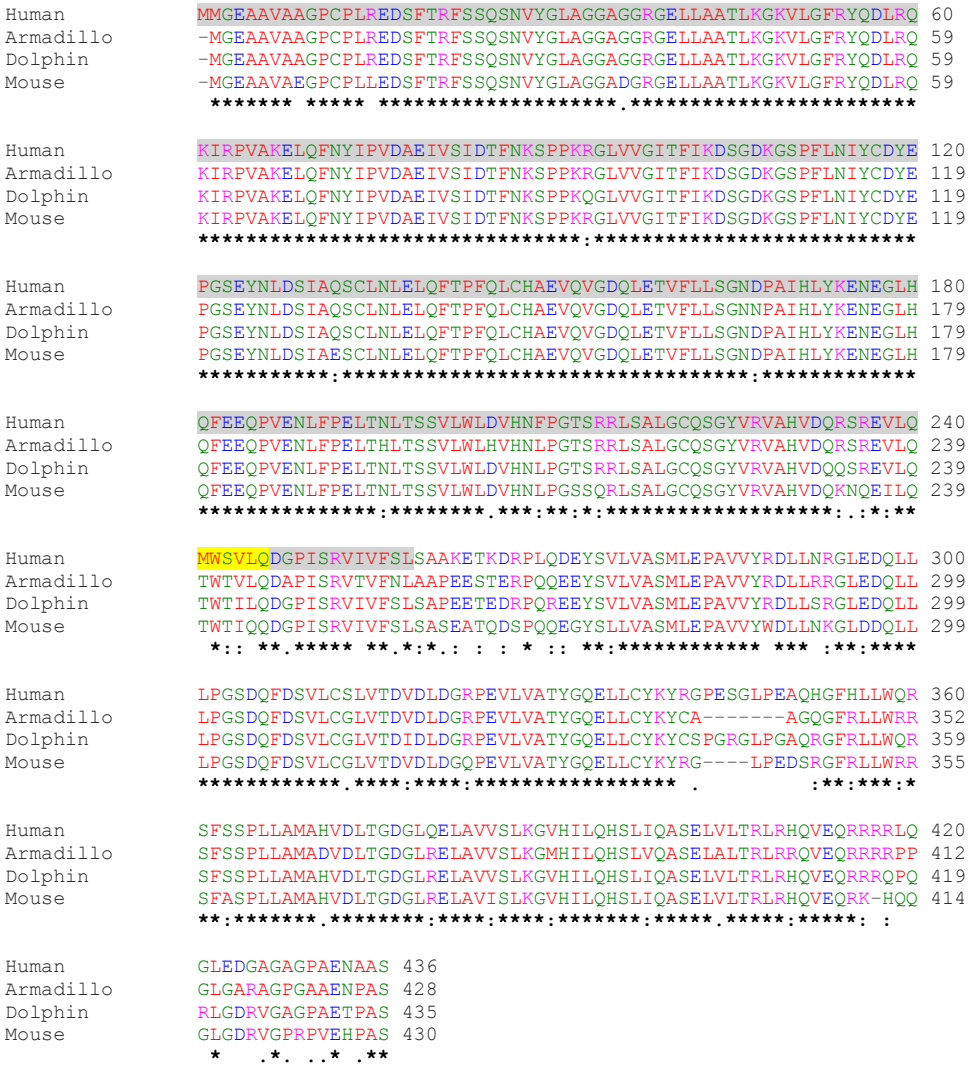


Figure 32: Alignment of kaptin proteins

The KPTN aminoacid sequence from diverse mammalian species including human (gi|108936022|sp|Q9Y664.2), armadillo (gi|488582075|ref|XP_004476450), dolphin (gi|470599504|ref|XP_004311953) and mouse (gi|148710164|gb|EDL42110.1) were aligned, revealing an almost complete sequence conservation of the N-terminal half of kaptin and an also very high conservation of the C-terminal half. Residues corresponding to kaptin amino acids 1-258 are shaded in grey, residues duplicated in the p.Met241_Gln246dup mutant are highlighted in yellow.

Kaptin was originally isolated from human blood platelets but subsequently found to be expressed in fibroblasts as well as intestinal and sensory epithelia (Bearer and Abraham, 1999). A previous study of this molecule suggested a role at stereocilia tips, and so *KPTN* was proposed as a candidate gene for hearing loss (Bearer et al., 2000). However, although no formal hearing assessments have been undertaken the information available for the affected individuals described in this study contains no evidence of hearing deficits. Bearer and Abraham showed that kaptin binds to filamentous F-actin by F-actin affinity chromatography and that it could be eluted from F-actin affinity columns (Bearer and Abraham, 1999). Given this evidence, and the importance of the actin cytoskeleton to neuromorphogenesis, we hypothesised that the likely underlying cause of this neurodevelopmental disorder was a disruption of the interaction between KPTN and F-actin and a consequential reduction in dendritic arborisation.

4.4.5 Kaptin colocalises with F-actin rich sites in primary neuronal cells

As no cellular material from affected individuals was initially available, the majority of the functional studies were undertaken in a primary rat hippocampal neuronal cell line. Initial investigations involved the expression and localisation of kaptin in neurons. In order to do so, human kaptin was first cloned from HEK293 cell cDNA. Primary rat hippocampal neurons, identifiable by anti-MAP2 immunostaining (Sigma, Abcam), were transfected with wild type Flag-tagged kaptin (Flag-kaptin) at DIV3 and imaged two days later. Wild type Flag-kaptin was observed to be localised at F-actin-rich foci in close proximity to the cell bodies as well as in growth cones (Figure 33A and B, arrow heads). At later stages of development, when neurons established synapses (DIV14), wild type Flag-kaptin again colocalised with F-actin-rich sites. Along dendrites, these sites appeared mainly to represent F-actin-rich postsynapses, as almost all kaptin-enriched puncta were contacted by bassoon (a marker for presynaptic active zones) containing presynaptic structures (Figure 33C). To be able to undertake immunolabeling experiments of endogenous kaptin, polyclonal rabbit anti-kaptin antibodies (Sigma) first were first characterized in COS-7 cells. Cells expressing wild type Flag-kaptin were highlighted by anti-kaptin immunolabelling (Figure 34A). Co-immunostaining of primary rat hippocampal neurons at DIV24 with anti-kaptin and anti-ProSAP1/Shank2 (Neuromab) antibodies demonstrated that

the dendritic accumulations of tagged kaptin at F-actin-rich puncta are indeed of physiological relevance and reflect the localisation of endogenous kaptin at postsynapses. The vast majority of anti-kaptin immunolabeled puncta were not only enriched for F-actin, but additionally were immunopositive for ProSAP1/Shank2, a postsynaptic scaffold protein interacting with F-actin binding proteins (Hering and Sheng, 2001, Qualmann et al., 2004); (Figure 33D). Kaptin thus appears to be associated with dynamic actin cytoskeletal structures of neuronal cells. Consistent with this, wild type Flag-kaptin accumulated at lamellipodia of COS-7 cells (Figure 34B), the subcellular regions of mobile fibroblastic cells marked by dense arrays of dynamic actin filaments. In neurons, the cortical actin cytoskeleton is known to be important for proper neuronal network formation during development. Our observations thus support our hypothesis of a significant role of kaptin in neuromorphogenesis.

4.4.6 *KPTN* c.776C>A and c.714_731dup likely result in kaptin loss of function

The c.776C>A sequence variant is predicted by to introduce a premature stop codon and result in loss of function due to degradation of the mutated transcript by mRNA surveillance cell mechanisms, although due to lack of patient material we have been unable to confirm whether a truncated protein (lacking the C-terminal amino acids 257 to 436) is produced. In contrast, the in-frame c.711_728dup mutation is likely to result in the insertion of 6 amino acids (Met-Trp-Ser-Val-Leu-Gln) into full-length kaptin. In order to investigate the functional outcome of this mutation, an *in silico* analysis of the secondary structural elements of wild type and mutant kaptin, was undertaken. This revealed that the N-terminal half of kaptin is likely to comprise a series of relatively densely organized β-sheets, interspersed by only three α-helices, and becomes more α-helical starting with α–helix 4 (spanning amino acids 234 to 245; Figure 35A). Strikingly, duplication of the 6 amino acids, Met-Trp-Ser-Val-Leu-Gln (241-246) is predicted to disrupt the α-helix 4 and result in its conversion into an extended β-sheet (Figure 35C), and so is likely to have a profound effect on kaptin function.

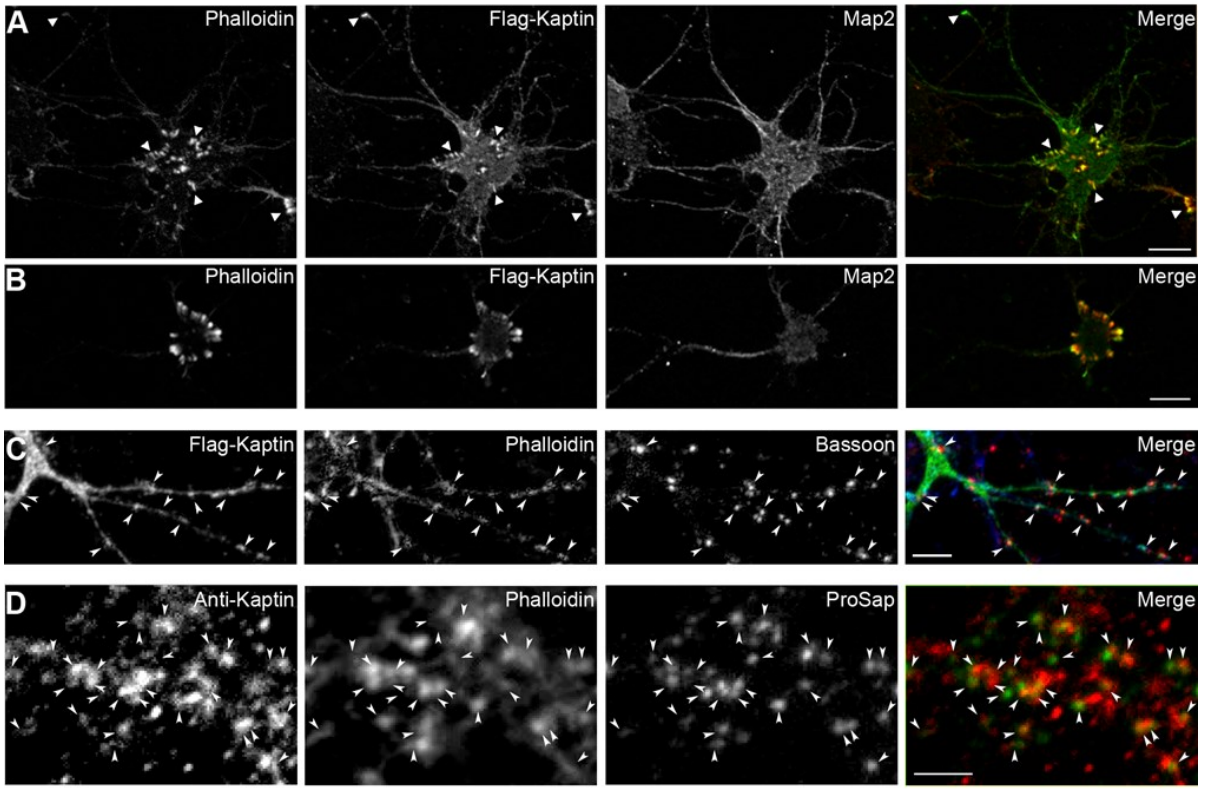


Figure 33: Kaptin immunolocalisation studies in neuronal cell lines

(**A** and **B**) Flag-kaptin colocalised with F-actin-rich foci at the cell body and in growth cones (examples of both are marked by arrow heads in **A**) in DIV5 rat hippocampal neurons transfected at DIV3. For clarity, the anti-MAP2 immunostaining was omitted from the merged images. Scale bars represent 10 mm. (**C** and **D**) Endogenous kaptin immunostained together with the presynaptic marker bassoon at DIV14 (**C**) and with the postsynaptic marker Shank2 at DIV24 (**D**). Puncta enriched with anti-kaptin immunoreactivity (marked by arrow heads) were rich in F-actin, as shown by fluorescently labelled phalloidin. (**C**) Furthermore, they were usually contacted by presynapses. The scale bar represents 5 mm. (**D**) Postsynapses marked by Shank2 were largely positive for both F-actin and anti-kaptin immunolabelling (arrow heads). For clarity, the anti-F-actin staining was omitted from the merged image in (**D**). High-magnification images are shown. The scale bar represents 2.5 mm. Postsynapses marked by ProSAP1/Shank2 to a large extend are positive for both F-actin and anti-kaptin immunolabelling (arrow heads). For clarity, the anti-F-actin staining has been omitted from the merge in D.High magnification images, Bar, 2.5 µm. Data courtesy of Dr Michael Kessels.

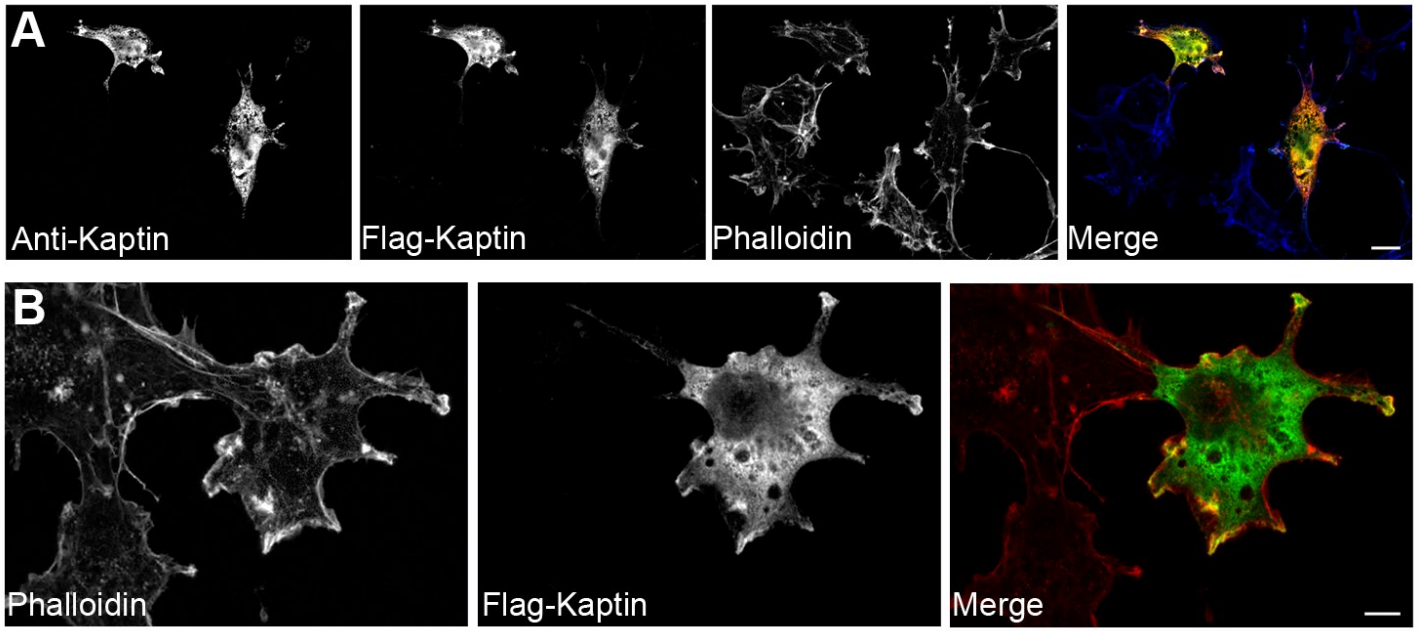


Figure 34: Kaptin immunolocalisation studies in COS-7 cells

(**A**) Flag-kaptin (green in merge) in COS-7 cells is enriched in F-actin-rich lamellipodia (red in merge) and additionally shows a cytoplasmic localisation. (**B**), Anti-kaptin antibody characterisation using Flag-kaptin overexpression in COS-7 cells. Note that transfected cells were recognised by anti-kaptin antibodies, that untransfected cells showed no significant immunolabelling and that there was a good spatial overlap of Flag-kaptin and anti-kaptin detection. Bars, 10 µm. Data courtesy of Dr Michael Kessels.

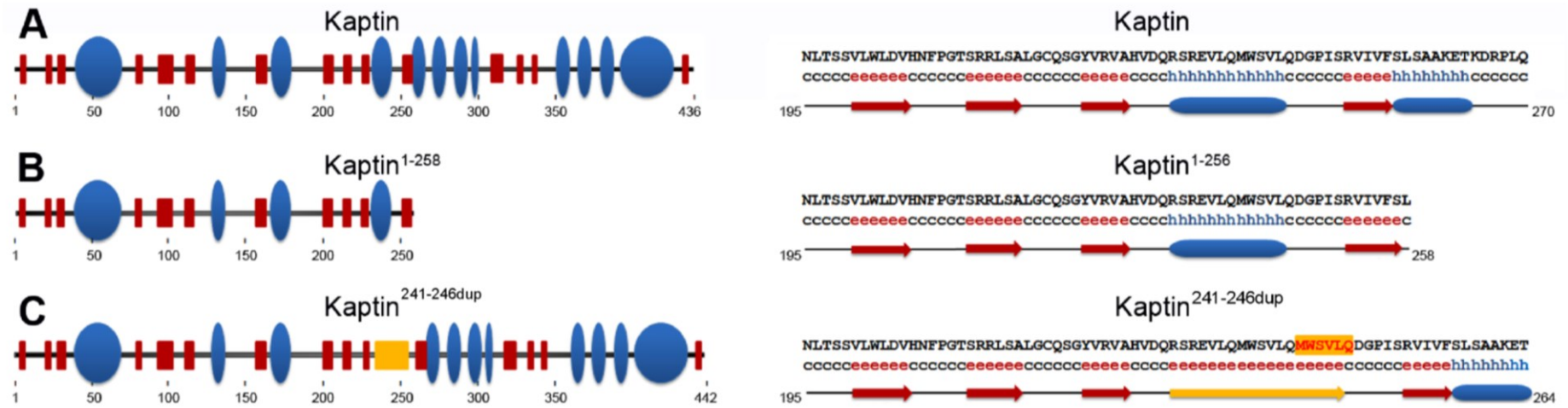


Figure 35: Schematic representation of wild-type human kaptin

(**A**), the truncation mutant S259* (c.1240>T; GFP-kaptin$^{1-258}$) **(B)** and the duplication mutant (GFP-kaptin^{Met241_Gln246dup}). **(C)**. On the left is a graphic overview of secondary structures with beta-sheets as red boxes and alpha-helices as blue ellipsoids. On the right are amino acids and secondary structure elements around amino acid 230. α-helix 4 is predicted to be converted into a β-sheet by the insertion of amino acids 241-246, as shown in yellow.

4.4.7 Mutant KPTN p.Ser259* and p.Met241_Gln246dup do not associate with F-actin

In order to explore experimentally the functionality of any protein arising from translated mutant transcripts, both disease-associated GFP-tagged mutants were generated. Whereas wild type GFP-tagged kaptin (GFP-kaptin) was found to localise at F-actin-rich lamellipodia of COS-7 cells, both kaptin mutants displayed no F-actin association but instead accumulated at irregular perinuclear sites (Figure 36A-C). GFP-Kaptin$^{1-258}$ showed a more pronounced tendency to form such accumulations with almost all cells being marked with little additional cytoplasmic staining outside of these foci, while GFP-kaptin$^{p.Met241_Gln246dup}$ typically displayed fewer and slightly smaller accumulations with slightly higher levels of cytoplasmatic staining outside of these foci (Figure 36A-C). This indicates that both proteins are likely to be non-functional, although it is not possible to exclude the possibility that misfolded mutant protein may accumulate in neurons of affected individuals leading to dominant-negative effects on other neuronal proteins or cell processes.

Finally, as KPTN mutations result primarily in a form of neurodevelopmental disease, we also analysed the behaviour of the kaptin mutants in primary hippocampal neurons during early development. Whereas wild type GFP-kaptin was again found to colocalise with dynamic F-actin in growth cones and foci at the cell body, both GFP-kaptin$^{1-258}$ and GFP-kaptin$^{p.Met241-Gln246dup}$ were found to accumulate in a manner reminiscent of the COS-7 studies in the cell body or at perinuclear sites (Figure 37A-C).

4.4.8 *KPTN*$^{-/-}$ mice are macrocephalic and exhibit impaired cognition

A *Kptn* knockout mouse model (Kptn$^{tm1a(EUCOMM)Wtsi}$) was identified on the Wellcome Trust Sanger Institute mouse resources portal. The preliminary *Kptn*$^{-/-}$ mouse phenotype described was primarily a metabolic phenotype. Female homozygotes had a significantly increased body weight related to an increase in adipose tissue when compared to WT control mice. In addition they exhibited impaired glucose tolerance following a high fat diet. Although the cohort of affected individuals from Ohio was small, no significant differences in weight between the sexes had been observed.

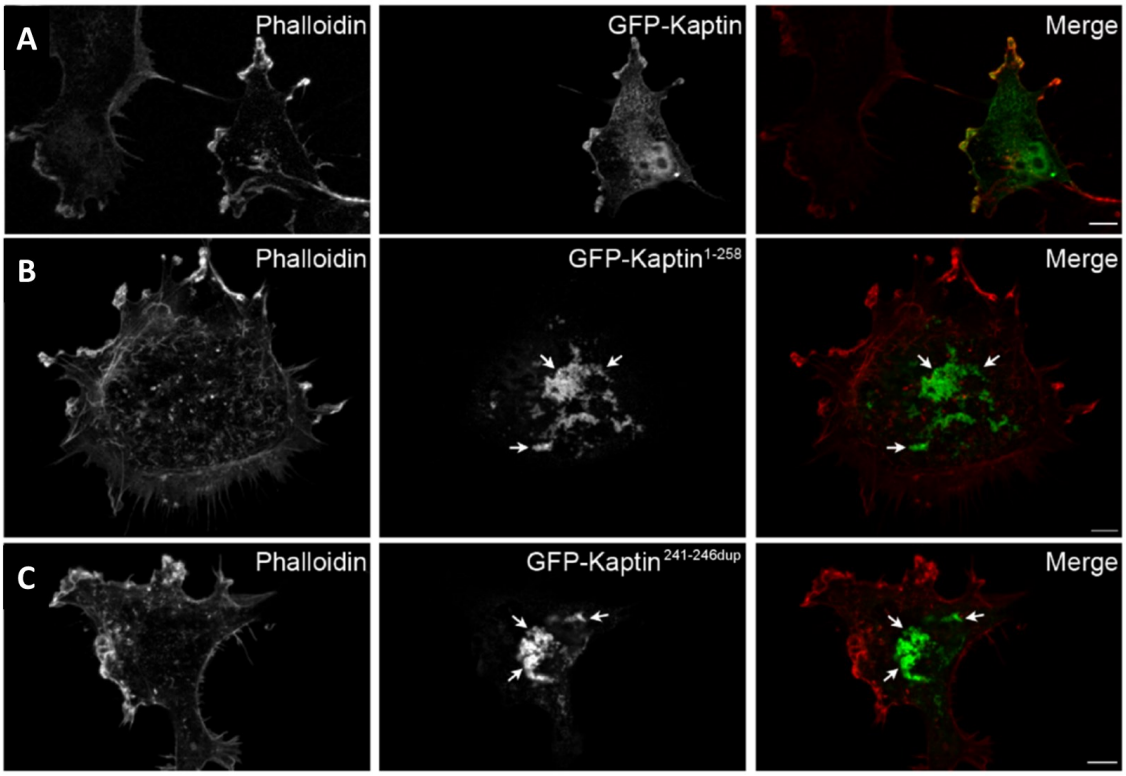


Figure 36:Immunolocalisation studies of altered kaptin in COS-7 cells

Localisations of GFP-kaptin, GFP-kaptin$^{1-258}$ and GFP-kaptin$^{Met241-Gln246dup}$ (green in merges) in COS-7 cells counterstained with phalloidin (red in merged images). Note that whereas wild-type kaptin is distributed in the cytosol and accumulated at F-actin-rich lamellipodia (**A**), both mutants show accumulations at perinuclear regions (**B** and **C**). Data courtesy of Dr Michael Kessels.

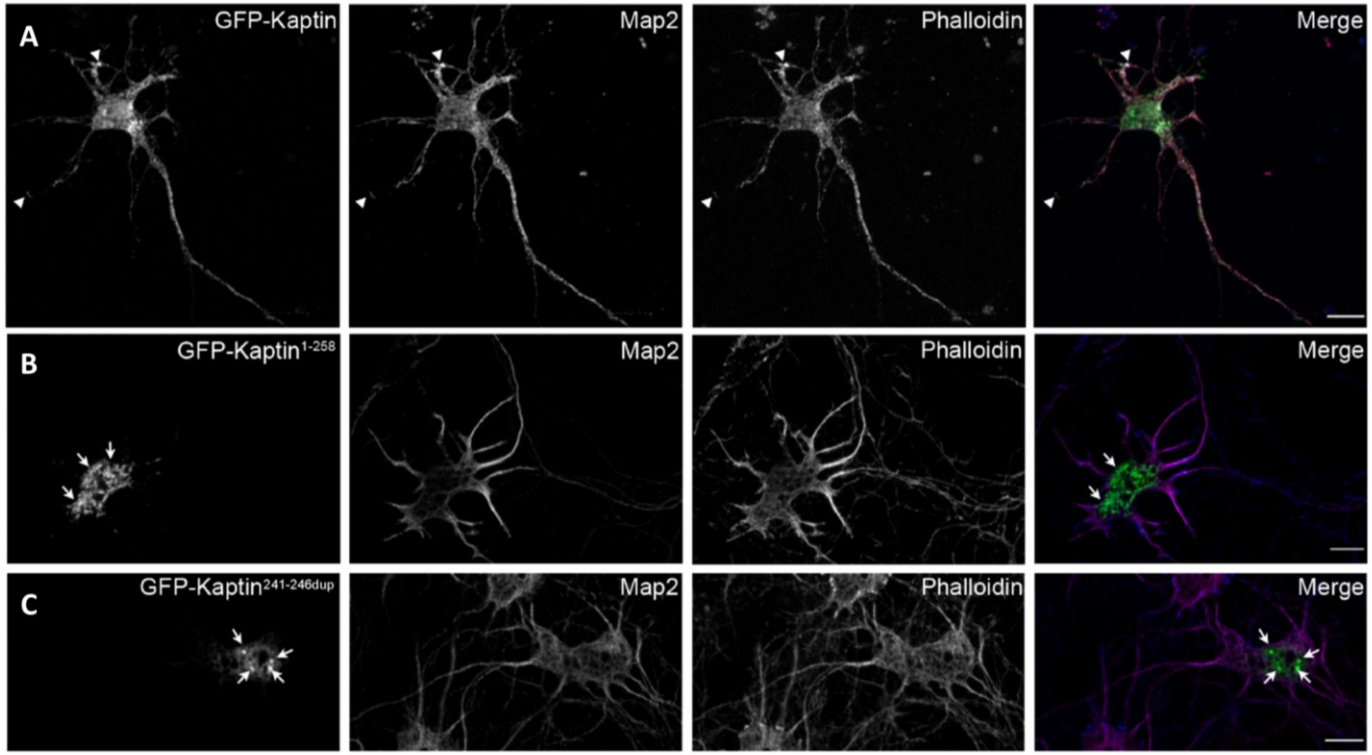


Figure 37: Immunolocallisation studies of altered kaptin in rat hippocampal neurones

Transfection of DIV3 rat hippocampal neurons with wild-type kaptin and mutants thereof fixed and imaged at DIV5. Wild-type kaptin is found throughout the cell with accumulations at growth cones and at F-actin rich sites of the cell body (**A**; arrow heads). In contrast, both mutants accumulate in areas of the cell body (**B** and **C**; arrows). Merged images show WT kaptin and altered kaptin in green, MAP2 in red and phalloidin in blue. Scale bar represents, 10 μm. Data courtesy of Dr Michael Kessels.

In order to explore the correlation between the human and mouse phenotypes further, the *Kptn*$^{-/-}$ mouse was re-examined. *KPTN*$^{-/-}$ mice had an increased skull circumference when compared to WT mice. Analysis of brain pathology revealed megalencephaly but otherwise unremarkable brain histopathology (Figure 38).

The social recognition assay which tests long term memory, was used to make an assessment of cognition in *Kptn*$^{-/-}$ mice. Female *Kptn*$^{-/-}$ mice were presented with an anaesthetised stimulus animal over 4 trials. After 24hrs, the mice were presented with two anaesthetised animals: 1) a novel anaesthetised mouse, and 2) the stimulus animal from the previous day. As expected the WT but not the *Kptn*$^{-/-}$ mice investigated the novel animal more than the stimulus (Figure 39).

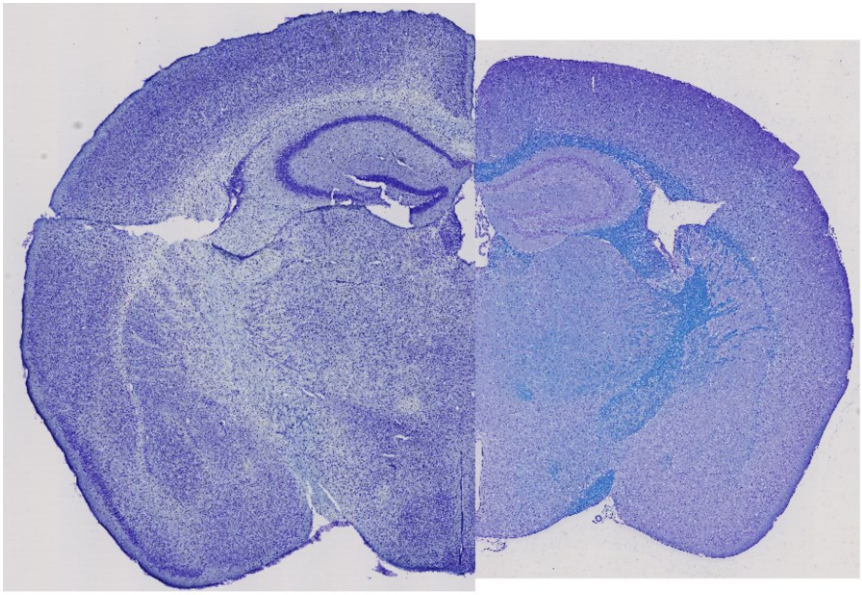


Figure 38: The kptn$^{-/-}$ mouse is megalencephalic.

Shown in the left panel is a half coronal section of Kptn -/- mouse (M00663346) Nissl stained. On the right is a half coronal section of a matched control mouse (M00526682) stained with Luxol Nissl. Data courtesy of Dr Christopher Lelliott, Wellcome Trust Sanger Institute.

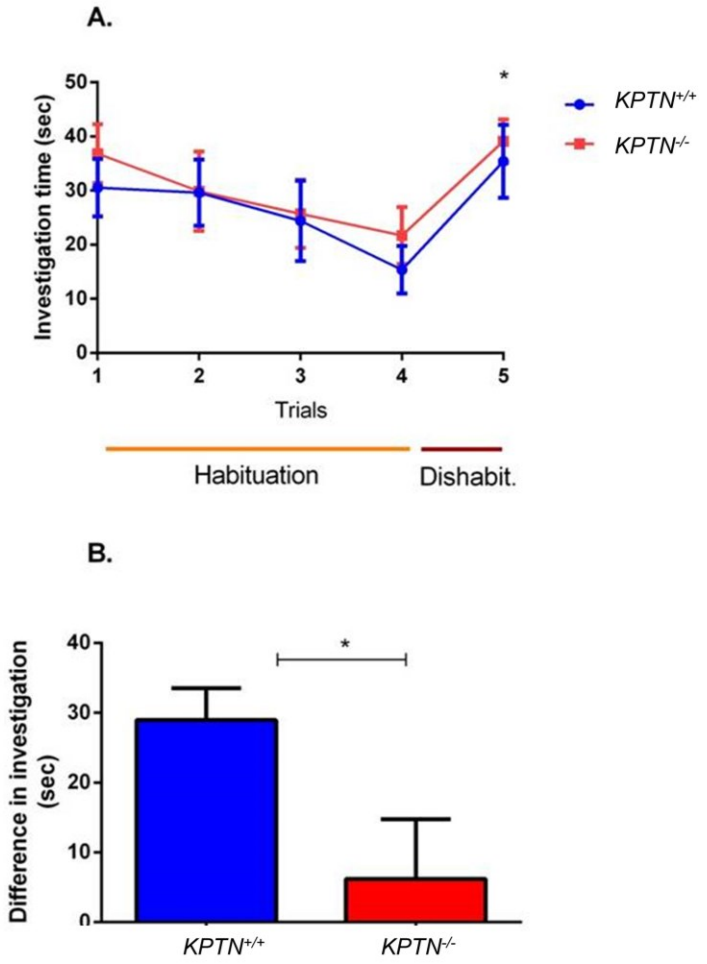


Figure 39: Long term memory impairment in *Kptn-/-* female mice

(**A**) Habituation-dishabituation test. Both *Kptn-/-* (n=9) and *Kptn+/+* (*Kptn* WT, n=9) mice recognise a stimulus animal (mouse A) repetitively presented to them over the course of four trials, as shown by a decline in the investigation time over trials 1-4. Both mutant and WT mice display an increase in the investigation time on trial 5 when presented with a novel stimulus animal (mouse B) (trial 4 vs trial 5, *p,0.05, *post-hoc* analysis after 2 way ANOVA). (**B**) Discrimination test (24hrs after the habituation-dishabituation test). The WT (+/+, n=9), but not the mutant mice (-/-), show a significant increase in time spent investigating the unfamiliar stimulus (mouse C) versus the familiar (mouse A), suggesting that *Kptn*$^{-/-}$ mice do not retain the memory of a familiar animal over 24hrs (P value = 0.0340*, t=2.348, df=16, two-tailed Student's t-test). The error bars represent the standard error of the mean. Experimental design was discussed with the author and analysis was observed. Data courtesy of Dr Darren Logan and Ms Maria Levitin, Wellcome Trust Sanger Institute.

4.5 Discussion

We investigated a number of families from the Anabaptist communities of Ohio, USA, with multiple individuals aged 1-30 years affected by a novel syndrome in which the cardinal features include macrocephaly, developmental delay, behavioural abnormalities and seizures. Our molecular studies determined that two distinct founder mutations affecting the same gene (*KPTN*, encoding kaptin) are responsible, both of which have become entrapped within the community. Those individuals found to be homozygous for the p.S259* nonsense mutation appeared to be more severely affected with a higher degree of intellectual impairment, compared with individuals found to be compound heterozygous for this as well as the p.Met241_Gln246dup mutation. This may indicate that the p.M241_Q246dup mutation retains a limited functionality *in vivo*, and perhaps consistent with this we observed that the p.Met241_Gln246dup mutation showed a slightly lesser tendency to form perinuclear accumulations in our transfection studies compared with the pS259X mutation. However, the sample cohort is currently too small to confidently determine any genotype-phenotype correlation.

Relatively little was known about kaptin other than a potential role at stereocilia tips, leading to the historical suggestion that *KPTN* may be a candidate gene for hearing loss (Bearer et al., 2000), not a feature seen in the affected individuals described here. The actin cytoskeleton plays a pivotal role in neuronal cell morphology and migration during development including the generation, protrusion and steering of growth cones as well as the formation of postsynapse and dendritic spines (Dent et al., 2011, Kessels et al., 2011, Svitkina et al., 2010). Our studies confirm kaptin expression in neuronal (MAP2-positive) cells. Since kaptin was found to localise to F-actin-rich structures, it is conceivable that loss of kaptin function may either directly, or indirectly, lead to impairment of the neuronal actin cytoskeleton required for dendritic arborisation and/or spine formation and result in the disease phenotype described. Support for this is provided by studies of Rab39B, a small GTPase associated with the Golgi apparatus (Chen et al., 2003), in which mutations lead to down regulation of Rab39B protein and a concomitant reduction in dendritic arborisation and synapse formation. This was associated with a disease phenotype comprising

intellectual disability, epilepsy and macrocephaly (Giannandrea et al., 2010, Tarpey et al., 2009), features which overlap with those described here arising from kaptin mutation. Similarly, deficiencies of Rho GTPases, which regulate the actin cytoskeleton by a growing variety of effector proteins, have been associated with spine structure defects and impaired cognition (Ba et al., 2013, Kasri et al., 2008, Nakayama and Luo, 2000, von Bohlen Und Halbach, 2010). Several other actin-associated proteins including drebrin A, Cortactin, and Abp1 have also been found to decrease spine density or formation (Haeckel et al., 2008, Tada and Sheng, 2006, Takahashi et al., 2003), and a growing body of evidence supports a role for the Arp2/3 complex and directly and indirectly associated proteins in postsynapse formation and proper development of neuronal morphology (Ahuja et al., 2007, Djagaeva and Doronkin, 2009, Irie and Yamaguchi, 2002, Kim et al., 2013, Koch et al., 2012, Martinez-Quiles et al., 2004, Michaelsen et al., 2010, Nakamura et al., 2011, Pinyol et al., 2007, Soderling et al., 2007, Tada and Sheng, 2006, Weed et al., 2000, Wegner et al., 2008).

Our analyses reveal that wild type kaptin is enriched in neuronal growth cones and at discrete cortical sites of neurons at early developmental stages. Furthermore, wild type kaptin was found to accumulate at postsynapses of neurons undergoing synaptogenesis (DIV14), as indicated by spatial correlation along dendrites of kaptin accumulations with phalloidin and synaptic marker immunostaining. Consistent with this, kaptin also was found to be present in the postsynapses of mature neurons. The sites demarcated by kaptin localisation represent areas of high F-actin content and high actin dynamics. An association of kaptin with dynamic F-actin was also indicated by the observed accumulations of tagged-kaptin at the dynamic lamellipodia, as opposed to the more static stress fibers in COS-7 cells. These observations are consistent with the suggestion of a lamellipodial localisation of kaptin in chicken embryonic fibroblasts as well as with the original isolation of kaptin from blood cells using F-actin columns (Bearer and Abraham, 1999).

The genetic and functional studies presented in this chapter indicate that both of the *KPTN* mutations identified are likely to result in loss of function of kaptin, either by degradation of the mutant transcript via mRNA surveillance

mechanisms (c.776C>A) or by the production of mislocalised and/or non-functional protein products. These *KPTN* mutations result in a distinctive clinical syndrome, and the presence of macrocephaly combined with global developmental delay/intellectual disability should prompt the diagnostic analysis of the *KPTN* gene in patients from Anabaptist communities. The benefits of early diagnosis in this condition is suggested by the improvement in developmental markers in the two youngest affected individuals in this study, both of whom received early and intensive developmental interventions, although the lack of seizures in these individuals may also have been beneficial. Finally our identification of two *KPTN* founder mutations within this Anabaptist population parallels the situation seen for a number of other genes with multiple mutations already known in the community which also commonly cause inherited diseases globally (e.g. connexin-26 mutations in inherited hearing loss, and ATM mutations in ataxia telangiectasia), indicating that the kaptin neurodevelopmental disorder may be similarly widespread.

