Investigations into the genetic causes of liver disease using molecular genetic technologies

by

Kirsten Elizabeth McKay Bounford

A thesis submitted to the University of Birmingham for the degree of Doctor of Philosophy

School of Clinical and Experimental Medicine
College of Medical and Dental Sciences
University of Birmingham
June 2016

Abstract

DNA sequencing technologies have developed quickly in the last decade, and new methodologies have moved into clinical practice. These can been used to investigate genetic causes of neonatal cholestasis. Neonatal cholestasis can be life-threatening and has a varied etiology.

In chapter 3, a targeted next generation sequencing (tNGS) assay was designed and assessed for suitability for detection of known mutations in genes associated with cholestasis. In chapter 4, this was used to screen over 200 infants presenting with liver disease for mutations in the *ATP8B1*, *ABCB11*, *ABCB4*, *NPC1*, *NPC2* and *SLC25A13* genes. Diagnoses were made in 9% and single heterozygous mutations were in 9% of cases. In chapter 5, patients suspected of PFIC-related disease were tested for mutations in the *ATP8B1*, *ABCB11* and *ABCB4* genes. This study uncovered 27 novel sequence variants, including 22 in UK patients, expanding the known mutation spectrum of these disorders. In chapter 6, patients suspected of NPC and were tested for mutations in *NPC1* and *NPC2*, or SLC25A13, respectively. These studies have identified 134 novel NPC mutations and 4 novel CD mutations.

Current and future DNA sequencing methods are discussed, as are new diagnostic strategies for genetically heterogeneous conditions like infantile liver disease.

Acknowledgements

I would like to thank my supervisor Professor Paul Gissen for giving me the opportunity to do this PhD and for the support he has given me whilst working on this project. I would also like to thank Professor Eamonn Maher and Dr Ferenc Mueller for their co-supervision over the course of my research. I would like to thank Dr Christopher Bruce for his training on microarray sequencing and Dr Nicola Ruth for her enthusiastic approach to my questions.

I am indebted to Dr Fiona Macdonald for her continuous and invaluable support for my work and for reading and commenting on this thesis. I also would like to thank Eleanor Baker (nee Rattenberry) for her moral support, scientific advice and loyal friendship. Thank you to the rest of the staff at the West Midlands Regional Genetics Laboratory, many of whom are also dear friends, for their support and encouragement throughout.

I also would like to thank the clinicians who contributed patient samples and clinical information, particularly those at Birmingham Women's Hospital, Birmingham Children's Hospital and Queen Elizabeth Hospital Birmingham whom I have corresponded with most frequently. I also owe gratitude to the patients and families involved whose information has contributed to this work.

Importantly I would like to thank my parents, family and friends for their encouragement and faith in me. Finally I would to give special thanks to my patient and caring husband Tom, who has been my rock throughout and without whom I would not have reached this point.

Contents listings

Table of Contents

Chapter 4

List of Figures

List of tables

Abbreviations

1KG	1000 Genomes Project
AA	48.48 Access Array
ACCE	Analytical validity, Clinical validity, Clinical utility and Ethical, social and legal issues
aCGH	array comparative genome hybridisation
AGS	Alagille syndrome
ALF	acute liver failure
ALL	acute lymphoid leukaemia
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AML	acute myeloid leukaemia
ARC	arthrogryposis, renal abnormalities and cholestasis syndrome
ARPKD	autosomal recessive polycystic kidney disease
AST	aspartate aminotransferase
BA	biliary atresia
BRUM1	Birmingham resequencing microarray 1
AATD	alpha-1 anti-trypsin deficiency
BASM	biliary atresia with splenic malformations
BRIC	benign recurrent intrahepatic cholestasis
BSEP	bile salt export protein
CD	citrin deficiency
cDNA	complementary DNA
CESD	cholesterol ester storage disease
CF	cystic fibrosis
COPD	chronic obstructive pulmonary disease
CTLN2	citrullinaemia type II
ddNTP	dideoxy nucleotide triphosphate
DNA	deoxyribose nucleic acid
EDTA	ethylenediaminetetraacetic acid
emPCR	emulsion PCR
ESP	Exome Variant Server project
FISH	fluorescent *in situ* hybridisation
FLP	fragment library preparation
FS	fluorescent capillary Sanger sequencing
FTTDCD	failure to thrive and dyslipidaemia caused by citrin deficiency
FXR	farnesoid X receptor
GD	Gaucher disease
GGT	gamma-glutamyl transferase
gDNA	genomic DNA
HGVS	Human Genome Variation Society
ICP	intrahepatic cholestasis of pregnancy
IgG	immunoglobulin G
LALD	lysosomal acid lipase deficiency
MID	molecular indices
MLPA	multiplex ligation-dependent probe amplification
mRNA	messenger RNA
MS	microarray sequencing
NAIC	North American Indian childhood cirrhosis
NGS	Next generation sequencing
NHC	neonatal haemochromatosis
NICCD	neonatal intrahepatic cholestasis caused by citrin deficiency
NISCH	neonatal icthyosis-sclerosing cholangitis syndrome
NMD	nonsense-mediated decay
PBD	peroxisomal biogenesis disorder
PCR	polymerase chain reaction
PEBD	partial external biliary diversion
PFIC	progressive familial intrahepatic cholestasis
PN	parenteral nutrition

PS	phosphatidylserine
PTP	picotiter plate
RNA	ribonucleic acid
PTC	premature termination codon
ROI	region of interest
RS	Rotor syndrome
SNP	single or small nucleotide polymorphism
SGS	second generation sequencing
TEM	target enrichment method
TNC	transient neonatal cholestasis
tNGS	targetted next generation sequencing
TORCH	toxoplasmosis, rubella, cytomegalovirus and herpes simplex viruses
TSCA	TruSeq custom amplicon
TYRSN1	tyrosinaemia type 1
UCDA	ursodeoxycholic acid
UK	United Kingdom
VACTERL	vertebral anomalies, anal atresia, cardiac defects, tracheoesophageal fistula and/or esophageal atresia, renal & radial anomalies and limb defects
vcf	variant call file
VSGP	vertical supranuclear gaze palsy
VUS	variant of uncertain clinical significance
ZS	Zellweger syndrome

Chapter 1 - INTRODUCTION

1.1. Liver anatomy and function

1.1.1. Anatomy of the Liver

The liver is the largest visceral organ of the human body and is located in the upper right hand side of the abdominal cavity, inferior to the diaphragm and ribcage and superior to the stomach, intestines and colon. Superficially, the liver is divided into left and right lobes, where the right is six times larger than the left (see Figure 1.1). In addition, the right lobe includes the posterior caudate lobe and the inferior quadrate lobe. The left and right lobes are separated by the falciform ligament and fissures for ligamentum verosum and ligamentum teres. Alternative subdivisioning of the liver, termed the 'functional anatomy' is based upon the structure of the vascular and biliary structures (Dooley and Sherlock, 2011). A major fissure called porta hepatis is the site of entry and exit for the vasculature, lymphatic system, hepatic ducts and nervous infrastructure of the liver. Blood enters the liver from the portal vein and the hepatic artery and exits via the hepatic vein to the inferior vena cava, close to the junction with the right atrium of the heart. The biliary system includes the left and right hepatic ducts, which meet to form the common hepatic duct (Figure 1.1). The cystic duct from the gallbladder meets the common hepatic duct to form the common bile duct. The common bile duct meets the pancreatic duct near the duodenum and forms the Ampulla of Vater, regulated by the sphincter of Oddi (Dooley and Sherlock, 2011).

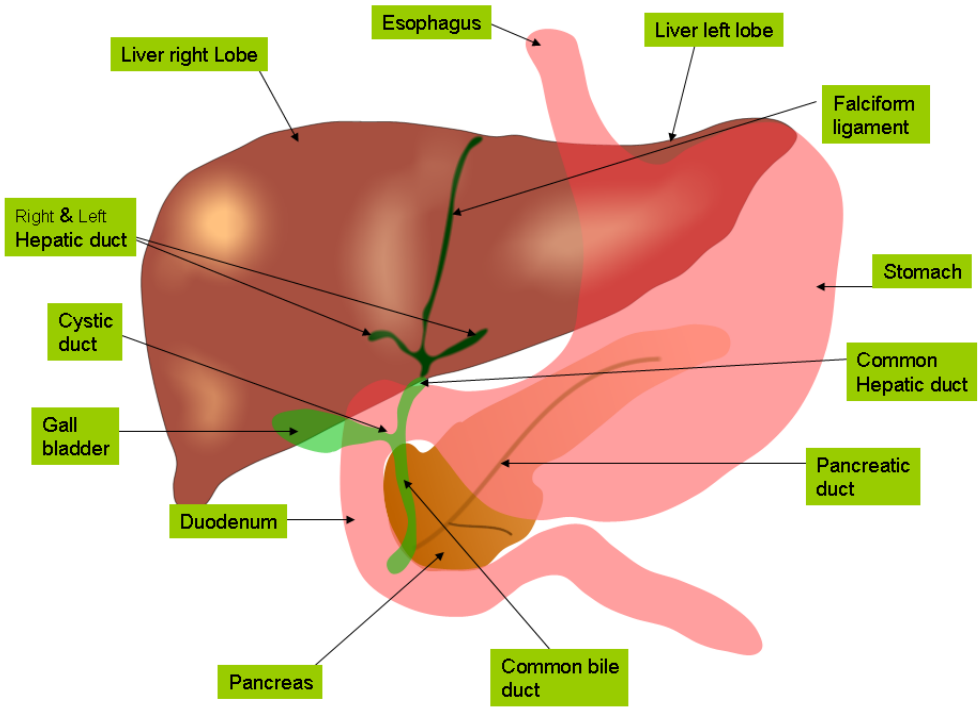


Figure 1.1: A basic diagram of the liver and biliary system.
Image taken from Wikimedia Commons. By Jiju Kurian Punnoose (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC-BY-SA-3.0-2.5-2.0-1.0 (http://creativecommons.org/licenses/by-sa/3.0)].

The internal structure of the liver tissue can be described as either a 'lobule'; a hexagonal arrangement of cells centred on the hepatic vein, or as an 'acinus'; a diamond-shaped region between two central vein branches (Dooley and Sherlock, 2011). Figure 1.2 is a diagram illustrating the structures of the lobule and the acinus. Both functional units define zones of hepatocytes which have different metabolic properties as well as different susceptibilities to damage. Zone 1 cells are closest to the arterioles and portal vein branches and therefore oxygen and nutrient supplies, but are more susceptible to injury from toxins (Dooley and Sherlock, 2011). Zone 3 cells are closest to the central veins and are therefore more susceptible to injury due to lack of oxygen (Dooley and Sherlock, 2011).

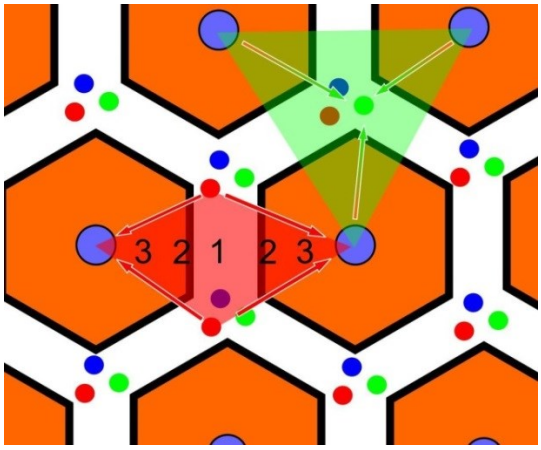


Figure 1.2: A diagram of the internal structure of the liver.
This shows the 'lobule' in orange with the portal vein in the centre. The red arrows outline the 'acinus' and show the flow of oxygen from the arteriole (red) to the portal vein past zone 1, 2 and 3 hepatocytes. The green arrows show the flow of bile from portal vein and hepatocytes towards the bile duct.

Image taken from Wikimedia Commons. "Liver scheme1". Licensed under Creative Commons Attribution-Share Alike 3.0 via - http://commons.wikimedia.org/wiki/File:Liver_scheme1.jpg#mediaviewer/File:Liver_scheme1.jpg

The liver tissue consists of a variety of cell types including hepatocytes (parenchymal cells), endothelial cells lining the sinusoids (vascular system), cholangiocytes (endothelial cells lining the bile ductules and ducts), Kupffer cells (phagocytic cells), pit cells (natural killer cells) and hepatic stellate cells (fat and retinoid storing cells) (Dooley and Sherlock, 2011). In addition, blood cells and blood vessels account for a significant proportion of cells. Figure 1.3 is a diagram showing the arrangement of cells in the liver in detail.

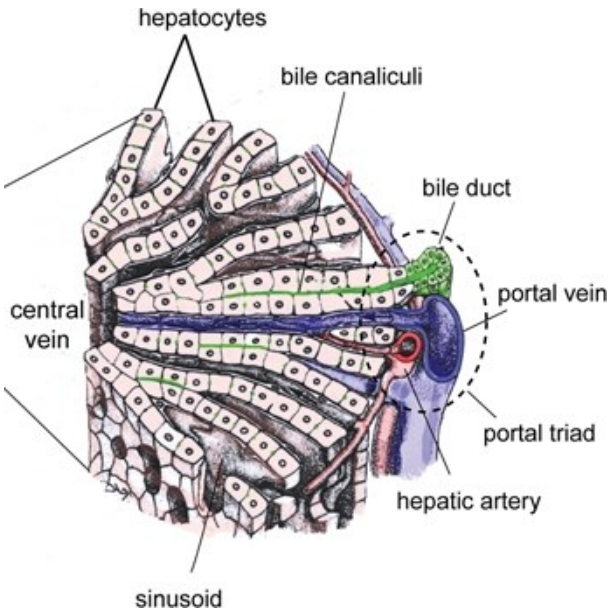


Figure 1.3 A diagram showing the cellular structure of the liver.

Image taken from Wikimedia Commons. By Zorn, A.M., Liver development (October 31, 2008), StemBook, ed. The Stem Cell Research Community, StemBook, doi/10.3824/stembook.1.25.1, http://www.stembook.org. [CC-BY-3.0 (http://creativecommons.org/licenses/by/3.0)].

1.1.2. Functions of the liver

The functions of the liver are carried out by the hepatocytes and there are up 500 separate functions known (Dooley and Sherlock, 2011). Life is not possible without the liver as it is not currently possible to artificially replace all of these functions, and thus severe liver damage or disease may only be treated by liver transplantation (Wang et al., 2013). Roles include synthesis of amino acids, carbohydrates, lipids, coagulation factors, red blood cells in the fetus, bile, insulin-like growth factor, albumin and thrombopoietin (Dooley and Sherlock, 2011). Metabolism of hormones including insulin, bilirubin to bile, ammonia to urea, and of drugs by the cytochrome P450 proteins also occurs (Dooley and Sherlock, 2011). Other functions of the liver include storage of glycogen, vitamins A, D, B12, K, iron and copper (Dooley and Sherlock, 2011).

1.1.2.1. Bile production and secretion

Bile is a dark green fluid made by the liver, temporarily stored in the gallbladder, and then released into the duodenum (Dooley and Sherlock, 2011). Bile is composed of water, bile acids, bilirubin, cholesterol, phospholipids and electrolytes, and its two main functions are to aid the digestion of in food and to allow excretion of waste products via faeces. Cholesterol is generally insoluble, however inside hepatocytes cholesterol is oxidised to form the bile acids cholic acid and chenodeoxycholic acid. Each of these can be conjugated with either taurine or glycine to produce a total of four types of conjugated bile acid. Conjugated bile acids are soluble and are secreted into bile (via the bile salt export protein; BSEP) where they aid digestion by facilitating the emulsification of lipids. Bile allows the uptake of lipids and lipid-soluble vitamins from the GI tract into the bloodstream. Bile also functions as the route by which bilirubin, a by-product of the breakdown of the haem molecule from red blood

cells, is excreted from the body. In hepatocytes, bilirubin becomes conjugated with glucuronic acid and it is in this conjugated form that is excreted via bile (via the MRP2 protein). Bile acids are reabsorbed by the ileum into the portal bloodstream and then are very efficiently recycled by the hepatocytes (Dooley and Sherlock, 2011).

1.2. Infantile liver disease

1.2.1. Overview of liver disease in infants

Abnormal liver function in the neonatal period is initially suspected by the signs and symptoms of jaundice, dark coloured urine, pale coloured stools, and an enlarged liver. Neonatal jaundice (or physiological jaundice) is common in neonates (approximately 60% of term babies and 80% of preterm babies) and is usually caused by unconjugated bilirubinaemia (NICE Guidelines, 2010). Conjugated hyperbilirubinaemia is suspected if jaundice persists beyond 2 to 3 weeks; this occurs in around 1 in 2,500 births and indicates underlying liver disease (Mckiernan, 2002). Blood tests can determine the levels of bilirubin in serum, as well as differentiating between conjugated and unconjugated species. Discolouration of urine and stools occurs because of disruption of bilirubin metabolism and transportation. Conjugated hyperbilirubinaemia may be caused by cholestasis; defined as impaired bile flow. Cholestasis can be associated with hepatomegaly and sometimes also splenomegaly. Some babies with cholestasis recover with time, however all cases should be thoroughly investigated due to the possibility of a serious condition such as biliary atresia (see 1.2.3.1). In addition, chronic liver failure may occur if persistent cholestasis is left untreated, due to fibrosis and

cirrhosis of the liver. Acute liver failure (ALF) in infants is defined as severe impairment of liver function without a recognised chronic liver disease. Infantile ALF is often a multi-system disorder and can occur with or without encephalopathy. The underlying causes of infantile liver disease, and in particular cholestasis, are many and varied and will be discussed further (see 1.2.2).

Routine investigations in infants with conjugated hyperbilirubinaemia aim to identify the underlying cause as this will influence management and likely prognosis. Such investigations include blood tests to measure liver enzymes such as ALP, ALT, AST and GGT, screens for infectious causes, coagulation studies which will determine the level of liver function, and imaging techniques such as ultrasound to look for signs of structural abnormalities, such as biliary atresia or gallstones. Liver biopsy is then considered and can uncover various abnormalities such as fibrosis, bile plugs, and hepatocyte or bile ducts abnormalities.

Treatments for neonatal liver diseases vary depending on the underlying pathophysiology but will include promotion of growth and nutrition, such as vitamin supplementation. Administration of ursodeoxycholic acid (UDCA) to reduce the effects of cholestasis is used as a treatment for gallstones and many types of cholestasis. The Kasai hepatoportoenteropathy is a surgical intervention used in cases of biliary atresia. In other cases with acute or chronic liver failure, liver transplantation is required.

1.2.2. Cholestasis in Infants

Infantile cholestasis may be ‘obstructive’ or ‘metabolic’ in nature. Obstructive cholestasis includes conditions such as biliary atresia and Alagille syndrome whereas metabolic cholestasis includes progressive familial intrahepatic cholestasis.

The most important causes of conjugated bilirubinaemia in neonates are listed in Table 1.1 and discussed below.

Table 1.1: Known causes of infantile cholestasis.

KEY; ALL=acute lymphoid leukaemia, AML=acute myeloid leukaemia, NHC=neonatal haemochromatosis.

Cause	Proportion of Neonatal Cholestasis
Biliary Atresia	~35% (Mieli-Vergani et al., 1989)
Metabolic and inherited disorders, including:	Up to 55% (De Bruyne et al., 2011a)
Alpha-1 anti-trypsin deficiency	5-15% (De Bruyne et al., 2011a)
Inherited forms of cholestasis	10-20% (De Bruyne et al., 2011a)
Inborn errors of metabolism	20% (De Bruyne et al., 2011a)
Infectious disease	5% (De Bruyne et al., 2011a)
Prematurity and/or parenteral nutrition	Unknown
Idiopathic cases (unspecified cause)	Up to 15%
Others	
Chromosomal disorders, i.e. Trisomy 21, Trisomy 18, Turner's syndrome	Rare
Leukaemia, i.e. ALL or AML	Rare
Immune system disorders i.e. NHC	Rare

1.2.3. Non-genetic causes

1.2.3.1. Biliary atresia

Biliary atresia (BA) is a rare disease affecting infants, which leads to progressive liver disease, liver failure and early death if untreated. BA is caused by the absence of adequate biliary tree structures, leading to cholestasis, progressive fibrosis and portal hypertension. It occurs in approximately 1 in 20,000 infants worldwide, but is more common in Asian populations, for example in 1 in 10,000 infants in Japan, or 1 in 3,000 in French Polynesia (Moreira et al., 2012).

There are three forms of BA; the first is isolated BA, also termed 'perinatal' or 'non-syndromic', and accounts for 80-90% of cases (Moreira et al., 2012). The second form is called 'early', 'embryonic' or 'syndromic', accounts for 10-20% of cases, and

is associated with other congenital malformations involving the spleen, intestines and heart (Moreira et al., 2012). The term 'biliary atresia with splenic malformations' (BASM) is sometimes used. The third form is called cystic biliary atresia and accounts for only a very small proportion of syndromic BA cases. There is no known cause for BA however, it is considered likely that non-syndromic and syndromic/cystic BA have different aetiologies. The presence of other congenital malformations in syndromic BA suggests it may be a developmental disorder (Nakamura and Tanoue, 2013). In comparison, bile duct remnants are detectable in patients with non-syndromic BA, indicating bile ducts had formed but have been destroyed and replaced by fibrous tissue. BA has therefore been classified as a progressive obliterative cholangiopathy (Nakamura and Tanoue, 2013). Many factors have been proposed to be contributors to BA, such as infection, aberrant immune-response and inflammation, and it is likely the answer will be multifactorial in nature (Hartley et al., 2013). It is very unlikely that BA is a Mendelian condition; however NOTCH signalling and left-right patterning are pathways which may play a role in at least the syndromic BA cases (Nakamura and Tanoue, 2013).

BA is suspected in infants with persistent neonatal jaundice, pale stools, dark urine or hepatomegaly. Blood tests show conjugated hyperbilirubinaemia and elevated liver enzymes. Diagnosis can be made by imaging of the liver such as ultrasound, magnetic resonance imaging or cholangiography or by liver biopsy. Management of BA patients is different from other cholestatic infants as early surgical intervention is recommended in the form of the Kasai procedure or Kasai portoenterostomy. BA initially affects the extrahepatic biliary structures and the Kasai procedure involves using a portion of the small intestine to connect the liver to the duodenum, allowing bile to drain via any remaining intraheptic bile ducts. This procedure may be

successful in around 60% of cases (Hartley et al., 2013). As the disease progresses the intraheptic biliary tree may also be obliterated and liver transplant is then required.

1.2.3.2. Cholelithiasis

Biliary sludge and resulting cholelithiasis (more commonly known as gallstones) has been described in children infrequently, with a prevalence ranging from 0.01 to 2% depending on population studied and method of identification. As described by Svensson and Makin four types of gallstones have been observed in children (Svensson and Makin, 2011). Gallstones consist of precipitated components of bile, which can block bile ducts leading to cholestasis. Gallstones can result due to genetic defects in bile formation (see 1.3.1).

1.2.3.3. Parenteral nutrition

Parenteral nutrition (PN) is the administration of nutrients intraveneously, bypassing the processes of feeding and digestion. PN is used in infants who have undergone extensive gastrointestinal surgery or have congenital or acquired defects of the GI tract (Kelly, 2006). PN in infants can be associated with serious hepatic complications including cholestasis (40-60%), biliary sludge and cholelithiasis and contributing factors are known, including prematurity, low birth weight, duration and composition of PN (Kelly, 2006).

1.2.3.4. Congenital infection

Congenital TORCH infections (toxoplasmosis, rubella, cytomegalovirus and herpes simplex) can present similarly with jaundice, hepatomegaly, splenomegaly and acute liver failure amongst other features (Roberts, 2003).

1.2.3.5. Neonatal haemochromatosis

Neonatal haemochromatosis (NHC), also known as congenital alloimmune hepatitis, is a materno-fetal alloimmune disease which is characterised by severe liver disease in the neonate and extrahepatic siderosis (Whitington, 2007). Untreated the prognosis is very poor, however a study in women who had suffered pregnancy or neonatal loss due to NHC and were treated with IgG from 18 weeks gestation, showed greatly improved outcomes for affected babies (Whitington and Kelly, 2008).

1.3. Important genetic causes of cholestasis

1.3.1. Progressive familial intrahepatic cholestasis (PFIC)

Progressive familial intrahepatic cholestasis is a group of heterogeneous childhood liver diseases, which present with cholestasis of hepatic origin and death from liver failure in the first year or up to adolescence. The incidence is estimated to be 1/50,000 to 1/100,000 live births and account for 10-15% of neonatal cholestasis cases and 10-15% of childhood liver transplant indications (Davit-Spraul et al., 2009). PFIC affects all populations and is inherited in an autosomal recessive fashion. PFIC is known to be caused by mutations in three genes; *ATP8B1*, *ABCB11* and *ABCB4* and so PFIC can be sub-classified according to the molecular diagnosis (see Table 1.2). PFIC types 1 and 2 are unusual in that they are characterised by normal serum GGT levels, whereas other metabolic caused of cholestasis (including PFIC 3) are associated with high serum GGT levels (Davit-Spraul et al., 2009).

Table 1.2: A summary of some of the key features of PFIC types 1, 2 and 3 (information drawn from Bull et al., 1998 and Davit-Spraul et al., 2009).