Assessment of the phenotype in *Kptn*$^{-/-}$ mice demonstrated that these animals are macrocephalic and have impaired long term memory which aligns well the clinical characteristics of the human disorder. It is interesting that none of the mice have developed seizures, although the numbers assessed have been small and it is not a consistent feature of the human disorder. Taken together, these results show that kaptin plays a crucial role in normal neurodevelopment in humans and in mice and that its disruption leads to a syndrome characterised by megalencephaly and impaired cognitive ability.

5

CHAPTER FIVE

AUTOSOMAL RECESSIVE MUTATIONS IN CLOSELY LINKED WDR73 AND WHAMM GENES UNDERLIE A NOVEL COMPOSITE NEPHROCEREBELLAR SYNDROME PHENOTYPE

5 AUTOSOMAL RECESSIVE MUTATIONS IN CLOSELY LINKED WDR73 AND WHAMM GENES UNDERLIE A NOVEL COMPOSITE NEPHROCEREBELLAR SYNDROME PHENOTYPE

5.1 Introduction

5.1.1 The Swartzentruber church

There are four main separate Amish affiliations in the Holmes County settlement (Figure 40A). The most conservative Amish subgroup, the "ultraconservative" Swartzentruber group has its origins in Holmes County. The Swartzentruber church group arose between 1913 -1920 following a particular interpretation of church discipline. Unlike other Amish affiliations the name does not originate from the surname of the original leader, Bishop Samuel E Yoder (1872-1932), instead it comes from the surname of the two bishops who's views shaped the group's formation following Yoder's death.

Shunning ex-members of the church is common to all Amish affiliations. Traditionally anyone who left the Amish church or who was excommunicated (Bann) for a major sin would be shunned until they repented. For those who did not, the shunning (Meidung) was life-long. Around 1900, many of the Old Order Amish Bishops wanted to relax these rules such that those who left the church to become a faithful member of a different Anabaptist church group would no longer be shunned. However, Samuel Yoder, felt strongly that strict shunning should persist no matter what the offence and opposed what he saw as any potential loopholes. His church district broke fellowship with the other nearby Amish church groups and mediators were unable to mend the schism, leading to the formation of the Swartzentruber group.

The Swartzentruber church are now known for their resistance to any innovation in household or farm technology and practices, such as indoor plumbing, milking machines and community phone booths (Figure 40B). Unlike other Amish groups who are permitted to travel in cars so long as they are not the drivers, the Swartzentruber group may only use this mode of transport in

“emergency” situations. The group has also refused to adopt what they call “fast-time” (daylight saving time). Despite efforts from law enforcement bodies in the USA, they have resisted the use of the orange “slow-moving vehicle” triangle sported on the rear of the buggies of Amish from other groups. The buggies of the Swartzentruber group are recognisable by their windowless and plain appearance and are only dimly lit at night by kerosene lanterns, as opposed to more modern and powerful electric lights which have been accepted by other Amish church groups. Whereas small business ownership is now common amongst the Amish, farming is still the main form of occupation for Swartzentruber men. The group have their own distinctive style of dress and worship rituals and remain very much separate from other Amish affiliations (Hurst and McConnell, 2010, Kraybill et al., 2013). Importantly, the Swartzentruber church districts in Holmes County do not contribute to the “Ohio crippled children fund”, meaning that any medical investigations or treatment needed by Swartzentruber children affected by an inherited disorder must be funded by their family. Consequently, affected individuals typically undergo only very limited clinical investigations and treatment (personal communication).

5.1.2 Genetic studies in the Swartzentruber church

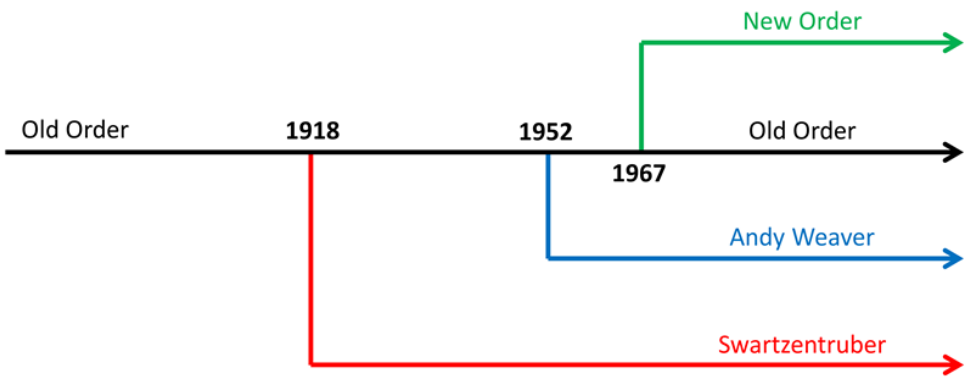
In 2012 Swartzentruber settlements could be found in a total of thirteen states and the group comprised around 7% of the total Amish population (Kraybill et al., 2013). Currently due to the particularly large numbers of children typical of families in this group, they remain the fastest growing affiliation. Parents in this group also tend to be more closely related than the 7-9 generation common ancestor typically seen in other groups. This is a significant figure to bear in mind given that the original founder population of this group was made up 107 already related individuals from two church districts. Consequently Swatrzentruber individuals typically harbour larger and more numerous autozygous genomic regions. While this may add an extra layer of complexity to genetic mapping studies of inherited disease amongst this group, this is offset at least in part by the beneficial effects of a larger family size.

As may be expected the structure and ancestry of these “ultraconservative” Amish has led to the expected enrichment of certain founder mutations within this group and an absence of other mutations commonly seen amongst the rest

of the Ohio Amish community. In 1967 the WOH co-founder, Professor Harold Cross published with his PhD supervisor Professor Victor Mckusick, a report describing three distinct and unique neurological conditions best classified as complex hereditary spastic paraplegias that he had identified amongst the Holmes County Amish. The syndromes were each named after the surname of the original nuclear family in which they were described (Cross and McKusick, 1967). Professors Cross and Crosby identified the molecular causes of two of these disorders, “Troyer syndrome” and “Mast Syndrome” in 2002 and 2003 respectively (Patel et al., 2002, Simpson et al., 2003b). The third condition named “Hershberger syndrome” appeared to be particularly prevalent amongst the Swartzentruber Amish in Holmes County. The syndrome had proven more difficult to study given the added clinical, genetic and social difficulties to investigating inherited disease amongst this ultraconservative church group.

The affected children from the original family described by Cross and Mckusick exhibited profound developmental delay, spasticity and irritability evident from 5 months of age. However, Hershberger syndrome appeared to be neurologically complex displaying significant phenotypic variability. This chapter describes the results of our clinical and molecular study of Hershberger syndrome. Many of the families recruited were directly related to the original family from the 1967 publication.

A



B

Figure 40: The Swartzentruber Amish church group

(**A**) Amish church group affiliations in the Holmes County settlement. The Swatrzentruber Amish affiliation in Holmes County, Ohio was formed after the earliest of the three main schisms from the Old Order Amish Church Group, adapted from (Hurst and McConnell, 2010). (**B**) Swartzentruber Amish barn raising in 2013. The men are wearing the typical wide brimmed hat sported by this church group. At the base of the barn the women of the family were quilting.

5.2 Methods

5.2.1 Exome sequencing

Kings College London

Sequencing of the consensus coding sequence (CCDS) defined exome was undertaken using the Agilent Sure Select All Exome Hybrid Capture kit and sequencing on the Illumina platform with 76bp paired end reads, in collaboration with Professor Richard Trembath and Dr Michael Simpson based at King's College London (KCL). Briefly, genomic DNA from was fragmented and enriched for exomic sequences with the Agilent SureSelect Whole Exome hybrid capture. The resulting enriched sequence library was sequenced across one lane of the Illumina GaIIx flowcell (v4 chemistry). The sequence reads were aligned to the reference genome (hg18) with the Novoalign aligner. Duplicate reads, resulting from PCR clonality or optical duplicates, and reads mapping to multiple locations were excluded from downstream analysis. Single nucleotide substitutions and small insertion deletions were identified and quality filtered within the SamTools software package. Filtering of variants for novelty was performed by comparison to dbSNP131, 1000 Genomes SNP calls (March 2010) and variants previously seen in the KCL cohort. The exome sequencing was performed by Dr Michael Simpson who also provided guidance and support with the analysis of the data.

Wellcome Trust Sanger Institute

Genomic DNA from venous blood was obtained and captured by SureSelect Target Enrichment V3 (Agilent).GA-II (Illumina, USA).was used to sequence the sample. Reads were mapped to the reference genome via BWA, duplicates where marked and removed using Picard (v1.27). Single nucleotide variants were called using GATK (v1.0.4269M) and Samtools (v0.1.7) while indels were called using Dindel (1.01) and Samtools. All variants were reported within bait regions +/- 100bp and the results merged in a single VCF file. The variant consequences on the protein structure was predicted using Ensembl (58) and annotated with dbSNP 130 and the minor allele frequencies from the 1000 genomes project to filter common variants (MAF >1%), to generate a profile of

rare sequence variants and variants not present in publically available databases. The exome sequencing was performed by Dr Saeed Al-Turki who also provided guidance and support with the analysis of the data.

5.2.2 Functional studies

Functional studies undertaken to understand the impact of the *WHAMM* variant identified in this study were carried out in collaboration with Dr Kenneth Campellone.

5.2.3 Genetic and clinical studies of other patient cohorts

Dr Erik Puffenberger planned and undertook the genetic studies for the Clinic for Special Children patient cohort. Exome sequencing was performed at the Broad Institute.

Dr Kevin Strauss and Dr Heng Wang phenotyped affected patients seen at the Lancaster Clinic for Special Children and DDC clinic, Geagua County respectively.

Analysis of the neuropathology was performed by Dr Brian Harding, Department of Neuropathology, University of Pennsylvania. The renal pathology is courtesy of Dr Kevin Strauss, Clinic for Special Children, Pennsylvania.

5.3 Results section A

5.3.1 Identification of a pathogenic Amish founder mutation in *SAMHD1*

5.3.1.1 Autozygosity mapping reveals a single notable autozygous region on chromosome 20q11.22-23 in affected individuals

A single nuclear Swartzentruber Amish family (Figure 41, family 1) were recruited through the WOH project and were originally seen by Dr Christos Proukakis (Consultant Neurologist, University College Hospital) and Professor Harold Cross in 2007. Three siblings were affected by what was felt likely to be a novel condition and to consist of significant short stature (greater than -2.5 SDS), hoarse voice, and skin sensitivity (the clinical phenotype is described in detail below). A further individual (VIII:X) from a closely related family (Figure 41, family 2), thought likely to be affected by the same condition, was ascertained through the maternal grandmother. Family 2 lived in a different region of the USA and so the information pertaining to that individual was obtained through his maternal great aunt. The family pedigree is illustrated in Figure 41.

Assuming that a founder mutation was responsible for this condition, autozygosity mapping studies were initially carried out for family 1. A genome wide microarray scan using Illumina Human CytoSNP-12v2.1 330K arrays was undertaken using DNA from the three affected brothers (VIII:I, VIII:III and VIII:IV) and one unaffected sibling (VIII:VIII). This identified a 5.4Mb region of chromosome 20q11.22-23 as the likely disease locus. The locus was delimited by markers rs2747539 and rs2009579 and contained 113 candidate genes.

5.3.1.2 Exome sequencing identifies a novel intronic variant in SAMHD1 within the splice-acceptor site of intron 12

In parallel with the mapping studies, given that the vast majority of disease causing mutations occur within coding regions and intronic splice junctions, whole exome second generation sequencing was undertaken. Sequencing of the consensus coding sequence (CCDS) defined exome was undertaken in a single affected individual (VIII:IV, Figure 41, family 1). The efficiency of the hybrid capture was 71.4% with 45576342 of the 63847126 uniquely mapped

sequence reads originating from the targeted exomic regions (74.6% to the targeted regions +150bp). This level of capture efficiency and sequence generation resulted in a mean coverage of the CCDS defined exome of 85.6%, with 91.6% of CCDS bases covered by at least 10 reads.

A single novel variant located within the essential splice site sequence of intron 12 of the *SAMHD1* gene (NM_015474.3 c.1411-2 A>G) was identified within the putative disease locus. The variant was analysed using the Alamut mutation interpretation software package (Interactive Biosoftware) which predicted the highly conserved splice acceptor site to be abolished. Primers were designed to flank the *SAMHD1* variant for further investigation (see Appendix for primer sequences). PCR amplification and automated Sanger sequencing of the 3 affected siblings, their mother and 5 unaffected siblings revealed that the mutation co-segregated with the disease phenotype (Figure 41). Homozygous loss of function mutations in *SAMHD1* have previously been found to cause Aicardi-Goutières syndrome associated with an intracerebral aneurysmal and occlusive intracerebral arteriopathy (Ramesh et al., 2010, Xin et al., 2011).

The *SAMHD1* c.1411-2 A>G variant identified in this Amish family introduces a novel MspI restriction site. This permits visualisation of the mutation by MspI digestion of the 400bp PCR product, producing fragment sizes of 209bp and 191bp. *SAMHD1* PCR products generated from 112 control chromosomes from unaffected individuals from the same Ohio Amish community and 106 from unaffected individuals of European descent were digested with MspI restriction endonuclease and subsequently visualised on agarose gels. While no carriers of the mutation were identified in European controls, 3 out of 56 healthy individuals from the same Ohio Amish community were found to be heterozygous for the sequence variant, indicating that this is a founder mutation with a relatively high carrier frequency within the community. No homozygotes for the mutation were identified.

5.3.2 *SAMHD1* c.1411-2 A>G results in skipping of exon 13

In order to establish the effect of the variant at the level of transcription, RNA was extracted from peripheral blood lymphocytes derived from the 3 affected siblings, their mother and a healthy control. RT-PCR was carried out.

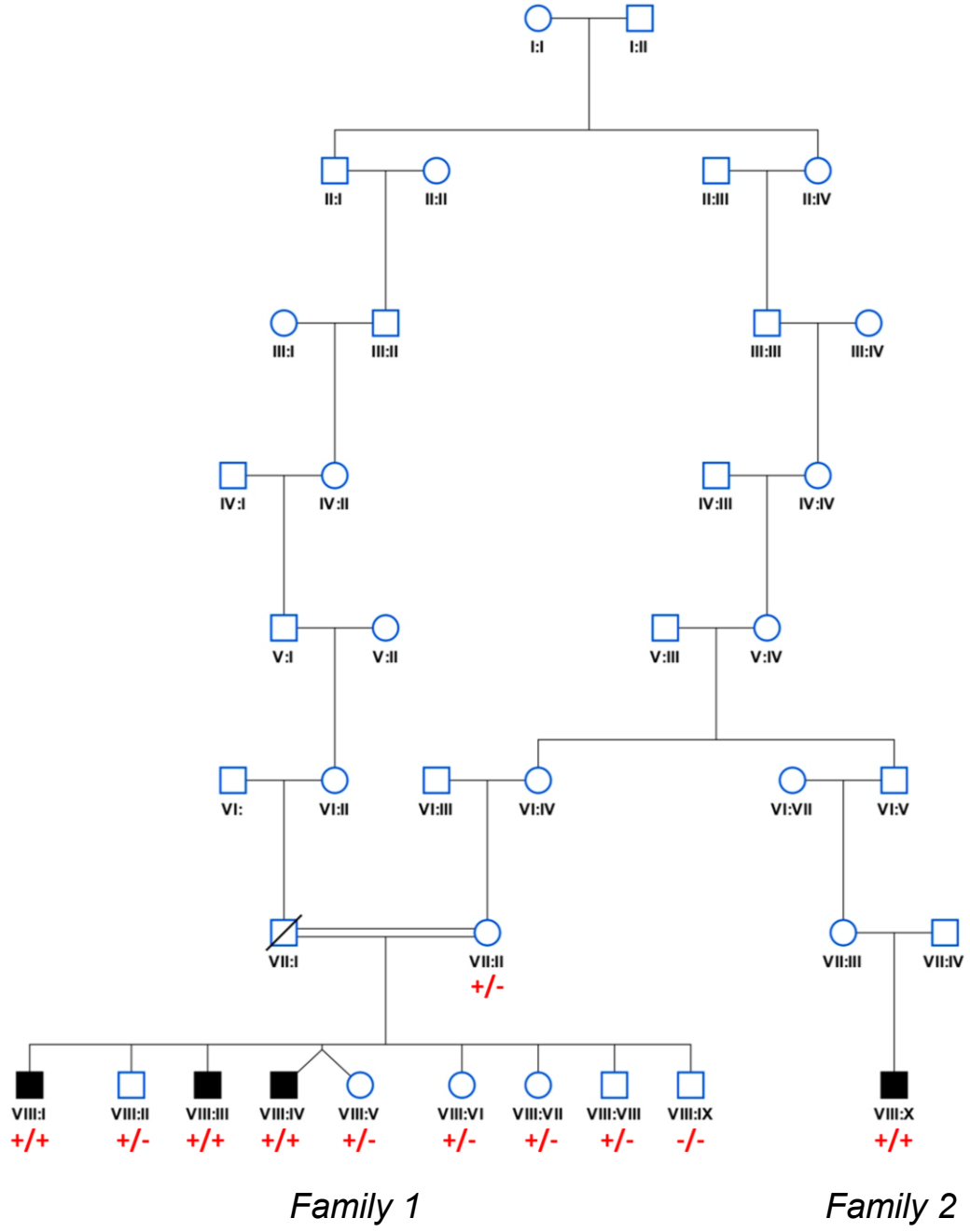


Figure 41: Simplified pedigree of the Amish family initially investigated

All affected individuals were subsequently shown to be homozygous for the *SAMHD1* (NM_015474.3 c.1411-2 A>G) variant (indicated). The parental sample was heterozygous and unaffected siblings were either wild type or heterozygous carriers. The genotype is shown in red under individuals in generations VII and VIII, with (+) denoting mutant and (-) wild type.

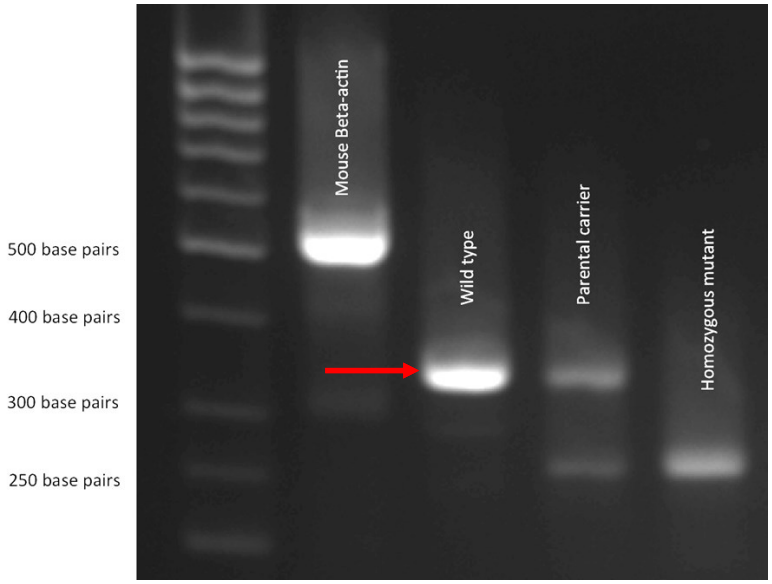
Forward and reverse primers were designed specific for sequences in exon 11 and 14 respectively; primers were designed to produce a product of size 356bp using wild type DNA. In accordance with this, amplification of control DNA produced a single product of expected size, whilst products from the affected individuals produced a 263bp fragment, consistent with the omission of exon 13. The mothers sample produced bands of both sizes 263bp and 356bp consistent with her carrier status, (Figure 42A).

The RT-PCR products from the healthy control, the mother and one affected individual were subsequently sequenced confirming the skipping of exon 13, that was predicted to result in an in-frame deletion of 31 amino acids (SAMHD1 p.Glu471_Asp501del), (Figure 42B).

5.3.3 Clinical findings of individuals homozygous *SAMHD1* c.1411-2 A>G

The identification of the Amish *SAMHD1* founder mutation permitted a more direct comparison of the affected Amish cases with previous reports of individuals homozygous for mutations in *SAMHD1.* Aicardi-Goutières syndrome, was initially described as a progressive familial early onset encephalopathy associated with intracranial calcifications and CSF lymphocytosis (Aicardi and Goutieres, 1984). The characteristic clinical picture associated with the condition is one of early onset subacute severe encephalopathy associated with intermittent sterile pyrexias, developmental regression and acquired microcephaly (Aicardi and Goutieres, 1984, Aicardi, 2002). The majority of affected individuals have abnormal neuroimaging findings include cerebral atrophy, intracranial calcification classically of the basal ganglia, and white matter changes most commonly affecting the frontotemporal region. In the early stages of the condition in addition to leukocytosis, raised levels of interferon alpha and neopterin are found in the CSF. Around 40% of affected individuals have been found to suffer chilblain lesions of the extremities (Depienne et al., 2007). Mutations in *TREX1*, *RNASEH2A, RNASEH2B, RNASEH2C SAMHD1, ADAR1* and most recently *IFIH1* (Crow et al., 2006a, Crow et al., 2006b, Rice et al., 2009, Rice et al., 2012, Rice et al., 2014), account for around 90% of cases with a classical presentation. Since the initial description of Aicardi-Goutières syndrome significant phenotypic variability has been recognised. In some

A



B

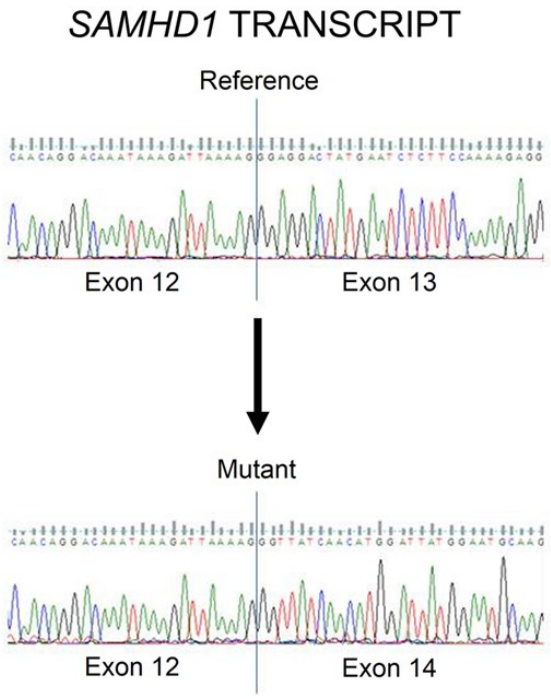


Figure 42: SAMHD1 RT-PCR products

(A) A single product of expected size 356bp (red arrow) is seen in the healthy control, the affected individual has a single product of size 263bp consistent with deletion of exon 13. The parent has two products, one of each size consistent with her carrier status. Positive control, mouse Beta-actin (500bp). (B**)** Electropherograms showing the sequence of *SAMHD1* RT-PCR products in a wild-type control (above) and a homozygous affected individual (below). Exon 13 has been spliced out in the affected individual.

individuals the condition is almost indistinguishable from congenital infection (Crow and Livingston, 2008) whereas others have a later onset of disease and very mild neurological manifestations (Depienne et al., 2007). Clinical overlap with systemic lupus erythematosus is also widely recognised (Lee-Kirsch et al., 2007, Ramantani et al., 2010), moreover heterozygous *TREX1* mutations have been identified in a subset of patients with systemic lupus erythematosus (Lee-Kirsch et al., 2007).

In light of the genetic results, further history was gathered and the family were re-examined, the case histories and clinical findings are summarised below. We were subsequently informed by the family that following the identification of the *SAMHD1* mutation as causative of the phenotype, the fourth affected child (

Figure 41, family 2) was tested locally and as expected, found to be homozygous for the *SAMHD1* variant.

Individual VIII:I: An 18 year old man was born at term weighing 2608g (-2 SDS) after an uncomplicated pregnancy. He grew appropriately until 3 months of age, when weight gain slowed. There was no history of regression or unexplained fever, although he was described to be more lethargic from the age of 6 months. He has a mild hearing loss, development was otherwise normal and he works on the family farm, neurological examination was unremarkable. From the age of 1 year he developed unexplained pruritic inflammatory lesions at extremities in the cold weather, particularly on the helices of his ears, nose, cheeks, hands and feet, resolution of the lesions left significant scaring. His cutaneous features had previously been unclassified but the appearance and history is consistent with chilblains. The lesions have become less frequent over the last 5 years. He experiences joint stiffness that worsens in winter making it difficult to climb up onto the buggy (the main Amish form of transport). A long history of recurrent severe mouth ulcers necessitating ENT investigation was reported and direct laryngoscopy demonstrated laryngeal oedema. Since the age of 5 years he has developed a progressively hoarse voice. At the age of 18 years he suffered a subarachnoid haemorrhage secondary to a ruptured middle cerebral artery aneurysm. Follow-up angiography detected a further berry aneurysm that was subsequently clipped. CT brain imaging revealed multiple

areas of hypoatenuation within the deep white matter of both cerebral hemispheres, but no intracranial calcification was reported.

Notable findings on examination included short stature (156.8cm, -2.9 SDS), OFC was within normal limits (56cm, -0.7 SDS). There was marked swelling of the interphalyngeal joints. Neurological examination was unremarkable. At age 18 he had no facial hair, genitalia were reportedly underdeveloped but understandably formal examination to establish pubertal status was refused. Additional findings included a left preauricular pit, said to also be present in other members of the family.

Individual VIII:III: A 13 year old boy born at term weighing 2920g (-1.3 SDS). He had short stature measuring 137cm at 13 years (-3 SDS), OFC was within normal limits (54.5cm, -0.8 SDS). Developmental progress was similar to his older brother with no reported problems at school. Oral ulceration had been less severe and the hoarseness of his voice was less pronounced. He also suffered with joint stiffness which worsened in the winter months. Examination revealed similar cutaneous lesions to individual VIII:I.

Individual VIII:IV: Is one of dizygotic twins aged 12 years. He weighed 2466g at birth (-2.3 SDS). He was short for his age at 122.5 cm (-3.8 SDS), OFC measured 52.5cm (-1.68 SDS). Development was essentially normal and he has no learning difficulties. He had joint stiffness, worst in cold weather and the most significant problems with oral ulceration. His voice was extremely hoarse. He had the same skin manifestations as his older brothers, although less severe.

Patient VIII:X: An 8 year old boy born at term, previously diagnosed with cerebral palsy. He suffers from chilblain lesions on his buttocks that worsen in severity during cold weather. This child was not available for review and no medical records were available.

Individual VII:I: The father of the 3 boys was reported to have had a hoarse voice and suffered with cutaneous lesions similar to his sons although much more mildly. He died at the age of 37 years from a presumed myocardial

infarction, no autopsy was performed as it is against Swartzentruber Amish beliefs. No DNA sample was available for molecular analysis, however one of his healthy children was homozygous for the wild-type *SAMHD1* sequence indicating that he was likely heterozygous for the c.1411-2 A>G sequence variant. Consequently it remains a possibility that the cause of death may be related to his *SAMHD1* carrier status.

The diagnosis of Aicardi-Goutières syndrome in this family was only considered in the light of the whole exome sequencing data, and subsequent thorough clinical evaluation that resulted in recognition of the cutaneous manifestations as chilblain lesions. Oral ulceration is commonly found in patients with systemic lupus erythematosus and is now a recognised feature of Aicardi-Goutières syndrome (Ramantani et al., 2010, Dale et al., 2010). In Aicardi-Goutières syndrome the ulcerative lesions are usually limited to the buccal mucosa although pharyngeal ulceration has been reported (Ramantani et al., 2010). Interestingly, laryngeal ulceration and oedema resulting in a hoarse voice have also been described in systemic lupus erythematosus (Teitel et al., 1992).

A specific association between Aicardi-Goutières syndrome caused by mutations in *SAMHD1* and features of arthritis has been reported (Ramantani et al., 2010, Dale et al., 2010), this combined with the inflammatory vascular features reported is strongly supportive of an underlying abnormal immune response. In patients affected with Aicardi-Goutières syndrome of known molecular cause higher expression of genes induced by type I interferon is seen leading to a sustained interferon signature (Crow et al., 2014) and (Dale et al., 2010, Rice et al., 2009).

While a manuscript reporting this work was in preparation by our group Xin *et al* published a paper entitled "Homozygous mutation in *SAMHD1* gene causes cerebral vasculopathy and early onset stroke" (Xin et al., 2011) in which they describe identifying the founder mutation in *SAMHD1* in 14 Old Order Amish individuals. Despite this report the work presented in this thesis is novel, and at the time of discovery had not previously been reported.

5.3.4 Hershberger syndrome is genetically heterogeneous, with *SAMHD1* Aicardi-Goutières syndrome accounting for the clinical presentation in some families

The phenotype of the affected Amish individuals reported by Xin *et al* varied from the more severe classical Aicardi-Goutières syndrome to milder presentations much like the family described here. The variability in clinical presentation of Aicardi-Goutières syndrome led to a review of the medical records of other individuals recruited to the WOH project and Sanger sequencing for the *SAMHD1* founder mutation in those with a phenotype suggestive of the condition. This molecular analysis revealed that *SAMHD1* likely accounted for the clinical presentation of four individuals from two nuclear Swartzentruber Amish families, all of whom were homozygous for the *SAMHD1* founder mutation. Interestingly, all four affected individuals had previously been diagnosed with Hershberger syndrome. Both families had moved from the Holmes County area since 2007 and so were unavailable for further clinical evaluation. However, all had been documented to have an Aicardi-Goutières syndrome compatible phenotype comprising of pre and post-natal growth failure, profound global developmental delay, spasticity, contractures, hyperreflexia and microcephaly.

Over the last four years through the WOH project, we have subsequently identified an increasing number of additional Amish children and adults affected with Aicardi-Goutières syndrome in Holmes County who are homozygous for *SAMHD1* c.1411-2 A>G, all are from the Swartzentruber church group.

5.4 Results section B

5.4.1 Hershberger syndrome is comprised of two phenotypically distinct groups

All the individuals from Holmes County originally described as having Hershberger syndrome that were WT for *SAMHD1* were subsequently reviewed. In total we identified and obtained DNA samples from 12 children from 6 interlinking nuclear families (Figure 43, families 3-5 and 7-9) all of whom were said to have Hershberger syndrome. In 5 of the families, there were additional children now deceased who had a history consistent with the same disorder. Although all were profoundly neurologically impaired and most were microcephalic, it was clear that the phenotype of these individuals was distinct from those that were homozygous for *SAMHD1*. Given the distinctive features of the condition described below, it was renamed nephrocerebellar syndrome. All of the children with nephrocerebellar syndrome recruited to the WOH project were from the Swartzentruber church group, consequently only limited medical investigations had been performed.

More detailed phenotyping revealed that at birth, affected children usually appeared normal. By 6 weeks of age, their parents report extreme irritability, poor feeding, and congenital roving nystagmus. Head size was normal at birth but brain growth was slow thereafter (Figure 44A). Psychomotor development was stagnant in all areas; affected children were non-communicative, had minimal purposeful hand use, and did not learn to sit or crawl. The only notable physical features were coarse facial features more evident with increasing age, microcephaly, receding forehead, fine slow growing hair and hypoplastic yellow teeth that were prone to decay (Figure 44B). All affected children had axial and limb hypotonia with preservation of deep tendon reflexes. Some developed a movement disorder in late infancy that manifested as a combination of axial dystonia, limb chorea, and restlessness. Abnormal movements were especially evident during agitation or emotional excitement. Brief, multifocal seizures were common and in some cases medically intractable. All affected children suffered from severe visual impairment that was attributed primarily to abnormal cortical processing, but some patients also had evidence of mild or moderate optic atrophy. The majority of patients had irregular sleep patterns and reversal of the

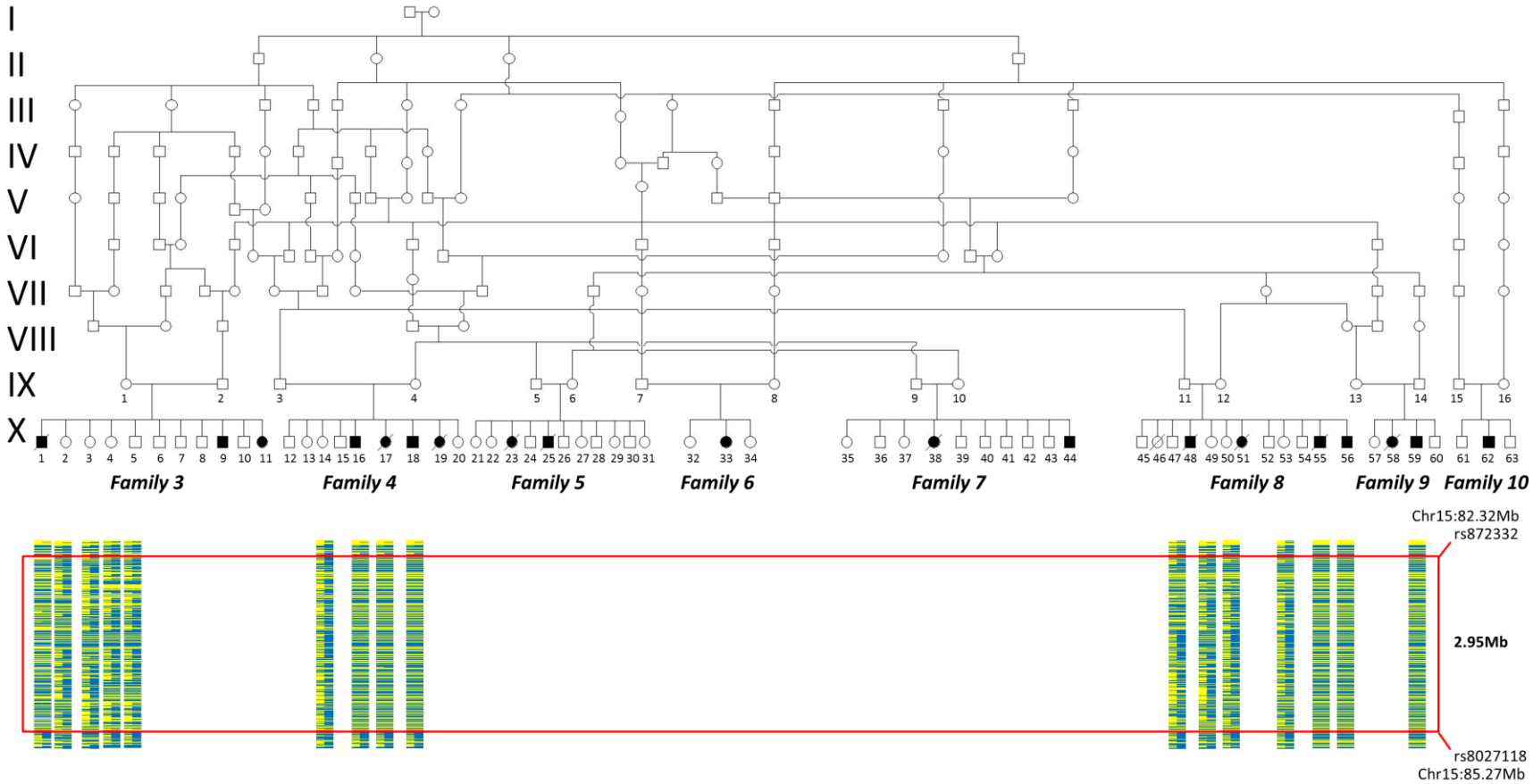


Figure 43: Nephrocerebellar syndrome family pedigree and gene mapping

Pedigree showing all of the Ohio families investigated, which interlink into a single extended family. Beneath is a pictorial representation of the SNP genotype data encompassing the chromosome 15 homozygous (red box) region in affected individual, healthy siblings investigated are heterozygous across this region.

normal sleep-wake cycle. Survival beyond teenage years was rare. The clinical presentations of individuals from Homes County affected with nephrocerebellar syndrome are summarised in Table 14.

In parallel with our studies in Holmes County, the Clinic for Special Children (CSC; Lancaster Pennsylvania) and the DDC Clinic (Geagua County) were studying a clinically similar condition to that under investigation by ourselves. Through discussions at the “Genomic Medicine and the Plain Populations of North America” conference (July 2013 at Franklin and Marshall College, Pennsylvania), it became clear that the same condition was likely under investigation by all groups. The Clinic for Special Children and DDC cohort included individuals from Ohio, Wisconsin, Virginia and Indiana, but most patients were primarily of Big Valley, Pennsylvania origin. All of these children were from a less conservative Amish background than the Swartzentruber church, which permitted more detailed clinical investigation allowing more detailed collaborative clinical studies.

Although the majority of the Holmes County Swartzentruber families underwent only minimal interactions with the modern medical system, a history of oedema with intercurrent illness was consistently given. All of the nephrocerebellar syndrome affected individuals who had had their renal function evaluated developed nephrotic syndrome during early childhood, which invariably progressed to fatal renal failure (Figure 45A). Proteinuria fluctuated over time from mild to heavy (mean urine protein:creatinine ratio 12.0 mg:mg, range 3.5 - 20.0 mg:mg; normal reference range 0.0-0.2 mg:mg) and did not appear to be steroid-responsive.

Neuroimaging revealed a spectrum of brain phenotypes. In some affected children, the myelination pattern appeared immature for age but the brain was otherwise morphologically normal (Figure 46A). In others, evidence of diffuse cerebral atrophy (most notably involving the frontal lobes), *ex vacuo* ventriculomegaly, hypomyelination, thin corpus callosum, and underdevelopment of the cerebellar vermis and hemispheres was identified (Figure 46B). At this stage, sufficient imaging data was not available to determine if this variability was age-dependent; i.e. if there was progressive

atrophy over time. Electroencephalograms were characterised by a slow and disorganized background, absent posterior (occipital) rhythm, poor wake-sleep transition, multifocal spikes and sharp waves, and rarely, high-voltage modified hypsarrhythmia.

5.4.2 Renal pathology

Two affected children had renal biopsies (Figure 45B-D), which showed enlarged and hypercellular glomeruli, narrowing of capillary lumina, lobulation of glomerular tufts, disorganised segmental sclerosis with endocapillary proliferation, and duplication of the glomerular basement membrane, producing a characteristic “tram-track” configuration. While components of both membranoproliferative glomerulonephritis type I and IgA nephropathy were present, the abundant number of glomeruli with disorganised segmental sclerosis and endocapillary proliferation were reflective of the rapid and intractable course of kidney disease observed in patients.

5.4.3 Neuropathology

An affected girl died at age 9.3 years and her brain was fixed for examination 27 hours post-mortem. Brain weight was 681grams (expected weight for age, 1100-1300 grams). The cerebral hemispheres were symmetric and showed no softening, atrophy, or mass effect. The leptomeninges were clear and transparent, there was no internal herniation, and the basal vessels and cranial nerves appeared normal. Gross coronal sections showed no malformation of the cortex, white matter, ventricular system or basal grey nuclei. Histological sections from the hexalaminar cortex, hippocampus, lentiform and caudate nuclei, internal capsule, hypothalamic nuclei, mammillary body, and the ependymal lining of the third ventricle were normal, and horizontal sections of midbrain and pons were unremarkable (the medulla was unavailable for examination). The white matter had scattered spongy changes but no cellular reaction; myelin was well preserved with no gliosis.

The cerebellum was markedly abnormal. Macroscopically the cerebellar vermis and hemispheres were firm and sclerotic (Figure 47A). Histologically the cerebellar hemispheres had sufficient branching, but folia were short and stubby and granule cells were markedly decreased (Figure 47B).

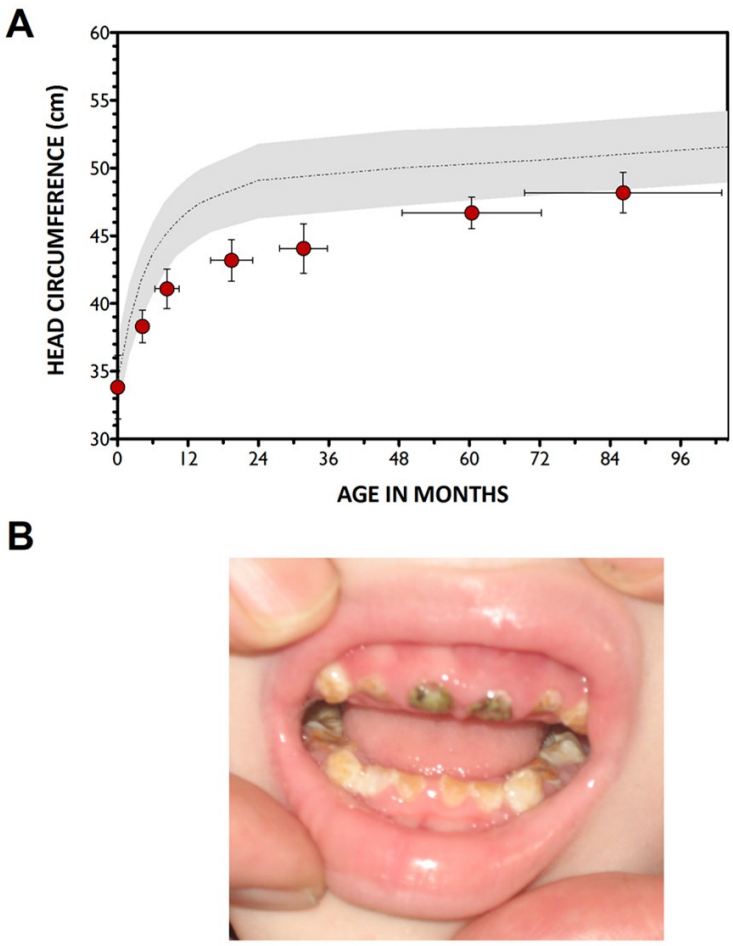


Figure 44: Clinical Features of nephrocerebellar Syndrome

(**A**) Children affected with nephrocerebellar syndrome (NCS) were typically born with a low-normal head circumference, but head growth decelerated by 6 months of age, reflecting decreased postnatal brain growth (gray shaded area = normal head circumference for age, mean ± 2SD; red circles = head circumference of patients, mean ± SD), data courtesy of Dr Kevin Strauss. (**B**) Primary and secondary dentition appeared hypoplastic and were prone to decay.

Table 14: Clinical findings of individuals with nephrocerebellar syndrome

Pedigree reference	X:1	X:9	X:11	X:16	X:17	X:18	X:19	X:25	X:33	X:44	X:55	X:56	X:59	X:62
Gender														
Age at evaluation (years)	†	5.9	0.8	9.8	†	6.4	2.9	†	10.3	2.1	11	8.29	1.3	X:7.8
Age of Death	13.7				3.3		3.2	2.7			12.3			
WDR73 Genotype	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M
WHAMM Genotype	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M	M/M
Development														
No independent sitting	1	1	1	1	1	1	1	1	1	1	1	1	1	1
No purposeful hand use		1			1		1	1		1	1	1	1	1
Neurology														
OFC SDS	N/K	-4	-3.3	-2.2	N/K	-2.2	-3.5	N/K	N/K	-6	-2.4	-3.1	-5	N/K
Non-communicative	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Central hypotonia	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Restless limb movements	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Roving eye movements	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Neuroimaging		CT-cerebellar hypoplasia	CT-cerebellar hypoplasia						CT-NAD				CT-NAD	
Epilepsy	1					1	1	1		1		Febrile	1	1
Renal														
Nephrotic syndrome	Severe Oedema prior to death	1	Unknown	Unknown	Symptomatic no evaluation	Unknown	Unknown	Oedema prior to death	1		Symptomatic no evaluation	Symptomatic no evaluation		1
Renal insufficiency/failure	Severe Oedema prior to death			N/K	Severe generalised oedema prior to death	N/K	N/K	Severe generalised oedema prior to death	1		N/K	N/K		
Physical anomalies														
Dental caries	1	1		1		1	1	1	1					1

†, Deceased; M, denotes the presence of the variant; 1, indicates the presence of a feature; OFC, occipitofrontal circumference; SDS, standard deviation scores; N/K, not known

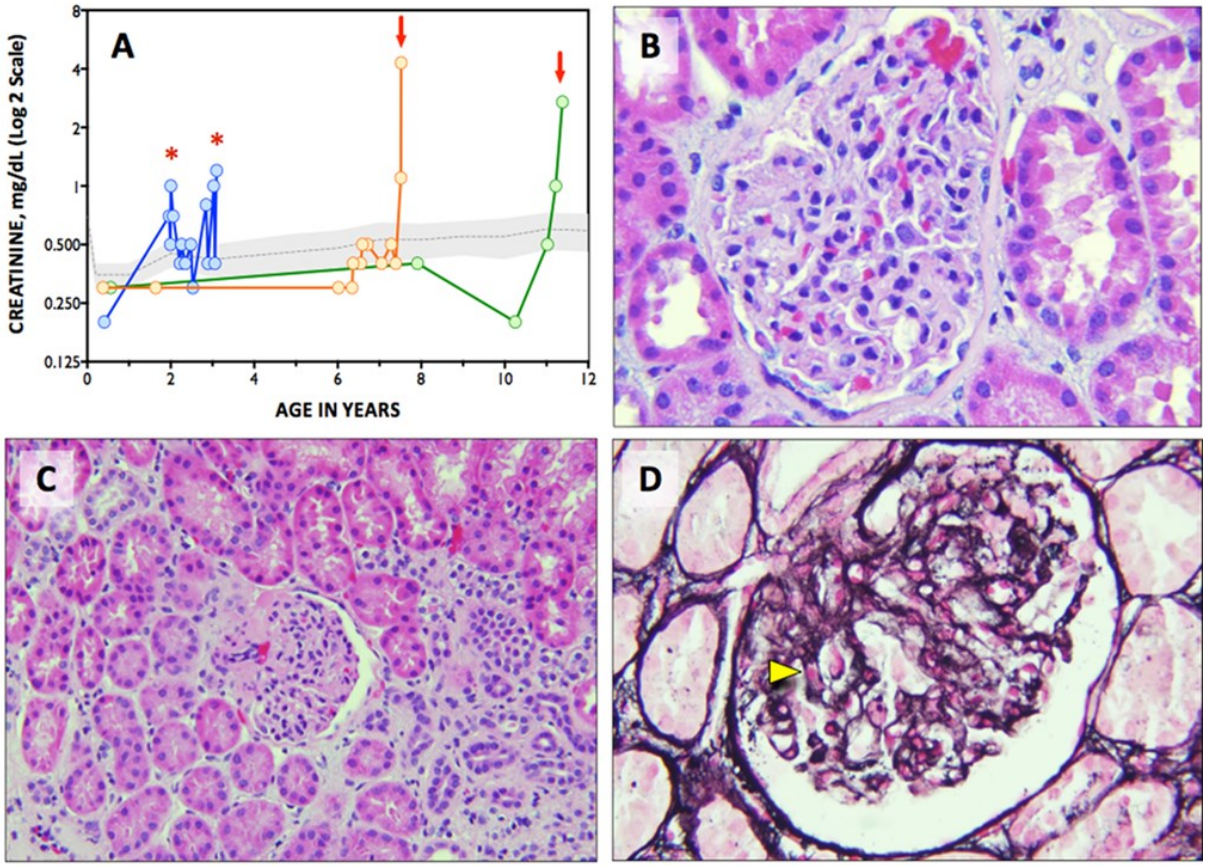


Figure 45: Nephrocerebellar syndrome renal pathology

(**A**) The pace of nephron loss varies among individual patients (indicated by blue, orange, and green symbols), and is punctuated by episodes of acute, reversible renal failure (red asterisks). Renal failure was the primary cause of mortality across all three cohorts (red arrow indicates time of death). Gray shaded area = normal serum creatinine for age, mean ± 2SD. (**B**) Glomeruli of affected patients are enlarged and hypercellular with narrowing of capillary lumina and lobulation of glomerular tufts. (H&E stain 40x magnification). (**C**) There is disorganised segmental sclerosis with endocapillary proliferation (H&E stain 20x magnification), and (**D**) duplication of the glomerular basement membrane produces the characteristic “tram-track “ configuration (yellow arrowhead; Jones stain, 40x magnification).

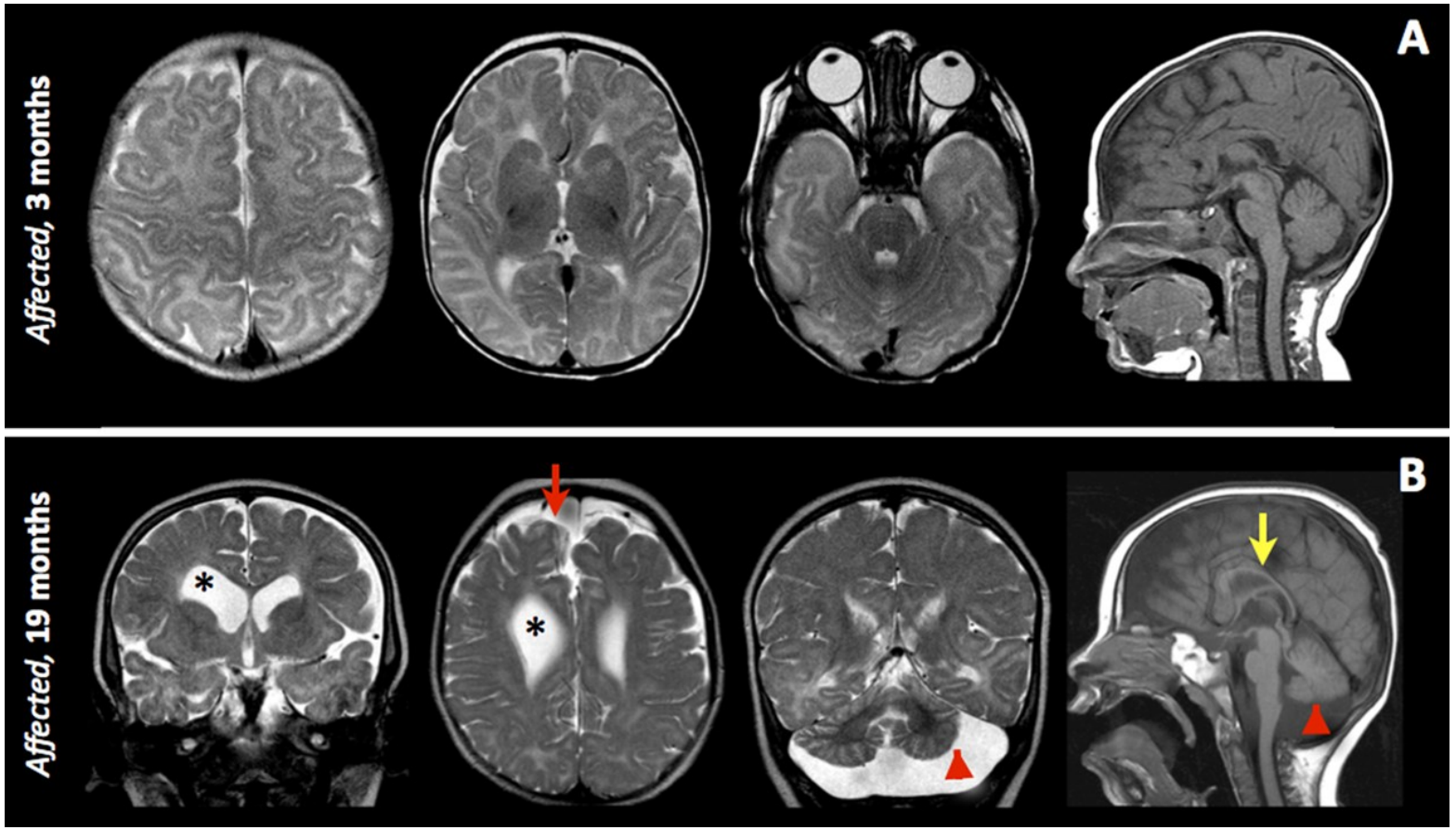


Figure 46: Neuroradiology findings in nephrocerebellar syndrome

(**A**) Some patients had small but apparently structurally normal brains (affected female, age 3 months). (**B**) Commonly observed abnormalities among other patients include enlarged, irregular ventricles (asterisks), indicating ex vacuo cerebral atrophy, hypoplastic frontal lobes (red arrow), underdevelopment and/or atrophy of the cerebellar vermis and hemispheres (red arrowheads), and thin corpus callosum (yellow arrow) (affected male, age 19 months; from left to right, coronal T2, transverse T2, coronal T2, sagittal T1). We did not have sufficient historical imaging to determine if this variability was age-dependent; i.e. reflective of postnatal degeneration or atrophy of brain tissue. Data courtesy of Dr Kevin Strauss, Clinic for Special Children, Pennsylvania

Purkinje cells were better preserved by comparison, but still decreased in number, crowded together, and surrounded by rich Bergman gliosis. Purkinje cells were overlaid by a thin molecular layer that was hypercellular: some of these cells were Bergman glia but others were misplaced Neu-N positive granule cells. Rare Purkinje cells were also present in the molecular layer. Globose swellings of Purkinje cell dendrites ('asteroid bodies') were covered in fine spikes, and dysmorphic Purkinje dendritic trees formed 'weeping willow' arrangements throughout the narrow molecular layer (Figure 47C and D).

Synaptophysin staining was strong throughout the molecular layer and present around Purkinje cell somata while punctate glomerular immunoreactivity was also present in the depleted granule layer. A neurofilament 211 stain demonstrated preservation of basket cells, fibers, and pericellular baskets, as well as occasional. Purkinje axonal swellings ("torpedos") (Figure 47C). The dentate nucleus was well formed, properly convoluted, and had a normal complement of neurons. Myelin of the amiculum was well preserved but there was a subtle decrease of myelin in the hilum where gliosis was visible with GFAP stain. No axon terminal swellings were detected in or around the dentate nuclei.

5.4.4 Disease gene mapping identifies a 2.95Mb homozygous region on chromosome 15q25.2-q25.3

Assuming that autosomal recessive inheritance of a founder mutation was likely causative of nephrocerebellar syndrome, we isolated DNA from the blood/buccal samples of 7 affected individuals, 9 healthy siblings and their parents (Figure 43). A genome wide microarray scan was performed using Illumina Human CytoSNP-12v2.1 330K arrays. Homozygosity analysis identified a single autozygous region shared by all affected individuals of 2.95Mb on chromosome 15q25.2-q25.3, delimited by markers rs872332 and rs8027118, chr15:82320166-85268752, likely to correspond to the disease locus and refining the locus above defined by our collaborators. No other notable regions of autozygosity were observed giving an estimated LOD_{max} of >15. The putative disease locus was predicted to contain 30 genes, 2 of which are hypothetical, and 2 of which are predicted pseudogenes.

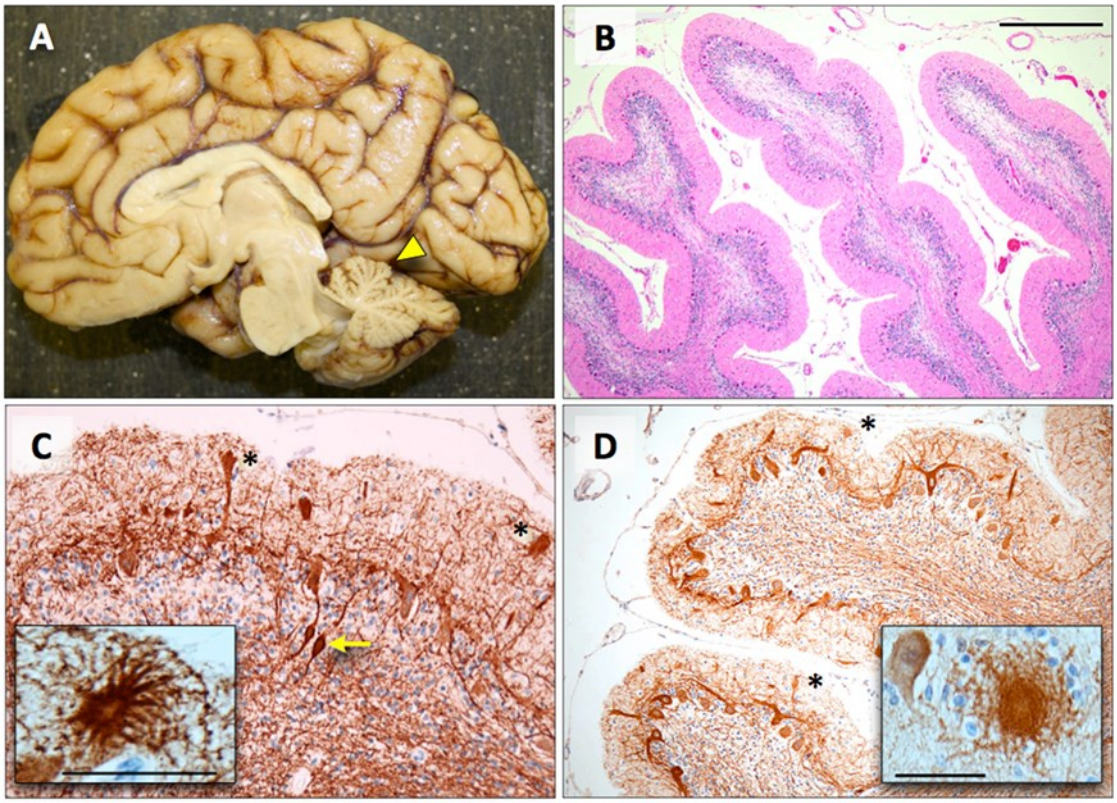


Figure 47: Nephrocerebellar syndrome neuropathology.

(A) Medial surface of the brain shows atrophy and sclerosis of cerebellar folia (yellow arrowhead), especially anteriorly. (**B**) Folia of the cerebellar cortex are short and stubby with severe loss of granule cells, crowded Purkinje neurons, and a hypercellular molecular layer (H&E, bar=500 µm). (**C**) Neurofilament stain shows disoriented Purkinje cell dendrites with irregular globose swellings, called asteroid bodies (inset, bar=100 µm), Purkinje torpedo axonal swellings (yellow arrow), and preservation of basket fibers with empty baskets. (**D**) Calbindin stain shows the fine structure of Purkinje neurons, which are decreased in number, crowded together, and surrounded by rich Bergman gliosis. The Purkinje cell layer is overlaid by a thin, hypercellular molecular layer composed of Bergman glia, some misplaced Neu-N positive granule cells, and rare Purkinje neurons. Dysmorphic Purkinje dendritic trees formed 'weeping willow' arrangements throughout the narrow molecular layer; abnormal, fine spikes were observed on the surface of some Purkinje cells within the molecular layer (asterisks; inset, bar=50 µm). The histological appearance suggests disruption of external granule layer development and survival.

5.4.5 Whole exome sequencing identifies two likely deleterious frameshift variants within the 15q25.2-q25.3 critical region

To identify the causative mutation, whole exome sequence analysis of a single affected individual was undertaken at the Wellcome Trust Sanger Institute, the sample generated 11.1Gb with a mean depth of 119. After appropriate filtering, two likely deleterious variants in two closely linked genes were identified within the 15q25.2-q25.3 critical region, both predicted to be novel homozygous frameshift mutations. The first was a seven base pair deletion (NM_001080435.2:c.1264_1270del) in exon 5 of the 10 exon *WHAMM* (WAS protein homolog associated with actin, golgi membranes and microtubules) gene. The proximal end of the *WHAMM* gene containing the putative mutation is highly repeated throughout the genome with multiple highly similar copies of the gene positioned throughout chromosome 15q, rendering sequence analysis problematical. The second mutation identified entailed a single base pair deletion (NM_032856.2:c.888del) in the final exon of *WDR73* (WD Repeat Domain 73) gene, which is not duplicated facilitating sequence analysis. Neither variant was listed in online sequencing project databases. Sanger sequencing confirmed the presence of the WDR73 frameshift variant (Figure 48A-B), which co-segregated perfectly with disease status, affected individuals were homozygous with respect to the mutation, all of the parents were carriers of the variant, and available unaffected siblings were either homozygous wild-type or heterozygous carriers (Appendix).

5.4.6 Optimising a strategy for confirming the *WHAMM* c.1264_1270del variant

The repeated nature of the WHAMM gene required the development of a sequencing strategy to confirm the presence of the c.1264_1270del putative mutation identified by exome sequencing. The Integrative Genomics Viewer (IGV, Broad Institute) plot for the variant did not indicate that the variant might be artifactual (Figure 49).

Given this, alignments were generated for *WHAMM* exon 5 containing the c.1264_1270del variant, and surrounding sequences. Primers were then manually positioned in regions in which at least a single nucleotide specific to the *WHAMM* gene was present.

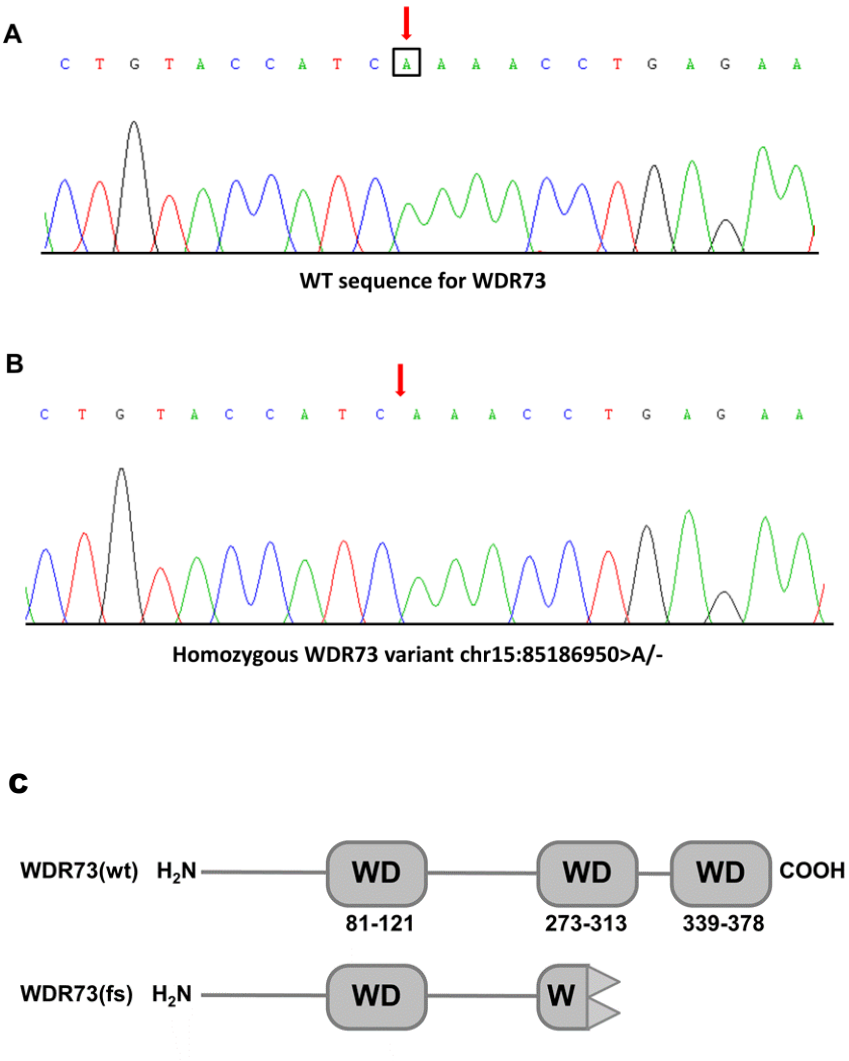


Figure 48: *WDR73* c.888del

(**A**) Electropherograms showing the *WDR73* c.888del sequence alteration in a WT control and (**B**) a homozygous affected individual. (**C**) Diagrams of the wild type (wt) WDR73 protein and frameshift p.Phe296LeuFs*26 (fs) variant are shown. The frameshift protein includes the N-terminal 295 amino acids followed by 26 aberrant residues, and lacks half of the second WD repeat domain and all of the third.

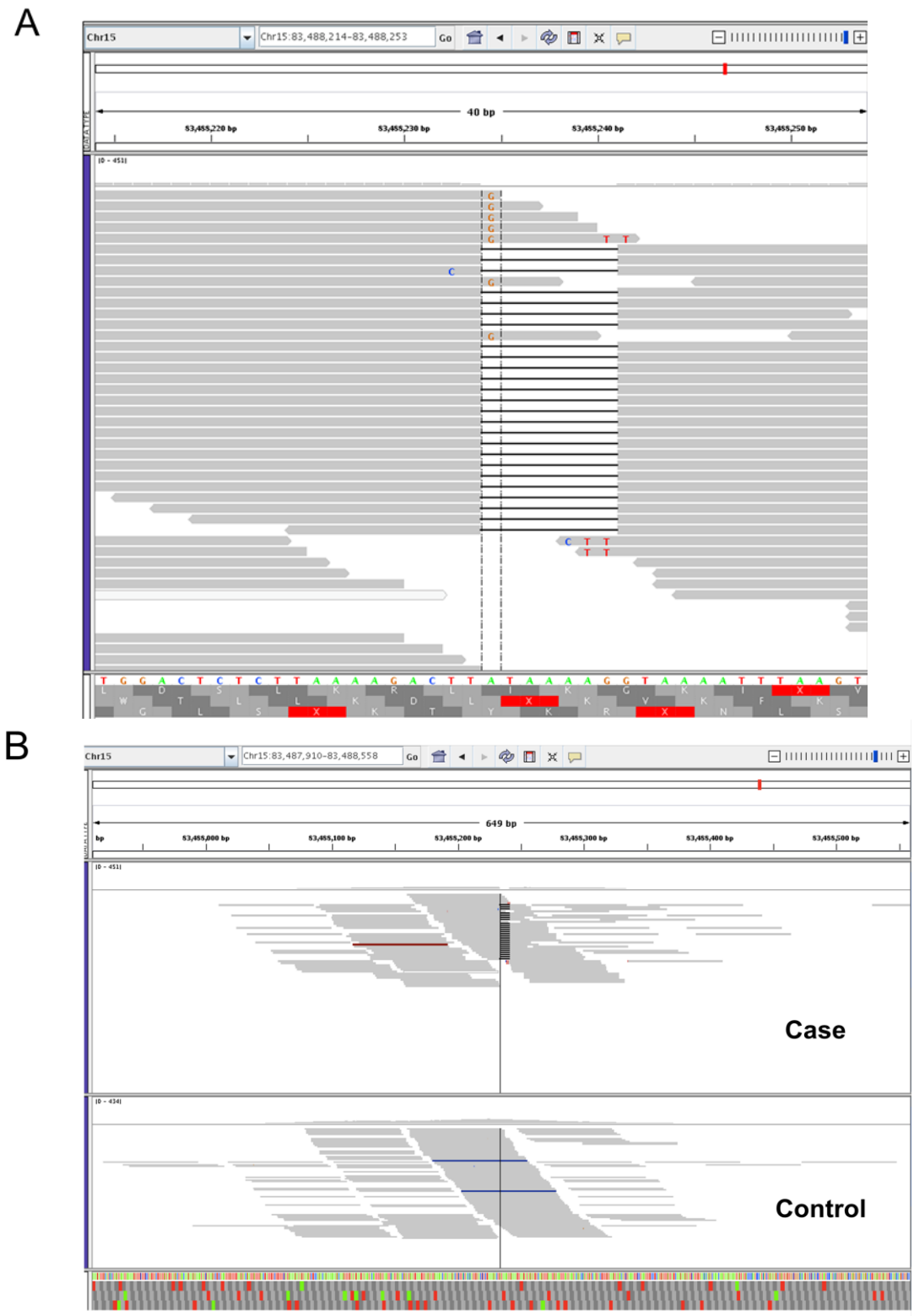


Figure 49: *WHAMM* Integrative Genomics Viewer plot

(**A**) Integrative Genomics (IGV) plot showing alignments of reads around the *WHAMM* c.1264_1270delATAAAAG variant called in an individual affected with nephrocerebellar syndrome. The variant is shown as present in all reads correctly aligned. (**B**) IGV plots of an affected individual and healthy unrelated control sequences are compared in parallel. The *WHAMM* variant is not seen in aligned reads derived from the control individual.

Using this approach PCR product corresponding specifically to the WHAMM gene was ultimately obtained, suitable for Sanger sequencing (Figure 50). This confirmed the presence of the variant which was found to cosegregate with the disease phenotype. All patients were thus confirmed to be doubly homozygous for both *WHAMM* and *WDR73* sequence variants, with all the parents carrying both variants and the available siblings either heterozygous or WT for both (Appendix). Sanger sequencing of 144 Holmes County, Ohio controls, identified 13 individuals who were heterozygous for both the *WDR73* and *WHAMM* variants, and a single individual who was heterozygous for *WDR73* while being wild type for *WHAMM*.

In parallel with our studies the Clinic for Special Children (CSC) and DDC clinic had SNP genotyped nephrocerebellar syndrome patients from Pennsylvania, Michigan and Ohio. A 3.96Mb homozygous region on 15q25.1-25.3 (chr15:81427307-85369577), was shared by all the affected individuals. The 2.95Mb locus identified by our studies fell within and refined this locus. Previously the CSC had performed sanger sequencing for the coding regions of 26 validated genes within the shared homozygous block including *WHAMM* and had identified no likely pathogenic variants. Exome sequencing and analysis of seven affected patients was undertaken by the CSC at the Broad Institute, USA, which identified only the *WDR73* variant. We shared our *WHAMM* sequencing strategy with the CSC, who confirmed the presence of this mutation in their cohort.

5.4.7 Nephrocerebellar syndrome is likely to comprise a composite phenotype arising from both *WHAMM* and *WDR73* mutations

The *WHAMM* c.1264_1270del variant was subsequently shown to cosegregate with the disease phenotype in a total of 29 patients. A three year old child proved to be the only exception. Given her Amish heritage and the lack of a unifying diagnosis, this child had been discussed with Dr Kevin Strauss, CSC. The neurological features described appeared consistent with nephrocerebellar syndrome and so DNA was sent to Dr Strauss for further investigation. The child was homozygous for the *WDR73* variant, but found to be only heterozygous for the *WHAMM* variant. Consistent with this, one parent was

heterozygous for both variants while the other was heterozygous for *WDR73* while being WT for *WHAMM.*

In light of the genetic findings described above, this child was subsequently reviewed at the CSC, at the age of three years her neurological features were entirely consistent with those of other nephrocerebellar syndrome affected children. She had notable microcephaly of -5 SDS, irritability, hypotonia, cortical visual impairment and a seizure disorder. Serial neuroimaging by MRI revealed cerebellar atrophy and degeneration. Significantly however, she had no evidence of renal disease. The age of onset of proteinuria and renal impairment is variable in nephrocerebellar syndrome (Figure 45A), thus a number of plausible possibilities remain; (i) The *WHAMM* variant underlies the renal phenotype of nephrocerebellar syndrome and *WDR73* the neurological component, (ii) *WDR73* is the sole cause of the nephrocerebellar syndrome phenotype and the child is yet to develop renal disease, (iii) the phenotype attributable to loss of function of WHAMM is unrecognisable beneath that of the more profoundly debilitating WDR73 phenotype or (iv) the child is compound heterozygous for *WHAMM.*

In order to investigate the fourth possibility further, the coding sequence and associated splice junctions of *WHAMM* was sequenced and primers checked against alignments of similar sequences to ensure unique amplification of *WHAMM* alone (primer sequences in Appendix). No potentially pathogenic second variant was identified. Amplification of the transcript from lymphocyte derived cDNA has proved problematical.

5.4.8 *In silico* modelling of the *WHAMM* and *WDR73* variants

Alamut and mutation taster insilico variant prediction software were used to predict the effects of the *WHAMM* and *WDR73* variants .The nature and position of the *WHAMM* variant lead to complexity in accurately predicting the effect. The canonical WHAMM gene encodes a “wild type” protein of 809 amino acids that consists of an N-terminal WHAMM membrane-interaction domain (WMD), a central coiled-coil (CC) domain that binds microtubules, and a

```
BC1117556       TTGTAGATGAATTAGAAATGCAATTTTATGAAATTCAGTTAGAACTATATGAAGTTAAAT 5381
segment4        TTGTAGATGAATTAGAAATGCAATTTTATGAAATTCAGTTAGAACTATATGAAGTTAAAT 5373
whamml2         TTGTAGATGAATTAGAAATACAATTTTATGAAATTCAGTTAGAACTATATGAAGTTAAAT 5590
whamml1         TTGTAGATGAATTAGCAATACAATTTTATGAAATTCAGTTAGAACTATATGAAGTTAAAT 5512
WHAMM           TTGTAGATGAATTAGAAATACAATTTTATGAAATTCAATTAGAACTATATGAAGTTAAAT 5110
                *************** *** ***************** ********************

BC1117556       TTGAGATATTAAAAAACGAAGAAATACTGCTTACTACACAGTTGGACTCTCTTAAAAGAC 5441
segment4        TTGAGATATTAAAAAACGAAGAAATACTGCTTACTACACAGTTGGACTCTCTTAAAAGAC 5433
whamml2         TTGAGATATTAAAAAACAAAGAAATACTGCTTACTACACAGTTGGACTCTCTTGAAAGAC 5650
whamml1         TTGAGATATTAAAAAACAAAGAAATACTGCTTACTACACAGTTGGACTCTCTTGAAAGAC 5572
WHAMM           TTGAGATATTAAAAAACGAAGAAATACTGCTTACTACACAGTTGGACTCTCTTAAAAGAC 5170
                ***************** ********************************** ******

BC1117556       TTATAAAAGGTAAAGTTTATATTTAAGTATATAGATAACAATGTTTA-----TAAATTTA 5496
segment4        TTATAAAAAGTAAAGTTTATATTTAAGTATATAGATTACAATGTTTA-----TAAATTTA 5488
whamml2         TTATAAAAGGTAAAGTTTTTATTTAAGTATATAGATTACAATGTTTA-----TAAATTTA 5705
whamml1         TTATAAAAGGTAAAGTTTTTATTTAAGTATATAGATTACAATGTTTA-----TAAATTTA 5627
WHAMM           TTATAAAAAGGTAAA------ATTTAAGTATATAGATTGCAATGTTTAAATTATAAATTTA 5224
                ******** *****      ***************  *********     ********

BC1117556       AGGAAATACAGACCATATTATCAATTACTTTTTGTAAACTGTAACATCTGAAAATTTCCT 5556
segment4        AGGAAATACAGACCATATTATCAATTACTTTTTGTAAACTGTAACATCTGAAAATTTCCT 5548
whamml2         AGGAAATACAGACCATATTATCAATTACTTTTTGTAAACTGTAACATCTGAAAATTTCCT 5765
whamml1         AGGAAATACAGACCATATTATCAATTACTTTTTGTAAACTGTAACATCTGAAAATTTCCT 5687
WHAMM           AGGAAATACAGACCATATTATCAATTACTTTTTGTAAATTATAACATCTGAAAATTTCCT 5284
                ************************************ * ******************

BC1117556       AAAGTTTTCCTTCAGTAGTTTATTATTCAAATAATATATTCATTGTTAGCACATAGCAAA 5616
segment4        AAAGTTTTCCTTCAGTAGTTTATTATTCAAATAATATATTCATTGTTAGCACATAGCAAA 5608
whamml2         AAAGTTTTCCTTCAGTGGTTTATTATTCAAATAATATATTCATTGTTAGCACATAGCAAA 5825
whamml1         AAAGTTTTCCTTCAGTGGTTTATTATTCAAATAATATATTCATTGTTAGCACATAGCAAA 5747
WHAMM           AAAGTTTTCCTTCAGTGGTTTATTATTCAAATAATATATTCATTGTTAACACATAGCAAA 5344
                *************** ******************************** ***********
```

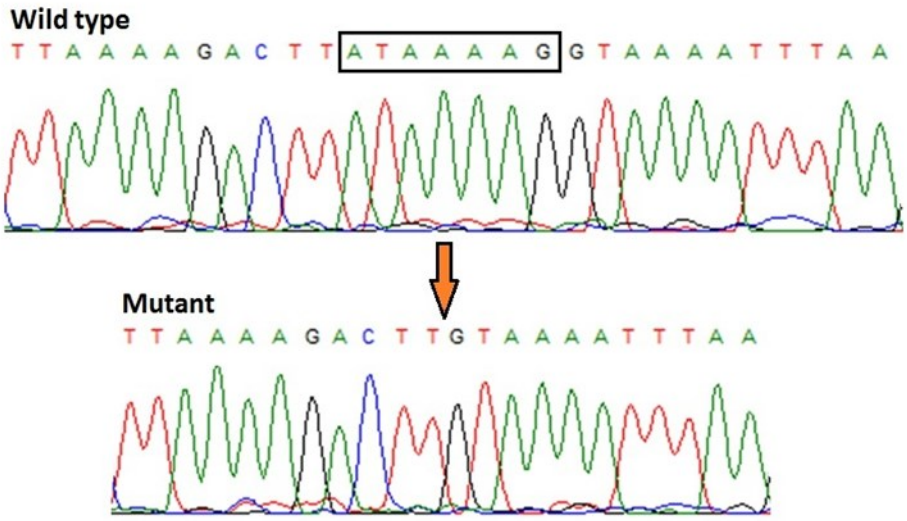


Figure 50 *WHAMM* c.1264_1270del variant

Alignment of WHAMM genomic sequence surrounding the c.1264_1270del variant (highlighted in red) is shown above. Primers sequences are highlighted in yellow with bases specific to the WHAMM sequence in bold and underlined. Electropherograms obtained using these primers show the DNA sequence at the position of the WHAMM variant in a WT control and a homozygous individual affected with nephrocerebellar syndrome.