Features	PFIC1	PFIC2	PFIC3
Cholestasis	In first few months, becoming permanent later	Becoming permanent in first few months	Rare in neonates, 1/3 of cases by 12 months, rest in infancy, childhood or adolescence.
Liver disease	Liver failure and cancer	Failure within first year. High risk of liver cancer by age 1 year.	Failure and transplant by average age of 7.5 years. No tumours associated.
Extrahepatic findings	Short stature, deafness, diarrhoea, pancreatitis.	None	None
Ductules visible on liver biopsy	Absent	Absent	Present
Serum GGT	Normal	Normal	High

PFIC1 is caused by loss of function mutations in *ATP8B1* (Bull et al., 1998), and it is allelic to the milder conditions benign recurrent intrahepatic cholestasis type 1 (BRIC-1), Greenland familial cholestasis and intrahepatic cholestasis in pregnancy (ICP) (Mullenbach, 2005). The gene encodes a P-type ATPase called FIC1, which is localised to the canalicular membrane in liver hepatocytes (Figure 1.4). FIC1 protein function is not completely characterised, however it is known to be responsible for the translocation of aminophospholipids (e.g. phosphatidylserine, PS) from one side of the phospholipid bilayer of the plasma membrane to the other, causing an enrichment of PS on the inner layer. This asymmetry is thought to protect the liquid-ordered state of the outer leaflet of the membrane and thus protect against high bile salt concentrations in the canalicular lumen. Loss of FIC1 protein function is thought to result in cholestasis at least partly because of decreased BSEP activity (see next paragraph), potentially as a result of disruption of membrane integrity (Paulusma et

al., 2009), or by downregulation of a nuclear receptor called Farnesoid X receptor (FXR) which is important in the bile salt signalling and regulation pathway (Chen and Chang, 2004). It is likely that FIC1 has similar roles in maintaining membrane stability in other tissues in which it is expressed (inner ear, pancreas etc), loss of which should explain the extra-hepatic features involved in PFIC1 (Table 1.2). *ATP8B1* mutations reflect a spectrum of phenotypes from mild BRIC1 to severe PFIC1 however; genotype-phenotype correlations are difficult due to compound heterozygosity. It is likely that PFIC1 mutations completely disrupt the protein whereas BRIC1 mutations only partially disrupt the protein. Heterozygous mutations in *ATP8B1* have been identified in ICP patients (Mullenbach, 2005).

PFIC2 is caused by mutations in the *ABCB11* gene (Strautnieks et al., 1998), which encodes the ATP-dependent canalicular bile salt export protein (BSEP). BSEP is the major exporter of primary bile acids at the hepatocyte canalicular membrane and functions in extreme concentration gradients (Figure 1.4). Deficiency of BSEP directly leads to the accumulation of bile salts in hepatocytes and liver damage. BSEP deficiency also results in a spectrum of phenotypes from severe PFIC2 to mild BRIC2 (van Mil et al., 2004a). PFIC2 mutations tend to result in prematurely truncated proteins and immunohistochemistry shows absence of protein (Jacquemin, 2012). Some missense mutations also lead to PFIC2 but show presence of BSEP protein. These tend to be located in conserved regions of the gene; nucleotide binding fold containing the Walker A and B motifs and ABC signature ATP-binding motifs (Byrne et al., 2009). In BRIC2, missense mutations are common and tend to be in less conserved regions of the gene. Heterozygous mutations have also been detected in ICP2, drug induced cholestasis and transient neonatal cholestasis (Dixon et al., 2009).

PFIC3 differs from the other two phenotypes as it is associated with high serum GGT and ductule proliferation in the liver. In addition, the age of onset tends to be later and more variable (Davit-Spraul et al., 2010). PFIC3 is caused by mutations in the *ABCB4* gene (De Vree et al., 1998), a member of the ABC transporter protein family similar in structure to *ABCB11*. This encodes the MDR3 protein (Class III multidrug resistance P-glycoprotein) which is localised to canalicular membrane of hepatocytes (see Figure 1.4). Similarly to FIC1, MDR3 is a phospholipid translocator, moving biliary phosphaditylcholine (PC) from the inner leaflet to the outer leaflet of the canalicular membrane, facing the lumen. PC is then extracted from the membrane by bile salts forming micelles. Lack of PC leads to excess bile salts in the bile, which have a detergent effect on the canalicular membrane of hepatocytes and cholangiocytes, damaging the cells and resulting in cholestasis and cholangiopathy (Davit-Spraul et al., 2010). The phenotype of PFIC3 ranges from neonatal cholestasis to cirrhosis in young adults (Davit-Spraul et al., 2010). One third of mutations resulted in a truncated protein product with no detectable protein by immunostaining, presumably as a result of mRNA decay. PFIC3 missense mutations are in highly conserved regions like Walker A and B motifs (ATP-binding). In addition, MDR3 mutations are involved in ICP3, cholesterol gallstone disease, drug induced cholestasis and transient neonatal cholestasis (TNC) (Jacquemin, 2012; Mullenbach et al., 2003).

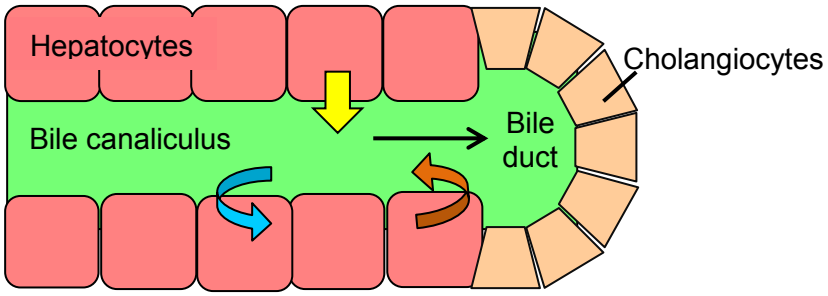


Figure 1.4: A schematic of the PFIC proteins at the cellular level within the liver.
The black arrow indicates bile flow.
The blue transporter is FIC1. It translocates aminophospholipids, e.g. PS from the outer to the inner leaflet of the canalicular plasma membrane of hepatocytes.
Yellow is BSEP which exports bile acids from hepatocytes into the canaliculus.
Orange is MDR3 which translocates phosphatidylcholine from the inner to the outer leaflet of the canalicular plasma membrane of hepatocytes.

1.3.2. Niemann Pick disease type C

Niemann-Pick disease type C (NPC) is caused by mutations in the genes *NPC1* (95% of cases) and *NPC2* (5% of cases) (Carstea, 1997; Naureckiene, 2000a). Both versions are autosomal recessively inherited. The overall frequency is 1 in 150,000 but this is likely to be an underestimate (Patterson, 2000). NPC may present at any age, however three characteristic periods of onset exist, and may relate to different mutations, the gene affected or even modifier genes (Imrie et al., 2007). Neonatal NPC in severe cases can sometimes be detected prenatally (Spiegel et al., 2009), and otherwise in the neonate with cholestatic liver disease, hepatomegaly and splenomegaly (Kelly et al., 1993). The liver disease can be progressive at this stage leading to liver failure and death. Those who survive are hypotonic (Kelly et al., 1993). Later on delayed motor development, neurological problems, enlarged liver and spleen, developmental delay and vertical supranuclear gaze palsy (VSGP) can also be observed (Kelly et al., 1993). Death usually occurs by age 5 (Imrie et al., 2007). Childhood-onset NPC is the classical presentation of NPC affecting the vast

majority of patients (Vanier, 2010). Initial symptoms are often clumsiness and gait disturbances and hepatomegaly or splenomegaly may be present upon examination (Vanier et al., 1988). Typically VSGP occurs, there is progressive cognitive impairment and dyspraxia, often ataxia and dystonia involving the limbs, deterioration of speech and ability to feed, seizures and gelastic cataplexy (Imrie et al., 2007; Vanier et al., 1988). Lifespan is variable and aspiration pneumonia is a common cause of death (Vanier, 2010). With adolescent or adult-onset NPC, the presentation can be similar to the neurological disease in childhood NPC but with a slower rate of progression (Sevin et al., 2006). Some adult-onset patients can present predominantly with psychiatric problems such as dementia, schizophrenia, delusions,hallucinations and paranoia (Sevin et al., 2006).

NPC is an unusual lysosomal storage disorder that is caused by defective intracellular lipid trafficking rather than enzyme deficiency (Vanier, 2015). The primary defect is accumulation of cholesterol from the late endosome/lysosome compartment. Sphingomyelin accumulation also occurs; a feature shared with Niemann-Pick disease types A and B explaining the original grouping of these diseases together. Accumulation of lipid-filled cells in the lungs leads to the lung impairment, in the liver the hepatomegaly, and in the neurons the neurological problems. Purkinje cells in the cerebellum seem to be particularly affected by the accumulation of lipids in NPC (Ko et al., 2003); however it is not entirely clear how this is harmful to these cells (Vanier, 2015).

The *NPC1* gene (located on chromosome 18q11-q12) encodes the NPC1 protein; an integral membrane protein with 13 transmembrane domains that is targeted to the late endosome and lysosome (Ioannou, 2000). It is known that NPC1 is involved in trafficking of cholesterol from the lysosome to the plasma membrane as part of the

usual distribution and cycling of lipids by the cell, how this happens is not fully understood. It is likely that NPC1 has other functions or at least other downstream effects as defective cholesterol trafficking does not fully explain the pathogenesis of NPC, for example in the brain (Ioannou, 2000). The *NPC2* gene (located on chromosome 14q24.3) encodes the NPC2 protein (Naureckiene, 2000a). NPC2 is widely expressed and is a soluble glycoprotein in the lumen of the late endosome and the lysosome (Naureckiene, 2000a). NPC2 binds cholesterol, transporting it from internal vesicle membranes within the endosome to NPC1 on the endosome membrane (Ko et al., 2003). High cholesterol levels are also likely to be responsible for reduced trafficking of other lipids, including sphingomyelin, leading to secondary accumulations (Vanier, 2010). Miglustat (Zavesca® from Actelion Pharmaceuticals) has been shown to be an effective therapy for slowing the progression of the neurodegenerative effects of NPC in children and adults (Lyseng-Williamson, 2014).

1.3.3. Citrin deficiency

Neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD), failure to thrive and dyslipidaemia caused by citrin deficiency (FTTDCD) and citrullinaemia type II (CTLN2) are all caused by mutations in the *SLC25A13* gene; also known as Citrin (Saheki and Kobayashi, 2002). Previously it was thought that NICCD and CTLN2 were allelic conditions, however with the discovery of FTTDCD, it is now considered that these are all probably the same condition (termed Citrin Deficiency) with distinct symptomatic periods and clinical presentations. It seems that not all patients with NICCD will go on to have CTLN2, and not all patients with CTLN2 will have had NICCD (Woo et al., 2014).

The symptoms of NICCD are prolonged intrahepatic cholestasis and metabolic abnormalities during the neonatal period, including amino acidaemia (especially citrulline) and galactosuria (Saheki and Kobayashi, 2002). NICCD is usually relatively mild and symptoms disappear within a year with basic treatment; this might include lactose-free formula milk and fat-soluble vitamin supplements to combat the effects of prolonged cholestasis (Saheki and Kobayashi, 2002; Woo et al., 2014). There are infrequent NICCD cases with severe liver dysfunction and where liver transplants have been required in childhood (Tamamori et al., 2002).

An apparently healthy phase was thought to follow the resolution of NICCD symptoms, however a new distinct phenotype has recently been described which includes failure to thrive, dyslipidaemia, echinosystosis and metabolic abnormalities (Song et al., 2011). CTLN2 is characterised by recurrent periods of hyperammonaemia and neurological symptoms in adulthood. The neurological symptoms are severe and usually sudden; these can include nocturnal delirium, behaviour changes, delusion, disorientation, drowsiness, loss of memory, tremor, seizures, coma and death (Kimura et al., 2013). In addition, CTLN2 (and FTTDCD) patients are often thin or underweight, with a preference for lipid-rich and protein-rich foods and an avoidance of carbohydrate-rich foods (Saheki and Kobayashi, 2002). The most successful treatment for CTLN2 is liver transplant at the onset of symptoms (Kimura et al., 2013).

The *SLC25A13* gene is located at 7q21.3 and encodes a protein called citrin, specifically in the liver (Kobayashi et al., 1999). The protein has a C-terminal region of the protein consists of 6 transmembrane domains, and the N-terminal region has 4 calcium-binding EF-hand domains. The protein is localised to the inner mitochondrial membrane with the N-terminal region in the mitochondrial lumen

between the two membranes. The protein functions as part of the malate-aspartate NADH shuttle and is important for aerobic glycolysis and gluconeogenesis, as well as in the urea cycle and protein and nucleotide synthesis (see Figure 1.5). Almost all reports of mutations in *SLC25A13* have been reported in the Japanese population, where the frequency of carriers is estimated to be 1 in 65 and affected patients is 1 in 17,000 (Saheki and Kobayashi, 2002). Mutations have now been found in patients from other populations, including the UK (Hutchin et al., 2009), although the carrier frequency and mutation spectrum are likely to be different.

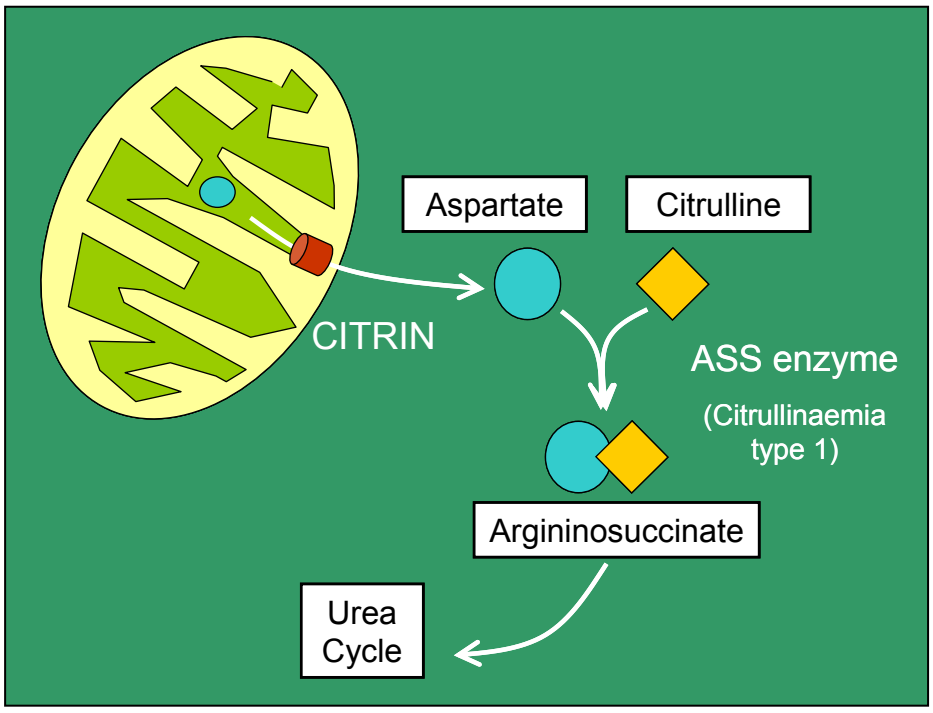


Figure 1.5: A diagram showing the subcellular location of the citrin protein. Citrin (in red) exports aspartate (blue) out of the mitochondria, where it feeds into the urea cycle.

1.3.4. Alagille syndrome

Alagille syndrome (AGS) is a multisystem disorder characterised by cholestasis in infancy, congenital heart defects, skeletal and eye malformations, renal abnormalities and distinctive facial features. The incidence of AGS is at least 1 in 70,000 births; however it is likely to be higher than this due to the widely variable expressivity of the condition (Turnpenny and Ellard, 2012). The most clinically significant feature of AGS is cholestasis in infancy caused by bile duct paucity, which occurs in over 90% of cases (Emerick et al., 1999). Left untreated, cholestasis in AGS may lead to fibrosis and liver failure in around 15% of cases (Emerick et al., 1999), however it is not yet possible to predict which children will require liver transplant and when. Cholestasis can be treated using ursodeoxycholic acid, partial external biliary diversion (PEBD) and liver transplant if required (Hartley et al., 2013). Cardiac abnormalities in AGS occur in over 90% of cases (Emerick et al., 1999). Most commonly these are outflow tract defects such as peripheral pulmonary stenosis, tertralogy of Fallot, atrial septal defect and ventricular septal defect (Turnpenny and Ellard, 2012). A distinctive feature of AGS is the presence of butterfly vertebrae, a congenital malformation of the spine which occurs in over 60% of cases with no associated morbidity (Emerick et al., 1999). Other skeletal complications of AGS can be clinically significant, such as craniosynostosis, osteoporosis and fractures (Turnpenny and Ellard, 2012). Posterior embryotoxon is an ocular feature which occurs in around 80% of AGS cases, but also in 15% of the normal population and up to 70% of patients with 22q11.2 deletion syndrome (McDonald-McGinn et al., 1999). Renal tubular acidosis may occur in around 70% of cases and structural abnormalities of the kidney are also common (Emerick et al., 1999). AGS is associated with distinctive facial features such as wide prominent

forehead, deep set eyes, long mandible and pointed chin, giving an overall appearance of an inverted triangular shape (Krantz et al., 1997).

AGS is caused by mutations in the *JAG1* gene in 90% of families and hundreds of different mutations have been described (Li et al., 1997; Oda et al., 1997). The majority are small-scale insertions, deletions or base substitutions, however whole gene deletions have been reported in around 7% of cases (Turnpenny and Ellard, 2012). DNA sequencing can be used to detect the sequence changes as these are spread throughout the gene. Multiplex ligation-dependent probe amplification (MLPA), fluorescent *in situ* hybridisation (FISH), array comparative genome hybridisation (aCGH) or other quantitative methods can be used to detect exonic deletions and duplications as well as whole gene deletions. Deletions of larger regions including this gene have been reported and are associated with extra phenotypes including mental retardation (Turnpenny and Ellard, 2012). A smaller group of AGS families have been shown to carry mutations in the NOTCH2 gene (McDaniell et al., 2006). These are clinical indistinguishable from *JAG1*-associated AGS patients, however there may be a higher incidence of renal disease in *NOTCH2*-AGS patients (McDaniell et al., 2006).

The genes *JAG1* and *NOTCH2* encode the Jagged-1 receptor and Notch-2 ligand proteins, respectively. AGS mutations in both genes are consistent with haploinsufficiency as they are heterozygous protein-truncating mutations, deletions or other predicted loss of function mutations (Li et al., 1997; Oda et al., 1997). Jag-1 and Notch-2 are components of the NOTCH signalling pathway, an essential cellular mechanism during development, explaining the multi-system features of AGS. Mouse and zebrafish models have shown NOTCH signalling to be essential for both the formation of the ductal plate and tubulogenesis processes during development of

the intrahepatic biliary tree, corresponding to the paucity of bile ducts phenotype observed in AGS patients (Lorent, 2004).

AGS is inherited in an autosomal dominant fashion with reduced penetrance and variable expressivity observed within families (Kamath et al., 2003). Many apparently healthy parents have been identified as a consequence of investigations into a more severely affected child. Often these individuals do not meet the clinical diagnostic criteria for AGS or have only mild clinical features; however they carry the same genetic lesion as their child (Kamath et al., 2003). A proportion of AGS mutations are have occurred *de novo* in the child (up to 60% of cases), or within the parental gonad and therefore germline mosaicism is a real possibility (Turnpenny and Ellard, 2012). These complex genetic issues can result in difficult decisions regarding prenatal diagnosis for future pregnancies. As a result genetic diagnosis of AGS and access to appropriate genetic counselling is recommended for these families.

1.3.5. Alpha-1 anti-trypsin deficiency (AATD)

AATD is characterised by lung disease in adults and liver disease in children and adults. Chronic obstructive pulmonary disease (COPD), specifically emphysema, is the most common manifestation of AATD in the lung. Presenting features include breathlessness, coughing and wheezing, typically for a significant period of time before diagnosis (American Thoracic Society/European Respiratory Society, 2003). Cigarette smoking is the biggest lifestyle factor influencing lung disease in AATD, with the age of onset being 30-50 years in smokers, and over age 50, or sometimes not at all, in never-smokers (American Thoracic Society/European Respiratory Society, 2003). Some individuals present with asthma or chronic bronchitis

(American Thoracic Society/European Respiratory Society, 2003). Liver disease in AATD has two clinical phases. Firstly, some children with AATD present with neonatal cholestasis and usually this resolves spontaneously (Sveger, 1976). Outcomes following neonatal presentation vary from some children having progressive liver damage, cirrhosis and requiring liver transplantation, to some being seemingly healthy into adulthood (Sveger and Eriksson, 1995). Secondly, adults with AATD present with fibrosis, cirrhosis and have an increased risk of hepatocellular carcinoma (Fairbanks and Tavill, 2008). Other rare features found in AATD include kidney disease, rheumatoid arthritis, panniculitis and necrotizing vasculitis (Fregonese and Stolk, 2008).

The prevalence of AATD is around 1 in 2,500 in Western Europe (Fregonese and Stolk, 2008), with the Z allele being mostly found in patients of Scandinavian descent and the S allele being more common in the Southern European population (see dbSNP:rs28929474, (Fregonese and Stolk, 2008)) . PI*ZZ homozygotes account for 95% of all AATD cases. AATD is inherited in an autosomal recessive fashion and shows both reduced penetrance and variable expressivity (Fregonese and Stolk, 2008).

AATD is caused by mutations in the *SERPINA1* gene which encodes the protease inhibitor Alpha-1 anti-trypsin (AAT) (Yamamoto et al., 1986). AAT is produced and secreted by hepatocytes into the bloodstream and is detectable in all tissues (Carlson et al., 1988). Deficiency of AAT leads to cell damage by protease activity, for example by neutrophil elastase, and the alveolar tissue of the lungs appear to be particularly affected. The most common mutation in *SERPINA1* is c.1096G>A p.Glu342Lys (NM000295.4) and results in the production of the Pi*Z protein variant (Jeppsson, 1976). This forms polymers and insoluble aggregates that are retained

within the endoplasmic reticulum of hepatocytes rather than being excreted (Fairbanks and Tavill, 2008). Other protein variants, such as S, may result in reduced serum AAT, but not the formation of polymers and damage to hepatocytes. Individuals with rarer null alleles are likely to have earlier onset of lung damage than individuals with S or Z alleles, which have some residual AAT activity (Ferrarotti et al., 2014).

1.3.6. Other genetic causes

1.3.6.1. Cystic fibrosis

Cystic fibrosis (CF) is the most common autosomal recessive condition in patients of Northern European origin and is caused by mutations in the *CFTR* gene (Riordan et al., 1989). It is characterised by meconium ileus, chronic obstruction of the lungs, recurrent lung infections, pancreatic insufficiency and infertility (Moskowitz et al., 1993). Some CF patients have chronic liver disease, and severe disease involving cirrhosis and portal hypertension occurs in 3 to 5% of CF patients (Lindblad et al., 1999). Liver involvement is a feature which varies between patients with the same CFTR genotype. Known genetic factors which contribute to this variation include mutations in the *MBL2* gene (Gabolde et al., 2001) and the Z-allele of the *SERPINA1* gene (Bartlett, 2009). CF is commonly diagnosed by a newborn screening programme in the UK.

1.3.6.2. Galactosaemia

Galactosaemia is a defect in galactose metabolism that usually presents in the first few weeks of life with jaundice, hepatomegaly, hepatic and renal dysfunction, and cataract (Bosch, 2006). The defect is in the *GALT* gene which encodes a crucial enzyme for galactose metabolism (Reichardt and Woo, 1991). Galactosaemia can

be diagnosed biochemically and may be treated by dietary restriction of galactose (Bosch, 2006).

1.3.6.3. Tyrosinaemia type 1

Tyrosinaemia type 1 (TYRSN1) is caused by mutations in the *FAH* gene which encodes the last enzyme in the tyrosine metabolism pathway; fumarylacetoacetase (Phaneuf et al., 1991). Acute cases of TYRSN1 will present in infancy with hepatomegaly, acute liver failure and cirrhosis (Laberge, 1969). TYRSN1 may be diagnosed using biochemical measurements.

1.3.6.4. Hereditary fructose intolerance

Hereditary fructose intolerance is caused by mutations in the *ALDOB* gene (Cross et al., 1988). In infants, this condition tends to present at weaning when fructose is introduced to the diet. Clinical features can include vomiting, failure to thrive and hepatomegaly (Baerlocher et al., 1978). Long term exposure to fructose in undiagnosed children can lead to cirrhosis (Ali et al., 1998).

1.3.6.5. Congenital bile acid synthesis defects

Congenital bile acid synthesis defects cause progressive liver disease, presenting with neonatal cholestatic jaundice, hepatomegaly, splenomegaly, and giant cell hepatitis and inflammation on biopsy (Cheng et al., 2003). Mutations have been identified in four genes known to be responsible for the different subtypes; *HSD3B7* (Schwarz et al., 2000), *AKR1D1* (Setchell et al., 1988), *CYP7B1* (Setchell et al., 1998) and *AMACR* (Ferdinandusse et al., 2000). The genes encode enzymes responsible for catalysing steps in the synthesis of bile acids. Deficiency of these enzymes leads to low levels of bile acids in the duodenum and associated problems

absorbing fats and fat-soluble vitamins in the gut. Patients generally respond well to bile acid replacement therapy; in comparison to PFIC patients (Cheng et al., 2003).

1.3.6.6. ARC syndrome

Arthrogryposis, renal abnormalities and cholestasis (ARC) syndrome is a rare disease characterised by the named features presenting in infancy. Liver-specific features include paucity of bile ducts, giant cell hepatitis and portal tract fibrosis (Gissen et al., 2006). It has been shown to be caused by mutations in one of two genes, *VPS33B* (Gissen et al., 2004) and *VIPAS39* (Cullinane et al., 2009).

1.3.6.7. NISCH syndrome

Neonatal icthyosis-sclerosing cholangitis syndrome (NISCH) is a very rare autosomal recessive condition caused by mutations in the *CLDN1* gene (Hadj-Rabia et al., 2004). It is characterised by hypotrichosis, scalp alopecia and scarring, sparse eyebrows, jaundice, pruritis, hepatomegaly and cholestasis (Hadj-Rabia et al., 2004). *CLDN1* encodes claudin-1; a tight junction protein.

1.3.6.8. Wolman disease

Lysosomal acid lipase deficiency (LALD) is a lysosomal storage disorder caused by mutations in the *LIPA* gene (Aslanidis et al., 1996). LALD is responsible for two phenotypes in humans; Wolman disease is a rare, infantile-onset form of LALD with feeding difficulties, diarrhoea and vomiting, hepatomegaly and splenomegaly. Cholesterol ester storage disease (CESD) is later-onset form which is more variable in presentation and progression. It is thought that Wolman disease results from complete loss of function mutations in *LIPA* whereas CESD mutations have some residual protein function (Aslanidis et al., 1996).

1.3.6.9. Caroli disease

Caroli disease is a congenital condition characterised by cystic dilation of intrahepatic bile ducts. Jandice, cholestasis, cholangitis, hepatomegaly and liver cirrhosis can occur. There is an association and overlap with autosomal recessive polycystic kidney disease (ARPKD). Mutations in the *PHKD1* gene cause ARPKD, congenital hepatic fibrosis and at least some cases of Caroli disease, in an autosomal recessive inheritance pattern (Adeva et al., 2006; Hao et al., 2014).

1.3.6.10. North American Indian childhood cirrhosis

North American Indian childhood cirrhosis (NAIC) is a disorder of isolated liver cirrhosis, predominantly affecting the bile ducts, found in Ojibway-Cree families in Quebec, Canada. Mutations in the *CIRHA1* gene encoding the cirhin protein have been identified as the cause of the disorder in these patients; however the incidence of NAIC in other ethnic groups is not known (Chagnon et al., 2002).

1.3.6.11. Familial hypercholanaemia

Familial hypercholanaemia is characterised by raised serum bile acid concentrations, fat malabsorption and pruritus. Originally associated with the Lancaster County Old Order Amish population, its incidence in other ethnic groups is not known. A missense mutation in the *TJP2* gene and a missense mutation in the *BAAT* gene were identified as the cause of disease in the majority of cases (Carlton et al., 2003).

1.3.6.12. Zellweger syndrome

Peroxisomal biogenesis disorders (PBDs) present as a spectrum of phenotypes including Zellweger syndrome (ZS). ZS is a neonatal-onset condition with jaundice and liver dysfunction as well as hypotonia, seizures and distinctive facial features. One of two common mutations in the *PEX1* gene is identified in about 80% of all

PBD patients (Waterham and Ebberink, 2012). Sequencing of the coding regions of *PEX1* will detect around 50% of all mutations, however the rest are distributed throughout at least 14 different genes; with *PEX6*, *PEX10*, *PEX12* and *PEX26* genes being the most common after *PEX1* (Waterham and Ebberink, 2012).

1.3.6.13. Gaucher disease

Gaucher disease (GD) is another autosomal recessive condition that may present as liver disease in infancy; typically hepatosplenomegaly (Jmoudiak and Futerman, 2005). Type I GD patients also have cytopenia, bone disease and lung disease, and type II and III patients also have CNS involvement. GD is a lysosomal storage disease cause by mutations in the *GBA* gene leading to deficiency of the glucocerebrosidase enzyme (Tsuji et al., 1987).

1.3.6.14. Mitochondrial DNA Depletion

Mitochondrial DNA Depletion syndrome is a group of disorders characterised by reduced copy number of mtDNA in one or more tissues. It can be caused by biallelic mutations in one of several nuclear genes; including *DGOUK*, *POLG*, *C10orf2*, *MPV17* and *RRMB2* (Nogueira et al., 2014). Patients with *DGOUK* mutations may present with an early-onset multisystem disease including cholestasis, progressive liver dysfunction and liver failure, or isolated liver disease later in childhood.

1.3.6.15. Dubin-Johnson syndrome

Dubin-Johnson syndrome is a relatively mild condition characterised by conjugated hyperbilirubinaemia and darkly stained deposits in an otherwise normal liver. It is caused by biallelic mutations in the *ABCC2* gene which encodes the MRP2 protein (Wada et al., 1998). MRP2 (also known as cMOAT) is expressed in the canalicular membrane of hepatocytes where is functions as an exporter of conjugated bilirubin.

1.3.6.16. Rotor syndrome

Rotor syndrome (RS) is a disease manifesting as mild conjugated and unconjugated bilirubinaemia. RS is an unusual autosomal recessive condition that requires complete loss of function mutations to be present on both alleles of two closely located genes; *SLCO1B1* and *SLCO1B3* (van de Steeg et al., 2012). These genes encode the proteins OATP1B1 and OATP1B3 respectively, which are hepatocyte membrane transporters responsible for uptake of bilirubin from sinusoidal blood.

1.3.6.17. Cholestasis-Lymphedema syndrome

Cholestasis-Lymphedema syndrome (also known as Aagenaes syndrome) is likely to be a genetic condition, with features of neonatal cholestasis and recurring cholestatic episodes. So far the gene responsible has not been identified but a region on chromosome 15 has been mapped (Bull et al., 2000).

1.4. Diagnostic genetic testing

Diagnostic genetics laboratories carry out genetic testing on samples from patients suspected of being affected by, or at risk of, or with a family history of, genetic disease. Various molecular or cytogenetic methods might be employed to look for specific genetic lesions that might be responsible. Ideal characteristics for a test to be used for diagnostic testing are that it should be accurate, robust, reproducible, amenable to up-scaling or automation, allow turnaround targets to be met and be cost effective whilst maintaining reasonable standards. These characteristics can be considered in the context of the ACCE framework for the evaluation of genetic tests; consisting of Analytical validity, Clinical validity, Clinical utility and Ethical, social and legal issues (Grosse and Khoury, 2006).

1.4.1. Inheritance patterns and pathogenic mechanisms

Autosomal recessive genetic conditions are those in which affected individuals inherit mutations on both alleles of a gene, and are typically caused by mutations that result in absence or reduction in protein or protein function; 'loss of function'. Those individuals with only a single loss of function allele are usually not affected by the condition and are often called 'carriers'. Autosomal dominant genetic conditions, in contrast, are those in which affected individuals have inherited a defective copy of a gene from one parent only, and this is sufficient to cause disease. The mechanism of pathogenicity for dominantly inherited conditions varies, and includes 'gain of function' and 'dominant negative' mutations; however loss of function mutations can also result in dominant disorders where there is 'haploinsufficiency' (T. Strachan, 2011). Haploinsufficiency results when one loss of function mutation leads to reduced levels of the protein product and this reduction is sufficient to cause the symptoms of the disease (T. Strachan, 2011). Not all individuals with one dominant mutation will manifest the disease. Often 'dominant' disorders show variability between individuals with mutations in the same gene; observed as variable expressivity, reduced penetrance, genetic predisposition effects, etc (T. Strachan, 2011). Mutations which have not been inherited from either parent are presumed to have occurred in the germline of a parent or during fetal development and are called *de novo* mutations (T. Strachan, 2011). Other modes of inheritance not discussed further include X-linked, mitochondrial and imprinting.

1.4.2. Types of mutation

There are a variety of types of mutations known to cause disease in humans and include small changes at the nucleotide level (such as base substitutions and

deletions, insertions and duplications of small numbers of bases), copy number changes (deletions and duplications ranging from a few hundred bases to millions of bases) and rearrangements (mutations which are copy number neutral but are aberrantly located in the genome and can result in disruption of gene sequences at the breakpoints). Other types of mutation include microsatellite repeat expansion mutations and mitochondrial DNA mutations. The vast majority of variation in the human genome does not cause disease (T. Strachan, 2011).

1.4.3. Testing strategy

In the six genes of interest; *NPC1*, *NPC2*, *ATP8B1*, *ABCB11*, *ABCB4* and *SLC25A13*, there is a large degree of allelic heterogeneity in terms of known disease-causing mutations. Variation in any part of a gene can result in loss of function alleles; however the vast majority of known disease-causing mutations in humans are located within the coding exons and the exon boundaries of the genes (T. Strachan, 2011). Therefore, an efficient testing strategy for these genes is mutation scanning of the coding exons and traditionally, this has been done using Sanger sequencing; dideoxy-sequencing (Sanger et al., 1977). Other mutation scanning methods are possible (for example high resolution melt curve analysis and denaturing high performance liquid chromatography); however the resources required were not available for this project and these techniques are not discussed further.

1.4.4. Sanger sequencing

Over the years since Sanger initially reported this method (Sanger et al., 1977), it has been developed to increase throughput and turnaround, utilising several commonly used modifications, robotics and sequencing analysis software (Wilson et

al., 1990). Fluorescent capillary Sanger sequencing (FS) has revolutionised diagnostic molecular genetics services and become the 'gold standard' method, however there are two major limitations. Firstly, costs are high for reagents, machinery and staff to carry out the work; it typically costs clinicians hundreds of pounds to sequence one gene in their patient, compared with much cheaper biochemical tests (De Leeneer et al., 2011). Secondly, throughput is limited; increasing demand for gene sequencing means that many diagnostic genetics laboratories are reaching saturation point in their FS capacity (De Leeneer et al., 2011). These two factors have a significant impact on the service that a typical laboratory can offer patients, for example a neonate presents with a severe but undefined phenotype whose management would benefit from a rapid molecular diagnosis but requires several genes to be sequenced. Even such an urgent sample might take a few months to be processed. Another example would be when a patient has been given a definite clinical diagnosis of an autosomal recessive condition, but only one mutation is identified by the FS test; the second mutation might be far into the intron or in the promoter region or indeed in another gene, however these are not often screened due to the constraints of finance and capacity. There are many further examples of how patient care can be improved by faster, cheaper and more extensive sequencing. When this work began in 2009, alternative methods for DNA sequencing other than FS were being considered by diagnostic labs.

1.4.5. Microarray sequencing

Microarray sequencing (MS) is a technology that has evolved from SNP genotyping arrays and array CGH (comparative genome hybridisation) (Hacia, 1999). DNA probes are designed to cover the target regions. In total, there are eight probes for

each base of target DNA on the array, which differ in their central base (A, C, G and T, forward and reverse strands). DNA from the test sample is labelled with a fluorescent dye and hybridised to the array. The array is then washed and scanned and the fluorescence measured. Bioinformatics is then used to interpret the pattern of hybridisation at each base and read the DNA sequence. The method relies on the principle that the test DNA sequence will hybridise most strongly to the version of the probe that is complementary (see example in Figure 1.6). For heterozygous base changes, hybridisation will occur at two probe sites. In this way, sequencing microarrays can detect single nucleotide variation in the heterozygous or homozygous state. Insertions, deletions and duplications are not well detected by this technology because the wild-type allele always gives a normal hybridisation pattern, and this is especially true for such mutations in the heterozygous state. Homozygous mutations of this type may be suspected by regions of low hybridisation (deletions) or apparent base substitutions (insertions). To improve the detection of these mutations, extra probes can be added to the design, which are complementary to DNA with known insertions and deletions.

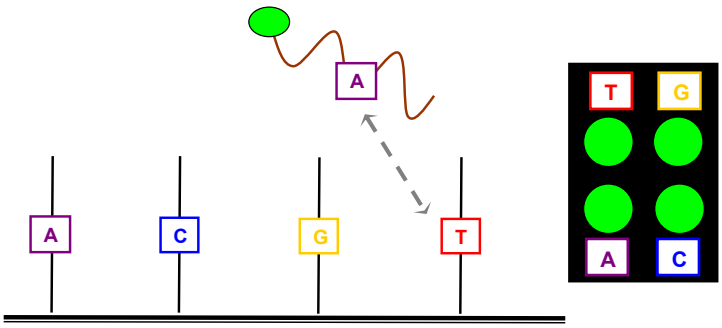


Figure 1.6: A diagram showing the principle involved in hybridisation-based sequencing. This sketch illustrates how the fluorescently tagged sample DNA hybridises most strongly to the complementary resequencing array probe, and this is translated into the brightness of the fluorescent signal emitted from each feature (spot) on the array.

1.4.6. Next Generation Sequencing

The last few years has seen the advent of Next Generation Sequencing (NGS; also known as massively parallel sequencing). This umbrella term encompasses several different methods and combinations of technologies from different companies; however it generally refers to the significant advancements made in increasing the capacity of sequencers and reducing the costs of sequencing per base (Metzker, 2010). Within NGS technology, second generation sequencing (SGS) refers to technologies which rely on a two-step sequencing (clonal amplification and then sequencing) whereas third generation sequencers are sensitive enough that they do not require the amplification step, and so are sometimes called Single Molecule Sequencers. As mentioned, SGS techniques have two steps; a PCR step to amplify the DNA fragment library, and then a combined sequencing and detection step (Metzker, 2010). Fragment library is a term used to describe the starting DNA that is to be sequenced and the production of this will be described later. 454 sequencing was the first method to be published and is the method used in this project. Illumina (Solexa) sequencing was next on the market and over-time has become the dominant method used for both research and diagnostic sequencing of human DNA. Life Technologies' SOLiD sequencing method was next to be launched but has largely failed to compete with the earlier established methods. Life Technologies hoped to regain its place in the DNA sequencing market by taking over Ion Torrent and acquiring its promising cheaper technology. These four technologies will be discussed in turn and criteria important to diagnostic laboratories are compared in Table 1.3.

Table 1.3: A comparison of capacity, read length and run time of commercially available second generation sequence platforms with Sanger sequencing.

Company	Method	Machine	Capacity	Read length	Run time
Applied Biosystems	Sanger	3730xl	96 samples ~86kb approx.	Up to 900bps	60-120 minutes
Roche	454	GS Junior+	70Mb	Up to 700bps	18 hours
		GS FLX+	700Mb	Up to 700 bps	23 hours
Illumina	Solexa	MiSeq (v3 kit)	3.3-15Gb	75-300bps	21-55 hours
		NextSeq (High output kit)	25-120Gb	75-150bps	11-29 hours
		HiSeq 2500 (High output mode)	128 Gb – 1Tb	36-125bps	29 hours - 6 days
Applied Biosystems	SOLiD	5500 W	120Gb	75bps	7 days
	Ion Torrent	Ion PGM	30Mb-2Gb	200-400bps	3-8 hours
		Ion Proton	Up to 10Gb	200 bps	2-4 hours

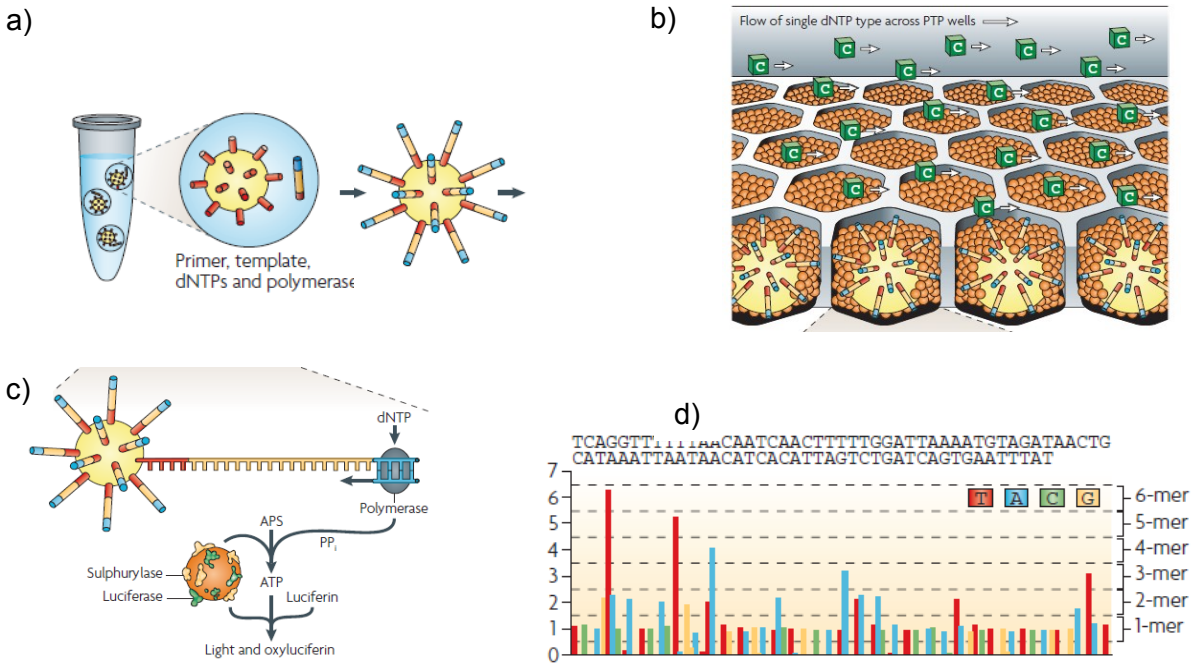
Capacity is a measure of the amount of bases sequenced in one run and should be divided by the depth of coverage required and the fragment length being sequenced, ie 70Mb from a GS Junior run = 100,000 reads of 700bps long at minimum 30x coverage = approximately 3,300 PCR fragments per run. This is not directly comparable for Sanger sequencing as each PCR fragment is sequenced individually.
Read length is important for sequence alignment with longer sequencing reads being more easily aligned, especially when aligning to the human genome reference.
Run time varies considerably between machines with Ion Torrent being the fastest chemistry, this is because it does not rely on capturing light released by the reactions, ie post-light chemistry. Sanger sequencing using the latest machines can allow 384-well setup and automated plate loading, however the amount of time to sequence a plate is limited by the number of capillaries on the machine and remains constant.

Numbers taken from Roche 454, Illumina and Applied Biosystems (Thermofisher) websites, June 2016.

1.4.6.1. 454 Sequencing

454 Sequencing uses emulsion PCR for its amplification step followed by pyrosequencing in its two platforms; GS FLX and GS Junior (Roche Diagnostics™). For the emulsion PCR (emPCR) the fragment library (see 1.4.6.5) is mixed with beads, which have DNA sequences complementary to the adaptor sequence attached, and a water-and-oil emulsion. These are mixed in the correct proportion to achieve one bead and one DNA molecule in one microreaction (or micelle; see Figure 1.7a). A PCR is carried out to clonally amplify the DNA fragment on the surface of the bead (see Figure 1.7a). The emulsion is broken and the beads are isolated and spread out on a picotitre plate so that one bead (with its clonally amplified DNA fragments attached) fits into each well on the plate (see Figure 1.7b). A massively parallel pyrosequencing reaction is then carried out and detected in real-time by the sequencer; nucleotides of known type are added to the plate and if incorporated emit light, and no light is emitted if the nucleotide is not incorporated (Figure 1.7c). The sequencer detects which areas of the plate emit light in response to addition of which nucleotides and the sequence is built up (Figure 1.7 d).

Figure 1.7: A diagram showing the steps involved in 454 sequencing.
a) emPCR b) beads in picotitre plate c) pyrosequencing reaction d) read is produced
Images taken from (Metzker, 2010)



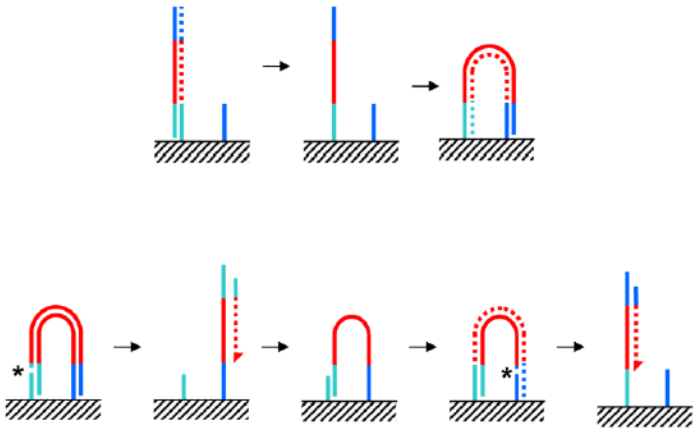
1.4.6.2. Illumina Sequencing

The strategy utilised by Illumina's various sequencing platforms involves bridge amplification PCR and reversible chain-terminating sequencing (Metzker, 2010). Bridge amplification is shown in Figure 1.8. Then is reversible chain-terminating sequencing, which uses four types of fluorescently-tagged chain terminating nucleotides (ddNTPs). A mix of fluorescent nucleotides is washed over the slide, one nucleotide is incorporated per DNA fragment, the fluorescence is detected by the machine and the fluorescent tag is cleaved off allowing extension from the 3' end possible again. This process is repeated many times and the nucleotide incorporated is identified by the colour emitted from each cluster.

Figure 1.8: A diagram showing the principles of bridge amplification PCR.

A fragment library is hybridised to a glass slide covered in a 'lawn' of adaptor primers (blue or green). DNA extension occurs across the bridge between the two fixed adapters (dotted line). Once ligated, the double-stranded DNA is melted into two strands, which are each free to fold over to hybridise to another adaptor. In this way clusters of clonally amplified DNA fragments are produced at a high density across the glass slide (flow cell).

Image taken from (Bentley et al., 2008)



1.4.6.3. SOLiD sequencing

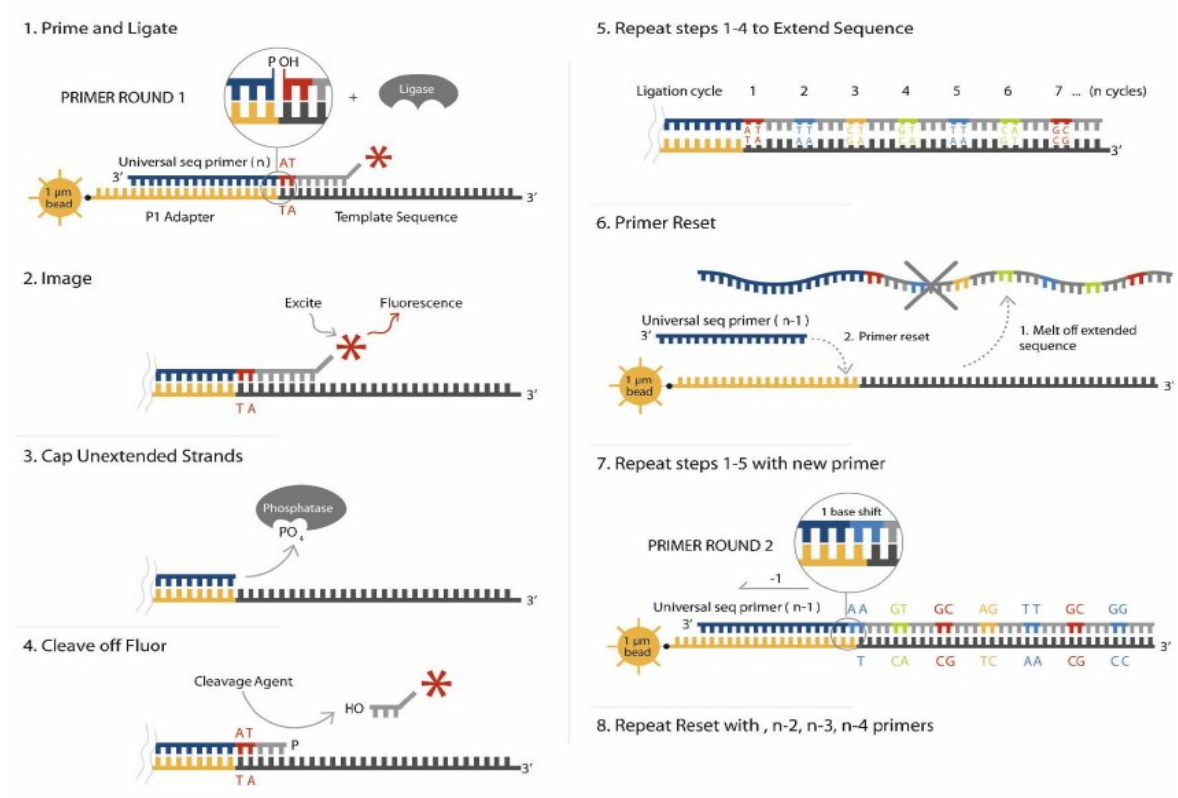
Applied Biosystems (ThermoFisher Scientific, previously Life Technologies) produce the SOLiD™ 4 system and the SOLiD™ PI system, which work using the same chemistries; emulsion PCR and fluorescent ligation-based sequencing. The emulsion PCR step is very similar to what happens with the Roche system (see Figure 1.7a). Later SOLiD machines utilise an alternative to emPCR called 'template walking' which has been packaged into the Wildfire™ chemistry and offers increased speed and automation.

Beads (with clonally amplified DNA fragments attached) are spread onto a glass slide. The sequencing reaction is complicated compared to the other chemistries as it requires several reads for each fragment to be overlaid or combined to obtain the sequence of the DNA, see Figures 1.9a and b (Metzker, 2010).

Figure 1.9: A diagram showing the principles of SOLiD sequencing chemistry.

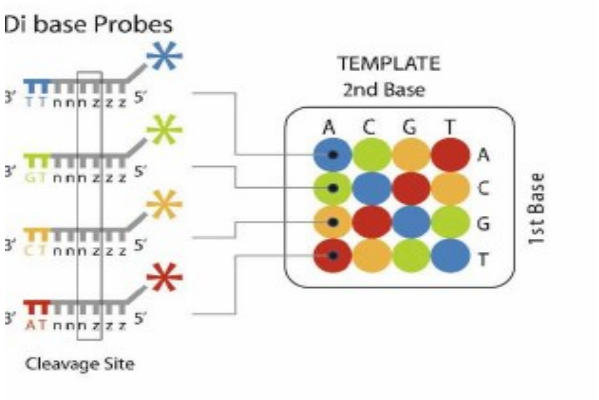
a) A sequencing primer is hybridised to the DNA fragment attached to the bead, then a set of fluorescently labelled probes are added. The probes are hybridised to the DNA fragment and then ligated (if next to the primer). Un-ligated primers are washed off and the fluorescence is detected across the slide by the machine. The DNA tail and fluorescent tag is cleaved off and the cycle is repeated many times. In this way, two bases at a time are interrogated, with three uninterrogated bases in between. The extended primer is removed and a second primer is added, this time n-1 for the adaptor sequence, and the whole process is repeated again. This is done for a total of 5 different primer starting points so that each base is interrogated twice.

Images taken from (Valouev et al., 2008).



b) The fluorescent label is red, green, yellow or blue and is determined by the first two bases (3'end) of the probe. The probe consists of two interrogating bases, followed by 3 degenerate bases, a non-complementary DNA tail and fluorescent label. The individual reads are then overlaid or combined by the software to determine the DNA sequence using the 2-base decoding algorithm.

Images taken from (Valouev et al., 2008).