C-terminal polyproline-WH2-connector-acidic domain that interacts with actin and Arp2/3 to promote actin nucleation (WHAMM WT;Figure 51A) (Campellone et al., 2008). The 7 nucleotide (ATAAAAG) deletion is predicted to potentially result in a translational frameshift, leading to premature termination. This *WHAMM* variant encodes a protein predicted to contain the first 421 amino acids of WHAMM followed by 34 aberrant residues after the frameshift mutation (WHAMM c.1264_1270del; WHAMM p.Ile422Lysfs*35 Figure 50A). Thus, any polypeptide produced from the mutated WHAMM allele would be predicted to retain its entire membrane-binding domain, half of its microtubule-binding domain, and completely lack the domain responsible for actin nucleation. The deleted bases c.1264_1270 of *WHAMM* comprise the last 7 base pairs of exon 5 of the gene. The last three bases of an exon form part of the donor splice junction consensus sequence. While the position of the variant would predict that the donor splice site would be disrupted, it is difficult to be certain what the specific outcome of this may be. A suitable cryptic splice donor site could be used, and exon loss or retention are other possibilities, with nonsense mediated mRNA decay seeming overall to be the most probable outcome resulting from this variant. It remains a possibility that any splicing defect would be incomplete and that some transcripts would be spliced normally resulting in production of the mutant protein.

Little is known about WDR73 and the protein function remains essentially uncharacterised. The WDR gene family encodes a series of proteins with WD (Trp-Asp) repeat domains that fold to form a tertiary structural beta-propeller conformation, which consist of 4 to 8 blade-like structures arranged in a circle. These act as a stable platform for the transient interplay among other proteins (van Nocker and Ludwig, 2003) and the assembly and disassembly of large protein complexes (Haack et al., 2012, Hayflick et al., 2013). WDR proteins participate in an array of cellular functions, including cell division, autophagy, signal transduction, vesicle trafficking, and cystoskeletal dynamics (van Nocker and Ludwig, 2003, Cohen-Katsenelson et al., 2011, Gregory and Hayflick, 1993, Gross et al., 2001, Saitsu et al., 2013). A number of WDR proteins are expressed in mammalian brain (Cohen-Katsenelson et al., 2011, Traka et al., 2013, Castets et al., 1996) and recently, mutations of *WDR45* (Haack et al., 2012, Saitsu et al., 2013) *WDR62* (Bilguvar et al., 2010, Memon et al., 2013),

and *WDR81* (Gulsuner et al., 2011, Tan, 2006), have been associated with severe neurodevelopmental disability in humans. The single nucleotide deletion (T) is within the last exon of the gene and is expected to result in a translational frameshift leading to premature termination, the *WDR73* c.888del variant encodes a protein which is predicted to contain the first 295 aminoacids of the protein followed by 27 aberrant residues after the frameshift (Figure 48C).The position of the variant within the last exon of the gene, may indicate that mutant transcript could escape nonsense mediated mRNA decay surveillance mechanisms.

To confirm whether WHAMM and WDR73 protein expression are altered by the mutations, skin fibroblasts were collected and cultured from two individuals with nephrocerebellar syndrome, both of whom were homozygous for both the *WHAMM* and *WDR73* variants, as well as from healthy individuals in the same community. In contrast to the two healthy controls, the fibroblasts from the two nephrocerebellar syndrome affected individuals were difficult to maintain in culture growing very slowly, halting proliferation after only a few cell divisions. Thus far, we have been able to maintain cultures from only one affected individual for several months. Similarly lymphoblastoid cell lines were established from nephrocerebellar syndrome affected individuals, heterozygous carriers of both the *WHAMM* and *WDR73* variants and related wild type individuals. No significant difference in growth rates between lymphoblastoid cell lines was observed.

Investigation of the functional consequences of the WDR73 variant are primarily being undertaken by the CSC and the results of these studies are not yet available. However functional assessments of the WHAMM mutation were undertaken in collaboration with Dr Ken Campellone.

5.4.9 Nephrocerebellar syndrome fibroblasts express a truncated WHAMM protein

WHAMM expression was examined by immunoblotting using extracts from nephrocerebellar syndrome primary fibroblasts and a rabbit polyclonal antibody raised to the whole protein, but preferentially recognising the C-terminal WCA domain of WHAMM previously generated by Dr Campellone

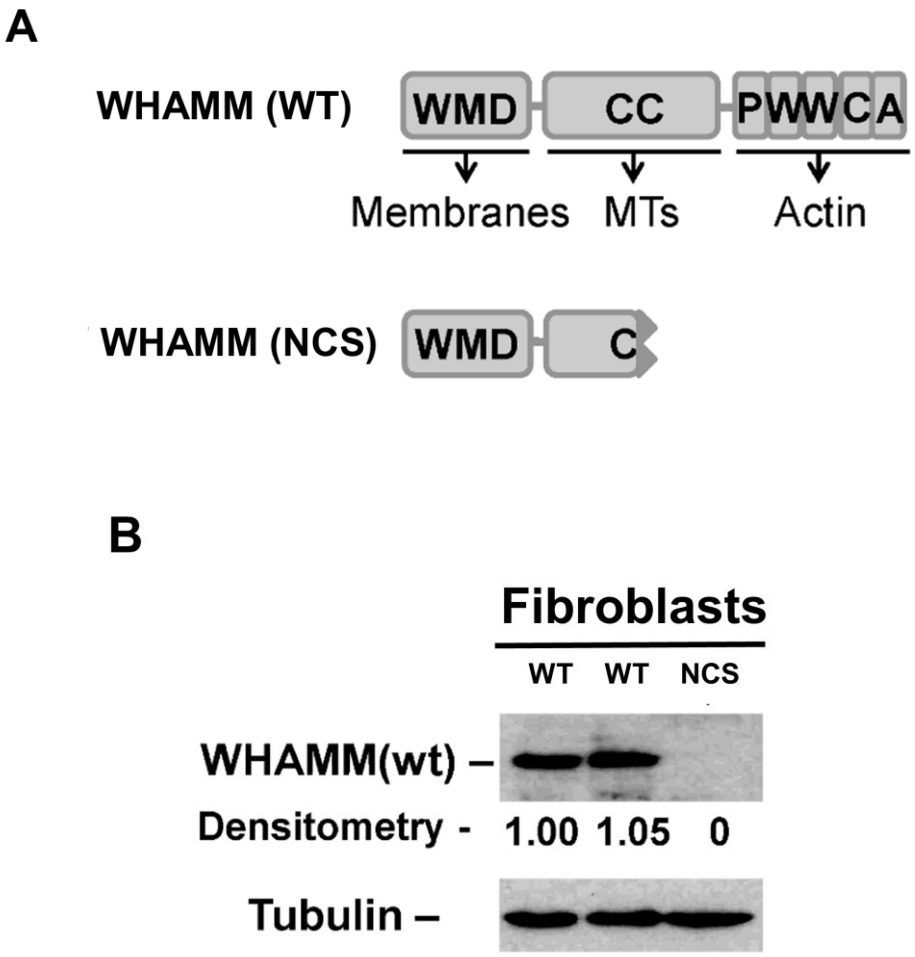


Figure 51: Nephrocerebellar syndrome fibroblasts do not express full length WHAMM protein

(**A**) Diagrams of the wild type (WT) WHAMM protein and frameshift p.Ile422Lysfs*35 (fs) variant. The 809-residue WT protein includes an N-terminal WHAMM membrane-interacting domain (WMD) that binds phospholipids, a central coiled-coil (CC) domain that binds microtubules, and a C-terminal polyproline-WH2-WH2-connector-acidic (PWWCA) segment promotes actin nucleation. The frameshift variant includes the N-terminal 421 residues of WHAMM followed by 34 aberrant residues after the frameshift mutation, and lacks half of the CC domain and all of the PWWCA segment. (**B**) Primary skin fibroblasts from WT, or homozygous mutant (NCS) individuals were subjected to immunoblotting with anti-WHAMM or anti-tubulin antibodies. Relative levels of the WT WHAMM protein were determined by densitometry. Data in B is courtesy of Dr Ken Campellone, University of Connecticut.

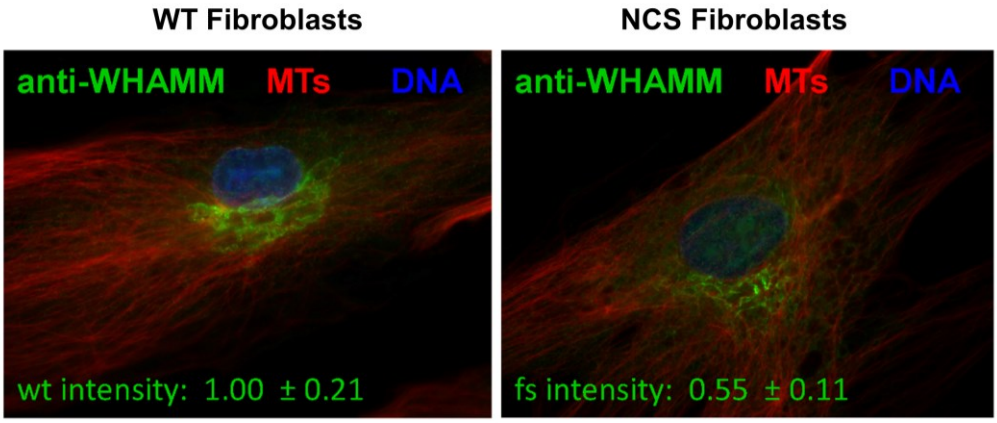


Figure 52: Nephrocerebellar syndrome fibroblast cells express a truncated WHAMM variant with an altered localisation pattern

WT or nephrocerebellar syndrome (NCS) fibroblasts were fixed and stained with antibodies to visualise WHAMM (green) or microtubules (red), and with DAPI to label DNA (blue). In contrast to the compact perinuclear WHAMM staining of in WT cells, the nephrocerebellar syndrome cells exhibited a more reticular WHAMM localisation pattern. The mean fluorescence intensity of WHAMM staining was quantified and found to be lower in nephrocerebellar syndrome cells. Data shown is courtesy of Dr Ken Campellone.

(Campellone et al., 2008). As expected, full-length WHAMM was found not to be expressed in fibroblasts (Figure 51B). However, utilising a guinea pig monoclonal antibody raised against the coil-coil domain of WHAMM detected a WHAMM protein with an abnormal and unusual intracellular localisation pattern in primary fibroblast cell lines from nephrocerebellar syndrome affected individuals (Figure 52).

5.4.10 Nephrocerebellar syndrome fibroblasts have reduced F-actin staining and altered organelle morphology

Among WASP-family proteins, WHAMM is the only molecule whose main function is known to lie in remodelling membranes of the secretory system as RNA interference approaches have demonstrated that WHAMM organises the Golgi apparatus near the centrosome promoting tubulation of endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC) membranes (Campellone et al., 2008). WHAMM is also the only WASP-family member known to physically interact with microtubules (Shen et al., 2012), and its membrane tubulation function has been shown to require both its microtubule-binding and actin nucleation activities (Campellone et al., 2008). To determine if nephrocerebellar syndrome fibroblasts have altered cytoskeletal or membrane organisation, we stained cultured fibroblasts with probes to visualise the microtubule and actin cytoskeletons and with antibodies to organelles of the traditional secretory pathway, including the ER (GRP94), ER-exit sites (Sec16), ERGIC (ERGIC-53), cis-Golgi (GM130), and trans-Golgi (TGN46). Nephrocerebellar syndrome fibroblasts were found not to contain any gross alterations in microtubule organisation or in ER morphology. However, nephrocerebellar syndrome fibroblasts were shown to contain fewer punctate F-actin structures in their cytoplasm when compared with control cells, unusual orthogonal bundles of stress fibres and F-Actin aggregates were also seen (Figure 53). Given that the nephrocerebellar syndrome affected individual from whom these cells originate is homozygous for both the *WHAMM* and *WDR73* sequence variants, it is not possible to attribute these intracellular abnormalities to either (or both) mutations. However, these observations would be consistent with a lack of WHAMM actin nucleation function in cells, moreover, again in accordance with a known role for WHAMM in organising membranes of the intermediate compartment and Golgi apparatus, the ER-Golgi Intermediate

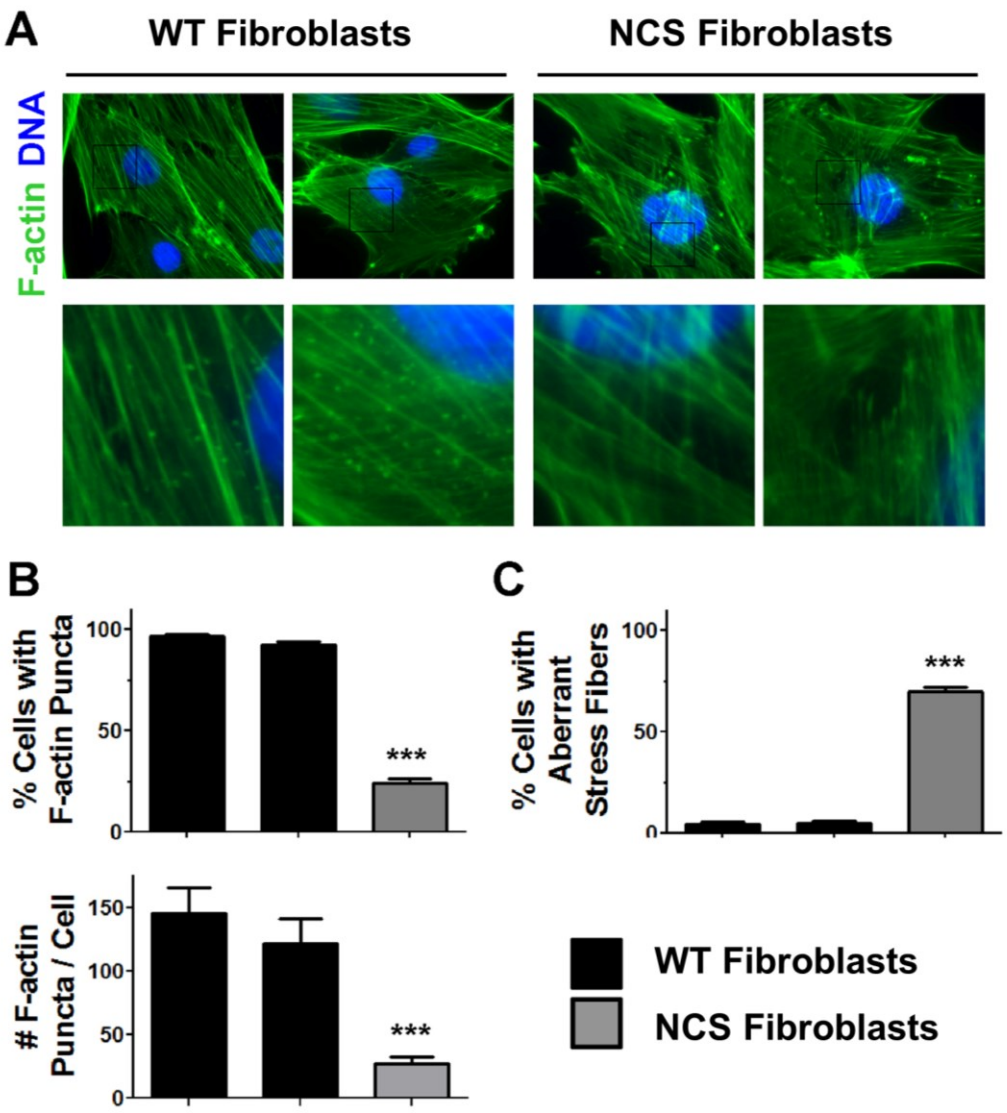


Figure 53: NCS fibroblasts have abnormalities in their actin cytoskeleton

(**A**) WT or nephrocerebellar syndrome (NCS) fibroblasts were fixed and stained with phalloidin to visualise F-actin (green) and DAPI to label DNA (blue). NCS cells contained unusual orthogonal bundles of stress fibers and F-actin aggregates. A lack cytoplasmic F-actin puncta in NCS cells is highlighted in the magnified images. (**B**) The % of cells with cytoplasmic F-actin puncta in the perinuclear region was quantified. Each bar represents the mean (+/- SEM) from analyses of 350 cells. The number of F-actin puncta per cell was also quantified. Each bar represents the mean (+/- SEM) from analyses of 11 cells. ***, $p<0.001$. (**C**) The % of cells with aberrant (orthogonal or aggregated) F-actin stress fibers was quantified. Each bar represents the mean (+/- SEM) from analyses of 350 cells. ***, $p<0.001$. Data shown is courtesy of Dr Ken Campellone.

compartment (ERCIG), cis-Golgi, and trans-Golgi were larger in nephrocerebellar syndrome fibroblasts when compared to control fibroblasts (Figure 54).

5.4.11 Nephrocerebellar syndrome fibroblasts exhibit defective autophagosome biogenesis

Other aspects of membrane remodelling beyond secretory organelle morphology might be also altered in nephrocerebellar syndrome fibroblasts. One such aspect of cell function that requires dramatic changes in membrane shape is macroautophagy, the endogenous degradation of the intracellular components by the cell (hereafter called autophagy). Since the ER is one major source of autophagosomal membranes (Hayashi-Nishino et al., 2009, Zoppino et al., 2010), patient fibroblasts were investigated to determine whether they contained normal autophagosomes by staining them with an antibody to LC3A, a general marker of these organelles. Interestingly, while fibroblasts from healthy individuals contained a consistent basal of LC3A-positive structures and circular autophagosomes under normal growth conditions, nephrocerebellar syndrome fibroblasts were found to be devoid of LC3A-associated membrane structures (Figure 55).

To test whether fibroblasts from nephrocerebellar syndrome affected individuals are able to form autophagosomes under conditions in which autophagic flux was disrupted, cells were treated with chloroquine, a drug which increases lysosomal pH and causes autophagosomes to accumulate in the cytoplasm. Strikingly, while control fibroblasts exhibited a buildup of LC3A-positive autophagosomes under these conditions (Figure 55), chloroquine treatment caused a significant amount of detachment (and presumably death) of nephrocerebellar syndrome fibroblasts. The remaining adherent cells showed a disorganised and diffuse LC3A staining pattern and notably their cytoplasm was filled with vacuoles (Figure 55). Importantly, re-introduction of WT WHAMM rescues the autophagy defects (Figure 56). Collectively, these results indicate that WHAMM plays a previously-unrecognised role in cell survival and autophagy, and that an essential role of WHAMM in human development is likely.

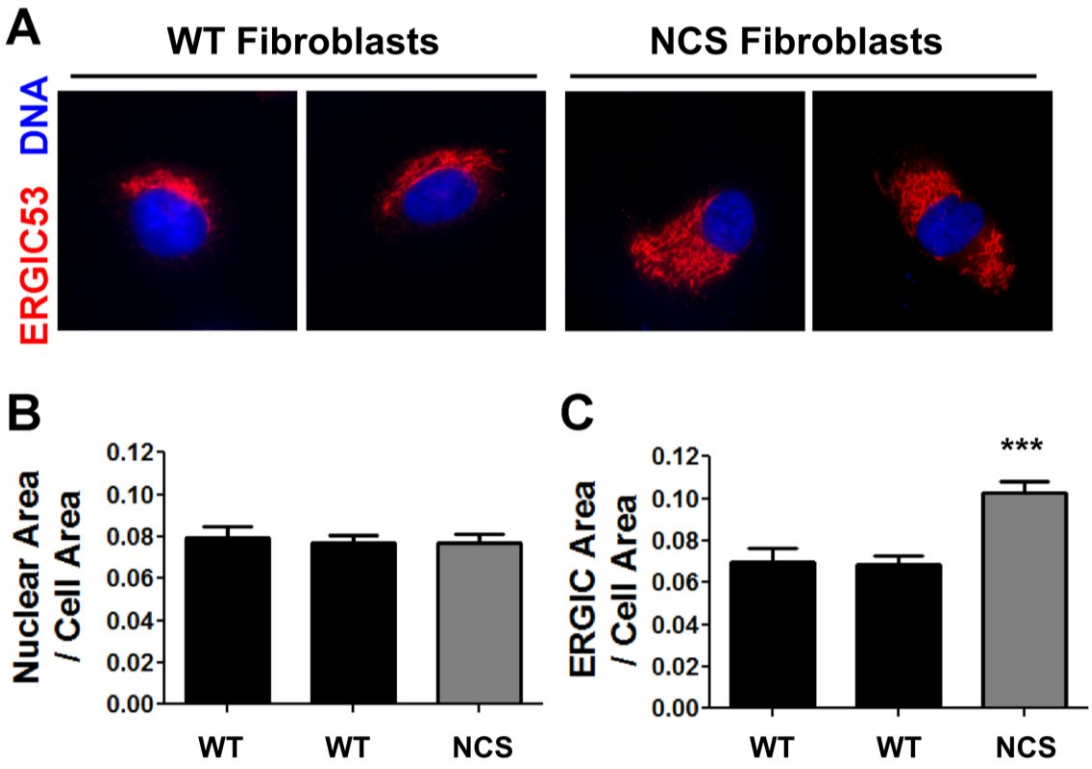


Figure 54: Nephrocerebellar syndrome fibroblasts have abnormalities in their ER-Golgi intermediate compartment (ERGIC).

(**A**) WT or nephrocerebellar syndrome (NCS) fibroblasts were fixed and stained with antibodies to detect ERGIC-53 (red) and with DAPI to label DNA (blue). The ERGIC appeared to be enlarged, elongated, or more dispersed in NCS cells. A similar phenotype was observed when cells were stained with antibodies to GM130, a marker of the cis-Golgi (data not shown). (**B-C**). The area occupied by the entire cell (determined by F-actin staining), the ERGIC (detected using antibodies to ERGIC-53), and the nucleus (identified by DAPI staining) were calculated using ImageJ and converted to a ratio of organelle area/cell area. Each bar represents the mean (+/- SEM) from analyses of 25 cells. ***, $p<0.001$. Data shown is courtesy of Dr Ken Campellone.

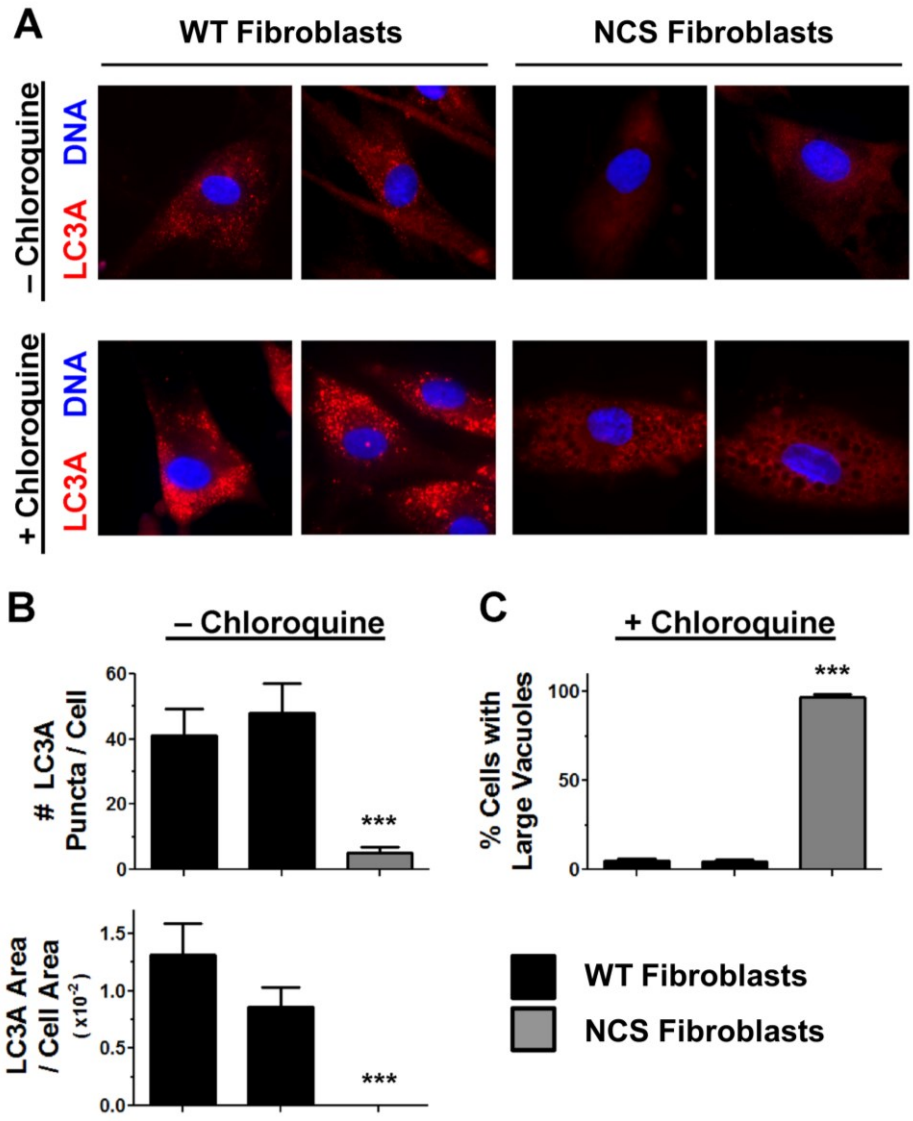


Figure 55: Nephrocerebellar syndrome fibroblasts have severe defects in autophagosome biogenesis

(**A**) WT or Nephrocerebellar syndrome (NCS) fibroblasts were grown in the absence or presence of chloroquine, fixed, and stained with antibodies to detect LC3A (red) and with DAPI to label DNA (blue). Compared to WT cells, NCS cells were virtually devoid of LC3A-positive (pre-) autophagosomal structures at steady-state (top row). When cells were grown in the presence of chloroquine to block autophagosome flux (bottom row), WT fibroblasts accumulated large numbers of mature autophagosomes (as expected), whereas NCS fibroblasts became filled with large vacuoles that did not stain for autophagosomal or lysosomal markers. (**B**) The number of LC3A puncta per cell, and the relative area within each cell that contained autophagosomes was quantified. The areas occupied by the entire cell (determined by F-actin staining) and autophagosomes (identified by LC3A staining) were calculated using ImageJ and converted to a ratio. Each bar represents the mean (+/- SEM) from analyses of 11 cells. ***, $p<0.001$. (**C**). The % of cells with large cytoplasmic vacuoles was quantified. Each bar represents the mean (+/- SEM) from analyses of 250 cells. ***, $p<0.001$. Data shown is courtesy of Dr Ken Campellone.

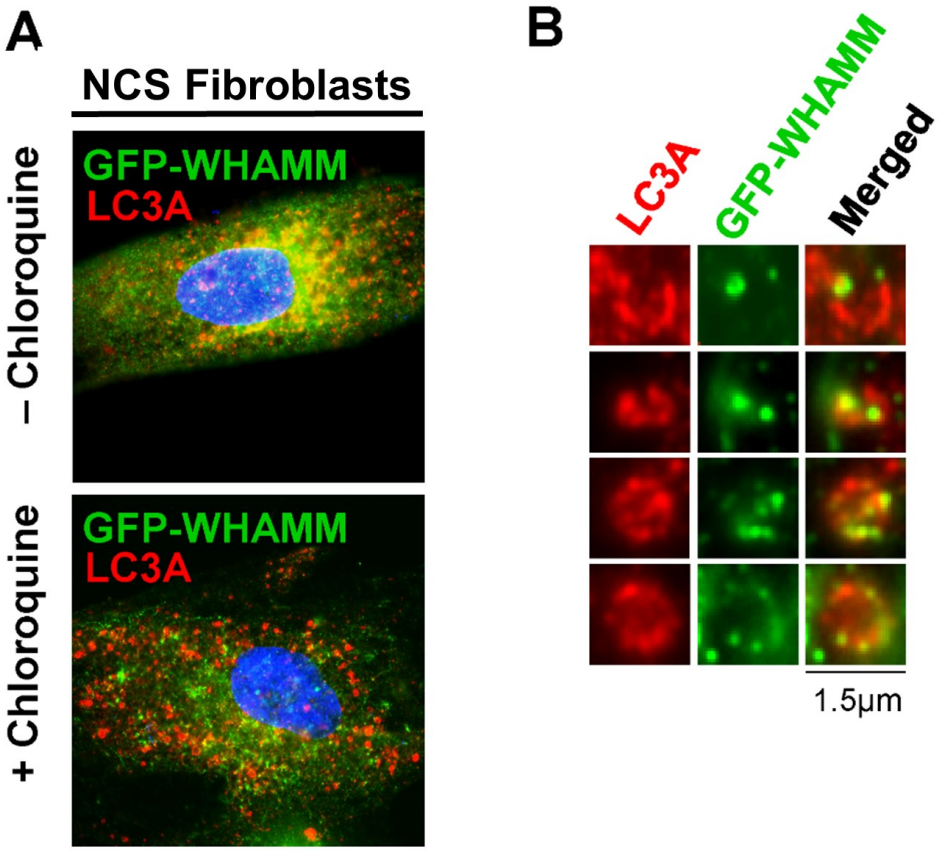


Figure 56: Re-introduction of wild type WHAMM into nephrocerebellar syndrome fibroblasts rescues the defects in autophagy

(**A**) Nephrocerebellar syndrome (NCS) fibroblasts were grown in the absence or presence of chloroquine, transfected with plasmids encoding GFP-WHAMM, fixed, and stained with antibodies to detect LC3A (red) and with DAPI to label DNA (blue). GFP-WHAMM localised to cytoplasmic membranes and rescued the defects in autophagosome biogenesis. (**B**) NCS fibroblasts were transfected with a plasmid encoding GFP-WHAMM, fixed, and stained as in part A. Individual autophagosome structures are highlighted in these composites. GFP-WHAMM localised to regions of autophagosomal membrane growth or curvature due to its ability to remodel membranes during a variety of intracellular trafficking processes. Data shown is courtesy of Dr Ken Campellone.

5.5 Discussion

A large number of interlinking families from the Swartzentruber Amish community in Ohio were investigated. Multiple individuals aged 9 months-12 years were found to be affected by a novel complex disorder associated with abnormal brain development and progressive renal failure that we termed nephrocerebellar syndrome. The cardinal features of this condition include progressive microcephaly of post-natal onset, profound global developmental delay, irritability, central hypotonia, episodic dystonic movements, roving nystagmus, cortical visual impairment, predisposition to develop multifocal seizures, poorly formed caries prone dentition and nephrotic syndrome. In parallel with our studies, others had independently identified the same condition amongst Amish communities throughout the USA. A collaborative approach was taken to identify the underlying molecular cause of this disorder. Collectively our molecular studies determined that two distinct founder mutations affecting two closely linked genes (*WHAMM* and *WDR73*) both of which have become entrapped within the community and act in autosomal recessive fashion, cosegregate with the disease phenotype in all but a single individual. However, this child lacks the renal aspect of the phenotype, leading to the intriguing possibility that this is a composite phenotype and that the *WHAMM* variant might underlie the nephrotic syndrome in this condition and the *WDR73* variant the neurological component. Although, as shown in Figure 45A the onset of renal disease in the other nephrocerebellar syndrome affected children shows significant variability, and lack of proteinuria or evidence of renal impairment at age 3 years, does not preclude later development of these.

The analysis of the cellular functions in nephrocerebellar syndrome fibroblasts revealed a defect in autophagy which was rescued by ectopic expression of WHAMM protein. Autophagy is an essential catabolic process in which organelles and soluble and aggregated cellular components are enveloped in double membrane structures called autophagosomes which eventually fuse with lysosomes, leading to the degradation and recycling of the autophagosome contents (Settembre and Ballabio, 2014). It is difficult to see how such a fundamental cellular defect could not result in a clinical outcome. However, little is known about the level of redundancy of this process cells can tolerate and it

is plausible that there would be no detectable adverse consequences of this defect.

The actin cytoskeleton controls membrane organisation, remodelling, and trafficking. To polymerise actin monomers into filaments, human cells rely on the activities of molecules called nucleators. Several classes of nucleator proteins promote actin assembly into linear unbranched filaments, including 15 formins (Chesarone et al., 2010) and multiple monomer clustering proteins (Dominguez, 2009). But only one nucleator, the Arp2/3 complex, is known to create branched filament networks (Campellone and Welch, 2010). At the molecular level, the Arp2/3 complex nucleates actin while bound to existing filaments and engaged by proteins called nucleation factors from the WASP family [Rottner, 2010]. The founding member, WASP, gained its name because the gene that encodes it is mutated in patients with Wiskott-Aldrich Syndrome, an X-linked immune disorder (Derry et al., 1994). WHAMM is a member of the WASP-family of proteins. These 8 WASP-family nucleation factors are each thought to assemble actin during a distinct cellular function. While WASP controls actin dynamics and signalling in cells of the immune system, N-WASP remodels the plasma membrane during receptor-mediated endocytosis [Benesch, 2005; Bu, 2009; Innocenti, 2005] and some types of protrusion [Yamaguchi, 2005; Legg, 2007]. The WAVE proteins also drive extensions of the plasma membrane, but in the lamellipodia and ruffles that accompany cell migration (Innocenti et al., 2005, Suetsugu et al., 2003, Ide et al., 2003). Similarly, JMY can be recruited to the plasma membrane to enhance epithelial cell motility (Zuchero et al., 2009). In contrast, WASH localises to several endosome populations where it controls membrane shape and receptor trafficking (Derivery et al., 2009, Duleh and Welch, 2010, Gomez and Billadeau, 2009, Zech et al., 2011). Finally, WHAMM is found at organelles of the secretory pathway where it promotes membrane tubulation and transport (Campellone et al., 2008, Shen et al., 2012).

WASP is expressed only in hematopoietic cells, whereas the subsequently described family members N-WASP (Miki et al., 1996), WAVE1-3 (Miki et al., 1998), WASH (Linardopoulou et al., 2007), WHAMM (Campellone et al., 2008), and JMY (Zuchero et al., 2009) are expressed broadly, but highest in neuronal

tissue. Due to their central role in important cellular processes, this protein family are also thought play a key role in mammalian development. Genetic studies have revealed that WASP knockout mice (Snapper et al., 1998) share many features with WASP-deficient humans, including defective T-cell signalling (Thrasher, 2009). Moreover, targeted disruptions of the murine genes encoding N-WASP (Lommel et al., 2001, Snapper et al., 2001), WAVE2 (Yamazaki et al., 2003, Yan et al., 2003), and WASH (Gomez et al., 2012) are embryonic lethal, and WAVE1-null mice have reduced viability (Dahl et al., 2003). These ubiquitously-expressed family members appear to be essential for neurological and cardiovascular development. However, it remains to be determined if the other newly-discovered nucleation factors also contribute to neurological and cardiovascular function. Surprisingly, aside from a variety of human mutations in WASP that result in Wiskott-Aldrich syndrome, thrombocytopenia, and neutropenia, mutations in other WASP-family members have not been implicated in human disease.

The genetic evidence for WDR73 involvement in the neurological component of nephrocerebellar syndrome is strong and suggestive of a fundamental role for this molecule in the survival and activity of cortical neurons and development of the cerebellar external granule layer. In the absence of additional mutations in *WDR73* associated with a similar phenotype to nephrocerebellar syndrome, functional evaluation of the effects of the Amish founder mutation in *WDR73* are important to conclusively establish pathogenicity of the variant. It also remains to be seen whether it is simply a lack of WT protein, or loss of function of the mutant protein, that causes the neurological phenotype of nephrocerebellar syndrome and neuropathological abnormalities observed. Aggregates of abnormally folded protein might also be a contributing factor, as has been suggested for a number of other neurological diseases, such as Alzheimer's disease, Huntington's disease and Amyotrophic lateral sclerosis (Kesidou et al., 2013). Much is left to learn about the largely uncharacterised WDR73 protein and the potential role played in human development. The human brain transcriptome (http://hbatlas.org) shows globally high levels of *WDR73* gene expression in the brain with the highest levels early in embryonic development but persisting throughout adult life. The Amish nephrocerebellar syndrome provides a unique and unparalleled opportunity to study this critically important

molecule. Protein localisation, expression and interaction studies currently being undertaken by the CSC should give some indication of likely protein function.

The Amish nephrocerebellar syndrome study also highlights one of the difficulties associated with the investigation of inherited diseases which occur within founder populations such as the Amish, and the importance of sequencing all of the coding regions within a putative disease locus. Prior to the advent of next generation sequencing technologies, the identification of a homozygous frameshift variant cosegregating with the phenotype, absent in the homozygous state in regional controls and predicted to significantly impair protein function such as is the case for both the *WHAMM* and *WDR73* alterations described here, was commonly considered acceptable evidence of pathogenicity. Therefore it is probable that a number of gene variants previously thought to be causal of autosomal recessive conditions (identified through mapping studies in endogamous communities and consanguineous unions) will turn out to be innocent passengers in linkage disequilibrium with an unidentified causative variant located elsewhere in the disease locus.

The FORGE Canada project, a large scale collaborative whole exome sequencing project seeking to identify the cause of rare paediatric onset single gene disorders, highlighted the potential confusion that composite phenotypes can cause in such studies. One of the conditions investigated by this group was the eponymously named Fitzsimmons syndrome (OMIM 270710), associated with a combination of spastic parapalegia and brachydactyly. All available cases of Fitzsimmons syndrome that were analysed were found to be composite phenotypes, a combination of two or more rare genetic disorders inherited in parallel (Beaulieu et al., 2014). This also seems to be the most plausible explanation for the Amish nephrocerebellar syndrome described in this chapter.

Regardless of the role of WDR73 in nephrocerebellar syndrome, in the light of the results of our cellular studies it is impossible to discount any involvement of the *WHAMM* mutation in this condition, which in the very least is likely to have a role in modifying the phenotypical outcome of the more severe *WDR73*

mutation. Taken together with the genetic data and well established role of WHAMM in cellular function these results are highly suggestive of WHAMM involvement in the nephrocerebellar syndrome phenotype. Reduced F-actin staining and altered organelle morphology which would both be compatible with absence of WHAMM induced actin nucleation, was observed in cells from individuals homozygous for both the *WDR73* and *WHAMM* variants. Additionally, our studies revealed previously-unrecognised roles for WHAMM in autophagy, a fundamental intracellular process. Also as of yet, no individual has been identified that is homozygous for the Amish *WHAMM* variant alone and it is difficult to predict the nature of the phenotype that might be expected and hence what patient population to look in.

Finally, further understanding of the basis of the Amish nephrocerebellar syndrome will shed important new light on the function of the nephron and development of the cerebellum, potentially leading to the identification of novel therapeutic targets that might benefit those affected by this devastating disorder.

6

CHAPTER SIX

FINAL DISCUSSION AND FUTURE WORK

6 FINAL DISCUSSION AND FUTURE WORK

In this thesis I describe three new disorders affecting brain growth and development identified amongst the Ohio Amish community. Individuals recruited to the Windows of Hope Amish inherited disease developmental delay and intellectual disability subproject underwent detailed clinical phenotyping, leading to the recognition and clinical delineation of novel syndromes. Mapping, traditional and next generation sequencing technologies were used to identify the genes and mutations responsible for these conditions. The causative mutations were verified by cosegregation and regional/non-regional (non-Amish) control analysis. Following identification of the causative genetic variants, tissue distribution analysis, protein localisation studies and pathway analysis were used to investigate the molecular processes perturbed by these aberrations. The functional biology techniques used included western blotting for assessment of protein levels and stability, fluorescence activated cell sorting, surface plasmon resonance, immunoprecipitation and *in vitro* assays to assess protein interactions. Rescue of abnormal cellular phenotypes by endogenous expression of the native protein was used to provide definitive proof of involvement of the defective molecule. The results of this work underscore the power of new genotyping technologies combined with population-specific genetic knowledge and robust clinical phenotyping.