1.4.6.4. Ion Torrent Sequencing

Ion Torrent sequencing marketed by Applied Biosystems (ThermoFisher Scientific, previously Life Technologies) also uses emulsion PCR as a first step (see Figure 1.7a) but the sequencing technology is different again (see Figure 1.10).

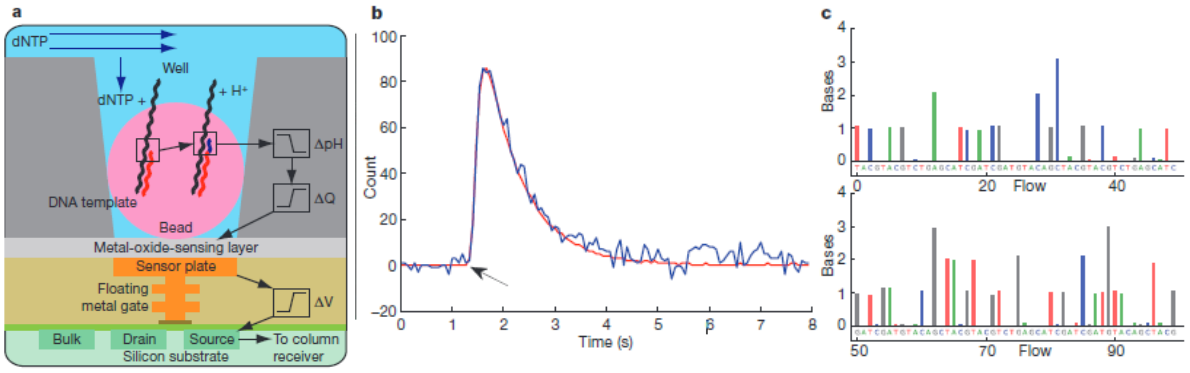


Figure 1.10: A diagram showing the steps involved in Ion Torrent sequencing.

a) The sequencing reactions are carried out on a microchip containing a high density array of wells. In each well, the incorporation of a nucleotide to a growing DNA strand fixed to a bead releases a hydrogen ion (H+).
b) This change in pH is detected by a sensor below the well.
c) One type of base is flooded onto the chip and the change in pH for each well is measured before the process is repeated with another type of base. In this way the DNA sequence in each well is recorded.

Images taken from (Rothberg et al., 2011).

1.4.6.5. Fragment library preparation

Fragment library preparation (FLP) is the DNA template preparation stage which is used for all of the SGS technologies. Sequencing of the whole genome or the whole transcriptome is perversely a more straightforward workflow that that required for targeted sequencing. The whole gDNA or cDNA sample is fragmented to the required length for PCR amplification and sequencing, and then adapter sequences added on to the fragments to allow binding to beads of plates. Alternatively, particular regions of the genome may be targeted or selected for sequencing. The strategy employed for this can vary and factors influencing the choice include the size of the target to be amplified, the number of samples to be tested in a batch, the time for the process, the scope for automation and the cost (see 1.4.6.6).

1.4.6.6. Target enrichment methods

Traditionally in diagnostic genetics, DNA sequencing is targeted to regions of interest (ROI), which is the gene or genes known to be involved in a particular phenotype or syndrome, most usually by polymerase chain reaction (PCR). Straightforward PCR is possible for SGS technologies, however thousands or more individual PCR reactions are required to fill one such run because of the machines' enormous capacities. To minimise the number of PCRs performed long PCR or multiplex PCR can be used. Other strategies are available from commercial companies that aim to solve this 'PCR bottleneck' and are collectively known as target enrichment methods (TEM). TEMs that have been developed are based upon PCR, hybridisation or a combination of both strategies (Mamanova et al., 2010).

Array capture technology is available from commercial companies including Agilent (SureSelect™ products) and Roche Nimblegen (SeqEZ™ products). Long oligonucleotide probes complementary to the regions of interest are synthesised

attached to a glass slide. These probes are then used to hybridise to, and draw out, the target regions from a solution containing fragmented sample DNA. This can either be performed in solution or on the glass slide.

Automated PCR strategies have been developed by Fluidigm Corporation and RainDance Technologies. Fluidigm produce an array system that allows the setup of 48 different PCRs for 48 different samples, a total of 2,304 PCR reactions, in nanolitre volumes. This open platform allows incorporation of adaptor sequences and molecular barcodes, and is flexible in terms of which sequences are to be amplified. RainDance Technologies utilise the ability to produce thermodynamically stable microdroplets. They synthesise a custom designed primer library covering the target regions, this is a pool of primer sets in a tube, each microdroplet containing one set. They also produce a machine onto which the primer library and the samples are loaded and the PCR reactions are carried out. The machine produces microdroplets from the sample DNA and then fuses them with primer microdroplets, creating picolitre sized PCR reactions. This technology allows the production of up to 2 million PCR reactions and amplification using up to 20,000 different primers sets.

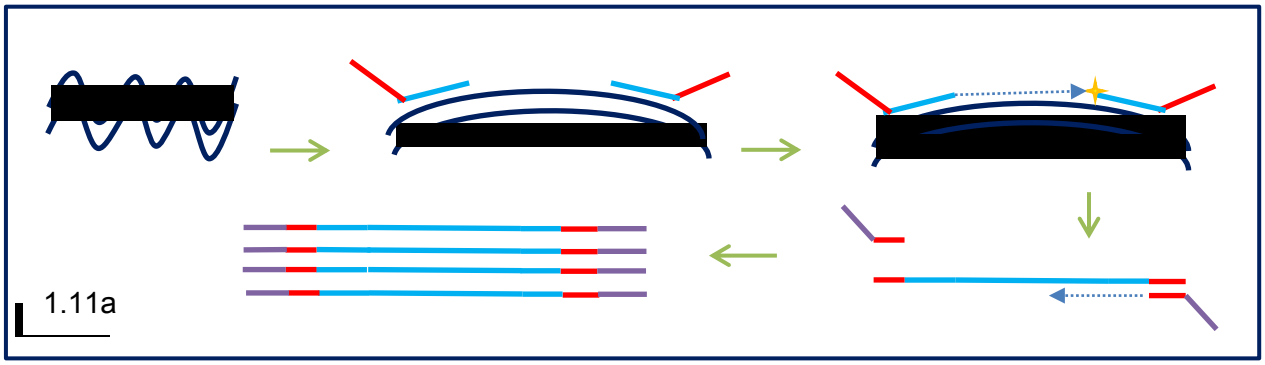
Technologies which combine PCR and hybridisation include Haloplex™ (Agilent Technologies) and TruSeq™ and TruSignt™ Rapid Capture methods (both by Illumina). Figure 1.11 illustrates the steps involved in these processes.

All TEMs involve selection of specific regions of DNA and therefore bias of selection towards wild type sequences is always possible. PCR in particular is sensitive to sequence changes under the primer binding sites, and similarly Haloplex™ is sensitive to sequence changes within restriction enzyme cleavage sites. The use of

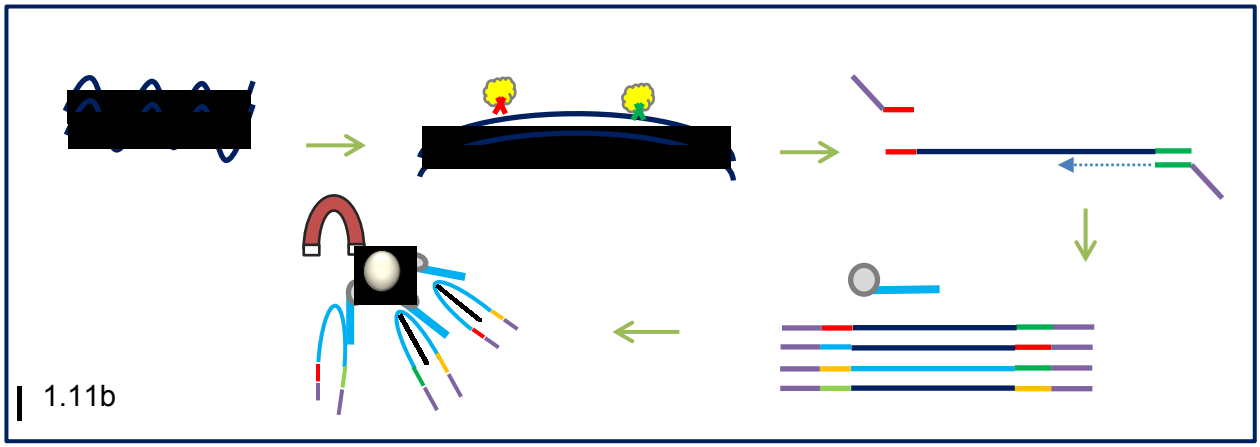
long probes and multiple probes can help to overcome some of these issues.

Figure 1.11: (next page) Target enrichment methods involving PCR and hybridisation.

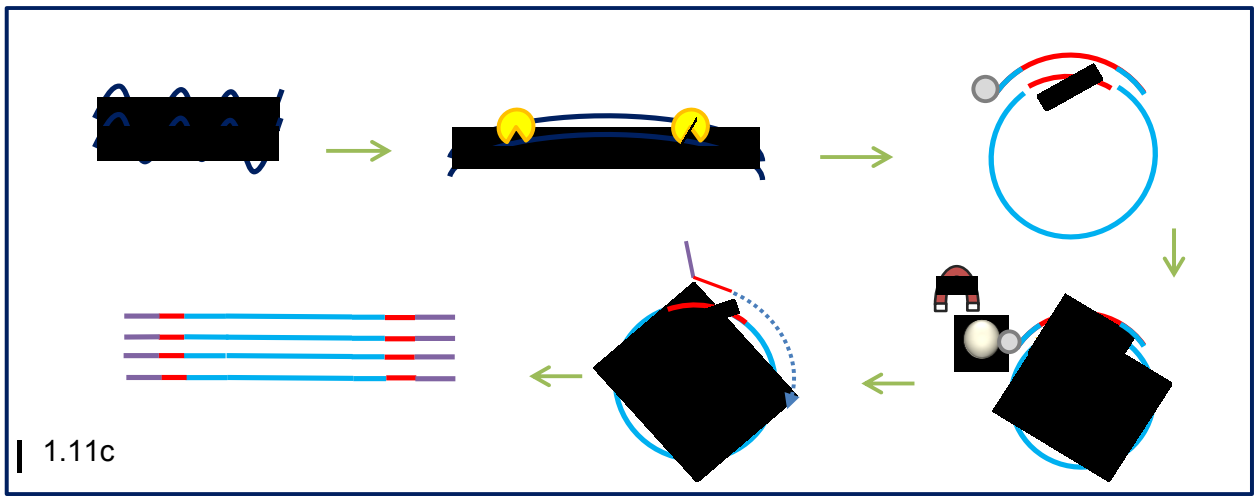
a) TruSeq™. Probes targeted to the regions of interest (ROI) are hybridised with genomic DNA. Extension of the bases between the probes, and ligation, occurs if both probes are hybridised. PCR using primers complementary to the probes is then used to amplify the ROIs.
b) TruSight™ Rapid Capture. Transposons are used to randomly fragment gDNA and add primer sequences. PCR is used to amplify all of the DNA fragments. Biotinylated probes are then used to pull out the fragments containing the ROIs using streptavidin coated magnetic beads.
c) Haloplex™. gDNA is digested using restriction enzymes. Biotinylated probes designed to be complementary to the digested ends of the ROIs are added. Hybridisation with the probes causes the DNA fragments to circularise, and then ligation occurs. Streptavidin coated magnetic beads are used to select the ROIs. PCR primers are added allowing amplification of the selected DNA fragments.



1.11a



1.11b



1.11c

1.4.6.7. Molecular barcodes

Many applications of SGS require the combining of several samples onto one sequencing run and therefore require the addition of molecular barcodes. Molecular barcodes are also known as molecular indices, tags or MIDs. These are short stretches of unique DNA sequence which are incorporated into the sample fragment library, either before, during or after the TEM. Barcodes are usually incorporated using PCR or ligation strategies. These barcodes are recognised by computer algorithms early in the bioinformatics process and allow reads, and therefore results, from individual samples to be separated out (Oliver et al., 2015).

1.4.6.8. Data processing

Data is outputted from NGS machines in various forms however the most basic and useful are the .fasta and .fastq files. These files contain hundreds of thousands to billions of reads of DNA sequence (.fasta) and with associated quality scores (.fastq). These files can then be processed in order to identify sequence variants in the samples and the steps are summarised in Figure 1.12. Usually reads of poor quality are trimmed to improve quality or discarded. The remaining good quality reads are aligned against a reference sequence; this may be for individual genes or the entire human genome. Good alignment is crucial to avoid mutant reads being mapped incorrectly and therefore missed. The aligned reads are then interrogated to identify bases which differ from the reference. Usually a minimum threshold is required in order to confidently call variants, as there is usually a low level of miscalled bases across the sequencing reads. Annotation of the variant calls is required in order to determine where the calls are in relation to genes and mRNA transcripts. A variant call file is produced called .vcf which is an annotated list of all the sequence changes called. Filtering strategies may be employed to the .vcf at this stage and may

include, for example, only listing variants within genes of interest, or only listing variants resulting in severe consequences to the protein.

These processes can be carried out as individual steps or sequentially in what is known as a bioinformatics pipeline (Oliver et al., 2015). Computer software options are available which can perform some or all of these steps (e.g. SoftGenetics NextGene).

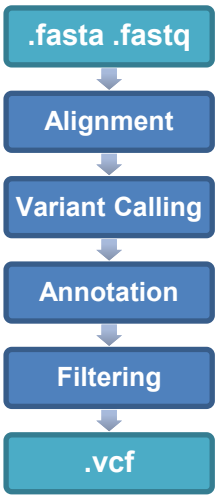


Figure 1.12: The basic steps involved in a bioinformatics pipeline for the processing of data from next generation sequencers.

1.4.7. Interpretation of sequence variants

There is variation in DNA sequence of humans compared to the reference human genome sequence, and much of this variation is 'benign'; that is it has no adverse impact on the health or fitness of those individuals. Therefore, it cannot be presumed that any sequence variant identified when testing a patient is definitely causative of their disease; 'pathogenic'. The first step to ascertaining the relevance of a sequence variant is to predict the effect on the protein product of the gene.

1.4.7.1. Categories of sequence variant

Variants that lead to premature termination codons (PTCs) are usually presumed to be pathogenic, because mRNA transcripts containing PTCs are subject to nonsense-

mediated decay or NMD (Miller and Pearce, 2014). These would be nonsense, frameshift and conserved splice-site mutations (nucleotides at positions -2, -1, +1 and +2 from the exon). Exceptions to this are PTCs which occur in the last exon or the last 30 bases of the penultimate exon, as these are likely to avoid the nonsense-mediated decay machinery (Miller and Pearce, 2014). In addition, the destruction of a splice-site would usually lead to a frameshift and therefore a PTC, however care with interpretation should be taken when there is a possibility of in-frame deletions or intron inclusions, especially when these consist of small numbers of nucleotides. In addition to PTCs, sequence variants that severely disrupt the gene would also be predicted to be pathogenic, for example whole exon deletions.

The predicted effect of missense changes are much more difficult to predict. Some missenses change crucial amino acids that are important for ligand binding or protein structure and are thus pathogenic, whereas others have no effect on the production and function of the protein and are thus benign. Similarly, silent variants (or non-synonymous variants) are not predicted to change the amino acid sequence but can affect the processing of the mRNA, therefore again these may be pathogenic or benign, depending upon their position in the gene. When sequence variants like these are detected in a patient, it is important to gather as much information as available to determine the clinical significance of the variant. Guidelines have been published by the Association for Clinical Genetic Science in the UK and the American College of Medical Genetics and Genomics in order to standardise the classification of sequence variants (Deans et al., 2015; Richards et al., 2015).

1.4.7.2. Methods for classification

Firstly, an important part of the classification process is to determine whether a sequence variant has been identified before, and then to establish whether this was

seen in a patient or in a healthy population. This can be done using literature and database searching. If identified in a patient population, it is important to ascertain whether any evidence other than co-occurrence with disease has been found. This evidence might be segregation with disease in a large family, mRNA, protein or functional studies to look at the downstream effects of the variant, or multiple reports in the literature by several different groups. In general, absence of a variant in a healthy population is evidence for pathogenicity; however care must be taken regarding the size and ethnic make-up of the healthy population tested. Datasets such as 1000Genomes and Exome Variant Server provide good sources of 'normal' variation however; absence of a variant from these datasets does not prove pathogenicity, as the variant may just be very rare, or even unique.

Secondly, *in silico* tools exist that can provide some information for classification of sequence variants. Tools to evaluate missense changes by analysing the conservation of nucleotides and amino acids between species, the physiochemical properties of the amino acids involved and the structure of the protein domains (Williams, 2012). The most commonly used are SIFT, PolyPhen, AlignGVGD and MutationTaster. The results produced can only be used as a guide as there are examples of well-known pathogenic missense changes and benign polymorphisms that give erroneous results using these tools.

There are also splicing prediction tools that can measure the effect of exonic and intronic changes on the strength of surrounding splice sites. Splicing tools have been shown to have a high degree of accuracy when used to assess the effect of variants on wild-type splice-sites in positions between +3 to +7 and -3 to at least -10 from the exon (Hellen, 2009), and similarly can predict the creation of alternative splice-sites at any position.

Thirdly, other circumstantial evidence might exist that can shed light on a particular variant. Examples might be a different change at the same amino acid position as the missense in question which has been causally linked to the disease. Another example could be a missense occurring in a particular mutation hotspot region in a gene, or within a known functional domain of the protein. A good example would be in Alport syndrome where mutations involving the glycine residues of the collagenous domain of the type α5 chain collagen subunit are always pathogenic (Hertz, 2009).

1.5. Aims of the thesis

- To investigate new sequencing methods for the diagnosis of genetic causes of cholestasis
- To determine to incidence of known genetic disorders in infants presenting with cholestasis
- To determine the mutation spectrum in the *ATP8B1*, *ABCB11*, *ABCB4*, *NPC1*, *NPC2* and *SLC25A13* genes in the UK population
- To investigate genotype-phenotype correlations in these genes, especially regarding novel genetic findings.

Chapter 2 – MATERIALS AND METHODS

2. Materials

2.1.1. Patient Material

The patient materials varied so are described within each results chapter. All references to blood can be assumed to be venous blood in EDTA tubes.

2.1.2. DNA extraction kits

Autopure extractions and Puregene manual extractions; RBC Lysis Solution, Cell Lysis Solution, Precipitation Solution, DNA Hydration Solution (all Gentra Puregene Bloodkit components from Qiagen).

QiaSymphony DSP DNA Midi kit (96) (Cat: 937255 Qiagen)

Phenol chloroform reagents; Phenol solution saturated with 100mM Tris-HCl (pH8.0), Phenol/Chloroform 5:1 solution, Chloroform/Isoamyl alcohol 24:1 mix, 100% Ethanol, 70% Ethanol, 3% Virkon, Salt/EDTA solution (4.39g of NaCl, and 50mls of 0.5M EDTA (pH 7.5) made up to 1 litre with distilled water), Proteinase K solution (10mg/ml), 10x Lysis buffer (41.45g ammonium chloride (NH_4CL), 4.6g potassium hydrogen carbonate ($KHCO_3$) and 20ml 0.5M EDTA (pH 7.5) made up to 1 litre with distilled water).

Distilled water (pure filtered water from Elga Veolia Water filtration system)

100% Isopropanol

70% Ethanol

2.1.3. PCR solutions

MegaMix PCR mastermix (Cat: 2MM-1 Microzone)

BioMix™ Red PCR mastermix (Cat: BIO-25006 Bioline)

Long PCR kit reagents; Long PCR Enzyme Mix (5 U/µL) and 10X Long PCR Buffer with 15 mM $MgCl_2$ (Fermentas).

FastStart HiFidelity PCR kit reagents; FastStart High Fidelity Enzyme Blend, (5 U/µl), FastStart High Fidelity Reaction Buffer 10x with 18 mM $MgCl_2$ (Roche).

dNTP 100mM stock (Cat: U1330 Promega)

1x Tris EDTA buffer (10nM Tris-HCL (pH 8.0) and 0.1 mM EDTA; ThermoFisher Scientific Cat: 12090-015).

PCR grade sterile water (Sigma-Aldrich Cat: 03315959001)

PCR primers are described in 2.1.12.

2.1.4. MLPA reagents

MLPA probe sets for *NPC1* and *NPC2* (P193-A2) and for *ABCB4* (P109-B1) (MRCHolland)

SALSA® MLPA® reagents (MRCHolland).

Mineral oil (Cat: M5904 Sigma)

Hi-Di™ Formamide (ThermoFisher Scientific Cat: 4311320)

GeneScan™ 600 LIZ® dye Size Standard v2.0 (Cat:4408399)

2.1.5. FS reagents

ExoSap IT (GE Healthcare)

BigDye 3.1 (Life Technologies Inc)

1x BigDye buffer (Life Technologies Inc)

Agencourt CleanSEQ magnetic beads (Beckman Coulter Inc)

2.1.6. Gel electrophoresis

Agarose Multipurpose (Cat: BIO-41025 Bioline)

Orange G loading dye (Sigma)

10 mg/ml ethidium bromide solution made from powdered stock (Cat: E8751 Sigma)

10x Tris borate EDTA buffer (Severn Biotech)

1kb plus DNA ladder (Invitrogen)

2.1.7. Capiliary electrophoresis

FS; DNA analyzer 3730xl (Life Technologies Inc)

MLPA products; DNA analyser 3130xl (Life Technologies Inc)

2.1.8. Picogreen reagents

100ng/ul λ standard DNA, 20X TE buffer and PicoGreen reagent from Quant-iT™ PicoGreen® dsDNA Assay Kit (Cat: P7589 Invitrogen)

2.1.9. MS reagents

All kits provided by Affymetrix for use with GeneChip® Resequencing chips.

In addition;

5 M Tetramethylammonium chloride solution (TMAC; Sigma)

Tween-20 (10% solution): Pierce, P/N 28320

Acetylated Bovine Serum Albumin (BSA) solution conc. 50µg/µL (Invitrogen)

Herring sperm DNA (Promega Corp)

Tris-HCl, 1 M, pH 7.8 (Sigma)

Molecular Biology Grade Water (Fisher Bioreagents BP2819-10)

SAPE (Molecular Probes)

Anti-streptavidin antibody (goat), biotinylated (Vector Labs)

20X SSPE solution (3 M NaCl, 0.2M NaH2 PO4, 0.02 M EDTA) (Cambrex)

Denhardt's Solution, 50X concentrate (Sigma)

MES hydrate (Sigma-Aldrich)

MES Sodium Salt (Sigma-Aldrich)

5 M NaCl, RNase-free, DNase-free (Ambion)

2.1.10. NGS kits

All kits provided by Roche 454 for use with the GS Junior sequencer.

Sodium hydroxide solution 50% (Merck Millipore Cat: 158793).

2.1.11. M13 primer tags

Forward primer tag sequence TGTAAAACGACGGCCAGT

Reverse primer tag sequence CAGGAAACAGCTATGACC

2.1.12. PCR primers

Primers were Invitrogen Custom DNA Oligos synthesised by Thermo Fisher Scientific. The standard synthesis was 50 nanomoles and they were rehydrated in Tris EDTA buffer to a stock concentration of 200 nanograms per micolitre.

2.1.12.1. PCR primer sequences

Table 2.1: PCR primer sequences.
Note these sequences include the M13 universal sequences.

Gene	Exon	Direction	Primer sequences	Other Info
ATP8B1	01	F	TGTAAAACGACGGCCAGTGCATGCAGGCAGTATTCAAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGACCGATCATCTTTGGCAC	Sanger + NGS
ATP8B1	02	F	TGTAAAACGACGGCCAGTCAACATTTGAATCTGGGGAAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGGCAGGTGTAGCATGAAGG	Sanger + NGS
ATP8B1	03	F	TGTAAAACGACGGCCAGTCTGTAAGCTGTGGGACTTGTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGTGTCTACAGCTTAAATGTTATCGAG	Sanger + NGS
ATP8B1	04_05	F	TGTAAAACGACGGCCAGTTTAGCAAGCTTAGAATTAGCAATAAG	Sanger only
		R	CAGGAAACAGCTATGACCTCTGAATGTGTTTCTAGGCAGAG	Sanger only
ATP8B1	04	F	TGTAAAACGACGGCCAGTTGATGACGGTGATGAGACTTG	NGS only
		R	CAGGAAACAGCTATGACCGAAGGAAGATGGGGAAATGC	NGS only
ATP8B1	05	F	TGTAAAACGACGGCCAGTAATCAAAGACCTGGTGGACG	NGS only
		R	CAGGAAACAGCTATGACCGAACCTGGGAGGTGGAAG	NGS only
ATP8B1	06_07	F	TGTAAAACGACGGCCAGTAAAGAATAGTAATTCCCTTGCCTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTCAGTGGAATGAATGTGCC	Sanger + NGS
ATP8B1	08_09	F	TGTAAAACGACGGCCAGTTCATGTCCAGGTATGGCTAATG	Sanger only
		R	CAGGAAACAGCTATGACCGACAGAAAAGCAATCCCCTC	Sanger + NGS
ATP8B1	08	F	TGTAAAACGACGGCCAGTCCAAACCTGAAGTTCCAAGG	NGS only
		R	CAGGAAACAGCTATGACCGGGTTCTTCACATTCAATAAAACC	NGS only
ATP8B1	09	F	TGTAAAACGACGGCCAGTTTTTAGTTGCTTGGTTCAATTCC	NGS only

ATP8B1	10	F	TGTAAAACGACGGCCAGTAGGCGTAAGCCACCATGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTCCACCTAAAGAGGTAAGATTTTG	Sanger + NGS
ATP8B1	11	F	TGTAAAACGACGGCCAGTGAAATGCAAGAGGTTGGAAATC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAATGAAGGCACTATGTTGGG	Sanger + NGS
ATP8B1	12	F	TGTAAAACGACGGCCAGTTCTCTACTACTTGCCACTCGTATCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGTAATGACCTGCACACGGC	Sanger + NGS
ATP8B1	13	F	TGTAAAACGACGGCCAGTAGCAAAGCCAGGTAAGGAGG	Sanger only
		R	CAGGAAACAGCTATGACCTGCTGGGACCCTGACATC	Sanger only
ATP8B1	13	F	TGTAAAACGACGGCCAGTAGACCAGGGGTCATCTAACG	NGS only
		R	CAGGAAACAGCTATGACCCCAAGTCTCCAAGCCAAGG	NGS only
ATP8B1	14	F	TGTAAAACGACGGCCAGTGAAAAGTTCATGTCATTTGTTAAGTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCATAAGCAGGAGTTACCTGG	Sanger + NGS
ATP8B1	15	F	TGTAAAACGACGGCCAGTGGCCTCAACAATGAGCTCTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCAATACAATGGGCACAAGC	Sanger + NGS
ATP8B1	16	F	TGTAAAACGACGGCCAGTTCTAGAGGAACTGATGGTTTTAAGC	Sanger only
		R	CAGGAAACAGCTATGACCTATCCCAGGAAGTCAGTGGC	Sanger only
ATP8B1	16	F	TGTAAAACGACGGCCAGTTGGACAACAGAGCAAGACCC	NGS only
		R	CAGGAAACAGCTATGACCGGTCAGCAGTTAACATACAGCTTTC	NGS only
ATP8B1	17	F	TGTAAAACGACGGCCAGTTCAAAGTAAAATTTCAGACTTGATCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCTTCCATTGTGCCAGTGTC	Sanger + NGS
ATP8B1	18	F	TGTAAAACGACGGCCAGTAGGAGAGCAGCAACCAGG	Sanger only
		R	CAGGAAACAGCTATGACCTCATCTTGGGCAAAGGAAAC	Sanger only
ATP8B1	19	F	TGTAAAACGACGGCCAGTCTAGGCAGTGGGAGTGAGATG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCTTCTACAGACAGTCTTGCATTTG	Sanger only
ATP8B1	19	R	CAGGAAACAGCTATGACCAAGTGACACATCGTAACCCC	NGS only
ATP8B1	20	F	TGTAAAACGACGGCCAGTAAAGTTATCTCAGAGTCAAGGGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCATCTAAAAGTGGCTCCAAATG	Sanger + NGS

ATP8B1	21	F	TGTAAAACGACGGCCAGTCTTGGGAATGGTACTCCTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCATTCCAGCCATTTCTCCTC	Sanger + NGS
ATP8B1	21A	R	CAGGAAACAGCTATGACCACCATGGCCTTCTGCTTG	NGS only
ATP8B1	21B	F	TGTAAAACGACGGCCAGTAAGAAAGACGGATGCGGAC	NGS only
ATP8B1	22	F	TGTAAAACGACGGCCAGTGATGTTTTGCCTCACAATCG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCTGTAAGGAGACACAGCCCC	Sanger + NGS
ATP8B1	23	F	TGTAAAACGACGGCCAGTGCTAACTAGGCTGGCATGTG	Sanger only
		R	CAGGAAACAGCTATGACCTTGAAACGTTTGCTTGGGAC	Sanger only
ATP8B1	23	F	TGTAAAACGACGGCCAGTGCTTAAGGCCAGGAGTTCG	NGS only
		R	CAGGAAACAGCTATGACCGACCCTTGATGCCTGACAAC	NGS only
ATP8B1	24	F	TGTAAAACGACGGCCAGTAACATATCTCTGACTGCTTTGACC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGTGCAGTGGGAGTCAGGTG	Sanger + NGS
ATP8B1	25	F	TGTAAAACGACGGCCAGTCAAGCCACATCATGCCTAAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCACTGTGCCCAGCCATTC	Sanger + NGS
ATP8B1	26	F	TGTAAAACGACGGCCAGTCCACACCTGGCGAAATATTAAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGAAACGCTTTGGTTTCTGTG	Sanger + NGS
ATP8B1	27	F	TGTAAAACGACGGCCAGTCCTCCTCCCTGGTGTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCACACACACAAAGTCCTGAGAG	Sanger + NGS
ABCB11	02	F	TGTAAAACGACGGCCAGTCAAATTGTTCTTTCGTTTGGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGCTCCTTGAAACTTGACCAG	Sanger + NGS
ABCB11	03	F	TGTAAAACGACGGCCAGTTGAGCAGGAAGAAAGAAAAGG	Sanger only;
		R	CAGGAAACAGCTATGACCCCTAGAAGGGATATTCCAAAAGG	Sanger only;
ABCB11	04	F	TGTAAAACGACGGCCAGTGCCAGTGGGGATTTTCTTTC	Sanger only
		R	CAGGAAACAGCTATGACCAACACTCCCCTCATGATCTAAAC	Sanger only
ABCB11	04	F	TGTAAAACGACGGCCAGTAGGTTTGGCCAGAATTTTCC	NGS only
		R	CAGGAAACAGCTATGACCGGATAGGATTTGTATGCTTTATGAGC	NGS only
ABCB11	05	F	TGTAAAACGACGGCCAGTAGTCCTCCTACCTCTCCTGC	Sanger + NGS

		R	CAGGAAACAGCTATGACCTCAGCCAGTAAAATCCCCTC	Sanger + NGS
ABCB11	06	F	TGTAAAACGACGGCCAGTAATCTCTGGTGGCTTGATCC	Sanger only
		R	CAGGAAACAGCTATGACCGTGGCAACACATTGCATCTC	Sanger only
ABCB11	07	F	TGTAAAACGACGGCCAGTCCCCTTTTCTCAACTGTTGTATTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCAATTTAGAAACAAGGGTTTTATTATCC	Sanger + NGS
ABCB11	08	F	TGTAAAACGACGGCCAGTGAGAGATGGGAATGTTTAAAAGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCAGGAAAAGGGACTCAAGC	Sanger + NGS
ABCB11	09	F	TGTAAAACGACGGCCAGTGACAGACTGACTTACCTAATTTCTTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCGCTTTGCACAAACTGAG	Sanger + NGS
ABCB11	10	F	TGTAAAACGACGGCCAGTGTAAAACCACTGCATCACGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCTGAAGGCACCAAAGTAATAAAC	Sanger + NGS
ABCB11	11	F	TGTAAAACGACGGCCAGTTGCGTTAACATGGAAGACCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGAGTTCATTCTGTGCCCCAC	Sanger + NGS
ABCB11	12	F	TGTAAAACGACGGCCAGTGCAGAGATACGCCAAAGATG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGAAACAGAGTCAGGCTTCAG	Sanger + NGS
ABCB11	13	F	TGTAAAACGACGGCCAGTAAGCATCTGCACCTGTAGCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCTGCCATTTGCACTTTACTG	Sanger + NGS
ABCB11	14	F	TGTAAAACGACGGCCAGTTGCCCATTGGTCAAGTATG	Sanger only
		R	CAGGAAACAGCTATGACCCTAAAACATGGCTTAAGAATTTAATG	Sanger only
ABCB11	14A	F	TGTAAAACGACGGCCAGTCCCATTGGTCAAGTATGAC	NGS only
		R	CAGGAAACAGCTATGACCAAGTTGTAGGCATTGGCCTC	NGS only
ABCB11	14B	F	TGTAAAACGACGGCCAGTTGGCTTAGAGATCAGATTGGG	NGS only
		R	CAGGAAACAGCTATGACCAATTTAATGACTTGGGAATCATAC	NGS only
ABCB11	15	F	TGTAAAACGACGGCCAGTTTATAGTGGATCACTGTCAGAAGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGCAGCACAAGCATTTCCAC	Sanger + NGS
ABCB11	16	F	TGTAAAACGACGGCCAGTTGATGCAAAGGTCAGTGTCAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCATAGAAAACCGTAAAGCACTATAGAC	Sanger + NGS

ABCB11	17	F	TGTAAAACGACGGCCAGTGTAAAGAATTCTACTTGGATATGGTTC	Sanger only
		R	CAGGAAACAGCTATGACCCAGAGTTTCCTTGTTGTACCTGAG	Sanger + NGS
ABCB11	17	F	TGTAAAACGACGGCCAGTGAATGTTTTGGCATTTGACATAG	NGS only
ABCB11	18	F	TGTAAAACGACGGCCAGTACACCAGTTGATCCTGCTCC	Sanger only
		R	CAGGAAACAGCTATGACCAAAGGGTACCCAACAGTCCC	Sanger only
ABCB11	18	F	TGTAAAACGACGGCCAGTCCATGTGACCTACCAAACATTTC	NGS only
		R	CAGGAAACAGCTATGACCTTTTAGTCTGACTTGAAACACTGC	NGS only
ABCB11	19	F	TGTAAAACGACGGCCAGTTGTGAATGCCAAAGGATCTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCATGAAAACAAAGAGCGGAC	Sanger + NGS
ABCB11	20_21	F	TGTAAAACGACGGCCAGTCCCACCAGAATGATACATTTCC	Sanger only
		R	CAGGAAACAGCTATGACCATCCCACTGGTCCCTATTCC	Sanger + NGS
ABCB11	20	F	TGTAAAACGACGGCCAGTAGGGTTCACTCTTGGTGTTG	NGS only
		R	CAGGAAACAGCTATGACCAGGAGCTCCCCAGATTTAGC	NGS only
ABCB11	21	F	TGTAAAACGACGGCCAGTTCTTTTCACCCAATTTCTACAGG	NGS only
ABCB11	22	F	TGTAAAACGACGGCCAGTTGGTAATTGGTAAAAGCGACTG	Sanger only
		R	CAGGAAACAGCTATGACCGGCTGACAGCTTCCTTCAGT	Sanger only
ABCB11	23	F	TGTAAAACGACGGCCAGTGCCACTGAAATGTCACGAAAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCAGAACCAGGCTATTCCTTCC	Sanger + NGS
ABCB11	24	F	TGTAAAACGACGGCCAGTATCACACCAACCACGCC	Sanger only
		R	CAGGAAACAGCTATGACCCAACCTTACCCCTCATCAATAC	Sanger only
ABCB11	24	F	TGTAAAACGACGGCCAGTTCTGCCTTCAGGTCATCACA	NGS only
		R	CAGGAAACAGCTATGACCCCCTGTGTCCATGTGTTCTG	NGS only
ABCB11	25	F	TGTAAAACGACGGCCAGTAAACTCAAGATTTAGGTGTGTTTTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGGGGTTGGAAATACTCTGC	Sanger + NGS
ABCB11	26	F	TGTAAAACGACGGCCAGTAAGCAAACCAAATGTCCTGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGCTCAACCTGTACACTCTGG	Sanger + NGS
ABCB11	27	F	TGTAAAACGACGGCCAGTGAGTTCAGTACAGCACAGGAGC	Sanger + NGS

		R	CAGGAAACAGCTATGACCTTGAAAATAGTGCCATTTTATTAAGG	Sanger + NGS
ABCB11	28	F	TGTAAAACGACGGCCAGTTTGTTATTCAGGTCGTGTTAACTG	Sanger only
		R	CAGGAAACAGCTATGACCTGGTGGCTGAGCTGCCACTTG	Sanger only;
ABCB4	02	F	TGTAAAACGACGGCCAGTCTGGCCTCGAAGGGAGAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAGGGTGAACTTAGTCCTGCTG	Sanger + NGS
ABCB4	03	F	TGTAAAACGACGGCCAGTATTTTGTGAACATCACTCTTATATTTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGATTACAAGTATGAGTCAGCTCG	Sanger + NGS
ABCB4	04	F	TGTAAAACGACGGCCAGTTGAAGGCCTCCTTTTCTAAGAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAATGAATAGCAAAATCAACTCCC	Sanger + NGS
ABCB4	05	F	TGTAAAACGACGGCCAGTACTTAACAATAGCATCTTCTAGCTTTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGGGTAAAGAGTACACGTTATTTG	Sanger + NGS
ABCB4	06	F	TGTAAAACGACGGCCAGTCTGAGATGGTGCCACTGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTTCCTTGACATATTTTCACACAG	Sanger + NGS
ABCB4	07	F	TGTAAAACGACGGCCAGTTTTGTTGGATGTCTACTTCATCTTTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGAACAGGTACAAGTACGAGACTTCTG	Sanger + NGS
ABCB4	08	F	TGTAAAACGACGGCCAGTAAAGGAGAGGGTTTGGGAAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGGTTAATATTAGGAAAGGGAAGG	Sanger + NGS
ABCB4	09	F	TGTAAAACGACGGCCAGTGTGTGACTCGGACTATGGATTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGCGATATCAAAGAAAAGAGAAGG	Sanger + NGS
ABCB4	10	F	TGTAAAACGACGGCCAGTGGATAAACCTAAACTTAATCCTTTATGTA	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAAAATATGCAAACTAAAGCCAGA	Sanger + NGS
ABCB4	11_12	F	TGTAAAACGACGGCCAGTGATTGTGACATTCCAGGTCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTACCAAAACTGGATTCACACG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCCCAAAGGAAAAGGCAC	NGS only
ABCB4	12	F	TGTAAAACGACGGCCAGTTTATGTGCCTTTTCCTTTGG	NGS only
ABCB4	13	F	TGTAAAACGACGGCCAGTTTCTTGCATATTGCTGTTTATTTC	Sanger only

		R	CAGGAAACAGCTATGACCCAGGGGACTTATCTAGCAAAGTTGGAC	Sanger only
ABCB4	13	F	TGTAAAACGACGGCCAGTTGAATGGTCCTGATACTTCAGC	NGS only
		R	CAGGAAACAGCTATGACCCAAAGTTGGACAATCTTGCATC	NGS only
ABCB4	14	F	TGTAAAACGACGGCCAGTCAAAGCTCCATGTTGTCTTTATG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCAGCCCAGACTCCGGAAGCAC	Sanger + NGS
ABCB4	15	F	TGTAAAACGACGGCCAGTTGCATGTTAATTGACAGTGTGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTTCTTCTTGCTCAGTATAGCATTC	Sanger + NGS
ABCB4	16	F	TGTAAAACGACGGCCAGTTCCTTGATTGAGAAGCAGTTAGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAAGAGTATGGCTCATAGTAGCAGTC	Sanger only
ABCB4	16A	R	CAGGAAACAGCTATGACCGGTTTCCACATCAAGGCTCT	NGS only
ABCB4	16B	F	TGTAAAACGACGGCCAGTTGTCTTGATATTCTTTCAGACATCAG	NGS only
		R	CAGGAAACAGCTATGACCAAAGAGTATGGCTCATAGTAGCAGTC	NGS only
ABCB4	17	F	TGTAAAACGACGGCCAGTATGGCCATGCCTTTTCTATG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGCTGCTTAATCCCAGAATG	Sanger + NGS
ABCB4	18	F	TGTAAAACGACGGCCAGTAAACATGTGACACTCAAGCCAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGCAGAGCCTTATGCCAATC	Sanger + NGS
ABCB4	19	F	TGTAAAACGACGGCCAGTGCTACATAAAAGTTGAACATGATTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGCAGGACTCTGCCCCATCC	Sanger + NGS
ABCB4	20	F	TGTAAAACGACGGCCAGTGCTCAGAGCAAGGCCAGGGAC	Sanger only
		R	CAGGAAACAGCTATGACCTCTGGAGGGGGCAGTGGGTC	Sanger only
ABCB4	20	F	TGTAAAACGACGGCCAGTAATTTCCTCAGCATTGGAGC	NGS only
		R	CAGGAAACAGCTATGACCCATCTTAACAAGTGTGGGTATGC	NGS only
ABCB4	21	F	TGTAAAACGACGGCCAGTGGAGCGCATGCATTTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGCTCAACGTTTTATGTTAATCACC	Sanger only
ABCB4	21A	R	CAGGAAACAGCTATGACCTCTCTTTTGGCATTTCCAGC	NGS only
ABCB4	21B	F	TGTAAAACGACGGCCAGTCAGGAACCAGGTTGGCTTTA	NGS only
		R	CAGGAAACAGCTATGACCAGTTGTAGTGGGCACAAACATT	NGS only

ABCB4	22	F	TGTAAAACGACGGCCAGTACCTTCAGGATACTTTTGACAGAGCCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCATATCATTGTTTGGAGCAGCAGAGC	Sanger + NGS
ABCB4	23	F	TGTAAAACGACGGCCAGTGGCCAGACTTAGAAGCCGTGCTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGACCTCATCTTTGGACACAGG	Sanger + NGS
ABCB4	24	F	TGTAAAACGACGGCCAGTGGGATGATTAAGGAGTAAAGGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCAATCTAGCAGACAGCAAACCTTAG	Sanger + NGS
ABCB4	25	F	TGTAAAACGACGGCCAGTCAGTCTTTGGTAAAGTTTCCTTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCATTATGACAATATTGGTTGGG	Sanger + NGS
ABCB4	26	F	TGTAAAACGACGGCCAGTGGAGGGTTGTTGTTTTGAAGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGTTGGCATAACTTTGGTAATTG	Sanger + NGS
ABCB4	27	F	TGTAAAACGACGGCCAGTTGTAAATAGAACTGTCAACTGTTAAGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCATGGTTGACAGCAAAATCCC	Sanger + NGS
ABCB4	28	F	TGTAAAACGACGGCCAGTAGCATGGGAACCCATTTGTG	Sanger only
		R	CAGGAAACAGCTATGACCAGGGAGAGCTAGCCTGTTGT	Sanger only
ABCB4	28A	R	CAGGAAACAGCTATGACCCCTGGACACTGACCATTGAA	NGS only
SLC25A13	01	F	TGTAAAACGACGGCCAGTGACTAGAAGTGAGCCGCCC	Sanger only
		R	CAGGAAACAGCTATGACCACCAACCCAGACACGTGAG	Sanger only
SLC25A13	02	F	TGTAAAACGACGGCCAGTCGGCTAGCTTGATTTCTCAGC	Sanger only
		R	CAGGAAACAGCTATGACCGGGCTGACACTTTGGGAC	Sanger + NGS
SLC25A13	02	F	TGTAAAACGACGGCCAGTTGGGGCAACATGGTATTTAAG	NGS only
SLC25A13	03	F	TGTAAAACGACGGCCAGTGAACTGTTGGGAGATAATGGTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGTCCTAAGAGATGGGAAGG	Sanger + NGS
SLC25A13	04	F	TGTAAAACGACGGCCAGTTTTGTATATGCTTCTGTTTTCCAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAATGCTCACACAAGTCCACA	Sanger only
SLC25A13	04	R	CAGGAAACAGCTATGACCTCAAAACTACTAAGCCAGAAACTTG	NGS only
SLC25A13	05	F	TGTAAAACGACGGCCAGTAAGCGATATCCATTCAATAACAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCATCTCCTGACCTCGTCATCC	Sanger + NGS

SLC25A13	06	F	TGTAAAACGACGGCCAGTGGCAGATGAGGGCTTGTTAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAAACACACTCTTTGTTTGAGTTTAG	Sanger + NGS
SLC25A13	06	F	TGTAAAACGACGGCCAGTAAACTCTCTTTCTTCAGTCTGATTTT	NGS only
		R	CAGGAAACAGCTATGACCCCAAAACACACTCTTTGTTTGA	NGS only
SLC25A13	07	F	TGTAAAACGACGGCCAGTTCATGATTAGTTGCAGTTGCTTC	Sanger only
		R	CAGGAAACAGCTATGACCTGGTGAAGATTGTTTTGTTTGC	Sanger only
SLC25A13	07	F	TGTAAAACGACGGCCAGTTCATGATTAGTTGCAGTTGCTTC	NGS only
		R	CAGGAAACAGCTATGACCACACCACAATACTTGCAGGC	NGS only
SLC25A13	08	F	TGTAAAACGACGGCCAGTTCACTCATTCCAGTGCCTTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCATTGCTGCCCTCCTCCTAAC	Sanger only
SLC25A13	08	R	CAGGAAACAGCTATGACCTTCAGTATAGCCTTCAGTTTGGC	NGS only
SLC25A13	09	F	TGTAAAACGACGGCCAGTGGCAGCAATCAGGAGAAAAA	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGAACAGGGTTGGGGTATC	Sanger + NGS
SLC25A13	10	F	TGTAAAACGACGGCCAGTCATGGATTTAGAACCCAATGAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGATAACTGGCATTGGGAAAG	Sanger + NGS
SLC25A13	11	F	TGTAAAACGACGGCCAGTTTGTCCTTGACTTTAACTACCTTCTTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCTTGCTAATTCATGTCAGGC	Sanger + NGS
SLC25A13	12	F	TGTAAAACGACGGCCAGTTTGCAGCAGAGATTAAGCAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCCCCTGCTTTCAAATTAGC	Sanger + NGS
SLC25A13	13	F	TGTAAAACGACGGCCAGTTTGACAGATGCATATTTATAAGTGTTT	Sanger + NGS
		R	CAGGAAACAGCTATGACCTTCCACCTCTGGAATGGTTC	Sanger + NGS
SLC25A13	14	F	TGTAAAACGACGGCCAGTTCATCTCTTCCTGACCCACC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTGTTGAAGAATAGTTTCTGCATTAGG	Sanger + NGS
SLC25A13	15	F	TGTAAAACGACGGCCAGTAGCTTGATCTCCTCAATGGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTAGCATGCAGCTAGGGAAGG	Sanger + NGS
SLC25A13	16_17	F	TGTAAAACGACGGCCAGTACGGCCAGCAGTTCAAAG	Sanger only
		R	CAGGAAACAGCTATGACCCCTTTCCCTACGACAACAGAG	Sanger only

SLC25A13	16	F	TGTAAAACGACGGCCAGTGAGGACAGCAAAATAGGCATT	NGS only
		R	CAGGAAACAGCTATGACCGGGGTGAGGATCGAAATACA	NGS only
SLC25A13	17	F	TGTAAAACGACGGCCAGTGCTGGTGGTATGGAAATAATGTG	NGS only
		R	CAGGAAACAGCTATGACCTGAGGTACCTTTCCCTACGAC	NGS only
SLC25A13	18	F	TGTAAAACGACGGCCAGTTGATGAGAATGTATCAACTCCTTTAC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCATTGCTTCATTCCCAGGAG	Sanger + NGS
NPC1	01	F	TGTAAAACGACGGCCAGTAACAGCCCGGGGAAGTAG	Sanger only;
		R	CAGGAAACAGCTATGACCCTCCATCGCCAGACCAAC	Sanger only;
NPC1	02	F	TGTAAAACGACGGCCAGTTGAGTGGGCACTTCTTGTTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCACCTCCACCCTGCAATAAC	Sanger + NGS
NPC1	03	F	TGTAAAACGACGGCCAGTTGAATGAATGTGTCTTAGTTCACTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCAAAGAATAAATGGAAAGCTGAG	Sanger + NGS
NPC1	04	F	TGTAAAACGACGGCCAGTTTTAAAATCGTTCTTGCTGGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCAATTTGCTCTGCTGTCCTG	Sanger + NGS
NPC1	05	F	TGTAAAACGACGGCCAGTTCTTGCCTCGTGAATTACAGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCACTGTGCCCAGCCAGTTC	Sanger + NGS
NPC1	06	F	TGTAAAACGACGGCCAGTGTTGTTTTATGTATTTCAGTGGGC	Sanger only
		R	CAGGAAACAGCTATGACCTGGAGGTATTTGTTTCTTGTCC	Sanger only
NPC1	06	F	TGTAAAACGACGGCCAGTTTCAGTGGGCTTTTCTTTGAG	NGS only
		R	CAGGAAACAGCTATGACCGAAAGCTCAAAGTGCCAGTG	NGS only
NPC1	07	F	TGTAAAACGACGGCCAGTGCCAGGAGGAGGAAGAAAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCACACCACCTCACCCACTG	Sanger + NGS
NPC1	08	F	TGTAAAACGACGGCCAGTACTTTCAGGAACGGCTTGG	Sanger only
		R	CAGGAAACAGCTATGACCCATGTAAAAGCCAGCAAACC	Sanger only
NPC1	08A	F	TGTAAAACGACGGCCAGTACGGCTTGGCTCTTAACCTC	NGS only
		R	CAGGAAACAGCTATGACCGAAGAAAGGCCCAAAGTGC	NGS only
NPC1	08B	F	TGTAAAACGACGGCCAGTTCATTACTGCGTGTTCGTCAG	NGS only

		R	CAGGAAACAGCTATGACCCAGCAAACCACAAGGTCATC	NGS only
NPC1	09	F	TGTAAAACGACGGCCAGTTGACCCTCAGGGCAATG	Sanger only
		R	CAGGAAACAGCTATGACCTTTGCTCACCTCTGGGTTATG	Sanger only
NPC1	09	F	TGTAAAACGACGGCCAGTACCCTCAGGGCAATGCTG	NGS only
		R	CAGGAAACAGCTATGACCTGCCCATGTACCCTAAGTCAG	NGS only
NPC1	10	F	TGTAAAACGACGGCCAGTAGGGCCCATGTTGTCCTTAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGTAAGAAATTAACAAAACTGCCC	Sanger + NGS
NPC1	11	F	TGTAAAACGACGGCCAGTGAGCCCAGAGATACAGTCCATAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCGTAACTCAGATCTGCCATTG	Sanger + NGS
NPC1	12	F	TGTAAAACGACGGCCAGTAAAACGTGGCCTTTGTATCG	Sanger only
		R	CAGGAAACAGCTATGACCTGAAGAAAATAGATGTAGGCAACAG	Sanger only
NPC1	13	F	TGTAAAACGACGGCCAGTGGAGGACCTTTTAGTAACAAGTGG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCACACTCACGAATGCGG	Sanger + NGS
NPC1	14	F	TGTAAAACGACGGCCAGTGCTTAGAAGACACTGCTAATCGTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAAAGGAAGCAACACAAAGGG	Sanger + NGS
NPC1	15_16	F	TGTAAAACGACGGCCAGTCATGAACATAAGACCTGCAGAGAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCTTAGAAGGCATGTGATAATCTG	Sanger only
NPC1	15	R	CAGGAAACAGCTATGACCATACCCGCTAGCTGCTTCC	NGS only
NPC1	16	F	TGTAAAACGACGGCCAGTAGAGGTAAGTTGGTGCCAGG	NGS only
		R	CAGGAAACAGCTATGACCAATCTCCTTCCCAGGCTGTC	NGS only
NPC1	17	F	TGTAAAACGACGGCCAGTGCCCTGTACTCCCTATTAGCC	Sanger + NGS
		R	CAGGAAACAGCTATGACCGGAAACCCTGTCACCATTTG	Sanger + NGS
NPC1	18	F	TGTAAAACGACGGCCAGTGGGAAGACTGGTGGTGTTAGG	Sanger only
		R	CAGGAAACAGCTATGACCTTTTCAGTGAGACATTTCAGGC	Sanger only
NPC1	18	F	TGTAAAACGACGGCCAGTGATCCTCGCCTTGCTTAGTTAC	NGS only
		R	CAGGAAACAGCTATGACCGACTGCCTGGCTGAGAGC	NGS only
NPC1	19	F	TGTAAAACGACGGCCAGTTGAAACTAAAGACTTCCTCCCTG	Sanger + NGS