This section provides an overview of the main findings in each chapter and relates these to the current literature. Further areas of future research are also highlighted.

6.1 Chapter 3: Hypomorphic PCNA mutation underlies a novel human DNA repair disorder

The identification of an Amish founder mutation in PCNA associated with a new neurodegenerative phenotype described in chapter 3 has provided important new insights into DNA repair and replication processes, including the definition of a novel human disease that may arise due to abnormalities in these cellular functions.

Orchestrating the multiple enzymatic activities that are required at the eukaryotic replication fork would appear to be a complex task. The proliferating cell nuclear antigen (PCNA) protein has long been recognised as a key contributor to this, binding a large diversity of partner proteins and recruiting them to their sites of action (Mailand et al., 2013). PCNA partners recruited to replication forks include DNA polymerases, DNA processing enzymes, DNA methyltransferases, cell cycle regulators, and chromatin assembly and modifying enzymes, each required at difference stages of chromosomal replication. The importance of PCNA for this most crucial of cellular functions is underlined by the fact that, until recently, no disease causing mutation in the PCNA protein had been reported in humans, suggesting that most sequence alterations that profoundly affect PCNA function would be incompatible with life.

Through the work of the Windows of Hope project, a novel neurodegenerative condition was identified amongst the Ohio Amish community. Affected individuals exhibited features similar to, but distinct from, known disorders of DNA repair. All of the affected individuals were found to be homozygous for a c.683G>T Amish founder mutation in *PCNA,* which results in the substitution of a stringently conserved serine residue to isoleucine at position 228 (p.Ser228Ile) of the protein. As individuals who are homozygous for the amino acid substitution survive, it may be predicted that this substitution does not completely destroy the function of the mutant PCNA molecule. Consistent with this, our studies did not detect any changes in DNA replication in cells from affected individuals. However altered parameters of DNA repair were identified, specifically within the nucleotide excision repair pathway which deals with damage to DNA generated by sunlight exposure (DiGiovanna and Kraemer, 2012). As PCNA is known to be important for reconstitution of nucleotide excision repair in vitro, this finding provides a clear molecular link between the disease causing genetic alteration and the cellular phenotypes observed.

While we were able to detect altered interaction between PCNA and three of its known binding partners (Fen1, Lig1 and XPG), these are unlikely to be represent the only outcomes of the sequence alteration. A systematic approach will be needed to assess the effect of the mutation on the binding to each known binding partner to ascertain whether other altered interactions may

contribute to the clinical manifestations of the disease. While the interactions between PCNA and XPG, Lig1 and Fen1 are mediated by 'PCNA interacting protein' (PIP)-boxes (Lehmann, 2011), the SILAC data suggests that not all PIP box containing proteins show altered affinity for PCNA p.Ser228Ile. For example the DNA methyltransferase DNMT1 binds both WT and PCNA p.Ser228Ile equivalently. Further work is needed to understand the structural changes triggered by the Ser228Ile amino acid alteration that result in such differential binding of some PCNA partners, but not others.

The separation of function outcome caused by this mutation is, at first sight, remarkable. PCNA's main role is during chromosomal replication, and yet the Amish individuals affected with this syndrome reach adulthood, and lymphoblastoid and primary fibroblast cells proliferate normally in culture with no apparent checkpoint activation or genomic instability. Thus the defective interactions observed, which are with key components of the DNA replication fork, appear not sufficient to significantly perturb replication while they do appear to significantly impair DNA repair. This presents a conundrum as the part of the nucleotide excision repair process that is presumed to be PCNA-dependent, DNA synthesis, is enzymatically the same in both situations. How can this be reconciled? This may possibly be explained by the nuclear organisation of the two processes (replication and repair). Nucleotide excision repair is performed independently at each damage site, which are distributed randomly across the genome. In contrast, DNA replication processes occur within "replication factories", transiently forming structures in the nucleus containing high local concentrations of DNA replication components, including Fen1 and Lig1 (Warbrick, 1998). A single factory can contain multiple replication forks, and at each fork there are likely more than two PCNA trimers, as PCNA is retained for some time on newly synthesised DNA (Sporbert et al., 2005). Such a high local concentration of PCNA at replication factories may serve to allow PCNA's efficient interaction with partner proteins even if the affinity of such interactions is reduced, as with the p.Ser228Ile alteration (Figure 57). In contrast, during nucleotide excision repair there is no process that results in local concentration of PCNA, and as such the very same interactions may become sensitive to Ser228Ile mediated changes in interaction affinity. This

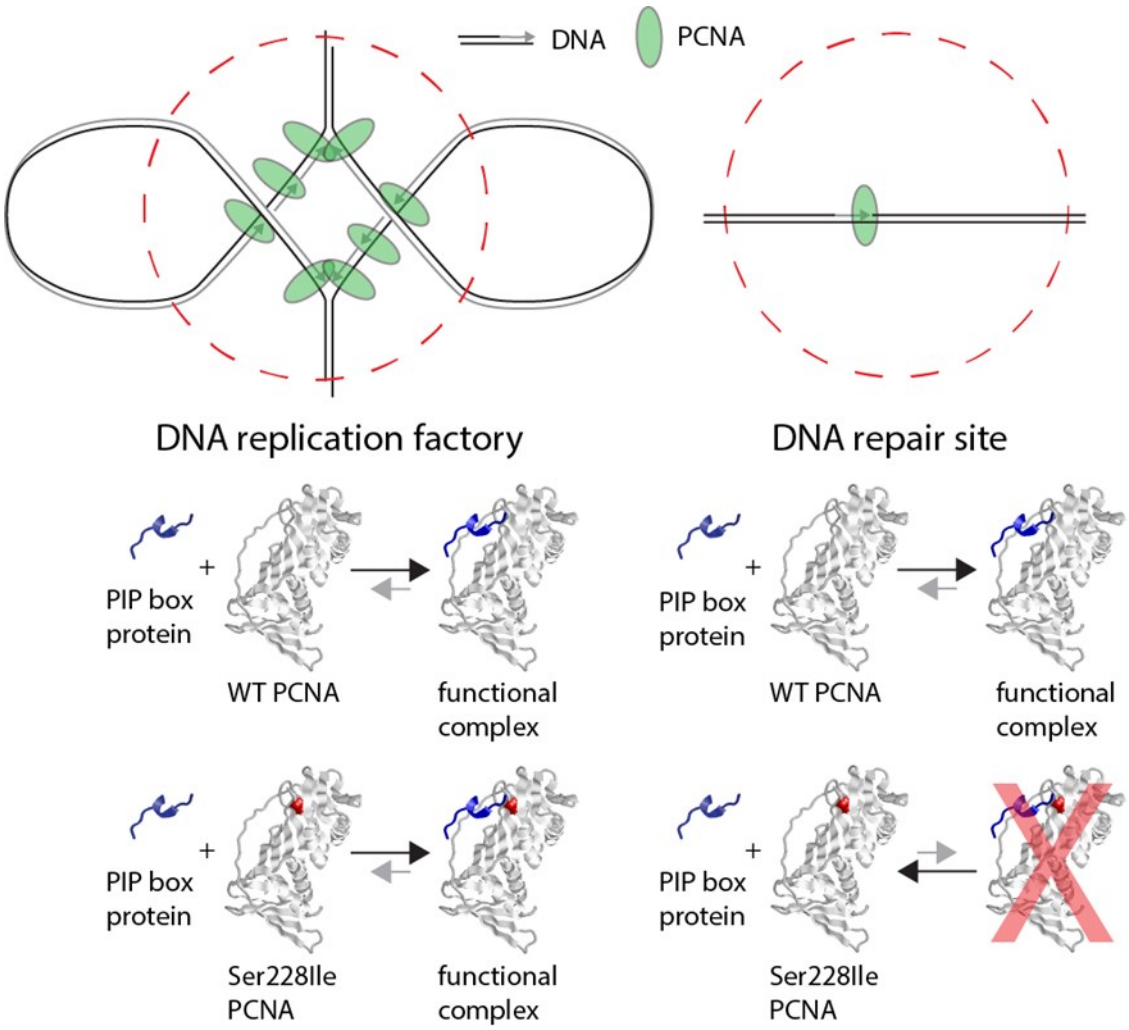


Figure 57: Hypomorphic effects of the PCNA p.Ser228Ile mutation

Top: Schematic of DNA replication and repair sites in vivo. Replication factories processing Okazaki fragments generate high local concentrations of PCNA, not seen at repair sites. **Bottom:** The Ser228Ile alteration in PCNA destabilises its interaction with PIP-box containing partner proteins, increasing the K_d for crucial binding reactions. This will have a greater biological effect when PCNA concentrations are low, i.e. at repair sites.

raises the intriguing possibility that the organisation of replication into these factory structures may contribute to the stability of the replicative process, as the locally induced high concentrations of factors induced by such a physical organisation make the system inherently robust against any temporary stresses. This could be predicted to be a useful property for a process in which accuracy is ultimately required to ensure the stability of the genome. PCNA has many known cellular functions, and thus the PCNA p.Ser228Ile mutation could affect a diverse set of cellular activities. PCNA is implicated in many aspects of chromosomal replication and also other types of DNA repair. It will be important to assess the sensitivity of cells from affected individuals to DNA damage inducers other than UV such as hydrogen peroxide, and replication fork stalling agents such as aphidocolin. A suitable approach would be to use a clonogenic survival assay similar to those shown in Figure 14B.

The cellular defects in nucleotide excision repair provide some explanation for the clinical overlap observed between individuals homozygous for PCNA p.Ser228Ile and those seen in Cockayne syndrome and xeroderma pigmentosa. The overlap with ataxia telangiectasia is more difficult to explain. The ocular and cutaneous telangiectasia and cerebellar atrophy seen in the affected individuals are remarkably reminiscent of ataxia telangiectasia. The defective molecule in ataxia telangiectasia is ATM, which is a crucial part of the DNA double strand break repair pathway (Perlman et al., 2003) as are other molecules implicated in the so called “ataxia telangiectasia – like disorders” such as MRE11 (Stewart et al., 1999). The hallmark of ataxia telangiectasia is profound radiosensitivity (Perlman et al., 2003). Although we did not find a cellular defect in the DNA double strand break repair pathway, the assays used were probably not sensitive enough to detect subtle abnormalities in this pathway. As with other specialised metabolic pathways for DNA repair, including nucleotide excision repair, base excision repair and mismatch repair, the repair synthesis step of this pathway is fundamentally dependent on PCNA and more sensitive assays such as those suggested below may well identify abnormalities in this pathway.

It seems reasonable to hypothesise that the clinical presentation of affected individuals, and the cellular phenotypes, arise because the mutated PCNA is

specifically deficient in particular protein interactions while others are maintained normally. To evaluate this further analysis of which PCNA partner proteins fail to bind normally to PCNA p.Ser228Ile under a selection of relevant conditions is required.

Analysing the effect of PCNA mutations on partner protein interactions is complicated by the trimeric nature of PCNA. Exogenously-expressed mutated PCNA will always form mixed trimers with endogenous WT PCNA, making it impossible to use standard purification strategies to identify interactions altered by the mutation. To overcome this the SILAC method was used to identify defective interactions of PCNA p.Ser228Ile partner proteins. This assay was used to quantitatively compare the full "interactome" of human PCNA, WT and p.Ser228Ile (Figure 21A). This experiment forms the foundation for a body of work, for which we were recently awarded MRC funding, to try to more fully explain why the p.Ser228Ile alteration causes the abnormal cellular phenotypes identified. This experiment has provided a candidate list of putatively altered interactions that are likely to be at least partially responsible for the cellular phenotypes seen in affected individuals, and in particular highlighted Fen1 and Lig1 as being potentially relevant. Each of the candidates identified by this approach now needs to be validated, and the extent of any defective interaction (either too tight or too weak) determined. Analysis of these PCNA interactions in a more physiological setting could be done by co-immunoprecipitation using extracts from PCNA$^{p.Ser228Ile/p.Ser228Ile}$ cells and control (WT or complemented) cells as was successfully done with Fen1. Alternatively if no suitable antibody can be identified, expression of tagged proteins via lentiviral transduction would be an alternative option. Binding assays using *E. coli*- or baculovirus-produced recombinant proteins by pull downs at different salt stringencies and by surface plasmon resonance will allow the kinetics of the binding reactions to be evaluated. Assessment of protein mislocalisation in PCNA$^{p.Ser228Ile/p.Ser228Ile}$ cells could also be examined by immunofluorescence. Once the defective nature of a partner interaction is verified, cell-based assays can be utilised to determine if that protein's functionality is compromised in the PCNA$^{p.Ser228Ile/p.Ser228Ile}$ cells.

It is known that some biologically important interactions on PCNA only occur after DNA damage or replication fork stalling (Lehmann et al., 2007). These

interactions would not have been recapitulated in the SILAC analysis to date. For example the PCNA–interacting proteins XPG and XPA were not identified by this analysis, although they may well be potentially responsible for some of the defective UV responses that were observed in PCNA$^{p.Ser228Ile/p.Ser228Ile}$ cells, and some of the phenotypic manifestations observed in affected individuals. The GST assay seen in Figure 28 which demonstrated no detectable interaction between PCNA p.Ser228Ile and the XPG PIP box provides preliminary evidence of this. The SILAC based procedure could be repeated using extracts from UV-irradiated cells, which should recapitulate the defective interaction observed between PCNA p.Ser228Ile and XPG and may well reveal such a defective interaction with XPA.

The identification of additional cellular processes that are perturbed by the PCNA p.Ser228Ile mutation, would also inform protein interaction studies. For example if cellular sensitivity to a particular agent such as gamma radiation were to be identified, the SILAC-assay could be used to determine the PCNA p.Ser228Ile interactions that are defective after this treatment. This strategy should lead to the identification of the PCNA partners that fail to interact normally with PCNA p.Ser228Ile after relevant damage or stress. An assay such as this can only identify relatively stable interactions and relatively abundant proteins. A candidate approach to assess the binding of known PCNA interactors (including XPA, p21, and others that are potentially involved in the cellular pathways) implicated by the results of cellular sensitivity assays, should reveal defective interactions in the case that they are not found in the SILAC screens.

Proteins that fail to bind properly to the mutated PCNA p.Ser228Ile are likely to be required for the PCNA functions that are perturbed in cells from affected individuals. It is therefore necessary to connect specific interaction defects to specific cellular effects. The Okazaki fragment processing proteins Fen1 and Lig1 are implicated by the SILAC analysis and subsequent validation studies. Therefore, the next step is to determine whether the PCNA p.Ser228Ile cells show normal Okazaki fragment maturation. Although gross parameters of DNA replication are unperturbed in PCNA$^{p.Ser228Ile/p.Ser228Ile}$ cells, recently published work suggests that even substantial disruption to Okazaki fragment maturation

does not in fact dramatically alter gross DNA synthesis parameters, but more subtly impacts on genome stability (Duxin et al., 2012). Comparing $PCNA^{p.Ser228Ile/p.Ser228Ile}$ cells and WT control cells for the parameters below would establish the effects of the PCNA p.Ser228Ile sequence alteration on Okazaki fragment maturation. Okazaki fragment size and quantity can be assessed utilising alkaline sucrose gradients and alkaline agarose gels. RNA persistence in the genome due to failure to remove primers would be a feature of perturbed Okazaki fragment maturation. This could be assessed by alkaline lysis and/or RNaseHII sensitivity of genomic DNA. If detected this phenotype could also be caused by a defective interaction between PCNA p.Ser228Ile and RNaseH, which interestingly was also suggested by the original SILAC analysis.

Fen1 and Lig1 are both required for the last stages of nucleotide excision repair: the reconstitution of complete DNA strands after excision of damage and re-synthesis. Thus these defective interactions could also be responsible for the UV damage sensitivities seen in the $PCNA^{p.Ser228Ile/p.Ser228Ile}$ cells. Analysis of nucleotide excision repair in more detail, by measuring the kinetics of UV lesion removal, repair protein recruitment and RNA polymerase stalling should determine if direct effects on the final stages of nucleotide excision repair can be identified. Evidence that these cellular defects are a direct result of altered protein interactions could be shown by engineering the partner proteins to restore normal binding to PCNA p.Ser228Ile, complementing $PCNA^{p.Ser228Ile/p.Ser228Ile}$ cells with these engineered proteins by lentiviral transduction, and assessing if the cellular phenotypes are suppressed.

Although much is known about the function of PCNA, there is unquestionably a great deal to learn about this fundamentally important molecule as well as the processes of DNA synthesis and repair. The further work detailed here would build upon the knowledge provided by the first disorder associated with PCNA dysfunction to enhance understanding about the role of PCNA in DNA synthesis and repair, and provide a molecular explanation for the intriguing clinical features of the novel syndrome that results. Such a complete understanding of the cellular defects associated with PCNA p.Ser228Ile will permit a more detailed investigation of the aberrant developmental and neurodegenerative processes associated with this disease.

It will be interesting to see if any additional pathogenic mutations in *PCNA* are detected in humans. Choosing a cohort of individuals in whom to screen for such mutations is somewhat difficult given that it is likely that mutation of only a very few residues of the molecule might be compatible with life and result in a similar or indeed dissimilar phenotype to that which was described in chapter 3. Individuals with an intermediate defect in recovery of RNA synthesis have been screened but none harboured any potentially pathogenic mutations in *PCNA*. Perhaps the next most likely place to look is in patients with clinical features suggestive of ataxia telangiectasia and minimal or complete absence of radiosensitivity. There are relatively few individuals who present in this way and collaborative screening for PCNA mutations in this cohort is planned.

I was awarded an MRC Centenary fund early career award to generate a mouse model of the PCNA p.Ser228Ile mutation. This work is being undertaken in collaboration with the Wellcome Trust Human Genetics unit in Oxford, using the TALEN approach. The first step will be to see in the homozygous mutant mouse recapitulates the human phenotype and if so neuronal cell lines from these animals can be used for downstream analyses similar to those described above. This animal model has the potential to illuminate the neurodegenerative processeses seen in the Amish PCNA disorder as well as other DNA repair disorders.

6.2 Chapter 4: Mutations in *KPTN* cause a neurodevelopmental disorder associated with macrocephaly and seizures

Chapter four describes the identification of two founder mutations in the Amish *KPTN,* a gene never previously associated with human disease. All of the affected individuals identified displayed intellectual disability and were macrocephalic. Most of the individuals who were phenotyped during the course of this study had a neurobehavioral phenotype with features suggestive of a pervasive developmental disorder. This clinical aspect due to KPTN loss of function will require focussed study with further formal intelligence quotient testing and diagnostic testing for autistic spectrum disorder. Although only limited neuroimaging data from three affected individuals was available, brain size was increased, but appeared to be morphologically grossly normal,

additional cases will be needed to firmly establish the absence of structural brain abnormalities in association with *KPTN* mutations.

At the time of the original study cellular material from affected individuals was not available, as such a primary rat neuronal cell line was used to model the effects of the Amish *KPTN* founder mutations. Cell studies revealed dynamic interactions between KPTN and components of the actin cytoskeleton which were lost upon introduction of the mutations. By their nature over expression studies have limited sensitivity and cannot be considered to be exhaustive. To gain further insight into KPTN function, recapitulation of the results of these studies in cell lines where the effect on endogenous KPTN can be examined are necessary. Primary fibroblasts and EBV transformed lymphoblastoid cell lines have been established from individuals homozygous for the *KPTN* c.776C>A nonsense mutation and compound heterozygotes for this and the 18bp duplication (c.714_731dup) as well as heterozygous and WT family members.

In silico analysis tools predict that the *KPTN* c.776C>A mutant transcript may well be degraded via mRNA-surveillance mechanisms. It will now be possible to carry out quantitative reverse transcription PCR to assess transcript levels in affected individuals homozygous and heterozygous for the mutation versus those of WT controls. Endogenous protein expression and localisation studies can also be performed to further confirm the likely outcome of both the *KPTN* variants identified in this study.

Given how little is known about the function of KPTN *in vivo* any comments on a potential mechanism of disease pathogenesis are largely speculative. A better understanding of the physiological role of KPTN in neurodevelopment will hinge on the further characterisation of the function of KPTN. The initial phenotyping of the recently generated *Kptn* knockout mouse, produced in collaboration with our collaborators at the Wellcome Trust Sanger Institute, revealed clinical features overlapping with those of humans including, megalencephaly and cognitive impairment. Once the mouse colony has been re-established, further studies are planned to fully characterise the phenotype of Kptn deficient mice. It will be particularly interesting to study the effects of Kptn deficiency on the

brain, firstly by more detailed phenotype comparisons between humans and mice, and secondly by establishing neuronal cell lines for protein localisation and downstream functional analyses. The planned MRI and micro-CT neuroimaging of these mice will help to further characterise the brain pathology. Gene expression analysis in brain tissue will be facilitated by the LacZ gene encoding β-galactosidase inserted into the targeted allele when the mice were generated, which will permit visualisation of the neurons that express Kptn through hydrolysis of X-Gal producing a blue precipitate that can be easily visualised under a microscope. Transcriptome analysis of the mouse brain tissues should help indicate downstream molecules whose expression is affected by KPTN loss of function, suggesting these as useful candidates for down-stream analysis. An alternative approach to this would be the SILAC assay used to study deficient PCNA interactions which could be used similarly to study the KPTN "interactome". In this way it may be possible to gain further insight into whether KPTN is involved in actin biology or whether neurodevelopment is affected through a distinct mechanism as yet unidentified.

Despite a relatively similar genome, the human neocortex has undergone enormous evolutionary modification with a threefold increase in size occurring over the last 3 million years. The complex gyrification of the human brain has resulted in a brain surface area 1000 x that of a mouse. The relative complexity of the human brain complicates animal modelling of disorders of human brain development. Mouse models of primary microcephaly have not always recapitulated the human phenotype. More recently a human pluripotent stem cell-derived 3D organoid culture system has been used to recapitulate human cortical organisation and successfully model primary microcephaly (Lancaster et al., 2013). Although this is still an imperfect system, it would be interesting to utilise this to model KPTN deficiency, which appears to be a pure global megalencephaly.

Human haploinsuffuiciency of the mTOR inhibitor PTEN is also associated with a global increase in brain size and high incidence of autistic features. The *Pten* haploinsufficient mouse is megalencephalic and shows impaired social behaviour (Clipperton-Allen and Page, 2014). *Kptn*$^{-/-}$ female mice showed evidence of impaired cognition, although the numbers in this preliminary

analysis were small. Further studies are planned to see if this result is recapitulated in a larger cohort of mice of both sexes and to establish whether or not social behavioural abnormalities are seen.

Splenomegaly and an apparent increased susceptibility to infection had been observed in a number of affected Amish individuals. It was notable that the primary phenotyping of *Kptn*$^{-/-}$ mice revealed a reduced level of IgD-positive mature B-cells and an increased susceptibility to salmonella infection. The youngest affected Amish individual in the original study has splenomegaly and has suffered repeated episodes of pneumonia. This individual is currently undergoing detailed immunological and haematological studies carried out by his clinicians in Ohio to establish whether there is any evidence of impaired immune function. In light of a possible human immunodeficiency phenotype, secondary immunophenotyping of the *Kptn*$^{-/-}$ mouse is planned when the next colony is established.

Although Wellcome Trust Sanger Institute in house data suggest that tissue obtained from mice homozygous for the Tm1a allele shows complete loss of KPTN function, it is theoretically possible that not all of the transcript is targeted for non-sense mediated mRNA decay. A Tm1b true KPTN knockout mouse model is currently being generated and it will be interesting to see if the phenotype matches that of the Tm1a homozygous mouse.

Loss-of-function mutations in other mTOR-inhibitory genes besides *PTEN* such as *TSC1, TSC2,* or *STRADA* and gain-of-function mutations in mTOR activators such as *PI3K* or *AKT3,* result in aberrant brain development with generalised or regional brain overgrowth, several are also autistic spectrum risk factors (Parker et al., 2013, Crino et al., 2006, Lee et al., 2012, Poduri et al., 2012, Clipperton-Allen and Page, 2014). Treatment with the mTOR inhibitor rapamycin has been shown to reduce seizure frequency and possibly improve receptive language and social interaction in individuals with Pretzel syndrome, a rare Amish disorder associated with a homozygous loss of function mutation in *STRADA* (Parker et al., 2013).

Autistic spectrum disorders are relatively common in the general population affecting approximately 1% of individuals. There appears to be a strong genetic component with a 90% concordance rate observed in monozygotic twins (Perou et al., 2013). This study has provided important information about the biological partnership between KPTN and the actin cytoskeleton, shedding further light onto the pathogenic processes that underlie megalencephaly, as well as highlighting mechanisms of potential relevance to autistic spectrum disorders. Mutations in *Rab39B* have been shown to cause an X-linked form of intellectual disability associated with seizures, macrocephaly and autistic features. Depletion of Rab39B in model cell systems was shown to disrupt the actin-cytoskeleton and result in a reduction in neurite growth cones and dendritic arborisation (Giannandrea et al., 2010). The identification of more than one gene involved in actin biology and associated with a neurodevelopmental disorder and megalencephaly provides evidence of a common pathway associated with these features and further studies may reveal potential therapeutic targets much like with the mTOR pathway group of disorders.

6.3 Chapter 5: Autosomal recessive mutations in closely linked WDR73 and WHAMM genes underlie a novel composite nephrocerebellar syndrome phenotype

Chapter 5 describes the clinical delineation of two separate disorders which were previously thought to possibly be the same entity, described as “Hershberger syndrome” in McKusick’s pioneering 1960s Amish studies. Both disorders cause profound neurological impairment and are present at extremely high frequency within the ultraconservative Swartzentruber Amish group. Molecular studies were able to help determine that one of these disorders was caused by homozygosity for a founder mutation in *SAMHD1.* This gene was previously associated with the highly genetically heterogeneous and phenotypically diverse type 1 interferonopathy Aicardi Goutières syndrome (Rice et al., 2009).

The second disorder described in Chapter 4 (nephrocerebellar syndrome) is new to medical science and appears to be a composite phenotype resulting from homozygous mutations in two closely linked genes *WHAMM* and *WDR73.*

All but one of the children identified are homozygous for founder mutations in both *WHAMM* and *WDR73,* and that child lacked the renal features of the condition. The two founder mutations are 1.7Mb apart and it is difficult to know if both arose in the same individual at the same time, or if one arose subsequently in cis with the other mutation. Despite the extensive screening of family members and Amish controls only two individuals have been identified as heterozygous for one (*WDR73*) and WT for the second (*WHAMM*) variant. The SNP haplotype of these individuals is distinct around the WT *WHAMM* sequence, suggesting that they have recombined the *WHAMM* mutation out. This evidence suggests that the two mutations occurred either in the same individual at the same time, or if not in very quick succession.

This work shows the benefit of access to expert next generation sequencing services, as the *WHAMM* mutation was missed by our collaborative partners at CSC (through the Broad Institute), but was picked up by the Wellcome Trust Sanger Institute. This likely reflects the complex and repetitive nature of WHAMM, which also rendered confirmation of this mutation by Sanger sequencing challenging.

Features that are common to both Aicardi Goutières syndrome and nephrocerebellar syndrome include profound psychomotor retardation, microcephaly of post-natal onset, a dystonic movement disorder, seizures, irritability, cortical visual impairment and abnormal eye movements. Distinguishing features include glaucoma and inflammatory skin and oral lesions in Aicardi Goutières syndrome and hypoplastic yellow coloured dentition and nephrotic syndrome in nephrocerebellar syndrome. Aicardi Goutières syndrome is associated with significant intrafamilial variability. In the classic form individuals are profoundly neurologically impaired, but some individuals may be only mildly affected, with essentially normal cognitive development and normal brain growth but features of arthropathy or chilblain lesions (Crow, 1993). All known cases of nephrocerebellar syndrome make little if any developmental progress. The predominating neurological finding in Aicardi Goutieres syndrome is peripheral spasticity and hyperreflexia, whereas all the children with nephrocerebellar syndrome are hypotonic with normal reflexes. The most frequent finding on neuroimaging from nephrocerebellar cases was

cerebellar hypoplasia and in Aicardi Goutières syndrome cases due to *SAMHD1,* was intracerebral calcification and large vessel vasculopathy (Ramesh et al., 2010).

Interestingly, the original description of “Hershberger syndrome” by Cross and Mckusick described it as a complex spastic paraplegia with onset of spasticity from 5 months of age (Cross and McKusick, 1967). This seems most compatible with Aicardi Goutières syndrome, where the majority of children have a period of apparently normal development followed by a subacute onset of severe encephalopathy and sterile pyrexia. Onset of neurological impairment follows the insult (Crow, 1993). The Swartzentruber Amish have historically had little interaction with modern medicine although that has been changing in some of the younger families. It is likely that this played a big part in the difficulty with separating these two disorders in the past.

The natural history of Aicardi Goutières syndrome presents a window of opportunity to treat affected individuals before brain disease progresses. Treatment trials with antiretroviral medications commonly used in the management of human immunodeficiency virus as well as other immunomodulatory therapies have been proposed (Crow et al., 2014). The Amish population represent the ideal group in which to carry out such clinical trials. The condition occurs at extremely high frequency particularly amongst the most conservative Amish group, the Swartzentruber Amish. Newborn screening for this condition is a viable option in this population and the potential benefits to affected children, their families and the wider Amish community of an effective treatment for SAMHD1 related Aicardi Goutières syndrome are enormous. SAMHD1 deficiency is associated with abnormalities of the cerebral vasculature and a concurrent high risk of stroke (Permanyer et al., 2010), it remains to be seen if any of the proposed therapies will be of benefit in reducing this risk. However even in the absence of a medication, the possibility of neonatal diagnosis in the Amish community presents the opportunity to monitor and intervene surgically in some cases.

The dental abnormalities and seizure predisposition in nephrocerebellar syndrome seemed somewhat reminiscent of those seen in Kohlschütter Tonz

syndrome, a condition associated with homozygous loss of function mutations in *ROGDI*. The neurological phenotypes of the two conditions though seem quite distinct, with spasticity the predominate neurological feature of Kohlschütter Tonz syndrome (Mory et al., 2014) *ROGD1* encodes a relatively uncharacterised protein expressed at high levels in the brain (Schossig et al., 2012) . It remains to be seen if one or other of the molecules implicated in nephrocerebellar syndrome interacts with this molecule.

Much is left to be learnt about the two molecules implicated in the Amish nephocerebellar syndrome. WHAMM is a well characterised molecule and the cellular data we acquired is highly suggestive of, at the very least, a disease modifying effect. It is interesting that despite the nature of the genetic mutation, the transcript appears to at least partially escape nonsense mediated mRNA decay and result in a protein that lacks the C-terminal of the molecule. Consequently it is possible that this molecule may retain some functionality, perhaps alleviating some of the clinical features that may result with complete loss of function of this molecule. Another possibility is that one of the number of *WHAMM* pseudogenes could be transcribed, which may also compensate for the loss of function of the molecule. A role in autophagy had not previously been ascribed to WHAMM, and this and the results of the other cellular studies presented in this thesis were used as preliminary data to support our recent successful collaborative NIH grant application to look at WHAMM function and its involvement in the nephrocerebellar syndrome in more detail.

Fibroblasts from individuals affected by nephrocerebellar syndrome lack autophagosomes at steady-state and undergo intracellular vacuolation when autophagic flux is blocked, but re-introduction of wild type WHAMM can restore normal autophagy functions to these cells. To test the hypothesis that WHAMM, microtubules, and the Arp2/3 complex function together in autophagosome biogenesis, the role of these cytoskeletal factors in autophagosome assembly and trafficking could be assessed by using WHAMM rescue constructs in nephrocerebellar syndrome fibroblasts and using RNA intereference in cultured neuronal cell lines. The activities of functional versus defective WHAMM derivatives in assays of actin nucleation, phospholipid interaction, and microtubule binding could also be assessed in vitro.

Under native cellular conditions, WASP-family members exist in multi-subunit protein complexes that control their activities. To test the hypothesis that WHAMM is also found in a complex that influences its function, a WHAMM-tagged derivative could be isolated from cultured cells using a tandem affinity approach, and native WHAMM could be extracted from brain tissue by conventional chromatography, it would then be possible to determine the composition of each 'WHAMM complex' by mass spectrometry. Measurement of the ability of each complex to nucleate actin, interact with phospholipids, and bind microtubules in vitro could then be undertaken. RNA interference could also be used to determine which members of the complex contribute to autophagosome remodeling in cells.

Completion of these analyses would provide key insights into cytoskeletal dynamics and autophagosome remodelling and help to establish the role of WHAMM in the nephrocerebellar syndrome phenotype.

Abnormal functioning of the WDR73 protein is almost certainly the predominate cause of the neurological phenotype seen in nephrocerebellar syndrome, but in what way these effects are mediated remains largely speculative. Nephrocerebellar syndrome fibroblast cells grow very poorly and have abnormal morphology which limits the possible functional assay that can be performed using these precious resources. Protein localisation studies will be carried out by transiently transfecting epitope-tagged WDR73 constructs into SH-SY5Y human neuronal cells and HEK-293 human embryonic kidney cells, and then examining the protein's subcellular localisation. The most likely outcome of the WDR73 mutation is loss of protein function, and therefore additional complimentary experiments are planned using siRNA knockdown to deplete the WDR73 protein in HeLa and primary fibroblast cell lines.

Mouse models of both the *WHAMM* and *WDR73* mutations as well as the double homozygote are currently being generated through our collaborators at the Wellcome Trust Sanger Institute. These models will be an invaluable resource in terms of trying to determine which molecule is responsible for which aspects of the nephrocerebellar phenotype. It will be particularly fascinating to compare the brain histology of these animals to the human histopathology that

we observed in one of the affected children. Lastly, mouse cell lines can be utilised for downstream studies of both WHAMM and WDR73 function.

6.4 Concluding remarks

6.4.1 Clinical benefits of community studies to the Amish community

This study stemmed from a long-running clinical-genetic project, called 'Windows of Hope' (WOH), based amongst the Amish communities of Ohio. The aims of the larger project are to determine the genetic basis of inherited diseases which occur in the Amish community and to translate the findings of this research into direct diagnostic and medical benefit, both within this community and the general population. The project has enjoyed considerable success and has to date discovered the genes for 16 novel conditions which occur in the community. These findings have formed the basis of an extensive molecular genetic testing and counselling program for the Amish. As widespread knowledge of these recently described disorders is still lacking, and the carrier frequency in the community is high, many of these children and their families undergo needless, costly and sometimes painful investigations. Improved awareness of the clinical features and incidence of these conditions amongst both the community and the local and specialist healthcare providers would circumvent this problem. The comprehensive genetic, molecular and clinical studies described in this thesis have empowered us with specialist knowledge of these syndromes which can now be used to facilitate appropriate medical care and follow-up for affected individuals. Knowledge of the full phenotype also permits screening for additional syndromic associations, and early intervention to prevent potentially serious complications such as the nephrotic syndrome associated with nephrocerebellar syndrome. All of the research results generated from the Windows of Hope project are fed back promptly to the community. This involves organisation of family support group meetings, community liaison and generation of disease specific and community appropriate information. Educational meetings for local health care professionals are also held, with the aim being to improve health care outcomes through improved knowledge of the different types of inherited conditions found within the community, effective management strategies and developing and increasing the availability of diagnostic testing for newly identified conditions. Recently, based on work within and related to this thesis I was awarded an Alexander Fleming dissemination award from the MRC, Medical Research Foundation. The purpose of this award is to support the dissemination of MRC

funded research results beyond the scientific peer reviewed press, to patients, participants and policy makers. This funding was given in recognition of the previous success of the community and healthcare worker engagement activities and to allow their further development.

The technological advances in genetics have rapidly increased the speed and scope of diagnostic genetic testing. The impact has been even more significant in the field of neurodevelopmental disorders, where conditions show significant phenotypic and genetic heterogeneity. Next generation sequencing technologies permit the screening of multiple potentially causative genes in parallel, which means that recognition of the precise phenotype of an individual is no longer completely essential. This has particularly application in communities like the Amish, where founder mutations are commonly causative, greatly facilitating the development of diagnostic gene panels for this type of disorder. Such panels can even be targeted to specific Amish church groups such as the Swartzentruber community, where specific inherited conditions are more prevalent. A longer term vision of the Windows of Hope project is to harness these technologies to work towards developing a newborn screening program, with the aim being to provide early diagnosis of conditions prevalent in the Ohio Amish community so that management strategies can be put in place promptly, resulting in improved outcomes for children and their families.

6.4.2 Benefits of community based studies to the wider society

The benefits to the wider society of these community based studies are numerous and wide ranging. Due to their marriage patterns and ancestry the Amish comprise an endogamous community, in which mutations responsible for certain genetic conditions have become enriched, greatly facilitating their identification. Importantly, almost all genetic conditions initially described and identified amongst the Amish have subsequently been found outside of the community, validating the importance of Amish genetic studies and widening the impact of such discoveries.

A good example is a recently published study which I was heavily involved in where mutations in the *B4GALNT1* gene were described in association with a variable condition best classified as complex form of spastic paraplegia (GM2

synthase deficiency). The disorder is present at high frequency in the Amish community, but we also identified it in Kuwaiti and Italian-Canadian families in parallel (Harlalka et al., 2013b). Cartilage hair hypoplasia is another well known example of an important genetic syndrome initially described in the Amish (McKusick et al., 1965). Certain conditions which are associated with a higher incidence outside of the Amish community have been found to be associated with the presence of more than one founder mutation within the community, notable examples include Cockayne syndrome and phenylketonuria (Xin and Wang, 2013, Strauss and Puffenberger, 2009). The identification of two founder mutations in *KPTN* in the same Amish community make it even more likely that this gene will be subsequently shown to be an important cause of non-Amish cases of neurodevelopmental disorder with macrocephaly.

Over the last 5-6 years, the advent of next generation sequencing technologies has led to rapid progress in identifying the genetic basis of disorders of brain development and intellectual disability. Consequently new pathways have been highlighted as crucial for normal human neuromorphogenesis. As mentioned earlier, disorders of neurodevelopment are genetically heterogeneous and distinct phenotypes can be hard to distinguish. Large scale general population based studies such as the Deciphering Developmental Disorders study based at the Wellcome Trust Sanger Institute and The Forge Canada study have been set-up to harness the power of new genetic technologies in an attempt to identify new genetic causes of abnormal human development, particularly those affecting the brain. These studies have been relatively successful, particularly in establishing a wider phenotypic spectrum associated with some genes than had previously been suspected. However, relatively few novel disorders have been identified through these studies to date, particularly given the number of individuals recruited, whereas studies within endogamous communities or of consanguineous pedigrees have been significantly more productive in this regard. Large scale targeted resequencing of exons within homozygous tracts in 136 consanguineous families with autosomal recessive intellectual disability identified 50 novel candidate genes emphasising the power of this approach (Najmabadi et al., 2011). Within the Amish community alone since starting this work in 2010, a total of ten new genetic causes of neurological impairment have been identified and published (Puffenberger et al., 2012, Cai et al., 2011, Xin et

al., 2010, Crosby et al., 2010, Harlalka et al., 2013a, Harlalka et al., 2013b, Lahiry et al., 2013), further demonstrating the scientific, clinical and financial advantages presented by such community based studies.