		R	CAGGAAACAGCTATGACCCAAATAGGTATAAACTGAGGCACG	Sanger + NGS
NPC1	20	F	TGTAAAACGACGGCCAGTAAGAAAGTAATGCCCCTCACTG	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCATGCAACTGTCTTAGCCC	Sanger + NGS
NPC1	21	F	TGTAAAACGACGGCCAGTTTTGCTTAGCCTCAAGTGCTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCACCCAGTGTAGGCCCTTTG	Sanger + NGS
NPC1	22	F	TGTAAAACGACGGCCAGTGTGACAGGATGAACACGCAG	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCTAAGACAGCCAATTCCCC	Sanger + NGS
NPC1	23	F	TGTAAAACGACGGCCAGTCTTGGCCTCCTCTAGCACC	Sanger + NGS
		R	CAGGAAACAGCTATGACCAGAAGCTGCTTTGTAAGTACAGG	Sanger + NGS
NPC1	24	F	TGTAAAACGACGGCCAGTctggggcaggagaatcactt	Sanger + NGS
		R	CAGGAAACAGCTATGACCtgaaaagaatgcctcaggataga	Sanger + NGS
NPC1	25	F	TGTAAAACGACGGCCAGTGCTGGTCTCAAGCAATTCTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCTCCGGTGTTCAACTTGGC	Sanger + NGS
NPC2	01	F	TGTAAAACGACGGCCAGTTGGTTACTGGTGACAGGTCG	Sanger only
		R	CAGGAAACAGCTATGACCCAGTTAGGTAGGGTCCAAGGC	Sanger only
NPC2	02	F	TGTAAAACGACGGCCAGTGGGAGAGCAGAGCACCTTC	Sanger + NGS
		R	CAGGAAACAGCTATGACCATTCATGACTGCCAATTCCC	Sanger + NGS
NPC2	03	F	TGTAAAACGACGGCCAGTGACCCTAGGAATGCTGTTGC	Sanger + NGS
		R	CAGGAAACAGCTATGACCCCCATCTCTGCTTCTTGCC	Sanger + NGS
NPC2	04	F	TGTAAAACGACGGCCAGTTAAGGCTGTAAGCTGTGCCC	Sanger only
		R	CAGGAAACAGCTATGACCCTCCACTGTCAGGGCAATAACCC	Sanger only
NPC2	04	F	TGTAAAACGACGGCCAGTAGGATTCATAGTTAAACCAATTATGG	NGS only
		R	CAGGAAACAGCTATGACCCAGCAGTCTTGAAATTCCTTTTATC	NGS only
NPC2	05	F	TGTAAAACGACGGCCAGTAGGAAGGTCCAGCCAGACAGGA	Sanger + NGS
		R	CAGGAAACAGCTATGACCAACCAGCCACCCGGAGCTCA	Sanger + NGS

2.1.12.2. MID primer sequences

Table 2.2 MID primer sequences

MID	Forward Primer	Reverse Primer
MID-1	CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGTGGTCATAGCTGTTTCCTG
MID-2	CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACAGGTCATAGCTGTTTCCTG
MID-3	CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTCACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGAGACGCACTCGGTCATAGCTGTTTCCTG
MID-5	CGTATCGCCTCCCTCGCGCCATCAGATCAGACACGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGATCAGACACGGGTCATAGCTGTTTCCTG
MID-6	CGTATCGCCTCCCTCGCGCCATCAGATATCGCGAGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGATATCGCGAGGGTCATAGCTGTTTCCTG
MID-7	CGTATCGCCTCCCTCGCGCCATCAGCGTGTCTCTAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCGTGTCTCTAGGTCATAGCTGTTTCCTG
MID-8	CGTATCGCCTCCCTCGCGCCATCAGCTCGCGTGTCACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCTCGCGTGTCGGTCATAGCTGTTTCCTG
MID-10	CGTATCGCCTCCCTCGCGCCATCAGTCTCTATGCGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCTCTATGCGGGTCATAGCTGTTTCCTG
MID-11	CGTATCGCCTCCCTCGCGCCATCAGTGATACGTCTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTGATACGTCTGGTCATAGCTGTTTCCTG
MID-13	CGTATCGCCTCCCTCGCGCCATCAGCATAGTAGTGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCATAGTAGTGGGTCATAGCTGTTTCCTG
MID-14	CGTATCGCCTCCCTCGCGCCATCAGCGAGAGATACACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCGAGAGATACGGTCATAGCTGTTTCCTG
MID-15	CGTATCGCCTCCCTCGCGCCATCAGATACGACGTAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGATACGACGTAGGTCATAGCTGTTTCCTG
MID-16	CGTATCGCCTCCCTCGCGCCATCAGTCACGTACTAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCACGTACTAGGTCATAGCTGTTTCCTG
MID-17	CGTATCGCCTCCCTCGCGCCATCAGCGTCTAGTACACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCGTCTAGTACGGTCATAGCTGTTTCCTG
MID-18	CGTATCGCCTCCCTCGCGCCATCAGTCTACGTAGCACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCTACGTAGCGGTCATAGCTGTTTCCTG
MID-19	CGTATCGCCTCCCTCGCGCCATCAGTGTACTACTCACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTGTACTACTCGGTCATAGCTGTTTCCTG
MID-20	CGTATCGCCTCCCTCGCGCCATCAGACGACTACAGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACGACTACAGGGTCATAGCTGTTTCCTG
MID-21	CGTATCGCCTCCCTCGCGCCATCAGCGTAGACTAGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCGTAGACTAGGGTCATAGCTGTTTCCTG
MID-22	CGTATCGCCTCCCTCGCGCCATCAGTACGAGTATGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACGAGTATGGGTCATAGCTGTTTCCTG
MID-23	CGTATCGCCTCCCTCGCGCCATCAGTACTCTCGTGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACTCTCGTGGGTCATAGCTGTTTCCTG
MID-24	CGTATCGCCTCCCTCGCGCCATCAGTAGAGACGAGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTAGAGACGAGGGTCATAGCTGTTTCCTG
MID-25	CGTATCGCCTCCCTCGCGCCATCAGTCGTCGCTCGACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCGTCGCTCGGGTCATAGCTGTTTCCTG
MID-26	CGTATCGCCTCCCTCGCGCCATCAGACATACGCGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACATACGCGTGGTCATAGCTGTTTCCTG

MID-27	CGTATCGCCTCCCTCGCGCCATCAGACGCGAGTATACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACGCGAGTATGGTCATAGCTGTTTCCTG
MID-28	CGTATCGCCTCCCTCGCGCCATCAGACTACTATGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACTACTATGTGGTCATAGCTGTTTCCTG
MID-29	CGTATCGCCTCCCTCGCGCCATCAGACTGTACAGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACTGTACAGTGGTCATAGCTGTTTCCTG
MID-30	CGTATCGCCTCCCTCGCGCCATCAGAGACTATACTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGAGACTATACTGGTCATAGCTGTTTCCTG
MID-31	CGTATCGCCTCCCTCGCGCCATCAGAGCGTCGTCTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGAGCGTCGTCTGGTCATAGCTGTTTCCTG
MID-32	CGTATCGCCTCCCTCGCGCCATCAGAGTACGCTATACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGAGTACGCTATGGTCATAGCTGTTTCCTG
MID-33	CGTATCGCCTCCCTCGCGCCATCAGATAGAGTACTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGATAGAGTACTGGTCATAGCTGTTTCCTG
MID-34	CGTATCGCCTCCCTCGCGCCATCAGCACGCTACGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCACGCTACGTGGTCATAGCTGTTTCCTG
MID-35	CGTATCGCCTCCCTCGCGCCATCAGCAGTAGACGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCAGTAGACGTGGTCATAGCTGTTTCCTG
MID-36	CGTATCGCCTCCCTCGCGCCATCAGCGACGTGACTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGCGACGTGACTGGTCATAGCTGTTTCCTG
MID-37	CGTATCGCCTCCCTCGCGCCATCAGTACACACACTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACACACACTGGTCATAGCTGTTTCCTG
MID-38	CGTATCGCCTCCCTCGCGCCATCAGTACACGTGATACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACACGTGATGGTCATAGCTGTTTCCTG
MID-39	CGTATCGCCTCCCTCGCGCCATCAGTACAGATCGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACAGATCGTGGTCATAGCTGTTTCCTG
MID-40	CGTATCGCCTCCCTCGCGCCATCAGTACGCTGTCTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTACGCTGTCTGGTCATAGCTGTTTCCTG
MID-42	CGTATCGCCTCCCTCGCGCCATCAGTCGATCACGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCGATCACGTGGTCATAGCTGTTTCCTG
MID-43	CGTATCGCCTCCCTCGCGCCATCAGTCGCACTAGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCGCACTAGTGGTCATAGCTGTTTCCTG
MID-44	CGTATCGCCTCCCTCGCGCCATCAGTCTAGCGACTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCTAGCGACTGGTCATAGCTGTTTCCTG
MID-45	CGTATCGCCTCCCTCGCGCCATCAGTCTATACTATACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTCTATACTATGGTCATAGCTGTTTCCTG
MID-46	CGTATCGCCTCCCTCGCGCCATCAGTGACGTATGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTGACGTATGTGGTCATAGCTGTTTCCTG
MID-47	CGTATCGCCTCCCTCGCGCCATCAGTGTGAGTAGTACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGTGTGAGTAGTGGTCATAGCTGTTTCCTG
MID-48	CGTATCGCCTCCCTCGCGCCATCAGACAGTATATAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACAGTATATAGGTCATAGCTGTTTCCTG
MID-49	CGTATCGCCTCCCTCGCGCCATCAGACGCGATCGAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACGCGATCGAGGTCATAGCTGTTTCCTG
MID-50	CGTATCGCCTCCCTCGCGCCATCAGACTAGCAGTAACTGGCCGTCGTTTTACA	CTATGCGCCTTGCCAGCCCGCTCAGACTAGCAGTAGGTCATAGCTGTTTCCTG

2.1.13. Software

Mutation Surveyor version 2.2 (SoftGenetics)

GeneMarker version 1.7 (SoftGenetics)

GSEQ version 4.1 (Affymetrix)

NextGENe version 2.16 (SoftGenetics)

Alamut v2.1 (Interactive Biosoftware)

2.1.14. On-line bioinformatics tools

Primer3 via Broad Institute http://bioinfo.ut.ee/primer3/

Exon-Primer via UCSC genome browser https://genome.ucsc.edu/

Primer Blast via NCBI genome browser

http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome

SNPCheck via NRGL Manchester https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm

PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/

1000 Genomes Project Browser http://browser.1000genomes.org/index.html

Exome Variant Server (browser for the NHLBI GO Exome Sequencing Project data)

http://evs.gs.washington.edu/EVS/

2.2. Methods

2.2.1. Extraction of genomic DNA from whole blood

2.2.1.1. Autopure method

This method was performed using the Autopure LS machine according to the manufacturer's instructions.

2.2.1.2. Qiasymphony SP method

This method was performed using the QiaSymphony SP machine according to the manufacturer's instructions.

2.2.1.3. Manual PureGene method

A minimum of 0.3ml of blood is required for a small-scale extraction or 3ml of blood for a medium-scale extraction. The blood is transferred to a 1.5ml Eppendorf tube (small) or a 15ml Falcon tube (medium). 0.9ml or 9ml of red cell lysis solution is added, respectively. The tube is incubated at room temperature for 10 minutes on a rotator then centrifuged (1 minute at 13,000rpm in a microfuge or 10 minutes at 3200 rpm in a bench top centrifuge) to pellet the white blood cells. The supernatant is poured off and cell pellet re-suspended in the remaining supernatant. 0.3ml or 3ml of cell lysis solution is added, respectively, and the tube is mixed thoroughly by vortexing. 2ul or 15ul of RNAse A solution is added and mixed by vortexing. Then 0.1ml or 1ml of protein precipitation solution and mixed thoroughly by vortexing. The tube is incubated for 15 minutes on a rotator to allow protein precipitation. Further protein precipitation solution can be added if required. The tube is centrifuged to pellet the protein (5 minutes at 13000 rpm in a microfuge or 20 minutes at 3600 rpm in a bench top centrifuge). The

supernatant is transferred to a 1.5 ml Eppendorf containing 300ul of 100% isopropanol or 15ml tube containing 3ml of 100% isopropanol. The tube is gently inverted to precipitate the DNA then centrifuged for 2 minutes at 13000 rpm in a microfuge or 5 minutes at 2000 rpm in a bench top centrifuge. The supernatant is poured off and the pellet washed using 70% ethanol (0.5ml or 2ml). After centrifuging again the supernatant is poured off, the pellet air dried at room temperature for a few minutes and then re-suspended in hydration solution.

2.2.1.4. Phenol Chloroform method

4ml of blood is transferred to a 15 ml falcon tube. An equal volume of cold 2x lysis buffer is added and the tube sealed with parafilm. This is left on ice for 10 minutes then mixed on a rotator rack for 10 minutes. Centrifuge at 3600 rpm for 20 minutes at 4°C to pellet the cell nuclei. The supernatant is poured off and the pellet is resuspended by gentle vortexing in 3ml of Salt/EDTA buffer. 300 µl of 10% SDS and 20 µl of 10mg /ml proteinase K is added before leaving overnight at 37°C or at 55°C for 3 hours. After incubation, 3 ml of phenol solution is added. This is mixed on a rotator rack for 10 minutes, centrifuged at 3600 rpm for 5 minutes at room temperature, and then the aqueous layer is carefully removed and transferred to a new 15 ml falcon tube. 3 ml of phenol/chloroform solution is added to the tube. This is mixed for on a rotator rack for 10 minutes, centrifuged at 3600 rpm for 5 minutes at room temperature, and then the aqueous layer is carefully removed and transferred to a new 15 ml falcon tube. 3 ml of chloroform solution is added to the tube. This is mixed on a rotator rack for 10 minutes, centrifuged at 3600 rpm for 5 minutes at room temperature, and then the aqueous layer is carefully removed and transferred to a new 15 ml falcon tube. Twice that volume of

chilled 100% ethanol is added to the aqueous layer, gently mixing by hand until DNA strands appear. The tube is centrifuged at 2000 rpm for 5 minutes at room temperature to pellet the DNA and the ethanol is poured off. The pellet is washed by adding 2 ml of 70% ethanol. The DNA is transferred into a 1.5 ml Eppendorf tube using a plastic inoculation loop, where it is allowed to air dry at room temperature for a few minutes. 100-400 ul of hydration solution is added, depending upon the size of the pellet. The tube is incubated at 65^{o}C on the Thriller Thermoshaker for 1 hour, then at 37^{o}C overnight to allow the DNA to be fully resuspended.

2.2.2. DNA quantification

2.2.2.1. Nanodrop method

All DNA samples were assessed for concentration and purity using a NanoDrop ND-µ000 machine and associated software. Using the software 'nucleic acid' was selected as the sample type. The pedestal was washed using sterile water and then 70% ethanol prior to use and between samples. Sterile water was used as a blank control, followed by TE buffer, followed by each DNA sample. For each measurement 2µl is added to the pedestal using a pipette, the measurement is taken, the sample is wiped off with a tissue and the pedestal washed as described before. The DNA concentration in ng/µl and 260/230and 260/280 ratios were recorded.

2.2.3. PCR amplification

Megamix, BioMix Red and Long PCR reactions were performed in 0.5 ml thin-walled Eppendorf tubes in a final volume of 25 µl.

2.2.3.1. Megamix method

The PCR reaction used 50 ng of DNA, 0.5 µM each primer and 23 µl of Megamix PCR buffer in a final volume of 25. The thermocycling conditions were as follows; 95°C for 5 minutes, 20 cycles of; 95°C for 1 minute, 65°C for 1 minute (reducing by 0.5°C per cycle), 72°C for 1 minute, followed by 10 cycles of; 95°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute, followed by 72°C for 10 minutes.

2.2.3.2. BioMix™ Red method

The PCR reaction used 50 ng of DNA, 0.5 µM each primer, 0.25 mM dNTPs and 2 units of Taq DNA polymerase in 1x PCR buffer. The thermocycling conditions were as follows; 95°C for 5 minutes, 30 cycles of 95 °C for 1 minute, 60 °C for 1 minute, 72 °C for 1 minute, and 72 °C for 10 minutes.

2.2.3.3. Long PCR method

The PCR reaction used 2µl of 50 ng/µl DNA, 1.25µl of each primer (stock at 20nM), 5µl of dNTP mix (each at 2 mM), 5 µl of 10x Long PCR buffer with 15 mM $MgCl_2$, 0.3 µl of Long PCR enzyme mix and made up to 50 µl with molecular grade water.

2.2.3.4. AA method

PCR products were generated using the 48.48 Access Array™ system (Fluidigm Inc.) according to the manufacturer's instructions.

Sample-specific pools containing 138 PCR products per sample were generated using three arrays and pooling of equal volumes of the products. All PCR primers were tagged at the 5' end with M13 sequences (see 2.1.1) to allow incorporation of MID primers during a second round of PCR (see 2.1.12.2). The second round PCR

components used were as follows; 2µl of 10x PCR buffer*, 3.6µl of 25mM $MgCl_2$*, 1µl of DMSO*, 0.4µl of 10mM dNTPs, 0.2µl of Taq*, 7.8µl of PCR grade sterile water, 4 µl of MID primer mix, and 1 µl of first round PCR product (diluted 1:1000); in 0.2ml thin-walled Eppendorf tubes with a final volume of 20 µl. The thermocycling conditions were as follows; 95°C for 10 minutes, 15 cycles of 95 °C for 15 seconds, 60 °C for 30 seconds, 72 °C for 1 minute, and 72 °C for 3 minutes. Equal volumes of second round PCR products, which had MID primers incorporated, were pooled.

2.2.4. PCR Cleanup

2.2.4.1. ExoSap method

5µl of PCR product was transferred to a new tube or 96-well plate. 0.5µl of ExoSAP-IT and 0.5µl of water per reaction were mixed and 1µl of the mix was added to each reaction. The reactions are then incubated on a PCR machine at 37°C for 15 mins followed by µ0ºC for 15 mins.

2.2.4.2. AMPure bead method

PCR cleanup of the library (final pool), was done using AMPure beads (Beckman Coulter Inc) adding the following components to an Eppendorf tube; 36µl of Agencourt AMPure beads, 24µl of 1xTE buffer and 12µl of library. The library was mixed and incubated at room temperature with the beads for 10 minutes, then placed on a Magnetic Particle Concentrator (MPC; Life Technologies Inc) for 5 minutes. The supernatant was removed and the pellet was twice washed with 170µl of 70% ethanol before returning the tube to the MPC for 5 minutes. The pellet was allowed to air dry at room temperature for a few minutes then resuspended in 40µl of TE buffer.

2.2.5. Fragment Library Quantification

Quantification of the library was done using the PicoGreen ® dsDNA quantitation assay. In a black 96-well flat bottom plate (obtained from ThermoFisher Scientific, catalogue number 611F96BK), a serial dilution of standard λ DNA 2ng/µl was done over wells A1 to A10 to generate a standard curve in a volume of 50µl per well. Well A11 was designated a 'no DNA' control and 50µl of water was added. 25µl of library was added to wells C1-C3, diluted 1 in 10, 1 in 20 and 1 in 40, respectively. The PicoGreen ® reagent was diluted 1 in 200 and 50µl was added to wells A1 to A11. After each addition, 50µl of mix was transferred to the row below B1 to B11. 25µl of diluted PicoGreen ® reagent was added to wells C1 to C3. The plate was transferred to the DTX 880 Multimode Detector (Beckman Coulter Inc). The fluorescence from each well was detected and used to determine the double stranded DNA concentration in well C1 to C3 by plotting the readings against the standard curve; an average of A1 to A10 and B1 to B11 serial dilutions. The DNA concentration and average amplicon length was used to calculate the number of molecules per µl. This allowed the volume of library to be added to the emulsion PCR reaction downstream to be calculated.

2.2.6. Sequencing methods

2.2.6.1. FS method

Sequencing primers (at a final concentration of 0.3µM) was added to 1µl of BigDye 3.1 and 1x BigDye buffer (Life Technologies Inc) in a final volume of 10µl. The sequencing programme was thermal cycling for 30 cycles of denaturing at 95^{o}C for 30 seconds,

annealing 50°C for 10 seconds and extension at 60°C for 4 minutes. Resulting products were purified using Agencourt CleanSEQ magnetic beads according to the maufacturer's instructions. Sequencing products were separated and detected using a DNA analyzer 3730xl (Life Technologies Inc).

2.2.6.2. MS method

PCR product quantitation by PicoGreen, pooling, fragmentation, labelling and hybridisation and was performed according to the GeneChip Custom Resequencing Array Protocol V2.1 (Affymetrix). Arrays were washed and stained using a FS450 fluidics station before being scanned with a GCS3000 7G scanner (Affymetrix). Fluorescence at each probe spot was measured using AGCC software (Command Console V1.0, Affymetrix) and intensity files were generated (CEL file).

2.2.6.3. NGS method

The emulsion PCR and sequencing reactions for the GS Junior system (Roche) were carried out according to the manufacturer's instructions. For the emPCR reaction a ratio of 1 molecule per bead was used ($5x10^6$ molecules of library were added to each of the A and B beads). Consistently, these libraries resulted in enrichment between 5% and 8%, by visual judgement. Around 5-6% was loaded onto the picotitre plate for sequencing.

2.2.7. MLPA

DNA is diluted to a final volume of 2.5µl in a 0.2ml Eppendorf tube using TE buffer and including 25ng to 125ng of DNA per reaction. A drop of mineral oil is added to each tube to prevent evaporation of the small volume. The tubes are heated to 98°C for 5

minutes then cooled to 25°C for 30 minutes using a PCR machine. 0.75µl of MLPA probe and 0.75µl per reaction are mixed together and aliquoted into each tube on the PCR machine below the mineral oil. Hybridisation occurs overnight on the PCR machine using the following program; 1 minute at 95°C, then 16 hrs at 60°C. Following hybridisation, the PCR machine temperature is changed to 54°C. 1.5µl of Ligase Buffer A, 1.5µl of Ligase Buffer B, 0.5µl of Ligase-65 enzyme and 12.5µl of water per reaction are mixed together and aliquoted into each PCR tube. The tubes are incubated for 10-15 minutes at 54°C, then 5 minutes at 98°C then stored 4°C until required. A PCR master mix is made consisting of 1µl of PCR primers, 0.25µl of Polymerase and 3.75µl of Water per reaction. At room temperature 5µl of the mix is added to each tube. The PCR program is as follows; 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute, for 33 cycles, then 72°C for 20 minutes.

1µl of PCR product, 1µl of size standard and 8µl of HiDi formamide are mixed together in a 96-well plate prior to analysis using a DNA analyzer 3130xl (Life Technologies Inc).

2.2.8. Data analysis

2.2.8.1. FS analysis

Analysis of the Sanger sequencing data was done using Mutation Surveyor™ version 3.2 (SoftGenetics®). .seq files were aligned against GenBank reference sequences and variants called. Sequence quality and variant calls were checked and manually recorded by two independent scorers.

2.2.8.2. MLPA analysis

MLPA analysis was done using GeneMarker software. The raw .fsa files for the run were loaded into a new GeneMarker project and run with pre-set parameters for the relevant MLPA kit. The peaks were sized and labelled automatically according to the kit instructions and the size standard peaks. The settings were as follows; AutoRange, Smooth, Peak Saturation, Baseline Subtraction, Pull-up Correction and Spike Removal were all selected, Size Call was Local Southern, Allele Call was Auto Range 120 to 1000, Peak Detection Threshold was >100 intensity, >1% of the maximum, >5% of the local maximum and with a maximum intensity of 30,000. The Plus A Filter and Stutter Peak Filter were selected and set to 25% for the Left and 25% for the Right. Population Normalisation was selected. The MLPA analysis module was used to perform a dosage quotient calculation for each probe, using internal control probes and three normal control samples on the run. The following settings were used; the Analysis Method was MLPA Ratio, the Quantification used Peak Height and the ratios were adjusted using the control probes. In addition to the normal controls, a sample with trisomy of chromosome 13 was included on the NPC runs to mimic a whole gene duplication. Ratios of 0.75 to 1.3 are considered in the normal range, below 0.75 indicates a heterozygous deletion, and above 1.3 indicates a heterozygous duplication. All positive results were repeated.

2.2.8.3. MS analysis

GeneChip Sequence Analysis Software (GSEQ 4.1, Affymetrix) was used to interrogate the CEL files for base calling and variant calling. Base calling assumed the diploid model and a quality score threshold of 2 (default settings were used for all other parameters). A list of software-generated variant calls were transferred into a Microsoft Excel

document, manually checked and decisions recorded. Any calls that were deemed not to be benign polymorphisms or sequencing artefacts were confirmed using FS.

2.2.8.4. tNGS analysis

Analysis of the NGS data for coverage and variant calling was done using NextGENe™ version 2.16 (SoftGenetics®). The GS Junior generated .sff file was converted to .fna using the Format Conversion tool and these settings; Median Score Threshold >=20, Called Base Number of Each Read >= 100, max # of uncalled bases <=2, Trim or Reject Read when >=3 Bases with Score <=16. The Barcode Sorting tool was used to interrogate the run .fna for reads corresponding to all the barcodes used in the laboratory to detect contamination. The AutoRun tool was used to perform three tasks; 1) splitting of the .fna file according to the barcodes used in the particular run to generate individual .fna files for each patient, 2) alignment of the reads against GenBank files and 3) variant calling using these settings (Matching requirement >=100 Bases and 85% homology, Remove Ambiguously Mapped Reads, Detect Large Indels, Rigorous Alignment, Hide Unmatched Ends, Mutation Percentage <=15%, SNP Allele <=3 Counts, Total Coverage <=15 reads, Except for Homozygous) . The Variant Comparison tool was used to generate a list of all the variants called by the software for all of the patients on the run. This was saved as a text file and imported into a Microsoft Excel template to allow scoring of the variants by two independent scorers. Mutation reports (detailing variant calls) and Expression reports (detailing Coverage) were generated for each patient from NextGene Viewer.

2.2.9. Mutation nomenclature

GenBank reference files were used for both sequencing methods and for the six genes were as follows; NM_000271 (*NPC1*), NM_006432.3 (*NPC2*), NM_005603.3 (*ATP8B1*), NM_003742.2 (*ABCB11*), NM_018849.2 (*ABCB4*) and NM_014251.2 (*SLC25A13*). Mutation nomenclature was done according to HGVS guidelines.

2.2.10. Variant interpretation

Variant interpretation was done using Alamut software. This allowed checking of variant nomenclature, literature searching, presence or absence on the HGMD database, and predictions of the effect on protein structure (using Align GVGD, SIFT and PolyPhen tools) and mRNA splicing (using SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and Human Splicing Finder tools). In addition, a search of dbSNP Build 138 was done and the 1000Genomes project data and Exome Sequencing Project data was used as a source of presumed normal variation (see 2.1.14). All sequence variants identified were investigated using this pipeline and classified into the following groups 'pathogenic', 'possibly pathogenic', 'uncertain significance', 'likely benign' and 'benign'. It is important to note that these classifications are based on available information, and may change in response to additional information, i.e. a new publication of mutation data or case studies.

Chapter 3 – EVALUATION OF MICROARRAY SEQUENCING AND TARGETTED NEXT GENERATATION SEQUENCING FOR THE GENETIC INVESTIGATION OF CHOLESTASIS

3.1. Introduction

3.1.1. Diagnostic genetic testing

Accurate genetic diagnosis can be essential to inform decisions on treatment and management of many genetic conditions. As an example, progressive familial intrahepatic cholestasis type 1 (PFIC type 1) caused by mutations in *ATP8B1* (FIC1 protein deficiency) and PFIC type 2 caused by mutations in *ABCB11* (BSEP deficiency) may present with the same clinical features, however liver transplantation is indicated for severe PFIC type 2 than and not for type 1 (Pawlikowska et al., 2010). Genetic testing provides definitive diagnoses and allows accurate assessment of the genetic risk of cholestasis in families. Fluorescent capillary Sanger sequencing (FS) has been the gold standard method for diagnostic laboratories for the detection of unknown mutations, because of its high specificity and developments in automation. However, FS is limited by machine capacity and relatively high consumables costs. In this chapter microarray sequencing (MS) is evaluated and a targeted next generation

sequencing (tNGS) assay is designed, optimised and evaluated, as alternative methods for the diagnosis of genetic cholestasis disorders.

3.1.2. Microarray sequencing

Initial testing of a 300kb custom-designed microarray (Birmingham ReseqUencing Microarray; BRUM1) was done previously using patients with known mutations (Bruce et al., 2010). BRUM1 includes probes to sequence *ATP8B1* and *ABCB11*. In this chapter further evaluation of the BRUM1 microarray using samples simultaneously tested by FS and MS in patients without previously known mutations is described and the advantages and disadvantages of MS for clinical laboratory use are discussed.

This work has been published and the paper is included in Appendix I: Mutation detection in cholestatic patients using microarray resequencing of *ATP8B1* and *ABCB11* (McKay et al., 2013).

3.1.3. Targeted Next Generation Sequencing

The combination of 48.48 Access Array (Fluidigm), GS Junior (Roche 454) and NextGene (SoftGenetics) analysis software were selected as the methods for a custom-designed tNGS assay to sequence six genes, *ATP8B1*, *ABCB11*, *ABCB4*, *NPC1*, *NPC2* and *SLC25A13*. This assay was to be transferred to clinical service at a future date therefore the considerations for assay design at the time were as follows;

- The combined size of the target regions was too small (23.7Kb) for commercially available TEMs, leaving PCR-based enrichment as the only option.
- Runs were required to be at the rate of one per month at least to ensure timely reporting of results.

- The target size and number of patients expected per month (12-20) meant that a small-scale sequencer was required. High capacity machines would not be cost effective to run.
- No bioinformatics expertise was readily available.

The main reasons for selecting the components of the assay were as follows;

- 48.48 Access Array was selected as the TEM as it offered automation, reproducibility and flexibility.
- The GS Junior was selected as the sequencer as it was compatible with the Access Array method, it produced long reads in order to make downstream data analysis simpler, and it had a short run time.
- NextGene was selected as the analysis software as it offered a simple user interface and automated sequence alignment and variant calling algorithms.

The design and optimisation of this assay will be described as well as an evaluation of the advantages and disadvantages of tNGS for use in clinical laboratories.

3.1.4. Chapter 3 Objectives

- To evaluate MS as an alternative sequencing method to FS for diagnosing genetic cholestasis conditions.
- To design and optimise a tNGS assay to sequence six genes involved in cholestasis
- To evaluate the tNGS assay as an alternative sequencing method to FS for diagnosing genetic cholestasis conditions.

3.2. Methods

3.2.1. Patients for evaluation of MS

DNA samples from twenty nine individuals with intrahepatic cholestasis, low-normal range GGT activity without ARC syndrome or defects in bile acid biosynthesis were used. ARC syndrome was excluded by clinical examination (Gissen et al., 2006) and bile acid biosynthesis disorders were excluded by determination of urinary bile acid profile by electrospray ionisation mass spectrometry. Thus there was a high likelihood of detecting mutations in *ATP8B1* and *ABCB11* in these samples.

3.2.2. Patients for evaluation of tNGS

Seventy four DNA samples from patients known to carry mutations and polymorphisms in the six genes of interest. Twenty three DNA samples from clinically unaffected controls which had previously been anonymised. Thirty nine DNA samples consented for research in which no mutations in the genes of interest had previously been found.

3.2.3. Fluorescent Sanger sequencing

PCR amplification of the coding exons and boundaries of the *ATP8B1, ABCB11, ABCB4, NPC1, NPC2 and SLC25A13* genes according to method described (2.2.3.1). PCR clean up, sequencing, sequencing clean up, capillary electrophoresis and data analysis was done according to methods described (2.2.4.1, 2.2.6.1 and 2.2.8.1).

3.2.4. Microarray sequencing

PCR amplification of the coding exons and boundaries of the *ATP8B1* and *ABCB11* genes according to the method described (2.2.3.2). Quantification and pooling of the PCR products was done according to the method described (2.2.5). Quantification and

pooling of the PCR products, the MS method and data analysis was done according to the methods described (2.2.6.2 and 2.2.8.3).

3.2.5. Targeted NGS

The methods for this process are discussed in the Results section.

3.3. Results

3.3.1. Evaluation of Microarray Sequencing (MS)

The coding regions of the *ATP8B1* and *ABCB11* genes were analysed using FS and MS in twenty nine patients with increased likelihood of carrying mutations. The sequence variants detected in thirteen of the samples (44.5% detection rate) are listed in Table 3.1. Eight variants had previously been described in association with PFIC and ten variants were novel. Four of the variants were nonsense mutations (22%), two were splice site changes (11%) and the rest were missense (67%). No insertions or deletions were detected using the specifically designed probes. No mutations were detected by either technique in the remainder of the patients and therefore the cause of 'low-normal GGT' cholestasis remains undetermined in these cases. The negative results in these cases may be explained by whole exon deletions or duplications, intronic mutations affecting splicing, or promoter region mutations in *ATP8B1* and *ABCB11*; as such mutations would not be detected by either strategy. In addition, it is possible that further genes are involved in the phenotype of neonatal cholestasis with low GGT.

In sample 17, MS detected a variant not confirmed by FS, constituting a false positive result. Two samples (6 and 22) had compound heterozygous changes in *ATP8B1* detected by FS but not by MS. Both samples had two variants in close proximity; therefore we speculate that these are in *cis* and that each sequence variant has impaired the hybridisation of surrounding probes. Our experience of testing patients for mutations in these genes suggests that false negative results arising in this manner are likely to be uncommon. However, in this cohort they occurred in 7% of samples.

Table 3.1: Mutations detection by MS and FS.

a - Variant was not confirmed by FS and is therefore a false positive finding. b - DNA changes were experimentally determined by sequencing. c - Protein changes are predicted rather than experimentally determined. d - Novel missense changes are of unclear pathogenicity.

Sample	Gene	DNA change [b]	Protein change [c]	Zygosity	Predicted effect	Previously reported?	Microarray Sequencing	Fluorescent Sanger sequencing
2	*ABCB11*	c.850G>C	p.V284L	Heterozygous	Missense	(Byrne et al., 2009)	Y	Y
4	*ABCB11*	c.2178+1G>T		Heterozygous	Altered splicing	(van Mil et al., 2004b)	Y	Y
6	*ATP8B1*	c.1010T>G	p.M337R	Heterozygous	Missense	Novel [d]	N	Y
	ATP8B1	c.1018A>G	p.M340V	Heterozygous	Missense	Novel [d]	N	Y
7	*ATP8B1*	c.208G>A	p.D70N	Heterozygous	Missense	(Klomp et al., 2004)	Y	Y
9	*ABCB11*	c.2170G>A	p.D724N	Heterozygous	Missense	Novel [d]	Y	Y
12	*ABCB11*	c.2611-2A>T		Heterozygous	Altered splicing	(Strautnieks et al., 2008)	Y	Y
17	*ATP8B1*	c.1660G>A	p.D554N	Homozygous	Missense	(Klomp et al., 2000)	Y	Y
	ATP8B1	c.1564G>A [a]	p.D522N [a]	Heterozygous [a]	Missense	Novel [d]	Y	N [a]
18	*ABCB11*	c.290T>G	p.L97*	Homozygous	Truncated protein	Novel	Y	Y
19	*ATP8B1*	c.3040C>T	p.R1014*	Heterozygous	Missense	(Klomp et al., 2004)	Y	Y
22	*ATP8B1*	c.1014C>G	p.N338K	Heterozygous	Missense	Novel [d]	N	Y
	ATP8B1	c.1018A>G	p.M340V	Heterozygous	Missense	Novel [d]	N	Y
	ABCB11	c.1636C>A	p.Q546K	Heterozygous	Missense	Novel [d]	Y	Y
25	*ABCB11*	c.499G>A	p.A167T	Heterozygous	Missense	(Liu et al., 2009)	Y	Y
	ABCB11	c.3458G>A	p.R1153H	Heterozygous	Missense	(Strautnieks et al., 2008)	Y	Y
26	*ABCB11*	c.3484G>T	p.E1162*	Homozygous	Truncated protein	Novel	Y	Y
29	*ABCB11*	c.483C>A	p.C161*	Homozygous	Truncated protein	Novel	Y	Y

3.3.2. tNGS Design

3.3.2.1. PCR design

According to the manufacturer's guidelines for assay design, the size range of the PCR products to be amplified in one assay was less than 150 base pairs, to avoid disparate levels of amplification between long and short products during the emPCR step of the GS Junior protocol. In addition, the maximum read length of the sequencer was around 500 bases and PCR products of this size or less would allow bi-directional sequencing of the fragments.

PCR primers covering the ROIs were already available in the laboratory for FS and included M13 tag sequences allowing universal sequencing primers to be used. The primer sequences and their product sizes are listed in chapter 2 (2.1.12.1, Table 2.1). The FS primers including M13 tags, amplified products ranging from 279 to 894 bases with an average of 574 bases, therefore some redesign of primers was required to achieve and optimal design. The size range of 364 to 514 bases was selected for the tNGS design to minimise the number of amplicons requiring redesign. 46 primers sets (25% of total) were redesigned to this size range using the Primer3 and Primer-BLAST tools and avoiding polymorphisms of more than 1% frequency using the SNPCheck tool (see 2.1.14).

3.3.2.2. Modified TEM - 3AA-2 step amplification

The capacity of one 48.48 Access Array (AA) was 48 PCR reactions and 47 samples (plus a negative control). Since 134 PCRs were required to cover the ROIs in the six genes of interest, three AAs were required to amplify all the fragments. These are shown in 96-well plate layout in Figure 3.1. The standard protocol for the AA method involved adding CS-tagged sequencing primers in with the DNA samples, each

containing a unique MID sequence for each DNA (2.1.12.2, Table 2.2). This allows PCR amplification of the ROIs and incorporation of the MID and sequencing primers into the PCR products in one-step. Since the M13 tags already incorporated in the primers were untested in this assay, a more flexible and less competitive two-step protocol was used, which also allowed for the potential of multiplexing PCR reactions at a later date. A modified protocol was devised that involved three separate AA set-ups, pooling of patient-specific products, followed by a second round of PCR incorporating the M13-tagged MID primers (see Figure 4.4).

3.3.2.3. PCR Clean-up and Quantification

The clean-up of the fragment library was done using AMPure beads as described (2.2.4.2). The quantification and dilution of the fragment library was done using the PicoGreen protocol previously described in Chapter 2 (2.2.5).

3.3.2.4. Sequencing method

The emPCR set-up, emPCR breaking, enrichment, clean-up, quantification, PTP loading and sequencing protocols are described in Chapter 2 (2.2.6.3).

Figure 3.1: The PCR primer layouts for the 3AA-2 step protocol.

These primers were used as the TEM for the tNGS assay. AA1 and AA2 contained the FS primers and AA3 contained the newly designed primers.

AA1 primers

	1	2	3	4	5	6
A	NPC1 Ex01	NPC1 Ex13	NPC1 Ex24	ATP8B1 Ex03	ATP8B1 Ex18	ABCB11 Ex03
B	NPC1 Ex02	NPC1 Ex14	NPC1 Ex25	ATP8B1 Ex06_07	ATP8B1 Ex20	ABCB11 Ex05
C	NPC1 Ex03	NPC1 Ex17	NPC2 Ex01	ATP8B1 Ex10	ATP8B1 Ex22	ABCB11 Ex07
D	NPC1 Ex04	NPC1 Ex19	NPC2 Ex02	ATP8B1 Ex11	ATP8B1 Ex24	ABCB11 Ex08
E	NPC1 Ex05	NPC1 Ex20	NPC2 Ex03	ATP8B1 Ex12	ATP8B1 Ex25	B l a n k
F	NPC1 Ex07	NPC1 Ex21	NPC2 Ex05	ATP8B1 Ex14	ATP8B1 Ex26	B l a n k
G	NPC1 Ex10	NPC1 Ex22	ATP8B1 Ex01	ATP8B1 Ex15	ATP8B1 Ex27	B l a n k
H	NPC1 Ex11	NPC1 Ex23	ATP8B1 Ex02	ATP8B1 Ex17	ABCB11 Ex02	B l a n k

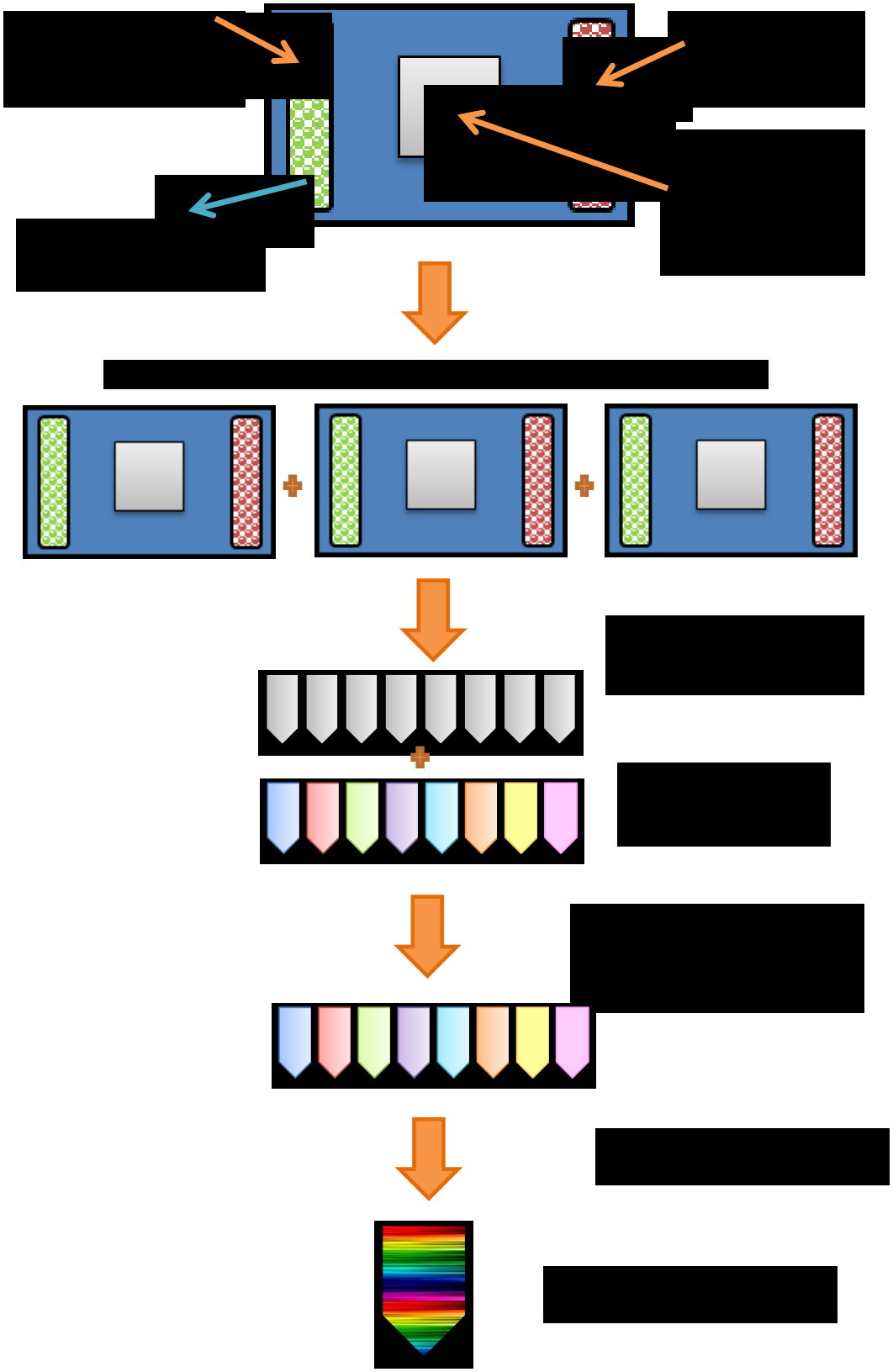
AA2 primers

	1	2	3	4	5	6
A	ABCB11 Ex09	ABCB11 Ex22	ABCB4 Ex05	ABCB4 Ex17	ABCB4 Ex27	SLC25A13 Ex13
B	ABCB11 Ex10	ABCB11 Ex23	ABCB4 Ex06	ABCB4 Ex18	SLC25A13 Ex01	SLC25A13 Ex14
C	ABCB11 Ex11	ABCB11 Ex25	ABCB4 Ex07	ABCB4 Ex19	SLC25A13 Ex03	SLC25A13 Ex15
D	ABCB11 Ex12	ABCB11 Ex26	ABCB4 Ex08	ABCB4 Ex22	SLC25A13 Ex05	SLC25A13 Ex18
E	ABCB11 Ex13	ABCB11 Ex27	ABCB4 Ex09	ABCB4 Ex23	SLC25A13 Ex09	B l a n k
F	ABCB11 Ex15	ABCB4 Ex02	ABCB4 Ex10	ABCB4 Ex24	SLC25A13 Ex10	B l a n k
G	ABCB11 Ex16	ABCB4 Ex03	ABCB4 Ex14	ABCB4 Ex25	SLC25A13 Ex11	B l a n k
H	ABCB11 Ex19	ABCB4 Ex04	ABCB4 Ex15	ABCB4 Ex26	SLC25A13 Ex12	B l a n k

AA3 primers

	1	2	3	4	5	6
A	NPC1 Exon06	NPC2 Exon04	ATP8B1 Exon21A	ABCB11 Exon18	ABCB4 Exon16A	SLC25A13 Exon04
B	NPC1 Exon08A	ATP8B1 Exon04	ATP8B1 Exon21B	ABCB11 Exon20	ABCB4 Exon16B	SLC25A13 Exon06
C	NPC1 Exon08B	ATP8B1 Exon05	ATP8B1 Exon23	ABCB11 Exon21	ABCB4 Exon20	SLC25A13 Exon07
D	NPC1 Exon09	ATP8B1 Exon08	ABCB11 Exon04	ABCB11 Exon24	ABCB4 Exon21A	SLC25A13 Exon08
E	NPC1 Exon12	ATP8B1 Exon09	ABCB11 Exon06	ABCB11 Exon28	ABCB4 Exon21B	SLC25A13 Exon16
F	NPC1 Exon15	ATP8B1 Exon13	ABCB11 Exon14A	ABCB4 Exon11	ABCB4 Exon28A	SLC25A13 Exon17
G	NPC1 Exon16	ATP8B1 Exon16	ABCB11 Exon14B	ABCB4 Exon12	ABCB4 Exon28B	B l a n k
H	NPC1 Exon18	ATP8B1 Exon19	ABCB11 Exon17	ABCB4 Exon13	SLC25A13 Exon02	B l a n k

Figure 3.2: Schematic showing the modified target enrichment protocol used for the tNGS assay (3AA-2-step protocol).



3.3.3. tNGS Optimisation

3.3.3.1. PCR validation

All PCR primer sets including the 46 redesigned sets were tested using the Roche Hi-Fidelity PCR kit and cycling conditions reflecting the 48.48 Access Array protocol (described in 2.2.3.4). The volumes were proportionally increased to a final volume of 8µl and set up in 96-well plates, as shown in Table 3.2. Please note the amplicon-specific primers were added upfront, instead of by the AA machine at a later point, and therefore the final volume of the PCR is increased from 4 µl to 8 µl.

Table 3.2: The recipe used for the PCR validation of the amplicon-specific primers using the Roche High-Fidelity kit. Those components with an asterisk were provided with the kit. The AA loading reagent was provided with the AA kit.

PCR Component	Validation PCR volume (µl)
10x Buffer (without $MgCl_2$)*	0.5
25mM $MgCl_2$*	0.9
DMSO*	0.25
10mM nucleotides*	0.1
Taq (5U/µl) *	0.05
20x AA loading reagent	0.25
Genomic DNA (50ng/µl)	0.8
Water	1.15
1µM M13 tagged amplicons-specific primers (F+R)	4
Total	8

PCR products were visualised using gel electrophoresis (see Chapter 2.1.6) and examples of gel pictures are shown in Figure 3.3. 128 of the 134 possible PCR products amplified well and only 6 showed potential problems for sequencing. For each gene a summary of the PCR results was as follows;

NPC1 - Exons 4 and 17 have weak products

ABCB11 – Exons 22 and 28 have weak products, Exon 6 has an extra band

ABCB4 - Ex28A has a weak product

NPC2, ATP8B1, SLC25A13 - All OK

Figure 3.3 Gel picture from PCR validation

A typical gel picture from PCR validation step showing a selection of *NPC1* exons. The vast majority of primer pairs generated single strong bands; however Ex17 shows a clearly weaker product in comparison.

3.3.3.2. Assay validation

It was not practical to validate each component of the assay individually, therefore the TEM, sequencing and data analysis methods were validated together by looking at the end point of the assay; the sequencing data. 4 validation runs were performed and optimisation continued on subsequent live runs whilst testing new research patients. Here is an outline of the runs;

1. AA1 primer sets and 46 DNA samples (9 anonymised 'normal' DNAs, 19 research samples with no known mutations and 18 with known mutations/polymorphisms from FS data). Done in duplicate.
2. AA2 primer sets and 46 DNA samples (4 normals, 10 research samples with no known mutations and 30 with known mutations/polymorphisms from FS data; 2 of these were run in duplicate).
3. AA3 primer sets and 46 DNA samples (10 normals, 10 research samples with no known mutations and 26 with known mutations/polymorphisms from FS data).

4. The products from 12 MIDs from Runs 1-3 were combined to produce pseudo-samples and sequenced.

5-7. 46 samples containing a mixture of samples run on Runs 1-3 and new research patients were enriched using the 3AA-2step protocol and the products were split into 3 GS Junior runs.

8. 'Live' run of new research patients.

The objectives of the optimisation were;

- to check performance of GS Junior sequencer
- to check the performance of the MIDs
- to test reproducibility of the test between and within runs
- to test PCR primers amplification within AA system and downstream sequencing
- to optimise the depth of coverage across ROIs
- to establish optimum number of patients on a run
- to check assay can detect all known sequence variants including polymorphisms.

3.3.3.2.1. GS Junior performance

Roche 454 provides various parameters against which data from the GS Junior may be measured. The most useful and relevant of those are shown in Table 3.3.