6.4.3 Scientific benefits of community based studies

The value of studying rare disorders is now widely recognised. Valuable insights into biological pathway are provided by such studies, often highlighting new candidate genes for related disorders. A good example of this is the study of autosomal recessive primary microcephaly. All of the genes associated with this condition to date have been identified through mapping studies of individuals born of consanguineous unions and all have been shown to be involved in mitosis usually through direct or indirect centrosomal interactions. Some of these genes are also implicated in primordial dwarfism, a group of disorders where both brain and body size are profoundly reduced.

An example relating more directly to the work in this thesis is the recent identification of an Amish founder mutation in the *HERC2* gene which when present in the homozygous form results in a neurodevelopmental disorder with striking clinical similarity to Angelman syndrome. This condition occurs at high frequency in the Old Order Amish. The mutation leads to a dramatic reduction in HERC2 protein levels, increased protein instability, and a dose dependent reduction in UBE3A activity, the primary molecule implicated in Angelman syndrome pathogenesis (Harlalka et al., 2013a). Although the *HERC2* and *UBE3A* genes are situated within close proximity to each other on chromosome 15, no functional link between the two encoded proteins and the Angelman syndrome phenotype had ever been proposed. This study provided important information about the biological partnership between UBE3A and HERC2, shedding further light onto the pathogenic processes that underlie Angelman syndrome, as well as highlighting mechanisms of relevance to other seizure and movement disorders.

As mentioned earlier, GM2 synthase deficiency is a disorder of ganglioside biosynthesis which occurs with high frequency amongst the Amish. Gangliosides are sialylated derivatives of glycosphingolipids. Although many glycosphingolipid catabolic defects have been defined only one inherited

disease arising from a defect in glycosphingolipid biosynthesis is known. This disease, arising due to defects in the first step of ganglioside biosynthesis (GM3 synthase), results in a severe epileptic disorder with profound intellectual impairment also found at high frequency amongst the Old Order Amish, but otherwise rare. Biochemical profiling of glycosphingolipid biosynthesis confirmed a lack of GM2 in affected subjects in association with a predictable increase in levels of its precursor, GM3 (Harlalka et al., 2013b). With the description of two neurological human diseases involving defects in two sequentially acting enzymes in ganglioside biosynthesis, there is the very real possibility that a previously unidentified family of ganglioside deficiency diseases exist. The study of patients and animal models of these disorders will pave the way for a greater understanding of the role gangliosides play in neuronal structure and function and may well provide insights into the development of effective treatment therapies.

Recently increasing numbers of genes implicated in disorders of human brain growth and development have been identified. These studies provide unprecedented opportunities to understand the molecules involved in the normal brain development and what goes wrong when their cellular functions are perturbed. The next challenge to be faced is to use this knowledge to develop novel and innovative therapies targeted directly to circumvent the identified cellular defect. The promise of genetics to enable deliverance of such treatments looks increasingly likely to transform the practice of medicine.

7 APPENDIX

7.1 APPENDIX CHAPTER 3

7.1.1 Chromosome 20 sequencing primers

c20orf30 Exon1F	…….	TAAAGAGACAAACGCTTCCG
c20orf30 Exon1R	…….	TGACTCGGAGCTTGATTTCG
c20orf30 Exon 2F	…….	GAACAGCTGGCATCTAACAAAG
c20orf30 Exon 2R	…….	TTTTCCTCCATAAAGTAACAAAGC
c20orf30 Exon 3F	…….	GAAGTGTAAATTGCCCATTGC
c20orf30 Exon 3R	…….	AATCCCCAAGGACAGTTAGC
c20orf30 Exon 4F	…….	TCACGCCATTGCTCTCTAGC
c20orf30 Exon 4R	…….	GGACGGACAAACACACACAC
c20orf30 Exon 5F	…….	TGGTGGTATACAACTCAAAGAAAAA
c20orf30 Exon 5R	…….	TCTGGCCATTATTTTCTTAAACA
PCNA Exon 2F	…….	GGCATTAAACGGTTGCAGG
PCNA Exon 2R	…….	CAGCCAATGAGGGCTAGG
PCNA Exon 3-4F	…….	ATGGGTGAGAGTGGTAACCC
PCNA Exon 3-4R	…….	CAGAGCCAAGACCCTGTCTG
PCNA Exon 5F	…….	CTCTTGTTCCCTGGATGGTG
PCNA Exon 5R	…….	TCGCAGATTTCAACAGTATCTC
PCNA Exon 6F	…….	TAGCTCCCTCCAAAGTGACC
PCNA Exon 6R	…….	TGTCCCATATCCGCAATTTT
PCNA Exon 7aF	…….	CGGTGACACTCAGTATGTCTGC
PCNA Exon 7aR	…….	GACCAGATCTGACTTTGGACTTT
PCNA Exon 7bF	…….	TGCGGATATGGGACACTTAAA
PCNA Exon 7bR	…….	TCAGGCTGCTCTAAACTGAAA
CDS2 Exon 1F	…….	TGCTAAGGGAACTGTGAGCC
CDS2 Exon 1R	…….	TCAGAGAACAAGACGACCCC
CDS2 Exon 2F	…….	TTCTTAGGCCCTGTTTGGAG
CDS2 Exon 2R	…….	AAGACCACGGGCTCTCTG
CDS2 Exon 3F	…….	TCACGGTGAATCTGGAAAGG
CDS2 Exon 3R	…….	GTCAATTCCCACAACCAACC
CDS2 Exon 4F	…….	ATGGGTTTTCTGGAAGGCTC
CDS2 Exon 4R	…….	TGCCTCTCCAGCATTTATCTC

CDS2 Exon 5F	…….	GCCCTCTGTCCCTTCCTC
CDS2 Exon 5R	…….	CCCCTGCATGGTCACAG
CDS2 Exon 6F	…….	GCTTTGACATTGTGCACTGG
CDS2 Exon 6R	…….	AACCTAAATGGGCATTCTTCC
CDS2 Exon 7F	…….	TGCTTCTCAGCAGTTTTCTATTTG
CDS2 Exon 7R	…….	TGCAGAGGATTCTGATGTGG
CDS2 Exon 8F	…….	TTTTATGGGTGATTTTGGTGC
CDS2 Exon 8R	…….	CAGCTAGATTTCTTCTGTCTGCC
CDS2 Exon 9F	…….	GGGGTTTTATGAAAGAAAGGTTC
CDS2 Exon 9R	…….	TGGCCTCCAAAGACACTCTC
CDS2 Exon 10F	…….	GCACCCTCAGGAACATGG
CDS2 Exon 10R	…….	TTACAGCTTTGGGAAGGAGG
CDS2 Exon 11F	…….	AGGCATCACCTCCCAGTG
CDS2 Exon 11R	…….	CAGCCTAGTATTTTGGAGGGG
CDS2 Exon 12F	…….	ATTGCCCTCAGCTACTCAGC
CDS2 Exon 12R	…….	TTTGGGATGCTTTTCACTGG
CDS2 Exon 13F	…….	TTCTGAGGCCTCCTGTCATC
CDS2 Exon 13R	…….	CAGTGAGTTGAAGCCTCGTC
C20orf196 Exon 2F	…….	AGCCTGGGCAACTTAGTGAG
C20orf196 Exon 2R	…….	CTCCTGAAAGCTTTGCTTAGG
C20orf196 Exon 3aF	…….	CCCATTGCCAGTTACTCAGG
C20orf196 Exon 3aR	…….	CCGGGCCATTTGAAAACT

7.1.2 Ataxia telangiectasia is excluded as a cause of the phenotype

Professor Malcolm Taylor excluded ataxia telangiectasia as a cause of the phenotype seen in the affected individuals described in chapter 3.

In the figure below, there are two lanes per lymphoblastoid cell lysate, one unirradiated (-) and one irradiated (+, activated ATM). In lanes 1 & 2 are cells from a normal positive control, showing phosphorylation of the targets SMC1 ser966 (panel C), KAP-1 ser824 (panel E), Nbn1 ser343 (panel G) and CREB ser121 (panel I). Also included, in panel A, is autophosphorylation of ATM at Ser 1981. There is a strong signal for each of these in lane 2 after activation. The total levels of SMC1, KAP-1, Nbn and CREB are also shown in panels D, F, H & J respectively. The negative control is cell lysate from a classical A-T affected individual is shown in lanes 3 & 4. There is no phosphorylation of these five targets in this individual's cells in lane 4 – consistent with having A-T; there is also absence of ATM protein (panel B).

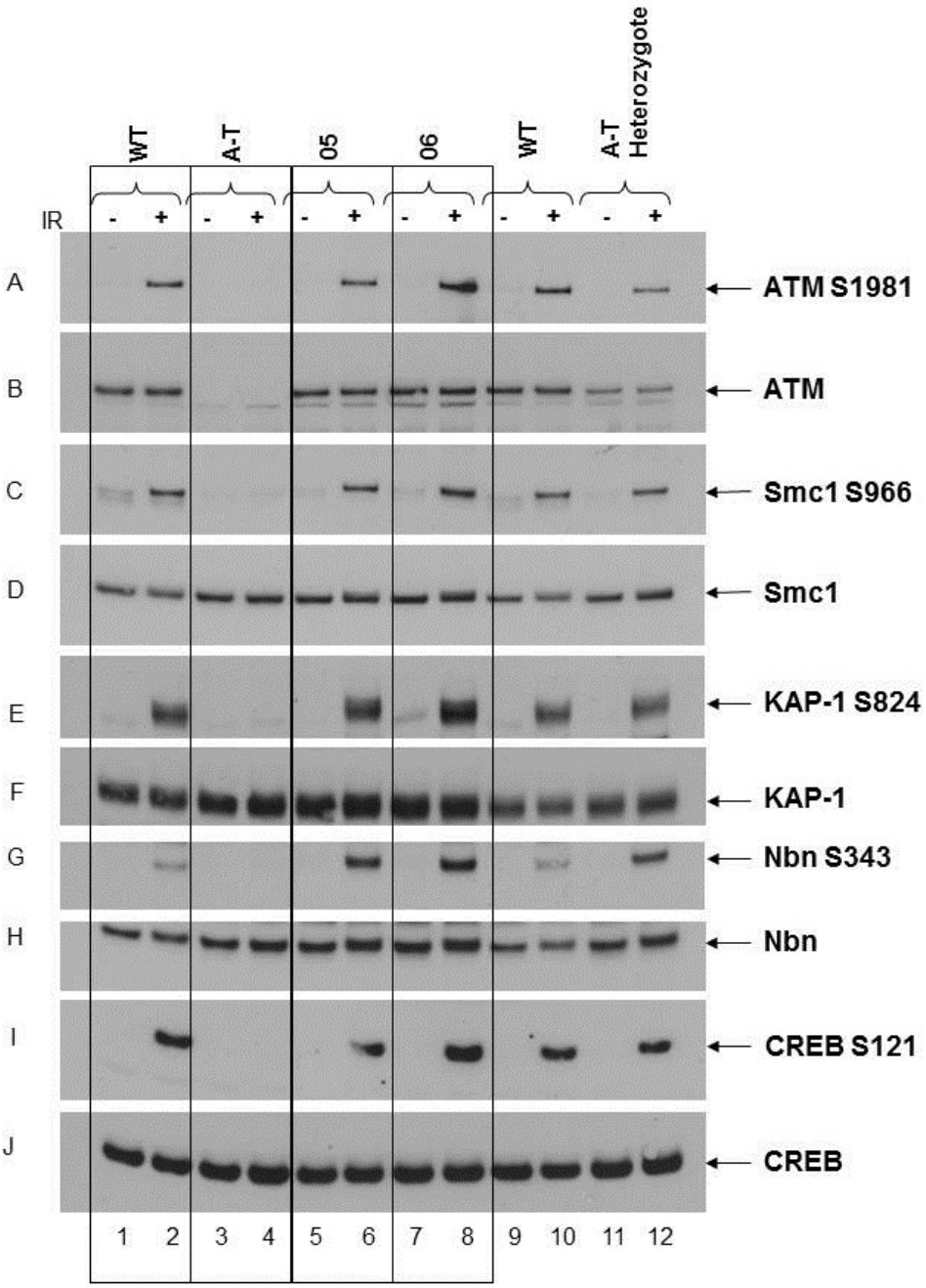
05 and 06 cell lysates are in lanes 5 & 6 and 7 & 8 respectively. Panel B shows that both lysates have a normal level of ATM protein. Just as with the normal control (lane 2) both 05 and 06 cells show normal ATM kinase activity, indicated by phosphorylation of SMC1 ser966, KAP-1 ser824, Nbn ser343 and CREB ser121 as well as autophosphorylation of ATM in panel A, in lanes 6 and 8 respectively. Also included in this analysis is a known *ATM* mutation carrier (+/-) in lanes 11 & 12. Even with half the level of ATM protein (panel B) there is normal ATM kinase activity.

Western blot was used to look for any loss of proteins associated with an AT-like phenotype. In Figure S2 below, lane 1 contains lysate of lymphoblastoid

cells from a normal individual, with a normal level of ATM and all the other proteins. In lane 2 is lysate from cells of a classical A-T patient, showing no ATM protein and in lane 3 lysate from an AOA2 patient with a reduced level of senataxin.

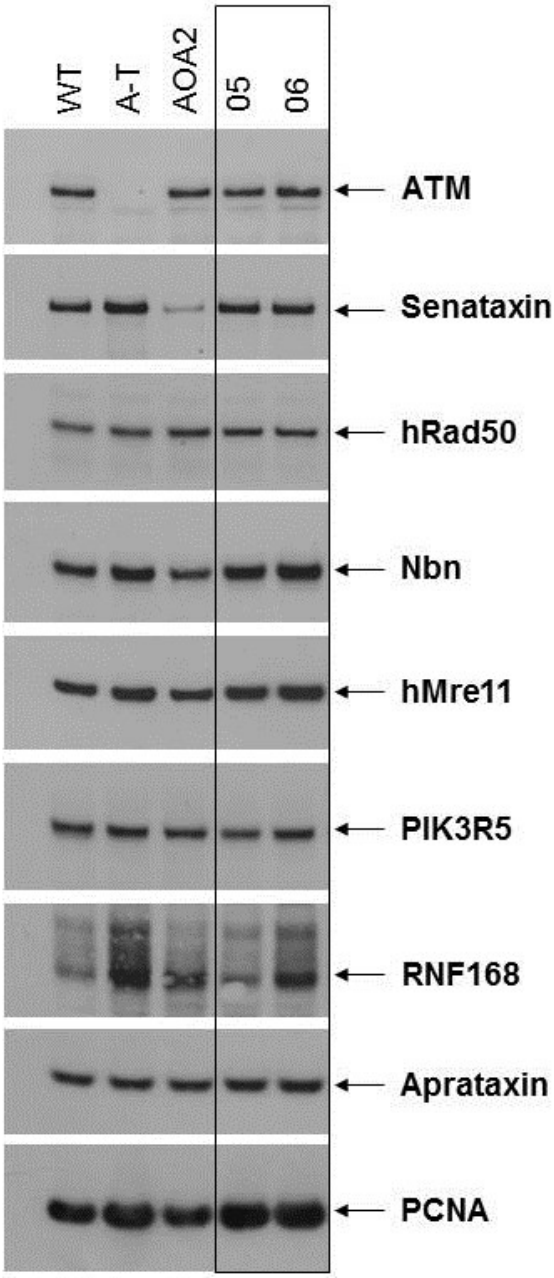
Given that the affected Amish individuals showed ataxic symptoms and cerebellar hypoplasia, similar to that seen in A-T, other A-T like disorders were also a possible cause of the phenotype. These include A-T like disorder (ATLD)(caused my *MRE11* mutation and reduction in Mre11 protein level), AOA2 (caused by reduction or loss of senataxin), AOA1 (caused by loss of aprataxin) and RIDDLE syndrome (caused by loss of RNF168). In lanes 4 & 5 is lysate from 05 and 06 lymphoblastoid cells showing normal levels of all the proteins tested.

There is thus no suggestion of a reduced level of ATM protein in 05 & 06 nor a reduced level of ATM kinase activity compared with normal. The normal level of ATM kinase activity is consistent with absence of pathogenic ATM mutations, comprehensively excluding AT as a cause of the phenotype. Normal levels of the proteins associated with ATLD, AOA1, AOA2 and RIDDLE syndrome in the context of the phenotype observed in the affected individuals effectively exclude these disorders as causative.



Cells homozygous for PCNA p.Ser228Ile have normal ATM kinase activity

ATM kinase levels in unirradiated (-) and irradiated (+) lymphoblastoid cell extracts. Lanes 1-2 and 9-10 are from a healthy control (WT) showing autophosphorylation of ATM (B) and phosphorylation of the targets SMC1 ser966 (C), KAP-1 ser824 (E), Nbn1 ser343 (G) and CREB ser121 (I). Lanes 3-4 and 11-12 show an absence of phosphorylation of these targets in AT cells (A-T). ATM kinase activity is unaffected in cells from a heterozygous ATM mutation carrier (lanes 11-12) and in 05 and 06 cells which are homozygous for PCNA p.Ser228Ile (lanes 5-8).



Cells homozygous for PCNA p.Ser228Ile have normal levels of proteins mutated in AT-like disorders

Western blot of extracts from EBV transformed control cells (WT), Ataxia telangiectasia cells (A-T), AOA2 cells (AOA2) and affected individuals homozygous for PCNA p.Ser228Ile (05 and 06). Normal levels of PCNA, ATM and additional proteins which when mutated cause AT-like disorders in cells derived from affected individuals.

7.1.3 PCNA functional studies

Immunofluorescence. EdU staining was carried out using the Click-iT EdU Alexa Fluor 555 Imaging Kit (Invitrogen). Cells were seeded onto glass coverslips so as to be subconfluent at the time of fixation. At least 24 h after plating EdU was added to the culture medium to a final concentration of 20 μM and cells returned to 37°C for 10 min. Soluble proteins were extracted by rinsing once in PBS, once in CSK buffer (100mM NaCl, 300mM sucrose, 10mM PIPES pH 7.0, 3mM MgCl2), and then soaking in CSK plus 0.2% Triton X-100 for 5 min on ice. After a further wash in CSK cells were fixed (ice-cold methanol for 20 min at -20°C), washed twice more in PBS, and then blocked (5% w/v bovine serum albumin (BSA) in PBST) for 30 min at room temperature. EdU staining was carried out according to manufacturer's instructions. Coverslips were then incubated for 1 h at room temperature in blocking buffer containing anti-PCNA monoclonal antibody (PC-10) diluted 1:1000, followed by 3 washes in PBST and incubation in Alexa Fluor 488 goat anti-rabbit immunoglobulin G (Invitrogen) diluted 1:1000 in blocking buffer for 45 min. Cells were rinsed a further 3 times in PBST, and coverslips mounted in Aqua-Poly/Mount (Polysciences) containing DAPI at a final concentration of 1.5μg/ml. Slides were visualized using a Leica SP5 confocal microscope.

DNA fibre analysis. Immunolabelling of DNA fiber spreads was carried out as described in (Unsal-Kacmaz et al., 2007), with minor modifications. Cells were incubated in IdU and CldU at a final concentration of 100 μM for 20 min each, and all antibody incubations were for 2 h at room temperature. Fibers were mounted in ProLong mounting medium (Invitrogen).

Cell studies. Unscheduled DNA synthesis (UDS), recovery of RNA synthesis (RRS), fibroblast survival assays and lentiviral mediated complementation were performed as described previously (Lehmann et al., 1993, Lehmann and Stevens, 1980, Arlett et al., 1992, Nakazawa et al., 2012, Nakazawa et al., 2010, Mayne and Lehmann, 1982).

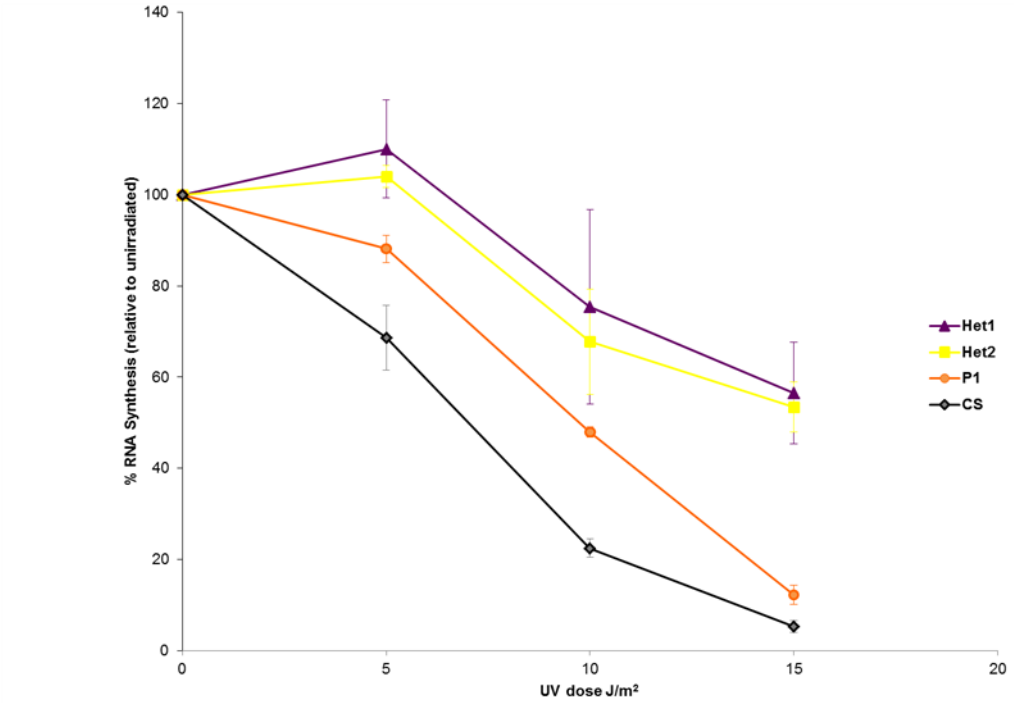
SILAC analysis. HeLa cells were grown in SILAC DMEM (Gibco) supplemented with 10% dialyzed FBS, 2mM $CaCl_2$, 1mM $MgSO_4$, 52mg/l L-

leucine and 100mg/l L-lysine and L-arginine (for “heavy” extracts [$^{13}C_6/^{15}N_2$] lysine and [$^{13}C_6/^{15}N_4$] arginine (CKgas) were substituted). Extracts were made by lysing cells in a low salt buffer containing benzonase (Merck), then adjusting to 150mM NaCl. All protein interaction assays were performed in interaction buffer (25mM Tris pH 7.5, 25mm NaCl, 10% glycerol, 0.01% Igepal, 1mM PMSF). Proteins eluted from PCNA affinity columns with 500 mM NaCl were combined before 10-fraction mass spectrometric (MS-MS) analysis (MS Bioworks). Data was analyzed, and H:L ratios determined using MaxQuant software.

MaxQuant outputs from the three SILAC repetitions for the previously characterized PCNA interactors

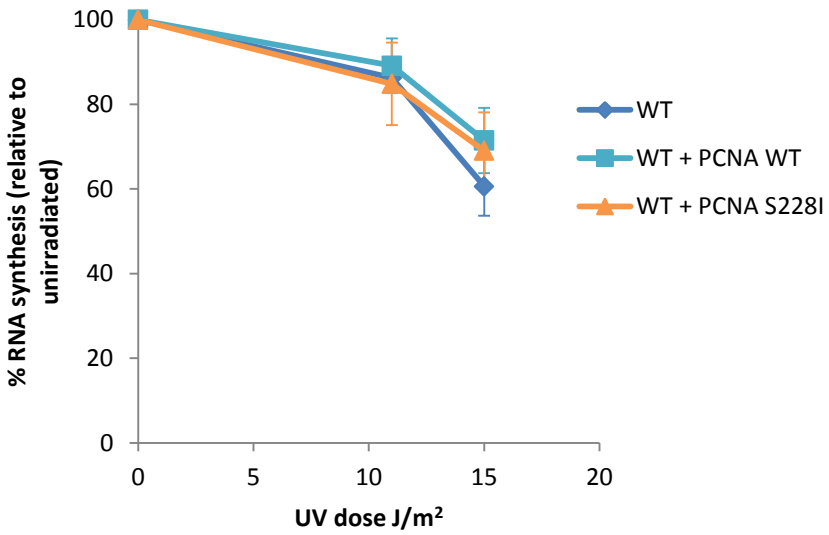
Protein	Total spectra	# unique peptides	Ratio h:l normalized	Total spectra	# unique peptides	Ratio h:l normalized	Total spectra	# unique peptides	Ratio h:l normalized
APEX1	141	28	0.11158	148	21	0.10287	28	15	1.5273
BUB3	29	14	1.0902	41	13	1.0567	59	17	0.7862
CDK1	12	8	1.0146	7	4	0.90092	7	6	1.0171
CHTF18	31	19	0.60947	30	14	0.63974	13	8	0.84334
DDB1	84	42	0.54328	38	24	0.50706	9	8	0.83215
DNMT1	210	72	0.88423	176	56	0.91885	115	61	0.83092
FEN1	43	21	11.544	63	18	18.617	20	14	0.10452
HIST1H1D	16	6	1.5639	13	5	2.8628	2	2	0.90141
LIG1	21	17	2.8494	22	11	3.2442	48	33	0.18195
MCM3	39	20	2.5048	28	21	1.7699	24	22	1.1822
MCM5	23	16	2.6787	21	9	2.0941	17	15	1.1049
MPG	19	11	0.72417	11	6	0.91538	13	10	0.24551
NAP1L1	27	8	0.80383	26	8	0.96409	71	14	0.23861
NAP1L4	33	11	0.79225	75	14	0.97825	93	18	0.25283
POLD1	161	58	0.41099	98	44	0.41942	79	46	2.5357
POLD2	41	19	0.46639	26	13	0.43519	18	13	2.1806
POLD3	33	13	0.42348	45	15	0.42695	24	12	1.7594
RBBP4	10	7	1.0121	6	3	0.76488	3	2	0.31712
RFC2	17	11	0.84726	17	11	0.87986	83	28	1.5911
RFC3	12	7	0.91678	5	5	0.86628	19	13	1.8282
RFC4	25	18	0.9077	29	16	0.92337	36	24	1.4823
RNASEH2A	7	7	0.22837	7	5	0.32916	8	6	1.8569
RNASEH2B	17	13	0.16996	17	8	0.20808	12	9	2.1405
RNASEH3C	14	8	0.15057	12	5	0.16212	12	9	2.7715
UNG	14	9	1.4155	12	5	2.0872	34	14	0.3147

A summary of relevant data from the mass spectrometric analysis of proteins eluted from PCNA affinity columns is presented. Repeats 1 and 2 were performed in the forward direction (extracts with “heavy” label on the p.Ser228Ile PCNA column, “light” extracts on the wild-type PCNA column). Repeat 3 was performed in the reverse direction (“heavy” extracts on the WT column, “light” extracts on the p.ser228Ile PCNA column). The number of spectra from each run assigned to each peptide by the MaxQuant software is given, as is the number of unique peptides making up those spectra. The “ratio h:l normalized” column gives the ratio of signals corresponding to peptides with “heavy” arginine or lysine over unlabeled peptides from the same protein. These ratios are normalized such the total h:l ratio in the entire sample is set to 1, to correct for slight differences in initial protein extract concentrations.



Deficient recovery of RNA synthesis in P1 cells

RNA synthesis measured by the incorporation of radiolabelled UTP 24 h after UVC-irradiation of non-dividing cells with indicated doses. Plotted is the mean ± SEM of three experiments. PCNA-P1 cells are defective in recovering RNA synthesis, to an extent approaching the defect seen in cells from a Cockayne syndrome (CS) patient, heterozygotes Het1 and 2 cell lines are normal.



Lentiviral mediated ectopic expression of PCNA has little effect on recovery of RNA synthesis in wild-type cells

Ectopic expression of wild-type or p.Ser228Ile PCNA was achieved by lentiviral transduction or GFP– or FLAG-tagged PCNA cDNAs. Virus infection was performed 48hours before RRS assays and viral infection efficiency (>85%) was confirmed by immunofluorescent staining. RNA synthesis was measured fluorescently by the incorporation of 5-ethynyluridine 12h after UVC-irradiation of non-dividing cells with indicated doses, revealed by copper-catalyzed fluorescent azide conjugation reaction (Click reaction). Plotted is the mean ± SEM of three experiments.

7.1.4 Site directed mutagenesis (SDM) primers used to generate PCNAp.Ser228Ile mutant constructs:

SDM oligonucleotides were designed following the QuikChange kit guidelines. The mutated base is highlighted in yellow, incorporation results in the Ser228Ile aminoacid substitution in PCNA.

Forward: 5'TTCAACGGTGACACTCATTATGTCTGCAGATGTACCC 3'

Reverse: 5'GTACATCTGCAGACATAATGAGTGTCACCGTTGAAGA 3'

7.2 APPENDIX CHAPTER 4

7.2.1 KPTN primer sequences

Primer Name	Forward	Reverse
KPTN1a	**CCTTGGTCCAGAAATGGAAA**	**CAGCCCGTACACATTGCTCT**
KPTN1b	**GCTTAACTGAGGGGCATGAT**	**GAGTCCTGACCACCATCACC**
KPTN2	**CTCAGACCAGAGGGCATGAT**	**GGGCTCGTAGTCGCAGTAAA**
KPTN3	**GCCCTCCTCCTTACACACAT**	**TGGAACGGAGTGAACTGGA**
KPTN4	**CCGGCTCTGAGTACAACCTT**	**CGGCTGGCATTTTAGTTTTC**
KPTN5	**GAACAGGGATTGGCAAACTT**	**GAAGTGCTGTTTTCTGTCTCCTT**
KPTN6	**TGAGGGCTCGTTTCTACAGTT**	**TAAGTCAGCCGTCAGCTTCA**
KPTN7	**CCTGTGGGATTCACTGTGG**	**CCCAATTCCTCTGAGGTCAA**
KPTN8	**GGCGGTGAGACCTTAGAGAA**	**AACCACTCTGACAGCCCAGA**
KPTN9	**GGGTGGTGAGAATTGACCTC**	**CTTATCCTCCTCCCCTACCC**
KPTN10 *	**GGGTGCAAATCTGGGTAGAA**	**GGCTATTGCTGGGTGAGGTA**
KPTN11	**GAGGCGGGGTTTTATCATGT**	**TCCAGCTCTGGGTAATCCAC**
KPTN12	**GCCTGAGAAGGGAAGCAGTA**	**GGGCCCCGGTACTTATAACA**
KPTN13	**GTGGGACCTGACAGTTGGAG**	**AGGAACTGCCTCCCTGAAAT**
KPTN14	**CAGGCTCTAGCGCAGTGA**	**TGGTGGCTCACACCTGTAAT**
KPTN15	**CTGTCACAGAGGCCCATTC**	**GGATCATCCCTGCTTTCAAAT**

7.2.2 RT-PCR *KPTN* cDNA primer sequences

Forward: CCAGAGCTGACGAACCTGACCAGTAGC

Reverse: CAGAGGACGCTGTCAAACTGGTCACTG

7.2.3 Methodology relating to the functional analysis of the Kaptin protein

7.2.3.1 Neuronal cell studies:

Human kaptin was cloned from HEK293 cell cDNA obtained by RNA isolation and reverse transcription PCR as previously described (Pinyol et al., 2007). Full-length kaptin (amino acids 1–436) was generated by PCR using primers:

BQ2046: AAGAATTCATGATGGGGGAGGCG

BQ2047: AAGGATCCTTAAGAGGCTGCATT

The PCR product generated was digested with *EcoRI* and *BamH1* and cloned in-frame into pEGFP-c2 and subcloned into pCMV-Tag2. Primary rat hippocampal cultures were prepared and cultured as described (Ahuja et al., 2007, Qualmann et al., 2004). Neurons were transfected with Lipofectamine 2000 (Invitrogen) at day 3, 13 and 23 in vitro, fixed in 4% PFA in PBS for 7 min at RT 24 and 48 h after transfection, respectively, and processed for immunofluorescence microscopy (Qualmann et al., 2004). Confocal imaging was performed using a Zeiss AxioObserver equipped with an ApoTome and Zeiss Plan-Apochromat 63 x/1.4 and 40 x/1.3 objectives and an AxioCam MRm CCD camera (Zeiss).

7.2.3.2 Generation of GFP tagged mutant Kaptin:

p.M241_Q246dup kaptin mutant (GFP-kaptin$^{M241\text{-}Q246dup}$) was generated by fusing an N-terminal portion carrying the mutation and a *SmaI* site introduced as a silent mutation (primers below).

BQ2046: AAGAATTCATGATGGGGGAGGCG

BQ2050:TCACCCGGGAGATGGGCCCGTCTTGCAACACGCTCCACATCTGC
AGGACCGACCAC

A C-terminal portion containing a *SmaI* site also introduced by silent mutation (primers below).

BQ2048: ATCTCCCGGGTGATTGTGTTCAG

BQ2047:AAGGATCCTTAAGAGGCTGCATT])

The p.S259X kaptin mutant (GFP-kaptin$^{1\text{-}258}$) was generated by PCR (primers below). PCR products were cloned in-frame into pEGFP-c2.

BQ2046:AAGAATTCATGATGGGGGAGGCG

BQ2070: AAGTCGACCtAGAGGCTGAACAC

Cartoon showing the construct used to generate the KPTN mouse

7.3 APPENDIX CHAPTER 5

WHAMM and WDR73 genotypes of the healthy siblings of nephrocerebellar syndrome affected individuals (pedigree shown in Figure 43)

Pedigree ID	*WHAMM* Genotype M denotes c.1264_1270delATAAAAG W denotes wildtype	*WDR73* Genotype M denotes c.888delT W denotes wildtype
X:2	M/W	M/W
X:3	M/W	M/W
X:4	W/W	W/W
X:5	W/W	W/W
X14	M/W	M/W
X:37	W/W	W/W
X:42	W/W	W/W
X:43	M/W	M/W
X:47	W/W	W/W
X:49	M/W	M/W
X:50	M/W	M/W
X:53	M/W	M/W

7.3.1 *SAMHD1* primers

FORWARD: CCTGGCCACTTTTTACAATTT

REVERSE: CATCTTTCTGTATTTTTCTAATTTGC

7.3.2 *SAMHD1* RT PCR primers

FORWARD: CAGCAATTGACGACATGGAAGCCTA

REVERSE: TGCTCTGTTGGGGGCAGTCTTACAA

7.3.3 *WDR73* primers

FORWARD: CAGGGACAAAACGCTACCAT

REVERSE: CATGGTGAAACCCCATCTCT

BIBLIOGRAPHY

ABECASIS, G. & WIGGINTON, J. 2005. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am J Hum Genet,* 77**,** 754-67.

AHUJA, R., PINYOL, R., REICHENBACH, N., CUSTER, L., KLINGENSMITH, J., KESSELS, M. M. & QUALMANN, B. 2007. Cordon-bleu is an actin nucleation factor and controls neuronal morphology. *Cell,* 131**,** 337-50.

AICARDI, J. 2002. Aicardi-Goutieres syndrome: special type early-onset encephalopathy. *Eur J Paediatr Neurol,* 6 Suppl A**,** A1-7; discussion A23-5, A77-86.

AICARDI, J. & GOUTIERES, F. 1984. A progressive familial encephalopathy in infancy with calcifications of the basal ganglia and chronic cerebrospinal fluid lymphocytosis. *Ann Neurol,* 15**,** 49-54.

ALKURAYA, F. S. 2010. Autozygome decoded. *Genet Med,* 12**,** 765-71.

ALKURAYA, F. S. 2013. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum Genet,* 132**,** 1197-211.

ALKURAYA, F. S., CAI, X., EMERY, C., MOCHIDA, G. H., AL-DOSARI, M. S., FELIE, J. M., HILL, R. S., BARRY, B. J., PARTLOW, J. N., GASCON, G. G., KENTAB, A., JAN, M., SHAHEEN, R., FENG, Y. & WALSH, C. A. 2011. Human mutations in NDE1 cause extreme microcephaly with lissencephaly [corrected]. *Am J Hum Genet,* 88**,** 536-47.

AMSELEM, S., DUQUESNOY, P., ATTREE, O., NOVELLI, G., BOUSNINA, S., POSTEL-VINAY, M. C. & GOOSSENS, M. 1989. Laron dwarfism and mutations of the growth hormone-receptor gene. *N Engl J Med,* 321**,** 989-95.

APA 2013. *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*.

ARAUJO, S. J., TIRODE, F., COIN, F., POSPIECH, H., SYVAOJA, J. E., STUCKI, M., HUBSCHER, U., EGLY, J. M. & WOOD, R. D. 2000. Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes Dev,* 14**,** 349-59.

ARLETT, C. F., HARCOURT, S. A., COLE, J., GREEN, M. H. & ANSTEY, A. V. 1992. A comparison of the response of unstimulated and stimulated T-lymphocytes and fibroblasts from normal, xeroderma pigmentosum and trichothiodystrophy donors to the lethal action of UV-C. *Mutat Res,* 273**,** 127-35.

ARNOLD, S. J., HUANG, G. J., CHEUNG, A. F., ERA, T., NISHIKAWA, S., BIKOFF, E. K., MOLNAR, Z., ROBERTSON, E. J. & GROSZER, M. 2008. The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. *Genes Dev,* 22**,** 2479-84.

BA, W., VAN DER RAADT, J. & NADIF KASRI, N. 2013. Rho GTPase signaling at the synapse: Implications for intellectual disability. *Experimental cell research*.

BARNES, D. E., TOMKINSON, A. E., LEHMANN, A. R., WEBSTER, A. D. & LINDAHL, T. 1992. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell,* 69**,** 495-503.

BEARER, E. L. & ABRAHAM, M. T. 1999. 2E4 (kaptin): a novel actin-associated protein from human blood platelets found in lamellipodia and the tips of the stereocilia of the inner ear. *European journal of cell biology,* 78**,** 117-26.

BEARER, E. L., CHEN, A. F., CHEN, A. H., LI, Z., MARK, H. F., SMITH, R. J. & JACKSON, C. L. 2000. 2E4/Kaptin (KPTN)--a candidate gene for the hearing loss locus, DFNA4. *Annals of Human Genetics,* 64**,** 189-96.

BEAULIEU, C. L., MAJEWSKI, J., SCHWARTZENTRUBER, J., SAMUELS, M. E., FERNANDEZ, B. A., BERNIER, F. P., BRUDNO, M., KNOPPERS, B., MARCADIER, J., DYMENT, D., ADAM, S., BULMAN, D. E., JONES, S. J., AVARD, D., NGUYEN, M. T., ROUSSEAU, F., MARSHALL, C., WINTLE, R. F., SHEN, Y., SCHERER, S. W., CONSORTIUM, F. C., FRIEDMAN, J. M., MICHAUD, J. L. & BOYCOTT, K. M. 2014. FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *Am J Hum Genet,* 94**,** 809-17.

BIACORE 1999. *BIACORE 3000 MANUAL*.

BICKNELL, L. S., BONGERS, E. M., LEITCH, A., BROWN, S., SCHOOTS, J., HARLEY, M. E., AFTIMOS, S., AL-AAMA, J. Y., BOBER, M., BROWN, P. A., VAN BOKHOVEN, H., DEAN, J., EDREES, A. Y., FEINGOLD, M., FRYER, A., HOEFSLOOT, L. H., KAU, N., KNOERS, N. V., MACKENZIE, J., OPITZ, J. M., SARDA, P., ROSS, A., TEMPLE, I. K., TOUTAIN, A., WISE, C. A., WRIGHT, M. & JACKSON, A. P. 2011a. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet,* 43**,** 356-9.

BICKNELL, L. S., WALKER, S., KLINGSEISEN, A., STIFF, T., LEITCH, A., KERZENDORFER, C., MARTIN, C. A., YEYATI, P., AL SANNA, N., BOBER, M., JOHNSON, D., WISE, C., JACKSON, A. P., O'DRISCOLL, M. & JEGGO, P. A. 2011b. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet,* 43**,** 350-5.

BILGUVAR, K., OZTURK, A. K., LOUVI, A., KWAN, K. Y., CHOI, M., TATLI, B., YALNIZOGLU, D., TUYSUZ, B., CAGLAYAN, A. O., GOKBEN, S., KAYMAKCALAN, H., BARAK, T., BAKIRCIOGLU, M., YASUNO, K., HO, W., SANDERS, S., ZHU, Y., YILMAZ, S., DINCER, A., JOHNSON, M. H., BRONEN, R. A., KOCER, N., PER, H., MANE, S., PAMIR, M. N., YALCINKAYA, C., KUMANDAS, S., TOPCU, M., OZMEN, M., SESTAN, N., LIFTON, R. P., STATE, M. W. & GUNEL, M. 2010. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature,* 467**,** 207-10.

BLOW, J. J. & DUTTA, A. 2005. Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol,* 6**,** 476-86.

BOND, J., ROBERTS, E., SPRINGELL, K., LIZARRAGA, S. B., SCOTT, S., HIGGINS, J., HAMPSHIRE, D. J., MORRISON, E. E., LEAL, G. F., SILVA, E. O., COSTA, S. M., BARALLE, D., RAPONI, M., KARBANI, G., RASHID, Y., JAFRI, H., BENNETT, C., CORRY, P., WALSH, C. A. & WOODS, C. G. 2005. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet,* 37**,** 353-5.

BRADFORD, P. T., GOLDSTEIN, A. M., TAMURA, D., KHAN, S. G., UEDA, T., BOYLE, J., OH, K. S., IMOTO, K., INUI, H., MORIWAKI, S., EMMERT, S., PIKE, K. M., RAZIUDDIN, A., PLONA, T. M., DIGIOVANNA, J. J., TUCKER, M. A. & KRAEMER, K. H. 2011. Cancer and neurologic

degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J Med Genet,* 48**,** 168-76.
BRUNING, J. B. & SHAMOO, Y. 2004. Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. *Structure,* 12**,** 2209-19.
BUSHBY, K. M., COLE, T., MATTHEWS, J. N. & GOODSHIP, J. A. 1992. Centiles for adult head circumference. *Arch Dis Child,* 67**,** 1286-7.
CAI, Q., WEN, W., QU, S., LI, G., EGAN, K. M., CHEN, K., DEMING, S. L., SHEN, H., SHEN, C. Y., GAMMON, M. D., BLOT, W. J., MATSUO, K., HAIMAN, C. A., KHOO, U. S., IWASAKI, M., SANTELLA, R. M., ZHANG, L., FAIR, A. M., HU, Z., WU, P. E., SIGNORELLO, L. B., TITUS-ERNSTOFF, L., TAJIMA, K., HENDERSON, B. E., CHAN, K. Y., KASUGA, Y., NEWCOMB, P. A., ZHENG, H., CUI, Y., WANG, F., SHIEH, Y. L., IWATA, H., LE MARCHAND, L., CHAN, S. Y., SHRUBSOLE, M. J., TRENTHAM-DIETZ, A., TSUGANE, S., GARCIA-CLOSAS, M., LONG, J., LI, C., SHI, J., HUANG, B., XIANG, Y. B., GAO, Y. T., LU, W., SHU, X. O. & ZHENG, W. 2011. Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of chinese, Japanese, and European ancestry. *Cancer Res,* 71**,** 1344-55.
CAMPBELL, K. & GOTZ, M. 2002. Radial glia: multi-purpose cells for vertebrate brain development. *Trends Neurosci,* 25**,** 235-8.
CAMPELLONE, K. G., WEBB, N. J., ZNAMEROSKI, E. A. & WELCH, M. D. 2008. WHAMM is an Arp2/3 complex activator that binds microtubules and functions in ER to Golgi transport. *Cell,* 134**,** 148-61.
CAMPELLONE, K. G. & WELCH, M. D. 2010. A nucleator arms race: cellular control of actin assembly. *Nat Rev Mol Cell Biol,* 11**,** 237-51.
CARR, I. M., BHASKAR, S., O'SULLIVAN, J., ALDAHMESH, M. A., SHAMSELDIN, H. E., MARKHAM, A. F., BONTHRON, D. T., BLACK, G. & ALKURAYA, F. S. 2013. Autozygosity mapping with exome sequence data. *Hum Mutat,* 34**,** 50-6.
CASTETS, F., BARTOLI, M., BARNIER, J. V., BAILLAT, G., SALIN, P., MOQRICH, A., BOURGEOIS, J. P., DENIZOT, F., ROUGON, G., CALOTHY, G. & MONNERON, A. 1996. A novel calmodulin-binding protein, belonging to the WD-repeat family, is localized in dendrites of a subset of CNS neurons. *J Cell Biol,* 134**,** 1051-62.
CHEN, T., HAN, Y., YANG, M., ZHANG, W., LI, N., WAN, T., GUO, J. & CAO, X. 2003. Rab39, a novel Golgi-associated Rab GTPase from human dendritic cells involved in cellular endocytosis. *Biochemical and biophysical research communications,* 303**,** 1114-20.
CHESARONE, M. A., DUPAGE, A. G. & GOODE, B. L. 2010. Unleashing formins to remodel the actin and microtubule cytoskeletons. *Nat Rev Mol Cell Biol,* 11**,** 62-74.
CLARKSON, S. G. 2003. The XPG story. *Biochimie,* 85**,** 1113-21.
CLEAVER, J. E. 2012. Photosensitivity syndrome brings to light a new transcription-coupled DNA repair cofactor. *Nat Genet,* 44**,** 477-8.
CLEAVER, J. E., LAM, E. T. & REVET, I. 2009. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet,* 10**,** 756-68.
CLIPPERTON-ALLEN, A. E. & PAGE, D. T. 2014. Pten haploinsufficient mice show broad brain overgrowth but selective impairments in autism-relevant behavioral tests. *Hum Mol Genet,* 23**,** 3490-505.

COHEN-KATSENELSON, K., WASSERMAN, T., KHATEB, S., WHITMARSH, A. J. & ARONHEIM, A. 2011. Docking interactions of the JNK scaffold protein WDR62. *Biochem J,* 439**,** 381-90.

COLE, T. J., FREEMAN, J. V. & PREECE, M. A. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med,* 17**,** 407-29.

CRINO, P. B., NATHANSON, K. L. & HENSKE, E. P. 2006. The tuberous sclerosis complex. *N Engl J Med,* 355**,** 1345-56.

CROSBY, A. H., PATEL, H., CHIOZA, B. A., PROUKAKIS, C., GURTZ, K., PATTON, M. A., SHARIFI, R., HARLALKA, G., SIMPSON, M. A., DICK, K., REED, J. A., AL-MEMAR, A., CHRZANOWSKA-LIGHTOWLERS, Z. M., CROSS, H. E. & LIGHTOWLERS, R. N. 2010. Defective mitochondrial mRNA maturation is associated with spastic ataxia. *Am J Hum Genet,* 87**,** 655-60.

CROSS, H. E. & MCKUSICK, V. A. 1967. A survey of neurological disorders in a genetic isolate. *Neurology,* 17**,** 743-51.

CROSSMAN, A. R. & NEARY, D. 2010. *Neuroanatomy*, Churchill Livingstone.

CROW, Y. J., HAYWARD, B. E., PARMAR, R., ROBINS, P., LEITCH, A., ALI, M., BLACK, D. N., VAN BOKHOVEN, H., BRUNNER, H. G., HAMEL, B. C., CORRY, P. C., COWAN, F. M., FRINTS, S. G., KLEPPER, J., LIVINGSTON, J. H., LYNCH, S. A., MASSEY, R. F., MERITET, J. F., MICHAUD, J. L., PONSOT, G., VOIT, T., LEBON, P., BONTHRON, D. T., JACKSON, A. P., BARNES, D. E. & LINDAHL, T. 2006a. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet,* 38**,** 917-20.

CROW, Y. J., LEITCH, A., HAYWARD, B. E., GARNER, A., PARMAR, R., GRIFFITH, E., ALI, M., SEMPLE, C., AICARDI, J., BABUL-HIRJI, R., BAUMANN, C., BAXTER, P., BERTINI, E., CHANDLER, K. E., CHITAYAT, D., CAU, D., DERY, C., FAZZI, E., GOIZET, C., KING, M. D., KLEPPER, J., LACOMBE, D., LANZI, G., LYALL, H., MARTINEZ-FRIAS, M. L., MATHIEU, M., MCKEOWN, C., MONIER, A., OADE, Y., QUARRELL, O. W., RITTEY, C. D., ROGERS, R. C., SANCHIS, A., STEPHENSON, J. B., TACKE, U., TILL, M., TOLMIE, J. L., TOMLIN, P., VOIT, T., WESCHKE, B., WOODS, C. G., LEBON, P., BONTHRON, D. T., PONTING, C. P. & JACKSON, A. P. 2006b. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet,* 38**,** 910-6.

CROW, Y. J. & LIVINGSTON, J. H. 2008. Aicardi-Goutieres syndrome: an important Mendelian mimic of congenital infection. *Dev Med Child Neurol,* 50**,** 410-6.

CROW, Y. J., VANDERVER, A., ORCESI, S., KUIJPERS, T. W. & RICE, G. I. 2014. Therapies in Aicardi-Goutieres syndrome. *Clin Exp Immunol,* 175**,** 1-8.

DAHL, J. P., WANG-DUNLOP, J., GONZALES, C., GOAD, M. E., MARK, R. J. & KWAK, S. P. 2003. Characterization of the WAVE1 knock-out mouse: implications for CNS development. *J Neurosci,* 23**,** 3343-52.

DALE, R. C., GORNALL, H., SINGH-GREWAL, D., ALCAUSIN, M., RICE, G. I. & CROW, Y. J. 2010. Familial Aicardi-Goutieres syndrome due to SAMHD1 mutations is associated with chronic arthropathy and contractures. *Am J Med Genet A,* 152A**,** 938-42.

DE LA CHAPELLE, A. 1993. Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet,* 30**,** 857-65.
DENT, E. W., MERRIAM, E. B. & HU, X. 2011. The dynamic cytoskeleton: backbone of dendritic spine plasticity. *Current opinion in neurobiology,* 21**,** 175-81.
DEPIENNE, C., STEVANIN, G., BRICE, A. & DURR, A. 2007. Hereditary spastic paraplegias: an update. *Curr Opin Neurol,* 20**,** 674-80.
DERIVERY, E., SOUSA, C., GAUTIER, J. J., LOMBARD, B., LOEW, D. & GAUTREAU, A. 2009. The Arp2/3 activator WASH controls the fission of endosomes through a large multiprotein complex. *Dev Cell,* 17**,** 712-23.
DERRY, J. M., OCHS, H. D. & FRANCKE, U. 1994. Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell,* 78**,** 635-44.
DIDERICH, K., ALANAZI, M. & HOEIJMAKERS, J. H. 2011. Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair (Amst),* 10**,** 772-80.
DIFFLEY, J. F. 2004. Regulation of early events in chromosome replication. *Curr Biol,* 14**,** R778-86.
DIGIOVANNA, J. J. & KRAEMER, K. H. 2012. Shining a light on xeroderma pigmentosum. *J Invest Dermatol,* 132**,** 785-96.
DJAGAEVA, I. & DORONKIN, S. 2009. COP9 limits dendritic branching via Cullin3-dependent degradation of the actin-crosslinking BTB-domain protein Kelch. *PLoS ONE,* 4**,** e7598.
DOMINGUEZ, R. 2009. Actin filament nucleation and elongation factors--structure-function relationships. *Crit Rev Biochem Mol Biol,* 44**,** 351-66.
DREVILLON, L., MEGARBANE, A., DEMEER, B., MATAR, C., BENIT, P., BRIAND-SULEAU, A., BODEREAU, V., GHOUMID, J., NASSER, M., DECROUY, X., DOCO-FENZY, M., RUSTIN, P., GAILLARD, D., GOOSSENS, M. & GIURGEA, I. 2013. KBP-cytoskeleton interactions underlie developmental anomalies in Goldberg-Shprintzen syndrome. *Hum Mol Genet,* 22**,** 2387-99.
DULEH, S. N. & WELCH, M. D. 2010. WASH and the Arp2/3 complex regulate endosome shape and trafficking. *Cytoskeleton (Hoboken),* 67**,** 193-206.
DUXIN, J. P., MOORE, H. R., SIDOROVA, J., KARANJA, K., HONAKER, Y., DAO, B., PIWNICA-WORMS, H., CAMPBELL, J. L., MONNAT, R. J., JR. & STEWART, S. A. 2012. Okazaki fragment processing-independent role for human Dna2 enzyme during DNA replication. *J Biol Chem,* 287**,** 21980-91.
EDERY, P., MARCAILLOU, C., SAHBATOU, M., LABALME, A., CHASTANG, J., TOURAINE, R., TUBACHER, E., SENNI, F., BOBER, M. B., NAMPOOTHIRI, S., JOUK, P. S., STEICHEN, E., BERLAND, S., TOUTAIN, A., WISE, C. A., SANLAVILLE, D., ROUSSEAU, F., CLERGET-DARPOUX, F. & LEUTENEGGER, A. L. 2011. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science,* 332**,** 240-3.
FALK, M. J., FEILER, H. S., NEILSON, D. E., MAXWELL, K., LEE, J. V., SEGALL, S. K., ROBIN, N. H., WILHELMSEN, K. C., TRASKELIN, A. L., KOLEHMAINEN, J., LEHESJOKI, A. E., WIZNITZER, M. & WARMAN, M. L. 2004. Cohen syndrome in the Ohio Amish. *Am J Med Genet A,* 128A**,** 23-8.
FENG, Y. & WALSH, C. A. 2004. Mitotic spindle regulation by Nde1 controls cerebral cortical size. *Neuron,* 44**,** 279-93.

FIRST, M. B. 2009. Harmonisation of ICD-11 and DSM-V: opportunities and challenges. *Br J Psychiatry,* 195**,** 382-90.

FRANCOMANO, C. A., MCKUSICK, V. A. & BIESECKER, L. G. 2003. Medical genetic studies in the Amish: historical perspective. *Am J Med Genet C Semin Med Genet,* 121C**,** 1-4.

FRANKENBURG, W. K., DODDS, J., ARCHER, P., SHAPIRO, H. & BRESNICK, B. 1992. The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics,* 89**,** 91-7.

GARY, R., LUDWIG, D. L., CORNELIUS, H. L., MACINNES, M. A. & PARK, M. S. 1997. The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem,* 272**,** 24522-9.

GIANNANDREA, M., BIANCHI, V., MIGNOGNA, M. L., SIRRI, A., CARRABINO, S., D'ELIA, E., VECELLIO, M., RUSSO, S., COGLIATI, F., LARIZZA, L., ROPERS, H. H., TZSCHACH, A., KALSCHEUER, V., OEHL-JASCHKOWITZ, B., SKINNER, C., SCHWARTZ, C. E., GECZ, J., VAN ESCH, H., RAYNAUD, M., CHELLY, J., DE BROUWER, A. P., TONIOLO, D. & D'ADAMO, P. 2010. Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *AMERICAN JOURNAL OF HUMAN GENETICS,* 86**,** 185-95.

GOMES, X. V. & BURGERS, P. M. 2000. Two modes of FEN1 binding to PCNA regulated by DNA. *EMBO J,* 19**,** 3811-21.

GOMEZ, T. S. & BILLADEAU, D. D. 2009. A FAM21-containing WASH complex regulates retromer-dependent sorting. *Dev Cell,* 17**,** 699-711.

GOMEZ, T. S., GORMAN, J. A., DE NARVAJAS, A. A., KOENIG, A. O. & BILLADEAU, D. D. 2012. Trafficking defects in WASH-knockout fibroblasts originate from collapsed endosomal and lysosomal networks. *Mol Biol Cell,* 23**,** 3215-28.

GORSKI, J. A., TALLEY, T., QIU, M., PUELLES, L., RUBENSTEIN, J. L. & JONES, K. R. 2002. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J Neurosci,* 22**,** 6309-14.

GREGORY, A. & HAYFLICK, S. 1993. Neurodegeneration with Brain Iron Accumulation Disorders Overview. *In:* PAGON, R. A., ADAM, M. P., ARDINGER, H. H., BIRD, T. D., DOLAN, C. R., FONG, C. T., SMITH, R. J. H. & STEPHENS, K. (eds.) *GeneReviews(R).* Seattle (WA).

GRIFFITH, E., WALKER, S., MARTIN, C. A., VAGNARELLI, P., STIFF, T., VERNAY, B., AL SANNA, N., SAGGAR, A., HAMEL, B., EARNSHAW, W. C., JEGGO, P. A., JACKSON, A. P. & O'DRISCOLL, M. 2008. Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nat Genet,* 40**,** 232-6.

GROSS, C., DE BAERE, E., LO, A., CHANG, W. & MESSIAEN, L. 2001. Cloning and characterization of human WDR10, a novel gene located at 3q21 encoding a WD-repeat protein that is highly expressed in pituitary and testis. *DNA Cell Biol,* 20**,** 41-52.

GUERNSEY, D. L., JIANG, H., HUSSIN, J., ARNOLD, M., BOUYAKDAN, K., PERRY, S., BABINEAU-STURK, T., BEIS, J., DUMAS, N., EVANS, S. C., FERGUSON, M., MATSUOKA, M., MACGILLIVRAY, C., NIGHTINGALE, M., PATRY, L., RIDEOUT, A. L., THOMAS, A., ORR, A.,

HOFFMANN, I., MICHAUD, J. L., AWADALLA, P., MEEK, D. C., LUDMAN, M. & SAMUELS, M. E. 2010. Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. *Am J Hum Genet,* 87**,** 40-51.

GULBIS, J. M., KELMAN, Z., HURWITZ, J., O'DONNELL, M. & KURIYAN, J. 1996. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell,* 87**,** 297-306.

GULSUNER, S., TEKINAY, A. B., DOERSCHNER, K., BOYACI, H., BILGUVAR, K., UNAL, H., ORS, A., ONAT, O. E., ATALAR, E., BASAK, A. N., TOPALOGLU, H., KANSU, T., TAN, M., TAN, U., GUNEL, M. & OZCELIK, T. 2011. Homozygosity mapping and targeted genomic sequencing reveal the gene responsible for cerebellar hypoplasia and quadrupedal locomotion in a consanguineous kindred. *Genome Res,* 21**,** 1995-2003.

GUO, Z., ZHENG, L., XU, H., DAI, H., ZHOU, M., PASCUA, M. R., CHEN, Q. M. & SHEN, B. 2010. Methylation of FEN1 suppresses nearby phosphorylation and facilitates PCNA binding. *Nat Chem Biol,* 6**,** 766-73.

GUVEN, A., GUNDUZ, A., BOZOGLU, T. M., YALCINKAYA, C. & TOLUN, A. 2012. Novel NDE1 homozygous mutation resulting in microhydranencephaly and not microlyssencephaly. *Neurogenetics,* 13**,** 189-94.

HAACK, T. B., HOGARTH, P., KRUER, M. C., GREGORY, A., WIELAND, T., SCHWARZMAYR, T., GRAF, E., SANFORD, L., MEYER, E., KARA, E., CUNO, S. M., HARIK, S. I., DANDU, V. H., NARDOCCI, N., ZORZI, G., DUNAWAY, T., TARNOPOLSKY, M., SKINNER, S., FRUCHT, S., HANSPAL, E., SCHRANDER-STUMPEL, C., HERON, D., MIGNOT, C., GARAVAGLIA, B., BHATIA, K., HARDY, J., STROM, T. M., BODDAERT, N., HOULDEN, H. H., KURIAN, M. A., MEITINGER, T., PROKISCH, H. & HAYFLICK, S. J. 2012. Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA. *Am J Hum Genet,* 91**,** 1144-9.

HAECKEL, A., AHUJA, R., GUNDELFINGER, E. D., QUALMANN, B. & KESSELS, M. M. 2008. The actin-binding protein Abp1 controls dendritic spine morphology and is important for spine head and synapse formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 28**,** 10031-44.

HARADA, Y. N., SHIOMI, N., KOIKE, M., IKAWA, M., OKABE, M., HIROTA, S., KITAMURA, Y., KITAGAWA, M., MATSUNAGA, T., NIKAIDO, O. & SHIOMI, T. 1999. Postnatal growth failure, short life span, and early onset of cellular senescence and subsequent immortalization in mice lacking the xeroderma pigmentosum group G gene. *Mol Cell Biol,* 19**,** 2366-72.

HARLALKA, G. V., BAPLE, E. L., CROSS, H., KUHNLE, S., CUBILLOS-ROJAS, M., MATENTZOGLU, K., PATTON, M. A., WAGNER, K., COBLENTZ, R., FORD, D. L., MACKAY, D. J., CHIOZA, B. A., SCHEFFNER, M., ROSA, J. L. & CROSBY, A. H. 2013a. Mutation of HERC2 causes developmental delay with Angelman-like features. *J Med Genet,* 50**,** 65-73.

HARLALKA, G. V., LEHMAN, A., CHIOZA, B., BAPLE, E. L., MAROOFIAN, R., CROSS, H., SREEKANTAN-NAIR, A., PRIESTMAN, D. A., AL-TURKI, S., MCENTAGART, M. E., PROUKAKIS, C., ROYLE, L., KOZAK, R. P., BASTAKI, L., PATTON, M., WAGNER, K., COBLENTZ, R., PRICE, J.,

MEZEI, M., SCHLADE-BARTUSIAK, K., PLATT, F. M., HURLES, M. E. & CROSBY, A. H. 2013b. Mutations in B4GALNT1 (GM2 synthase) underlie a new disorder of ganglioside biosynthesis. *Brain,* 136**,** 3618-24.

HAYASHI-NISHINO, M., FUJITA, N., NODA, T., YAMAGUCHI, A., YOSHIMORI, T. & YAMAMOTO, A. 2009. A subdomain of the endoplasmic reticulum forms a cradle for autophagosome formation. *Nat Cell Biol,* 11**,** 1433-7.

HAYFLICK, S. J., KRUER, M. C., GREGORY, A., HAACK, T. B., KURIAN, M. A., HOULDEN, H. H., ANDERSON, J., BODDAERT, N., SANFORD, L., HARIK, S. I., DANDU, V. H., NARDOCCI, N., ZORZI, G., DUNAWAY, T., TARNOPOLSKY, M., SKINNER, S., HOLDEN, K. R., FRUCHT, S., HANSPAL, E., SCHRANDER-STUMPEL, C., MIGNOT, C., HERON, D., SAUNDERS, D. E., KAMINSKA, M., LIN, J. P., LASCELLES, K., CUNO, S. M., MEYER, E., GARAVAGLIA, B., BHATIA, K., DE SILVA, R., CRISP, S., LUNT, P., CAREY, M., HARDY, J., MEITINGER, T., PROKISCH, H. & HOGARTH, P. 2013. beta-Propeller protein-associated neurodegeneration: a new X-linked dominant disorder with brain iron accumulation. *Brain,* 136**,** 1708-17.

HENDERSON, D. S., BANGA, S. S., GRIGLIATTI, T. A. & BOYD, J. B. 1994. Mutagen sensitivity and suppression of position-effect variegation result from mutations in mus209, the Drosophila gene encoding PCNA. *EMBO J,* 13**,** 1450-9.

HERING, H. & SHENG, M. 2001. Dendritic spines: structure, dynamics and regulation. *Nature reviews. Neuroscience,* 2**,** 880-8.

HOEK, M., MYERS, M. P. & STILLMAN, B. 2011. An analysis of CAF-1-interacting proteins reveals dynamic and direct interactions with the KU complex and 14-3-3 proteins. *J Biol Chem,* 286**,** 10876-87.

HURST, C. E. & MCCONNELL, D. L. 2010. *An Amish Paradox*.

IDE, M., ESAKI, Y., YAMAZAKI, K. & KATO, H. 2003. Newly developed T-wave inversion with cardiac wall-motion abnormality predominantly occurs in middle-aged or elderly women after noncardiac surgery. *J Anesth,* 17**,** 79-83.

INNOCENTI, M., GERBOTH, S., ROTTNER, K., LAI, F. P., HERTZOG, M., STRADAL, T. E., FRITTOLI, E., DIDRY, D., POLO, S., DISANZA, A., BENESCH, S., DI FIORE, P. P., CARLIER, M. F. & SCITA, G. 2005. Abi1 regulates the activity of N-WASP and WAVE in distinct actin-based processes. *Nat Cell Biol,* 7**,** 969-76.

IRIE, F. & YAMAGUCHI, Y. 2002. EphB receptors regulate dendritic spine development via intersectin, Cdc42 and N-WASP. *Nature neuroscience,* 5**,** 1117-8.

ITO, S., KURAOKA, I., CHYMKOWITCH, P., COMPE, E., TAKEDACHI, A., ISHIGAMI, C., COIN, F., EGLY, J. M. & TANAKA, K. 2007. XPG stabilizes TFIIH, allowing transactivation of nuclear receptors: implications for Cockayne syndrome in XP-G/CS patients. *Mol Cell,* 26**,** 231-43.

KAINDL, A. M., PASSEMARD, S., KUMAR, P., KRAEMER, N., ISSA, L., ZWIRNER, A., GERARD, B., VERLOES, A., MANI, S. & GRESSENS, P. 2010. Many roads lead to primary autosomal recessive microcephaly. *Prog Neurobiol,* 90**,** 363-83.

KALAY, E., YIGIT, G., ASLAN, Y., BROWN, K. E., POHL, E., BICKNELL, L. S., KAYSERILI, H., LI, Y., TUYSUZ, B., NURNBERG, G., KIESS, W., KOEGL, M., BAESSMANN, I., BURUK, K., TORAMAN, B., KAYIPMAZ,

S., KUL, S., IKBAL, M., TURNER, D. J., TAYLOR, M. S., AERTS, J., SCOTT, C., MILSTEIN, K., DOLLFUS, H., WIECZOREK, D., BRUNNER, H. G., HURLES, M., JACKSON, A. P., RAUCH, A., NURNBERG, P., KARAGUZEL, A. & WOLLNIK, B. 2011. CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat Genet,* 43**,** 23-6.

KAMILERI, I., KARAKASILIOTI, I. & GARINIS, G. A. 2012. Nucleotide excision repair: new tricks with old bricks. *Trends Genet,* 28**,** 566-73.

KANNER, L. 1968. Autistic disturbances of affective contact. *Acta Paedopsychiatr,* 35**,** 100-36.

KANNOUCHE, P. L., WING, J. & LEHMANN, A. R. 2004. Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell,* 14**,** 491-500.

KASRI, N. N., GOVEK, E. E. & VAN AELST, L. 2008. Characterization of oligophrenin-1, a RhoGAP lost in patients affected with mental retardation: lentiviral injection in organotypic brain slice cultures. *Methods in enzymology,* 439**,** 255-66.

KESIDOU, E., LAGOUDAKI, R., TOULOUMI, O., POULATSIDOU, K. N. & SIMEONIDOU, C. 2013. Autophagy and neurodegenerative disorders. *Neural Regen Res,* 8**,** 2275-83.

KESSELS, M. M., SCHWINTZER, L., SCHLOBINSKI, D. & QUALMANN, B. 2011. Controlling actin cytoskeletal organization and dynamics during neuronal morphogenesis. *European journal of cell biology,* 90**,** 926-33.

KIM, I. H., RACZ, B., WANG, H., BURIANEK, L., WEINBERG, R., YASUDA, R., WETSEL, W. C. & SODERLING, S. H. 2013. Disruption of Arp2/3 results in asymmetric structural plasticity of dendritic spines and progressive synaptic and behavioral abnormalities. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 33**,** 6081-92.

KLEIN, S., SHARIFI-HANNAUER, P. & MARTINEZ-AGOSTO, J. A. 2013. Macrocephaly as a Clinical Indicator of Genetic Subtypes in Autism. *Autism Research,* 6**,** 51-56.

KOCH, N., DHARMALINGAM, E., WESTERMANN, M., QUALMANN, B., THOMAS, U. & KESSELS, M. M. 2012. Abp1 utilizes the Arp2/3 complex activator Scar/WAVE in bristle development. *Journal of cell science,* 125**,** 3578-89.

KOOB, M., LAUGEL, V., DURAND, M., FOTHERGILL, H., DALLOZ, C., SAUVANAUD, F., DOLLFUS, H., NAMER, I. J. & DIETEMANN, J. L. 2010. Neuroimaging in Cockayne syndrome. *AJNR Am J Neuroradiol,* 31**,** 1623-30.

KRAEMER, K. H., PATRONAS, N. J., SCHIFFMANN, R., BROOKS, B. P., TAMURA, D. & DIGIOVANNA, J. J. 2007. Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience,* 145**,** 1388-96.

KRAYBILL, D. B., JOHNSON-WEINER, K. M. & NOLT, S. M. 2013. The Amish.

KRISHNA, T. S., KONG, X. P., GARY, S., BURGERS, P. M. & KURIYAN, J. 1994. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell,* 79**,** 1233-43.

KUIDA, K., HAYDAR, T. F., KUAN, C. Y., GU, Y., TAYA, C., KARASUYAMA, H., SU, M. S., RAKIC, P. & FLAVELL, R. A. 1998. Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell,* 94**,** 325-37.

LAHIRY, P., RACACHO, L., WANG, J., ROBINSON, J. F., GLOOR, G. B., RUPAR, C. A., SIU, V. M., BULMAN, D. E. & HEGELE, R. A. 2013. A mutation in the serine protease TMPRSS4 in a novel pediatric neurodegenerative disorder. *Orphanet J Rare Dis,* 8**,** 126.

LANCASTER, M. A., RENNER, M., MARTIN, C. A., WENZEL, D., BICKNELL, L. S., HURLES, M. E., HOMFRAY, T., PENNINGER, J. M., JACKSON, A. P. & KNOBLICH, J. A. 2013. Cerebral organoids model human brain development and microcephaly. *Nature,* 501**,** 373-9.

LANDER, E. S. & BOTSTEIN, D. 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science,* 236**,** 1567-70.

LANGERAK, P., KRIJGER, P. H., HEIDEMAN, M. R., VAN DEN BERK, P. C. & JACOBS, H. 2009. Somatic hypermutation of immunoglobulin genes: lessons from proliferating cell nuclear antigenK164R mutant mice. *Philos Trans R Soc Lond B Biol Sci,* 364**,** 621-9.

LANGERAK, P., NYGREN, A. O., KRIJGER, P. H., VAN DEN BERK, P. C. & JACOBS, H. 2007. A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J Exp Med,* 204**,** 1989-98.

LEE-KIRSCH, M. A., GONG, M., CHOWDHURY, D., SENENKO, L., ENGEL, K., LEE, Y. A., DE SILVA, U., BAILEY, S. L., WITTE, T., VYSE, T. J., KERE, J., PFEIFFER, C., HARVEY, S., WONG, A., KOSKENMIES, S., HUMMEL, O., ROHDE, K., SCHMIDT, R. E., DOMINICZAK, A. F., GAHR, M., HOLLIS, T., PERRINO, F. W., LIEBERMAN, J. & HUBNER, N. 2007. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet,* 39**,** 1065-7.

LEE, J. H., HUYNH, M., SILHAVY, J. L., KIM, S., DIXON-SALAZAR, T., HEIBERG, A., SCOTT, E., BAFNA, V., HILL, K. J., COLLAZO, A., FUNARI, V., RUSS, C., GABRIEL, S. B., MATHERN, G. W. & GLEESON, J. G. 2012. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet,* 44**,** 941-5.

LEHMANN, A. R. 2001. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev,* 15**,** 15-23.

LEHMANN, A. R. 2011. DNA polymerases and repair synthesis in NER in human cells. *DNA Repair (Amst),* 10**,** 730-3.

LEHMANN, A. R., MCGIBBON, D. & STEFANINI, M. 2011. Xeroderma pigmentosum. *Orphanet J Rare Dis,* 6**,** 70.

LEHMANN, A. R., NIIMI, A., OGI, T., BROWN, S., SABBIONEDA, S., WING, J. F., KANNOUCHE, P. L. & GREEN, C. M. 2007. Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst),* 6**,** 891-9.

LEHMANN, A. R. & STEVENS, S. 1980. A rapid procedure for measurement of DNA repair in human fibroblasts and for complementation analysis of xeroderma pigmentosum cells. *Mutat Res,* 69**,** 177-90.

LEHMANN, A. R., THOMPSON, A. F., HARCOURT, S. A., STEFANINI, M. & NORRIS, P. G. 1993. Cockayne's syndrome: correlation of clinical features with cellular sensitivity of RNA synthesis to UV irradiation. *J Med Genet,* 30**,** 679-82.

LIAN, G., LU, J., HU, J., ZHANG, J., CROSS, S. H., FERLAND, R. J. & SHEEN, V. L. 2012. Filamin a regulates neural progenitor proliferation and cortical

size through Wee1-dependent Cdk1 phosphorylation. *J Neurosci,* 32**,** 7672-84.

LINARDOPOULOU, E. V., PARGHI, S. S., FRIEDMAN, C., OSBORN, G. E., PARKHURST, S. M. & TRASK, B. J. 2007. Human subtelomeric WASH genes encode a new subclass of the WASP family. *PLoS Genet,* 3**,** e237.

LOMMEL, S., BENESCH, S., ROTTNER, K., FRANZ, T., WEHLAND, J. & KUHN, R. 2001. Actin pedestal formation by enteropathogenic Escherichia coli and intracellular motility of Shigella flexneri are abolished in N-WASP-defective cells. *EMBO Rep,* 2**,** 850-7.

MAILAND, N., GIBBS-SEYMOUR, I. & BEKKER-JENSEN, S. 2013. Regulation of PCNA-protein interactions for genome stability. *Nat Rev Mol Cell Biol,* 14**,** 269-82.

MARTINEZ-QUILES, N., HO, H. Y., KIRSCHNER, M. W., RAMESH, N. & GEHA, R. S. 2004. Erk/Src phosphorylation of cortactin acts as a switch on-switch off mechanism that controls its ability to activate N-WASP. *Molecular and cellular biology,* 24**,** 5269-80.

MAYNE, L. V. & LEHMANN, A. R. 1982. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res,* 42**,** 1473-8.

MCDONALD, L., RENNIE, A., TOLMIE, J., GALLOWAY, P. & MCWILLIAM, R. 2006. Investigation of global developmental delay. *Arch Dis Child,* 91**,** 701-5.

MCKUSICK, V. A. 2007. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet,* 80**,** 588-604.

MCKUSICK, V. A. & CROSS, H. E. 1966. Ataxia-telangiectasia and Swiss-type agammaglobulinemia. Two genetic disorders of the immune mechanism in related Amish sibships. *JAMA,* 195**,** 739-45.

MCKUSICK, V. A., ELDRIDGE, R., HOSTETLER, J. A., RUANGWIT, U. & EGELAND, J. A. 1965. Dwarfism in the Amish. Ii. Cartilage-Hair Hypoplasia. *Bull Johns Hopkins Hosp,* 116**,** 285-326.

MEMON, M. M., RAZA, S. I., BASIT, S., KOUSAR, R., AHMAD, W. & ANSAR, M. 2013. A novel WDR62 mutation causes primary microcephaly in a Pakistani family. *Mol Biol Rep,* 40**,** 591-5.

MICHAELSEN, K., MURK, K., ZAGREBELSKY, M., DREZNJAK, A., JOCKUSCH, B. M., ROTHKEGEL, M. & KORTE, M. 2010. Fine-tuning of neuronal architecture requires two profilin isoforms. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 15780-5.

MIKI, H., MIURA, K. & TAKENAWA, T. 1996. N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskeletal rearrangement in a PIP2-dependent manner downstream of tyrosine kinases. *EMBO J,* 15**,** 5326-35.

MIKI, H., SUETSUGU, S. & TAKENAWA, T. 1998. WAVE, a novel WASP-family protein involved in actin reorganization induced by Rac. *EMBO J,* 17**,** 6932-41.

MIRZAA, G. M. & PACIORKOWSKI, A. R. 2014. Introduction: Brain malformations. *Am J Med Genet C Semin Med Genet,* 166C**,** 117-23.

MIRZAA, G. M. & PODURI, A. 2014. Megalencephaly and hemimegalencephaly: breakthroughs in molecular etiology. *Am J Med Genet C Semin Med Genet,* 166C**,** 156-72.

MIRZAA, G. M., VITRE, B., CARPENTER, G., ABRAMOWICZ, I., GLEESON, J. G., PACIORKOWSKI, A. R., CLEVELAND, D. W., DOBYNS, W. B. & O'DRISCOLL, M. 2014. Mutations in CENPE define a novel kinetochore-centromeric mechanism for microcephalic primordial dwarfism. *Hum Genet,* 133**,** 1023-39.

MIYACHI, K., FRITZLER, M. J. & TAN, E. M. 1978. Autoantibody to a nuclear antigen in proliferating cells. *J Immunol,* 121**,** 2228-34.

MOLDOVAN, G. L., PFANDER, B. & JENTSCH, S. 2007. PCNA, the maestro of the replication fork. *Cell,* 129**,** 665-79.

MORTON, D. H., MORTON, C. S., STRAUSS, K. A., ROBINSON, D. L., PUFFENBERGER, E. G., HENDRICKSON, C. & KELLEY, R. I. 2003. Pediatric medicine and the genetic disorders of the Amish and Mennonite people of Pennsylvania. *Am J Med Genet C Semin Med Genet,* 121C**,** 5-17.

MORY, A., DAGAN, E., SHAHOR, I., MANDEL, H., ILLI, B., ZOLOTUSHKO, J., KUROLAP, A., CHECHIK, E., VALENTE, E. M., AMSELEM, S. & GERSHONI-BARUCH, R. 2014. Kohlschutter-Tonz syndrome: clinical and genetic insights gained from 16 cases deriving from a close-knit village in Northern Israel. *Pediatr Neurol,* 50**,** 421-6.

MURRAY, J. E., BICKNELL, L. S., YIGIT, G., DUKER, A. L., VAN KOGELENBERG, M., HAGHAYEGH, S., WIECZOREK, D., KAYSERILI, H., ALBERT, M. H., WISE, C. A., BRANDON, J., KLEEFSTRA, T., WARRIS, A., VAN DER FLIER, M., BAMFORTH, J. S., DOONANCO, K., ADES, L., MA, A., FIELD, M., JOHNSON, D., SHACKLEY, F., FIRTH, H., WOODS, C. G., NURNBERG, P., GATTI, R. A., HURLES, M., BOBER, M. B., WOLLNIK, B. & JACKSON, A. P. 2014. Extreme growth failure is a common presentation of ligase IV deficiency. *Hum Mutat,* 35**,** 76-85.

MURRAY, J. E. & JACKSON, A. P. 2012. Exploring microcephaly and human brain evolution. *Dev Med Child Neurol,* 54**,** 580-1.

MYSZKA, D. G. 1999. Improving biosensor analysis. *J Mol Recognit,* 12**,** 279-84.

NAJMABADI, H., HU, H., GARSHASBI, M., ZEMOJTEL, T., ABEDINI, S. S., CHEN, W., HOSSEINI, M., BEHJATI, F., HAAS, S., JAMALI, P., ZECHA, A., MOHSENI, M., PUTTMANN, L., VAHID, L. N., JENSEN, C., MOHEB, L. A., BIENEK, M., LARTI, F., MUELLER, I., WEISSMANN, R., DARVISH, H., WROGEMANN, K., HADAVI, V., LIPKOWITZ, B., ESMAEELI-NIEH, S., WIECZOREK, D., KARIMINEJAD, R., FIROUZABADI, S. G., COHEN, M., FATTAHI, Z., ROST, I., MOJAHEDI, F., HERTZBERG, C., DEHGHAN, A., RAJAB, A., BANAVANDI, M. J., HOFFER, J., FALAH, M., MUSANTE, L., KALSCHEUER, V., ULLMANN, R., KUSS, A. W., TZSCHACH, A., KAHRIZI, K. & ROPERS, H. H. 2011. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature,* 478**,** 57-63.

NAKABAYASHI, K., AMANN, D., REN, Y., SAARIALHO-KERE, U., AVIDAN, N., GENTLES, S., MACDONALD, J. R., PUFFENBERGER, E. G., CHRISTIANO, A. M., MARTINEZ-MIR, A., SALAS-ALANIS, J. C., RIZZO, R., VAMOS, E., RAAMS, A., LES, C., SEBOUN, E., JASPERS, N. G., BECKMANN, J. S., JACKSON, C. E. & SCHERER, S. W. 2005. Identification of C7orf11 (TTDN1) gene mutations and genetic heterogeneity in nonphotosensitive trichothiodystrophy. *Am J Hum Genet,* 76**,** 510-6.

NAKAMURA, Y., WOOD, C. L., PATTON, A. P., JAAFARI, N., HENLEY, J. M., MELLOR, J. R. & HANLEY, J. G. 2011. PICK1 inhibition of the Arp2/3 complex controls dendritic spine size and synaptic plasticity. *The EMBO journal,* 30**,** 719-30.

NAKAYAMA, A. Y. & LUO, L. 2000. Intracellular signaling pathways that regulate dendritic spine morphogenesis. *Hippocampus,* 10**,** 582-6.

NAKAZAWA, Y., SASAKI, K., MITSUTAKE, N., MATSUSE, M., SHIMADA, M., NARDO, T., TAKAHASHI, Y., OHYAMA, K., ITO, K., MISHIMA, H., NOMURA, M., KINOSHITA, A., ONO, S., TAKENAKA, K., MASUYAMA, R., KUDO, T., SLOR, H., UTANI, A., TATEISHI, S., YAMASHITA, S., STEFANINI, M., LEHMANN, A. R., YOSHIURA, K. & OGI, T. 2012. Mutations in UVSSA cause UV-sensitive syndrome and impair RNA polymerase IIo processing in transcription-coupled nucleotide-excision repair. *Nat Genet,* 44**,** 586-92.

NAKAZAWA, Y., YAMASHITA, S., LEHMANN, A. R. & OGI, T. 2010. A semi-automated non-radioactive system for measuring recovery of RNA synthesis and unscheduled DNA synthesis using ethynyluracil derivatives. *DNA Repair (Amst),* 9**,** 506-16.

NANCE, M. A. & BERRY, S. A. 1992. Cockayne syndrome: review of 140 cases. *Am J Med Genet,* 42**,** 68-84.

NATALE, V. & RAJAGOPALAN, A. 2014. Worldwide variation in human growth and the World Health Organization growth standards: a systematic review. *BMJ Open,* 4**,** e003735.

NEVO, Y., KRAMER, U., SHINNAR, S., LEITNER, Y., FATTAL-VALEVSKI, A., VILLA, Y. & HAREL, S. 2002. Macrocephaly in children with developmental disabilities. *Pediatr Neurol,* 27**,** 363-8.

NIIMI, A., BROWN, S., SABBIONEDA, S., KANNOUCHE, P. L., SCOTT, A., YASUI, A., GREEN, C. M. & LEHMANN, A. R. 2008. Regulation of proliferating cell nuclear antigen ubiquitination in mammalian cells. *Proc Natl Acad Sci U S A,* 105**,** 16125-30.

NOUSPIKEL, T., LALLE, P., LEADON, S. A., COOPER, P. K. & CLARKSON, S. G. 1997. A common mutational pattern in Cockayne syndrome patients from xeroderma pigmentosum group G: implications for a second XPG function. *Proc Natl Acad Sci U S A,* 94**,** 3116-21.

O'DRISCOLL, M., RUIZ-PEREZ, V. L., WOODS, C. G., JEGGO, P. A. & GOODSHIP, J. A. 2003. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nat Genet,* 33**,** 497-501.

OLNEY, A. H. 2007. Macrocephaly syndromes. *Semin Pediatr Neurol,* 14**,** 128-35.

PARKER, W. E., ORLOVA, K. A., PARKER, W. H., BIRNBAUM, J. F., KRYMSKAYA, V. P., GONCHAROV, D. A., BAYBIS, M., HELFFERICH, J., OKOCHI, K., STRAUSS, K. A. & CRINO, P. B. 2013. Rapamycin prevents seizures after depletion of STRADA in a rare neurodevelopmental disorder. *Sci Transl Med,* 5**,** 182ra53.

PATEL, H., CROSS, H., PROUKAKIS, C., HERSHBERGER, R., BORK, P., CICCARELLI, F. D., PATTON, M. A., MCKUSICK, V. A. & CROSBY, A. H. 2002. SPG20 is mutated in Troyer syndrome, an hereditary spastic paraplegia. *Nat Genet,* 31**,** 347-8.

PERLMAN, S., BECKER-CATANIA, S. & GATTI, R. A. 2003. Ataxia-telangiectasia: diagnosis and treatment. *Semin Pediatr Neurol,* 10**,** 173-82.

PERMANYER, J., NAVARRO, R., FRIEDMAN, J., POMARES, E., CASTRO-NAVARRO, J., MARFANY, G., SWAROOP, A. & GONZALEZ-DUARTE, R. 2010. Autosomal recessive retinitis pigmentosa with early macular affectation caused by premature truncation in PROM1. *Invest Ophthalmol Vis Sci.* United States.

PEROU, R., BITSKO, R. H., BLUMBERG, S. J., PASTOR, P., GHANDOUR, R. M., GFROERER, J. C., HEDDEN, S. L., CROSBY, A. E., VISSER, S. N., SCHIEVE, L. A., PARKS, S. E., HALL, J. E., BRODY, D., SIMILE, C. M., THOMPSON, W. W., BAIO, J., AVENEVOLI, S., KOGAN, M. D., HUANG, L. N., CENTERS FOR DISEASE, C. & PREVENTION 2013. Mental health surveillance among children--United States, 2005-2011. *MMWR Surveill Summ,* 62 Suppl 2**,** 1-35.

PINYOL, R., HAECKEL, A., RITTER, A., QUALMANN, B. & KESSELS, M. M. 2007. Regulation of N-WASP and the Arp2/3 complex by Abp1 controls neuronal morphology. *PLoS ONE,* 2**,** e400.

PODURI, A., EVRONY, G. D., CAI, X., ELHOSARY, P. C., BEROUKHIM, R., LEHTINEN, M. K., HILLS, L. B., HEINZEN, E. L., HILL, A., HILL, R. S., BARRY, B. J., BOURGEOIS, B. F., RIVIELLO, J. J., BARKOVICH, A. J., BLACK, P. M., LIGON, K. L. & WALSH, C. A. 2012. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron,* 74**,** 41-8.

POROLLO, A. & MELLER, J. 2007. Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics,* 8**,** 316.

PUFFENBERGER, E. G., JINKS, R. N., SOUGNEZ, C., CIBULSKIS, K., WILLERT, R. A., ACHILLY, N. P., CASSIDY, R. P., FIORENTINI, C. J., HEIKEN, K. F., LAWRENCE, J. J., MAHONEY, M. H., MILLER, C. J., NAIR, D. T., POLITI, K. A., WORCESTER, K. N., SETTON, R. A., DIPIAZZA, R., SHERMAN, E. A., EASTMAN, J. T., FRANCKLYN, C., ROBEY-BOND, S., RIDER, N. L., GABRIEL, S., MORTON, D. H. & STRAUSS, K. A. 2012. Genetic mapping and exome sequencing identify variants associated with five novel diseases. *PLoS One,* 7**,** e28936.

QUALMANN, B., BOECKERS, T. M., JEROMIN, M., GUNDELFINGER, E. D. & KESSELS, M. M. 2004. Linkage of the actin cytoskeleton to the postsynaptic density via direct interactions of Abp1 with the ProSAP/Shank family. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 24**,** 2481-95.

RAKIC, P. 1995. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci,* 18**,** 383-8.

RAKIC, P. 2007. The radial edifice of cortical architecture: from neuronal silhouettes to genetic engineering. *Brain Res Rev,* 55**,** 204-19.

RAKIC, P. & LOMBROSO, P. J. 1998. Development of the cerebral cortex: I. Forming the cortical structure. *J Am Acad Child Adolesc Psychiatry,* 37**,** 116-7.

RAMANTANI, G., KOHLHASE, J., HERTZBERG, C., INNES, A. M., ENGEL, K., HUNGER, S., BOROZDIN, W., MAH, J. K., UNGERATH, K., WALKENHORST, H., RICHARDT, H. H., BUCKARD, J., BEVOT, A., SIEGEL, C., VON STULPNAGEL, C., IKONOMIDOU, C., THOMAS, K., PROUD, V., NIEMANN, F., WIECZOREK, D., HAUSLER, M., NIGGEMANN, P., BALTACI, V., CONRAD, K., LEBON, P. & LEE-KIRSCH, M. A. 2010. Expanding the phenotypic spectrum of lupus

erythematosus in Aicardi-Goutieres syndrome. *Arthritis Rheum,* 62**,** 1469-77.

RAMESH, V., BERNARDI, B., STAFA, A., GARONE, C., FRANZONI, E., ABINUN, M., MITCHELL, P., MITRA, D., FRISWELL, M., NELSON, J., SHALEV, S. A., RICE, G. I., GORNALL, H., SZYNKIEWICZ, M., AYMARD, F., GANESAN, V., PRENDIVILLE, J., LIVINGSTON, J. H. & CROW, Y. J. 2010. Intracerebral large artery disease in Aicardi-Goutieres syndrome implicates SAMHD1 in vascular homeostasis. *Dev Med Child Neurol,* 52**,** 725-32.

RANSON, S. W. 1920. *The anatomy of the nervous system*, Saunders W B.

RICE, G. I., BOND, J., ASIPU, A., BRUNETTE, R. L., MANFIELD, I. W., CARR, I. M., FULLER, J. C., JACKSON, R. M., LAMB, T., BRIGGS, T. A., ALI, M., GORNALL, H., COUTHARD, L. R., AEBY, A., ATTARD-MONTALTO, S. P., BERTINI, E., BODEMER, C., BROCKMANN, K., BRUETON, L. A., CORRY, P. C., DESGUERRE, I., FAZZI, E., CAZORLA, A. G., GENER, B., HAMEL, B. C., HEIBERG, A., HUNTER, M., VAN DER KNAAP, M. S., KUMAR, R., LAGAE, L., LANDRIEU, P. G., LOURENCO, C. M., MAROM, D., MCDERMOTT, M. F., VAN DER MERWE, W., ORCESI, S., PRENDIVILLE, J. S., RASMUSSEN, M., SHALEV, S. A., SOLER, D. M., SHINAWI, M., SPIEGEL, R., TAN, T. Y., VANDERVER, A., WAKELING, E. L., WASSMER, E., WHITTAKER, E., LEBON, P., STETSON, D. B., BONTHRON, D. T. & CROW, Y. J. 2009. Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat Genet,* 41**,** 829-32.

RICE, G. I., DEL TORO DUANY, Y., JENKINSON, E. M., FORTE, G. M., ANDERSON, B. H., ARIAUDO, G., BADER-MEUNIER, B., BAILDAM, E. M., BATTINI, R., BERESFORD, M. W., CASARANO, M., CHOUCHANE, M., CIMAZ, R., COLLINS, A. E., CORDEIRO, N. J., DALE, R. C., DAVIDSON, J. E., DE WAELE, L., DESGUERRE, I., FAIVRE, L., FAZZI, E., ISIDOR, B., LAGAE, L., LATCHMAN, A. R., LEBON, P., LI, C., LIVINGSTON, J. H., LOURENCO, C. M., MANCARDI, M. M., MASUREL-PAULET, A., MCINNES, I. B., MENEZES, M. P., MIGNOT, C., O'SULLIVAN, J., ORCESI, S., PICCO, P. P., RIVA, E., ROBINSON, R. A., RODRIGUEZ, D., SALVATICI, E., SCOTT, C., SZYBOWSKA, M., TOLMIE, J. L., VANDERVER, A., VANHULLE, C., VIEIRA, J. P., WEBB, K., WHITNEY, R. N., WILLIAMS, S. G., WOLFE, L. A., ZUBERI, S. M., HUR, S. & CROW, Y. J. 2014. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat Genet,* 46**,** 503-9.

RICE, G. I., KASHER, P. R., FORTE, G. M., MANNION, N. M., GREENWOOD, S. M., SZYNKIEWICZ, M., DICKERSON, J. E., BHASKAR, S. S., ZAMPINI, M., BRIGGS, T. A., JENKINSON, E. M., BACINO, C. A., BATTINI, R., BERTINI, E., BROGAN, P. A., BRUETON, L. A., CARPANELLI, M., DE LAET, C., DE LONLAY, P., DEL TORO, M., DESGUERRE, I., FAZZI, E., GARCIA-CAZORLA, A., HEIBERG, A., KAWAGUCHI, M., KUMAR, R., LIN, J. P., LOURENCO, C. M., MALE, A. M., MARQUES, W., JR., MIGNOT, C., OLIVIERI, I., ORCESI, S., PRABHAKAR, P., RASMUSSEN, M., ROBINSON, R. A., ROZENBERG, F., SCHMIDT, J. L., STEINDL, K., TAN, T. Y., VAN DER MERWE, W. G., VANDERVER, A., VASSALLO, G., WAKELING, E. L., WASSMER, E., WHITTAKER, E., LIVINGSTON, J. H., LEBON, P., SUZUKI, T., MCLAUGHLIN, P. J., KEEGAN, L. P., O'CONNELL, M. A., LOVELL, S.

C. & CROW, Y. J. 2012. Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nat Genet,* 44**,** 1243-8.

RISS, T. L., MORAVEC, R. A., NILES, A. L., BENINK, H. A., WORZELLA, T. J. & MINOR, L. 2004. Cell Viability Assays. *In:* SITTAMPALAM, G. S., GAL-EDD, N., ARKIN, M., AULD, D., AUSTIN, C., BEJCEK, B., GLICKSMAN, M., INGLESE, J., LEMMON, V., LI, Z., MCGEE, J., MCMANUS, O., MINOR, L., NAPPER, A., RISS, T., TRASK, O. J. & WEIDNER, J. (eds.) *Assay Guidance Manual.* Bethesda (MD).

RIVIERE, J. B., VAN BON, B. W., HOISCHEN, A., KHOLMANSKIKH, S. S., O'ROAK, B. J., GILISSEN, C., GIJSEN, S., SULLIVAN, C. T., CHRISTIAN, S. L., ABDUL-RAHMAN, O. A., ATKIN, J. F., CHASSAING, N., DROUIN-GARRAUD, V., FRY, A. E., FRYNS, J. P., GRIPP, K. W., KEMPERS, M., KLEEFSTRA, T., MANCINI, G. M., NOWACZYK, M. J., VAN RAVENSWAAIJ-ARTS, C. M., ROSCIOLI, T., MARBLE, M., ROSENFELD, J. A., SIU, V. M., DE VRIES, B. B., SHENDURE, J., VERLOES, A., VELTMAN, J. A., BRUNNER, H. G., ROSS, M. E., PILZ, D. T. & DOBYNS, W. B. 2012. De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nature genetics,* 44**,** 440-4, S1-2.

ROLLINS, J. D., COLLINS, J. S. & HOLDEN, K. R. 2010. United States head circumference growth reference charts: birth to 21 years. *J Pediatr,* 156**,** 907-13, 913 e1-2.

ROMANI, M., MICALIZZI, A. & VALENTE, E. M. 2013. Joubert syndrome: congenital cerebellar ataxia with the molar tooth. *Lancet Neurol,* 12**,** 894-905.

ROPERS, F., DERIVERY, E., HU, H., GARSHASBI, M., KARBASIYAN, M., HEROLD, M., NURNBERG, G., ULLMANN, R., GAUTREAU, A., SPERLING, K., VARON, R. & RAJAB, A. 2011. Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the WASH complex member SWIP. *Hum Mol Genet,* 20**,** 2585-90.

RUIZ-PEREZ, V. L., IDE, S. E., STROM, T. M., LORENZ, B., WILSON, D., WOODS, K., KING, L., FRANCOMANO, C., FREISINGER, P., SPRANGER, S., MARINO, B., DALLAPICCOLA, B., WRIGHT, M., MEITINGER, T., POLYMEROPOULOS, M. H. & GOODSHIP, J. 2000. Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis. *Nat Genet,* 24**,** 283-6.

SAITSU, H., NISHIMURA, T., MURAMATSU, K., KODERA, H., KUMADA, S., SUGAI, K., KASAI-YOSHIDA, E., SAWAURA, N., NISHIDA, H., HOSHINO, A., RYUJIN, F., YOSHIOKA, S., NISHIYAMA, K., KONDO, Y., TSURUSAKI, Y., NAKASHIMA, M., MIYAKE, N., ARAKAWA, H., KATO, M., MIZUSHIMA, N. & MATSUMOTO, N. 2013. De novo mutations in the autophagy gene WDR45 cause static encephalopathy of childhood with neurodegeneration in adulthood. *Nat Genet,* 45**,** 445-9, 449e1.

SARKER, A. H., TSUTAKAWA, S. E., KOSTEK, S., NG, C., SHIN, D. S., PERIS, M., CAMPEAU, E., TAINER, J. A., NOGALES, E. & COOPER, P. K. 2005. Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell,* 20**,** 187-98.

SCHARER, O. D. 2008. XPG: its products and biological roles. *Adv Exp Med Biol,* 637**,** 83-92.

SCHOSSIG, A., WOLF, N. I., FISCHER, C., FISCHER, M., STOCKER, G., PABINGER, S., DANDER, A., STEINER, B., TONZ, O., KOTZOT, D., HABERLANDT, E., AMBERGER, A., BURWINKEL, B., WIMMER, K., FAUTH, C., GROND-GINSBACH, C., KOCH, M. J., DEICHMANN, A., VON KALLE, C., BARTRAM, C. R., KOHLSCHUTTER, A., TRAJANOSKI, Z. & ZSCHOCKE, J. 2012. Mutations in ROGDI Cause Kohlschutter-Tonz Syndrome. *Am J Hum Genet,* 90**,** 701-7.

SCHUMACHER, B. 2009. Transcription-blocking DNA damage in aging: a mechanism for hormesis. *Bioessays,* 31**,** 1347-56.

SESSA, A., MAO, C. A., COLASANTE, G., NINI, A., KLEIN, W. H. & BROCCOLI, V. 2010. Tbr2-positive intermediate (basal) neuronal progenitors safeguard cerebral cortex expansion by controlling amplification of pallial glutamatergic neurons and attraction of subpallial GABAergic interneurons. *Genes Dev,* 24**,** 1816-26.

SESSA, A., MAO, C. A., HADJANTONAKIS, A. K., KLEIN, W. H. & BROCCOLI, V. 2008. Tbr2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex. *Neuron,* 60**,** 56-69.

SETTEMBRE, C. & BALLABIO, A. 2014. Lysosome: regulator of lipid degradation pathways. *Trends Cell Biol*.

SHAHEEN, R., FAQEIH, E., SHAMSELDIN, H. E., NOCHE, R. R., SUNKER, A., ALSHAMMARI, M. J., AL-SHEDDI, T., ADLY, N., AL-DOSARI, M. S., MEGASON, S. G., AL-HUSAIN, M., AL-MOHANNA, F. & ALKURAYA, F. S. 2012. POC1A truncation mutation causes a ciliopathy in humans characterized by primordial dwarfism. *Am J Hum Genet,* 91**,** 330-6.

SHEN, Q. T., HSIUE, P. P., SINDELAR, C. V., WELCH, M. D., CAMPELLONE, K. G. & WANG, H. W. 2012. Structural insights into WHAMM-mediated cytoskeletal coordination during membrane remodeling. *J Cell Biol,* 199**,** 111-24.

SHEVELL, M. I., MAJNEMER, A., ROSENBAUM, P. & ABRAHAMOWICZ, M. 2000. Etiologic yield of subspecialists' evaluation of young children with global developmental delay. *J Pediatr,* 136**,** 593-8.

SHIVJI, K. K., KENNY, M. K. & WOOD, R. D. 1992. Proliferating cell nuclear antigen is required for DNA excision repair. *Cell,* 69**,** 367-74.

SILOVE, N., COLLINS, F. & ELLAWAY, C. 2013. Update on the investigation of children with delayed development. *J Paediatr Child Health,* 49**,** 519-25.

SIMPSON, M., CROSS, H., PROUKAKIS, C., PRYDE, A., HERSHBERGER, R., CHATONNET, A., PATTON, M. & CROSBY, A. 2003a. Maspardin is mutated in mast syndrome, a complicated form of hereditary spastic paraplegia associated with dementia. *Am J Hum Genet,* 73**,** 1147-56.

SIMPSON, M. A., CROSS, H., PROUKAKIS, C., PRYDE, A., HERSHBERGER, R., CHATONNET, A., PATTON, M. A. & CROSBY, A. H. 2003b. Maspardin is mutated in mast syndrome, a complicated form of hereditary spastic paraplegia associated with dementia. *Am J Hum Genet,* 73**,** 1147-56.

SNAPPER, S. B., ROSEN, F. S., MIZOGUCHI, E., COHEN, P., KHAN, W., LIU, C. H., HAGEMANN, T. L., KWAN, S. P., FERRINI, R., DAVIDSON, L., BHAN, A. K. & ALT, F. W. 1998. Wiskott-Aldrich syndrome protein-deficient mice reveal a role for WASP in T but not B cell activation. *Immunity,* 9**,** 81-91.

SNAPPER, S. B., TAKESHIMA, F., ANTON, I., LIU, C. H., THOMAS, S. M., NGUYEN, D., DUDLEY, D., FRASER, H., PURICH, D., LOPEZ-ILASACA, M., KLEIN, C., DAVIDSON, L., BRONSON, R., MULLIGAN, R. C., SOUTHWICK, F., GEHA, R., GOLDBERG, M. B., ROSEN, F. S., HARTWIG, J. H. & ALT, F. W. 2001. N-WASP deficiency reveals distinct pathways for cell surface projections and microbial actin-based motility. *Nat Cell Biol,* 3**,** 897-904.

SOBEL, E. & LANGE, K. 1996. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *AMERICAN JOURNAL OF HUMAN GENETICS,* 58**,** 1323-37.

SODERLING, S. H., GUIRE, E. S., KAECH, S., WHITE, J., ZHANG, F., SCHUTZ, K., LANGEBERG, L. K., BANKER, G., RABER, J. & SCOTT, J. D. 2007. A WAVE-1 and WRP signaling complex regulates spine density, synaptic plasticity, and memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 27**,** 355-65.

SPORBERT, A., DOMAING, P., LEONHARDT, H. & CARDOSO, M. C. 2005. PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res,* 33**,** 3521-8.

STEVENSON, R. E., SCHROER, R. J., SKINNER, C., FENDER, D. & SIMENSEN, R. J. 1997. Autism and macrocephaly. *Lancet,* 349**,** 1744-5.

STEWART, G. S., MASER, R. S., STANKOVIC, T., BRESSAN, D. A., KAPLAN, M. I., JASPERS, N. G., RAAMS, A., BYRD, P. J., PETRINI, J. H. & TAYLOR, A. M. 1999. The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell,* 99**,** 577-87.

STRACHAN, T. & READ, A. 2011. *Human Molecular Genetics, 4th Edition*.

STRAUSS, K. A., MORTON, D. H., PUFFENBERGER, E. G., HENDRICKSON, C., ROBINSON, D. L., WAGNER, C., STABLER, S. P., ALLEN, R. H., CHWATKO, G., JAKUBOWSKI, H., NICULESCU, M. D. & MUDD, S. H. 2007. Prevention of brain disease from severe 5,10-methylenetetrahydrofolate reductase deficiency. *Mol Genet Metab,* 91**,** 165-75.

STRAUSS, K. A. & PUFFENBERGER, E. G. 2009. Genetics, medicine, and the Plain people. *Annu Rev Genomics Hum Genet,* 10**,** 513-36.

SUETSUGU, S., YAMAZAKI, D., KURISU, S. & TAKENAWA, T. 2003. Differential roles of WAVE1 and WAVE2 in dorsal and peripheral ruffle formation for fibroblast cell migration. *Dev Cell,* 5**,** 595-609.

SUN, T. & HEVNER, R. F. 2014. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nat Rev Neurosci,* 15**,** 217-32.

SVITKINA, T., LIN, W. H., WEBB, D. J., YASUDA, R., WAYMAN, G. A., VAN AELST, L. & SODERLING, S. H. 2010. Regulation of the postsynaptic cytoskeleton: roles in development, plasticity, and disorders. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 30**,** 14937-42.

TADA, T. & SHENG, M. 2006. Molecular mechanisms of dendritic spine morphogenesis. *Current opinion in neurobiology,* 16**,** 95-101.

TAKAHASHI, H., SEKINO, Y., TANAKA, S., MIZUI, T., KISHI, S. & SHIRAO, T. 2003. Drebrin-dependent actin clustering in dendritic filopodia governs synaptic targeting of postsynaptic density-95 and dendritic spine morphogenesis. *The Journal of neuroscience : the official journal of the Society for Neuroscience,* 23**,** 6586-95.

TAN, U. 2006. A new syndrome with quadrupedal gait, primitive speech, and severe mental retardation as a live model for human evolution. *Int J Neurosci,* 116**,** 361-9.

TARPEY, P. S., SMITH, R., PLEASANCE, E., WHIBLEY, A., EDKINS, S., HARDY, C., O'MEARA, S., LATIMER, C., DICKS, E., MENZIES, A., STEPHENS, P., BLOW, M., GREENMAN, C., XUE, Y., TYLER-SMITH, C., THOMPSON, D., GRAY, K., ANDREWS, J., BARTHORPE, S., BUCK, G., COLE, J., DUNMORE, R., JONES, D., MADDISON, M., MIRONENKO, T., TURNER, R., TURRELL, K., VARIAN, J., WEST, S., WIDAA, S., WRAY, P., TEAGUE, J., BUTLER, A., JENKINSON, A., JIA, M., RICHARDSON, D., SHEPHERD, R., WOOSTER, R., TEJADA, M. I., MARTINEZ, F., CARVILL, G., GOLIATH, R., DE BROUWER, A. P., VAN BOKHOVEN, H., VAN ESCH, H., CHELLY, J., RAYNAUD, M., ROPERS, H. H., ABIDI, F. E., SRIVASTAVA, A. K., COX, J., LUO, Y., MALLYA, U., MOON, J., PARNAU, J., MOHAMMED, S., TOLMIE, J. L., SHOUBRIDGE, C., CORBETT, M., GARDNER, A., HAAN, E., RUJIRABANJERD, S., SHAW, M., VANDELEUR, L., FULLSTON, T., EASTON, D. F., BOYLE, J., PARTINGTON, M., HACKETT, A., FIELD, M., SKINNER, C., STEVENSON, R. E., BOBROW, M., TURNER, G., SCHWARTZ, C. E., GECZ, J., RAYMOND, F. L., FUTREAL, P. A. & STRATTON, M. R. 2009. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature genetics,* 41**,** 535-43.

TATTON-BROWN, K., SEAL, S., RUARK, E., HARMER, J., RAMSAY, E., DEL VECCHIO DUARTE, S., ZACHARIOU, A., HANKS, S., O'BRIEN, E., AKSGLAEDE, L., BARALLE, D., DABIR, T., GENER, B., GOUDIE, D., HOMFRAY, T., KUMAR, A., PILZ, D. T., SELICORNI, A., TEMPLE, I. K., VAN MALDERGEM, L., YACHELEVICH, N., CHILDHOOD OVERGROWTH, C., VAN MONTFORT, R. & RAHMAN, N. 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat Genet,* 46**,** 385-8.

TEITEL, A. D., MACKENZIE, C. R., STERN, R. & PAGET, S. A. 1992. Laryngeal involvement in systemic lupus erythematosus. *Semin Arthritis Rheum,* 22**,** 203-14.

THORNTON, G. K. & WOODS, C. G. 2009. Primary microcephaly: do all roads lead to Rome? *Trends Genet,* 25**,** 501-10.

THRASHER, A. J. 2009. New insights into the biology of Wiskott-Aldrich syndrome (WAS). *Hematology Am Soc Hematol Educ Program***,** 132-8.

TRAKA, M., MILLEN, K. J., COLLINS, D., ELBAZ, B., KIDD, G. J., GOMEZ, C. M. & POPKO, B. 2013. WDR81 is necessary for purkinje and photoreceptor cell survival. *J Neurosci,* 33**,** 6834-44.

TSURIMOTO, T. & STILLMAN, B. 1991. Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. *J Biol Chem,* 266**,** 1950-60.

UNSAL-KACMAZ, K., CHASTAIN, P. D., QU, P. P., MINOO, P., CORDEIRO-STONE, M., SANCAR, A. & KAUFMANN, W. K. 2007. The human Tim/Tipin complex coordinates an Intra-S checkpoint response to UV that slows replication fork displacement. *Mol Cell Biol,* 27**,** 3131-42.

VAN NOCKER, S. & LUDWIG, P. 2003. The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics,* 4**,** 50.

VERLOES, A., DRUNAT, S., GRESSENS, P. & PASSEMARD, S. 1993. Primary Autosomal Recessive Microcephalies and Seckel Syndrome Spectrum Disorders. *In:* PAGON, R. A., ADAM, M. P., ARDINGER, H. H., BIRD, T. D., DOLAN, C. R., FONG, C. T., SMITH, R. J. H. & STEPHENS, K. (eds.) *GeneReviews(R).* Seattle (WA).

VON BOHLEN UND HALBACH, O. 2010. Dendritic spine abnormalities in mental retardation. *Cell and tissue research,* 342**,** 317-23.

WARBRICK, E. 1998. PCNA binding through a conserved motif. *Bioessays,* 20**,** 195-9.

WASEEM, N. H., LABIB, K., NURSE, P. & LANE, D. P. 1992. Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA. *EMBO J,* 11**,** 5111-20.

WEED, S. A., KARGINOV, A. V., SCHAFER, D. A., WEAVER, A. M., KINLEY, A. W., COOPER, J. A. & PARSONS, J. T. 2000. Cortactin localization to sites of actin assembly in lamellipodia requires interactions with F-actin and the Arp2/3 complex. *The Journal of cell biology,* 151**,** 29-40.

WEGNER, A. M., NEBHAN, C. A., HU, L., MAJUMDAR, D., MEIER, K. M., WEAVER, A. M. & WEBB, D. J. 2008. N-wasp and the arp2/3 complex are critical regulators of actin in the development of dendritic spines and synapses. *The Journal of biological chemistry,* 283**,** 15912-20.

WOODS, K. A., CAMACHO-HUBNER, C., SAVAGE, M. O. & CLARK, A. J. 1996. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med,* 335**,** 1363-7.

WU, Z., LIN, Y., XU, H., DAI, H., ZHOU, M., TSAO, S., ZHENG, L. & SHEN, B. 2012. High risk of benzo[alpha]pyrene-induced lung cancer in E160D FEN1 mutant mice. *Mutat Res,* 731**,** 85-91.

XIN, B., JONES, S., PUFFENBERGER, E. G., HINZE, C., BRIGHT, A., TAN, H., ZHOU, A., WU, G., VARGUS-ADAMS, J., AGAMANOLIS, D. & WANG, H. 2011. Homozygous mutation in SAMHD1 gene causes cerebral vasculopathy and early onset stroke. *Proc Natl Acad Sci U S A,* 108**,** 5372-7.

XIN, B., PUFFENBERGER, E. G., TURBEN, S., TAN, H., ZHOU, A. & WANG, H. 2010. Homozygous frameshift mutation in TMCO1 causes a syndrome with craniofacial dysmorphism, skeletal anomalies, and mental retardation. *Proc Natl Acad Sci U S A,* 107**,** 258-63.

XIN, B. & WANG, H. 2013. Identification of Two Novel ERCC6 Mutations in Old Order Amish with Cockayne Syndrome. *Mol Syndromol,* 3**,** 288-90.

YAMAZAKI, D., SUETSUGU, S., MIKI, H., KATAOKA, Y., NISHIKAWA, S., FUJIWARA, T., YOSHIDA, N. & TAKENAWA, T. 2003. WAVE2 is required for directed cell migration and cardiovascular development. *Nature,* 424**,** 452-6.

YAN, C., MARTINEZ-QUILES, N., EDEN, S., SHIBATA, T., TAKESHIMA, F., SHINKURA, R., FUJIWARA, Y., BRONSON, R., SNAPPER, S. B., KIRSCHNER, M. W., GEHA, R., ROSEN, F. S. & ALT, F. W. 2003. WAVE2 deficiency reveals distinct roles in embryogenesis and Rac-mediated actin-based motility. *EMBO J,* 22**,** 3602-12.

YU, T. W., MOCHIDA, G. H., TISCHFIELD, D. J., SGAIER, S. K., FLORES-SARNAT, L., SERGI, C. M., TOPCU, M., MCDONALD, M. T., BARRY, B. J., FELIE, J. M., SUNU, C., DOBYNS, W. B., FOLKERTH, R. D., BARKOVICH, A. J. & WALSH, C. A. 2010. Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with

simplified gyri and abnormal cortical architecture. *Nat Genet,* 42**,** 1015-20.

ZECH, T., CALAMINUS, S. D., CASWELL, P., SPENCE, H. J., CARNELL, M., INSALL, R. H., NORMAN, J. & MACHESKY, L. M. 2011. The Arp2/3 activator WASH regulates alpha5beta1-integrin-mediated invasive migration. *J Cell Sci,* 124**,** 3753-9.

ZHANG, Z., SHIBAHARA, K. & STILLMAN, B. 2000. PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature,* 408**,** 221-5.

ZHENG, L., DAI, H., HEGDE, M. L., ZHOU, M., GUO, Z., WU, X., WU, J., SU, L., ZHONG, X., MITRA, S., HUANG, Q., KERNSTINE, K. H., PFEIFER, G. P. & SHEN, B. 2011. Fen1 mutations that specifically disrupt its interaction with PCNA cause aneuploidy-associated cancer. *Cell Res,* 21**,** 1052-67.

ZHENG, L., DAI, H., QIU, J., HUANG, Q. & SHEN, B. 2007. Disruption of the FEN-1/PCNA interaction results in DNA replication defects, pulmonary hypoplasia, pancytopenia, and newborn lethality in mice. *Mol Cell Biol,* 27**,** 3176-86.

ZOPPINO, F. C., MILITELLO, R. D., SLAVIN, I., ALVAREZ, C. & COLOMBO, M. I. 2010. Autophagosome formation depends on the small GTPase Rab1 and functional ER exit sites. *Traffic,* 11**,** 1246-61.

ZUCHERO, J. B., COUTTS, A. S., QUINLAN, M. E., THANGUE, N. B. & MULLINS, R. D. 2009. p53-cofactor JMY is a multifunctional actin nucleation factor. *Nat Cell Biol,* 11**,** 451-9.

PUBLICATIONS RELATING TO THE WORK IN THIS THESIS

PCNA mutation affects DNA repair not replication. (2014) Green CM, Baple EL, Crosby AH. *Cell Cycle*, 13:20, 3157-3158

Hypomorphic PCNA mutation underlies a human DNA repair disorder. (2014) Baple EL, Chambers H, Cross HE, Fawcett H, Nakazawa Y, Chioza BH, Harlalka G, Mansour S, Sreekantan-Nair A, Patton MA, Muggenthaler M, Rich P, Wagner K, Coblentz Ro, Stein C, Last J, Taylor M, Jackson A, Ogi T, Lehmann AR, Green C, Crosby A. *J Clin Invest*; 1;124(7):3137-46

Mutations in KPTN cause macrocephaly, neurodevelopmental delay, and seizures. (2014). **Baple EL**, Maroofian R, Chioza BA, Izadi M, Cross HE, Al-Turki S, Barwick K, Wagner K, Coblentz R, Zainy T, Patton MA, Qualmann B, Hurles M, Kessels MM, Crosby AH . *Am J Hum Genet*, 94(1):87-94

Mutations in beta-1, 4-N-acetyl-galactosaminyl transferase 1 (B4GALNT1; GM2 Synthase) Underlie a New Disorder of Ganglioside Biosynthesis. (2013). Harlalka GV, Baple EL*, Lehman A, Chioza BA, Maroofian R, Cross HE, Nair AS, Priestman DA, Al-Turki S, McEntagart ME, Proukakis C, Royle L, Kozak RP, Bastaki L, Patton MA, Wagner K, Coblentz R, Price J, Mezei M, Schlade-Bartusiak K, Platt FM, Hurles M, Crosby AH. *Brain*, 136(Pt 12):3618-24 ***Joint first author**

Mutation of HERC2 causes developmental delay with Angelman-like features. (2013). Harlalka GV, **Baple EL***, Cross H, Kühnle S, Cubillos-Rojas M, Matentzoglu K, Patton MA, Wagner K, Coblentz R, Ford DL, Mackay DJ, Chioza BA, Scheffner M, Rosa JL, Crosby AH. *J Med Genet* 50(2):65-73 ***Joint first author**