Table 3.3: GS Junior run metrics

Run	Total raw wells ≤ 250,000	Total key pass/Total raw well >90%	Total pass filtering/Total key pass 25-50%	Mix&Dots/ Total key pass <20%	Trimmed too short quality reads 30-55%	Number of passed filter (PF) reads 50-80,000
1A	218745	97.00%	44.00%	11.80%	44.30%	93305
1B	226761	96.30%	53.57%	8.26%	38.16%	116971
2	236377	97.88%	45.17%	17.37%	37.45%	104506
3	227232	97.41%	48.79%	6.84%	44.36%	107990
4	215435	97.38%	58.29%	7.18%	34.52%	122297
5	220307	97.40%	40.15%	14.56%	45.28%	86119
6	205250	96.88%	59.32%	5.01%	35.66%	117968
7	219522	97.37%	49.63%	5.25%	45.11%	106091
8	203877	96.50%	60.67%	4.74%	34.47%	119335

Run metrics from assay validation showing the numbers advised by Roche in the header and the numbers achieved on each validation run. The most important column is the number of passed filter reads (PF) as this is the number of reads useable for analysis. Note that the PF numbers achieved are significantly higher than the number Roche considered as an adequate run. This is because Roche are conservative in order to reduce customer expectations of machine capacity. In our hands with well-optimised assays the machine can routinely produce around 120,000 passed filter reads (60MB compared to the 40MB advertised for GS Junior with standard chemistry).

3.3.3.2.2. MID validation

It was important to exclude the MID primer pairs as variable factors in the assay and therefore the total number of reads per MID was compared against Runs 1-3 in which 3 different sets of sample DNAs were amplified. This found that whilst there were some poor samples on each run, no MID set performed badly in all runs and therefore the problems were caused by the DNA or the preparation rather than the MID primers.

In runs 5-8 the total number of reads per MID was compared across each run to look for over or under-representation of MIDs. The results for Run 5 are shown in Figure 3.4.

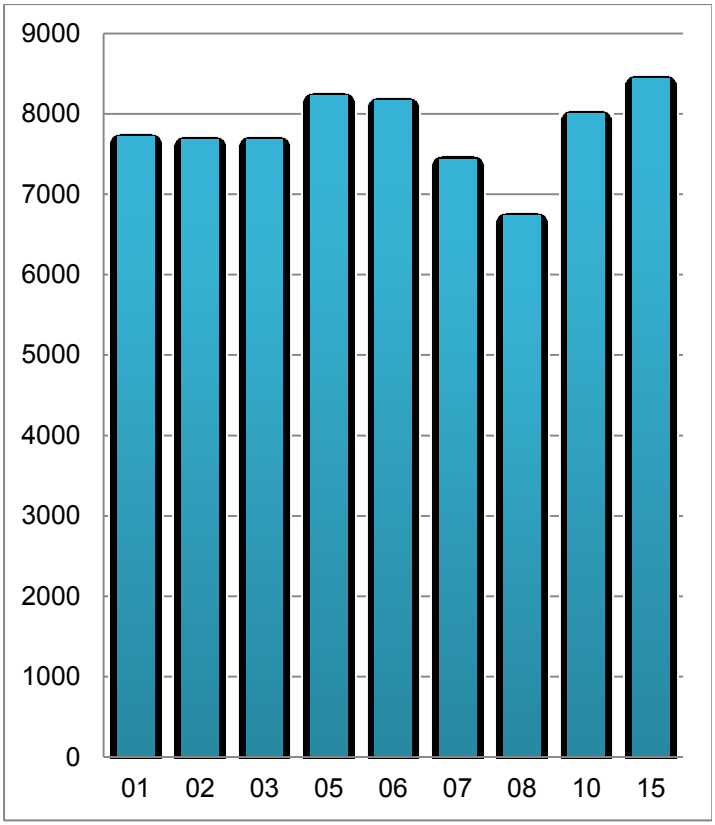


Figure 3.4: Total matched reads per MID

A graph showing the total number of matched reads for Run 5 by MID. The reads are evenly split between MIDs and a similar pattern was seen for Runs 6-8 where the 3AA-2step method was used. This shows that the AA TEM is effective in producing even PCR amplification between samples. In addition, it shows that the subsequent pooling of equal volumes of sample-specific second round PCR products is sufficient to prevent uneven distribution of reads between samples/MIDs.

3.3.3.2.3. Reproducibility

Run 1 was performed in duplicate by running the same fragment library on two different machines (1A and 1B). Run 1B was a better performing run than 1A and as a result average numbers of reads per amplicon were slightly higher for 1B. Overall the numbers of reads per amplicon and per sample correlated well and demonstrated that the sequencer will reproducibly sequence the material in the fragment library (see Figure 3.5).

In order to evaluate the reproducibility between runs, two samples were run in duplicate on Run 2 using two different MIDs. The comparison was measured using total number of reads and average reads per amplicon (see Table 3.4). Both measurements showed considerable variation and the effect of MIDs and TEM preparation was considered as possible reasons for this. However as discussed in

section MID validation, both factors were shown to have minimal effects when compared across the validation runs.

Table 3.4: Inter-run comparison for two samples

Inter-run comparison between samples 1 and 2 using two different MIDs. There was considerable variation between the samples for both measurements.

Sample	Total number of reads	Average reads per amplicon
1 MID 003	2950	61
1 MID 007	1465	31
2 MID 033	4151	86
2 MID 036	8465	176

3.3.3.2.4. Coverage of ROIs

In order to evaluate the ability of the AA TEM to enrich for the ROIs in the target genes, the total number of reads per amplicon in Runs 1-3 was measured, see Figure 3.6. There were 9 amplicons identified as fails (less than 10 reads on average per sample). Of these *ABCB11* exons 6, 22 and 28 were all identified as problem amplicons at the PCR validation stage. In addition, three amplicons had high GC content (*NPC1* exon 1 at 71%, *NPC2* exon 1 at 67%and *SLC25A13* exon 1 at 75%) and one had high AT content (*ABCB11* exon 3 at 70%), both of which are known to reduce efficiency of PCR reactions performed using standard conditions such as the emPCR step of the tNGS assay (Metzker, 2010). These amplicons were removed from the assay. Lastly, *ATP8B1* exon 18 and *NPC1* exon 12 produced very few reads however no obvious reason for this was identified. These amplicons were included in Runs 5-7 to rule out loading errors or random dropout, however the sequencing results did not improve therefore they were also removed from the assay at a later date.

3.3.3.2.5. Assay balancing

A minimum of 30 reads covering each ROI was the target for this assay in order to ensure a high probability of sequencing both alleles in a heterozygote (De Leeneer et al., 2011). As highlighted in Figure 3.7 several amplicons in Runs 1 to 3 were not reaching the minimum of 30 reads on average (these runs were based on 46 samples and 44 to 46 amplicons). In order to increase the number of reads per sample, the number of samples per run for runs 5 to 7 was reduced to ten samples, from the equivalent of sixteen samples per run in Runs 1 to 3. In addition, underperforming amplicons (10 to 30 reads on average) were added to the AA plate layout twice, in order to boost the amount of PCR product and increase the proportion of reads. In order to free up space on the 3 AAs, over-performing amplicons (more than 70 reads on average) were di-plexed together. Two rounds of assay balancing were performed; round one before Runs 5 to 7, and round two before Run 8, see Figure 3.7. Lastly, over the course of the optimisation of the assay the protocol for the GS Junior was optimised and this lead to an increase in the number of reads produced. Overall these measures increased the average number of reads per amplicon from 42 in Runs 1 to 3, to 114 in Run 8.

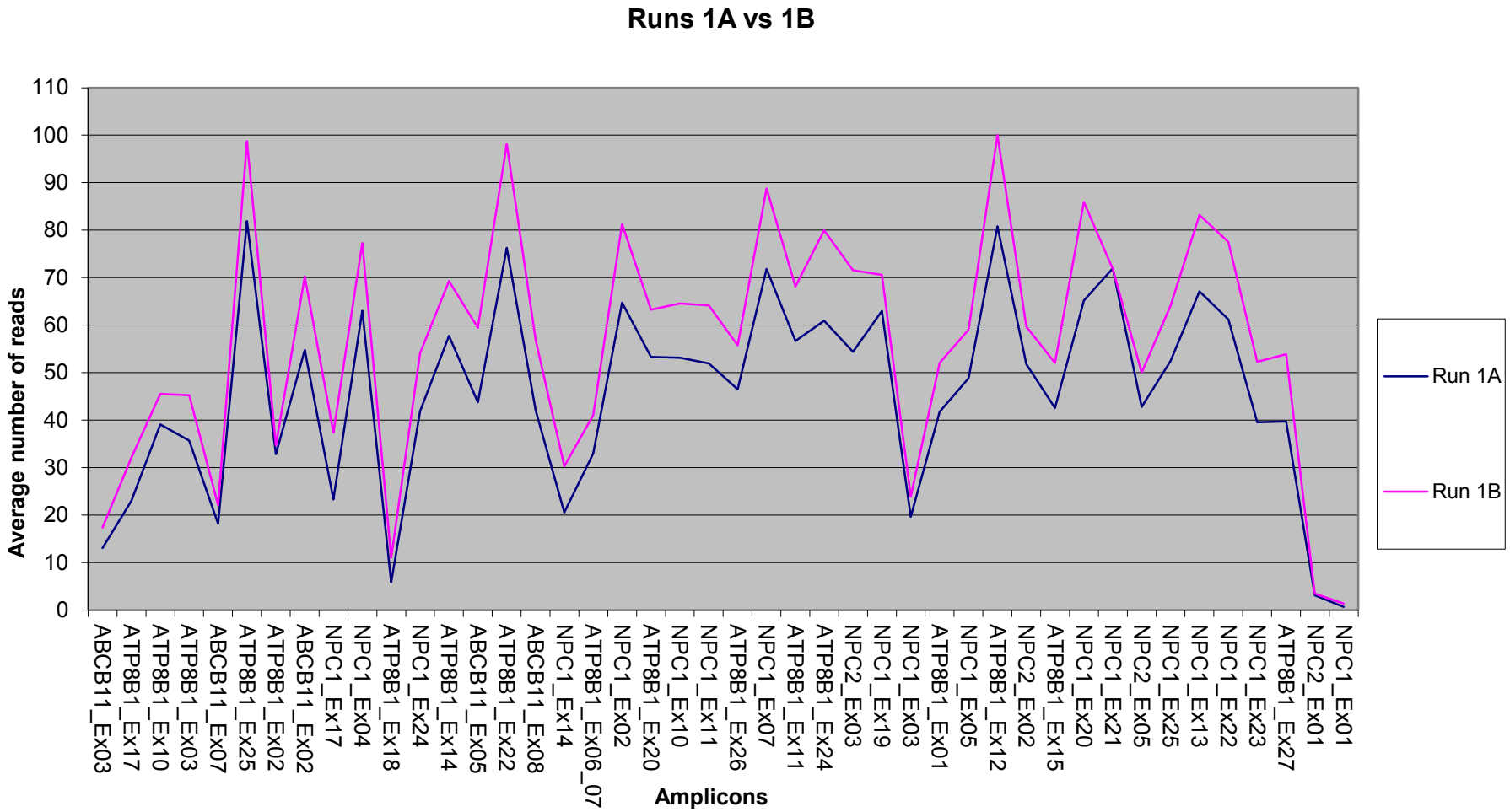


Figure 3.5 Comparison of Run 1A and Run 1B.

These runs were the same fragment library run on two different GS Junior machines. This graph compares the average number of reads for each amplicon and shows good correlation between the runs.

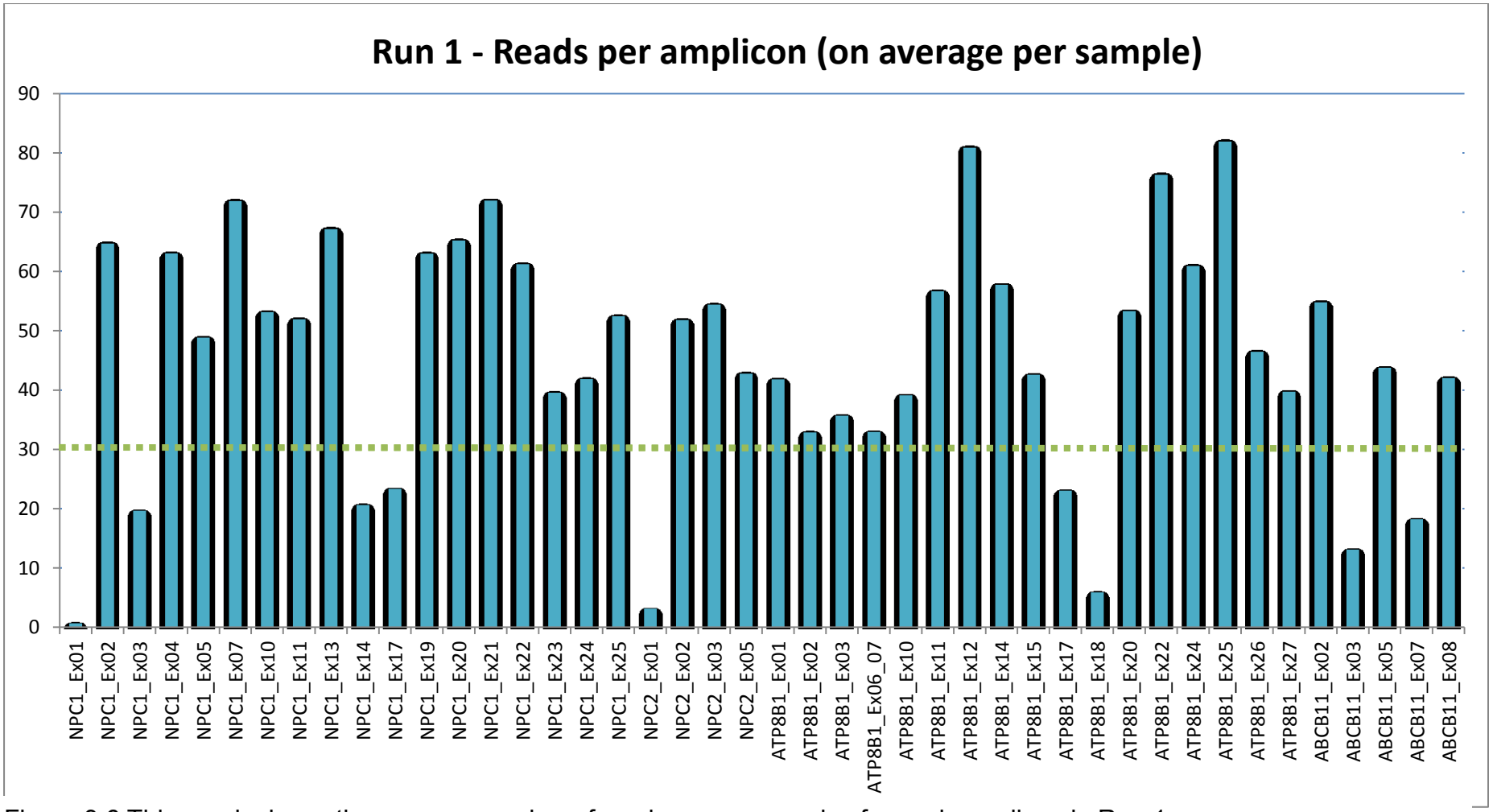


Figure 3.6 This graph shows the average number of reads across samples for each amplicon in Run 1.

It shows significant variation between amplicons, ie NPC1_Ex01 and NPC1_Ex02 for example. In addition, it shows several amplicons that are below 30 reads on average, and this was replicated across Runs 2 and 3.

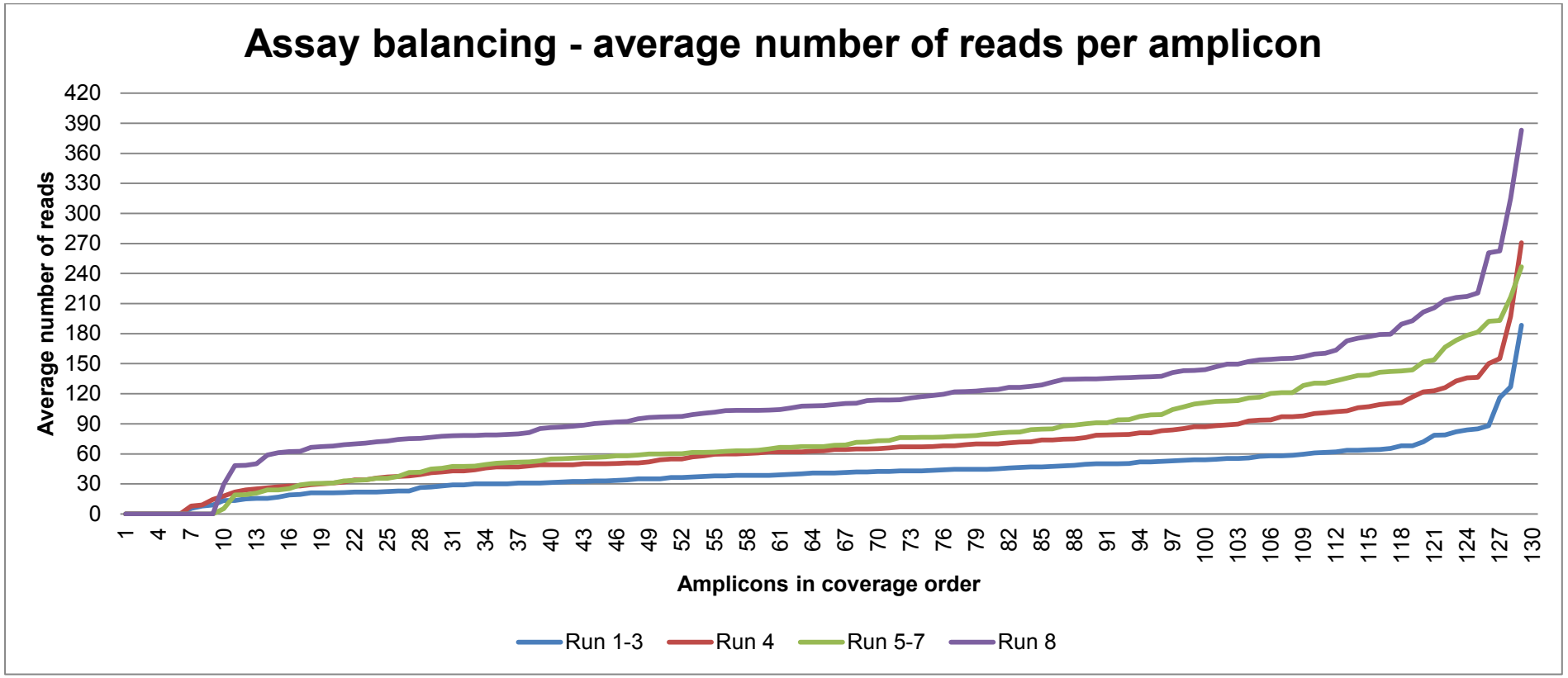


Figure 3.7 This graph shows the average number of reads for amplicons.

The results for Runs 1-3 were adjusted for total numbers of reads on the run. Runs 5-7 were averaged. The improvements between Run 1-3 and Run 4 were achieved by decreasing the number of patients on a run from 16 to 10. The improvement to Runs 5-7 were achieved by amplicon balancing as described in the text. The improvements seen in Run 8 were as a result of further amplicon balancing and increased reads from the GS Junior.

3.3.3.2.6. Mutation detection

Runs 1 to 3 were used to test the mutation detection capabilities of the tNGS assay. This was done by including samples with known mutations in the six genes. Since mutations were not available for all the amplicons in the assay, polymorphism data from samples previously tested using FS was also used to establish whether sequence changes were reliably detected. The majority of sequence variants not detected were in Run 3 because they were present in regions detected by the FS primers but were outside of the newly designed tNGS primers. The next most common reason was that the amplicon containing the variant failed to amplify (less than 10 reads) and some variants were not detected because of low coverage on Runs 1-3 (less than 30 reads). Excluding exon failure, sample failure and variants being outside of new primer sets, a total of 326 variants were tested (103 being unique). 321 of them were detected by the tNGS assay and 5 were not correctly called. Of the 5, 3 were the same polymorphism in *ATP8B1*, therefore there were 3 unique variants incorrectly called. A breakdown of the detection of previously known sequence variants across the 3 runs is shown in Table 3.5.

Table 3.5: Summary of the sequence variants detected.

A summary of the sequence variants during the evaluation of the tNGS assay. Three unique variants were not correctly identified by the assay.

Run	Number of unique variants	Total number of variants	Number correctly identified	Number not correctly identified	Comments
1	36	97	96	1	**Variant 1** was within a homopolymer tract.
2	25	89	89	0	
3	42	140	136	4	**Variant 2** was seen 3 times and was within a homopolymer tract. **Variant 3** was in an exon with two overlapping amplicons.
Total	103	326	321	5	

Variant 1 was a polymorphism present at one end of a homopolymer stretch of DNA; *ATP8B1* c.2931+59T>A. Sequencing of homopolymer stretches is a known problem for 454 sequencing (Loman 2012, see sections 3.3.3.3.6 and 3.3.3.3). The sample in Run 1 was previously shown to be homozygous for this polymorphism, however the allele was only detected in 83% of reads, and the remainder of the reads were showing a deletion of a T. This is because a deletion artefact was introduced when the homopolymer stretch increased from 5 adenines to 6 adenines (see Figure 3.8).

Variant 2 resulted from a similar scenario. It was a polymorphism present in a homopolymer stretch of DNA; *ATP8B1* c.1820-54T>A. c.1820-54T>A creates a poly-adenine stretch of 8 bases. 6 samples tested were known to have this polymorphism, 2 heterozyotes and 4 homozygotes. 3 homozygotes had low coverage, therefore the remaining 1 homozygote and 2 heterozygotes were incorrectly called. Variation was identified by the pipeline at this position for all 3 samples, although this was a deletion of a T rather than a T to A substitution.

Variant 3 was a homozygous missense mutation c.1766C>T p.Ser589Phe in *SLC25A13* which had been called as heterozygous by the software in one patient. This was present in an exon containing two overlapping PCR fragments and was located within a primer sequence. Therefore the primer sequence was contributing wild-type bases to the number of bases at that position and resulting in a heterozygous rather than a homozygous call. The risk was another such mutation could be missed in the future, see 3.3.3.3.7.

Although there were no true misses, as Variants 1 to 3 were called incorrectly rather than missed, there were three possible risks to missing a sequence variant identified by this analysis, and these were;

1. variant is not covered adequately by assay,
2. variant is in region overlapped by primer sequence,
3. and, variant is located within a homopolymer stretch.

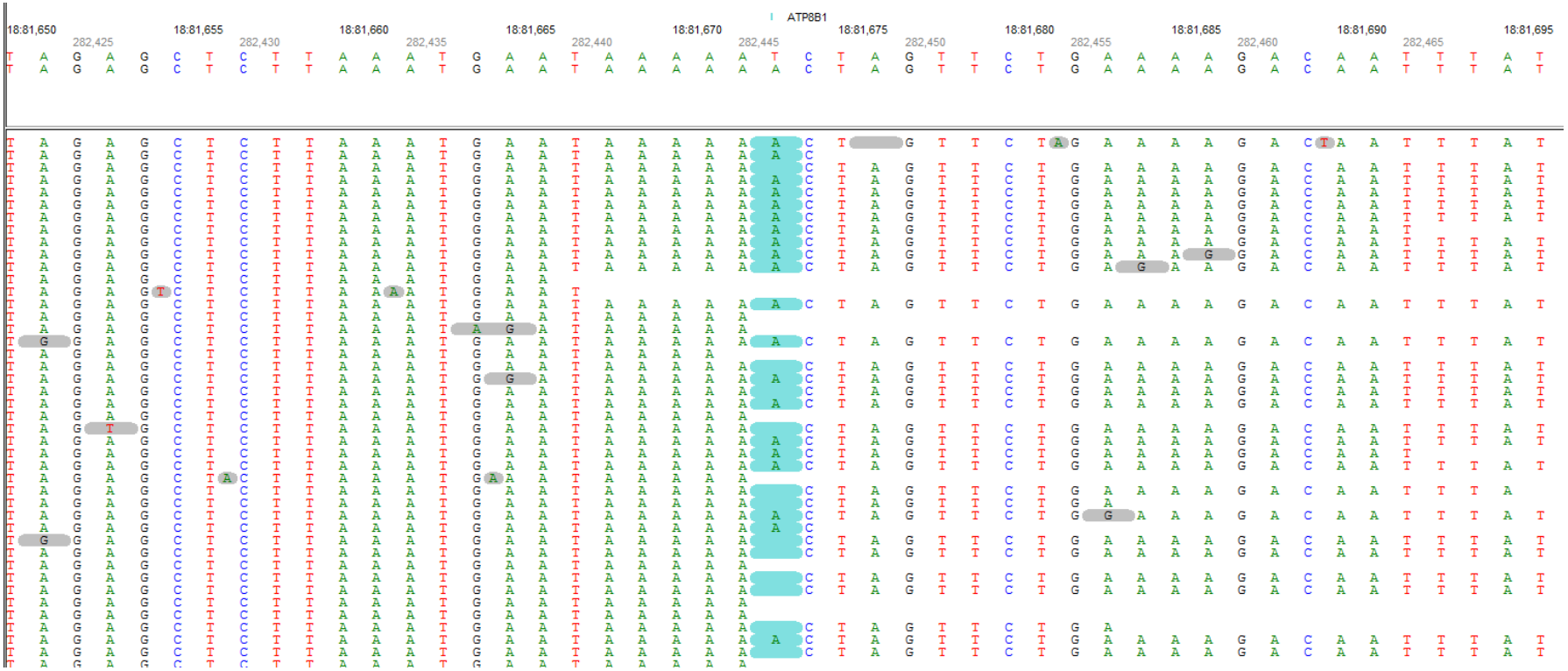


Figure 3.8 This is a screen shot of NextGene v1.6 Viewer, showing Variant 1.

The top panel shows the reference sequence (top line) and the consensus sequence for the sample (second line) the homozygous T>A call is seen at the end of stretch of 5 adenines. The bottom panel shows the individual reads for the sample and the base with difference from the reference sequence is highlighted in light blue. The vast majority of the bases are adenines, however a small proportion are showing a 1 base deletion.

In addition, this mutation detection analysis was used to determine the optimum mutation filter threshold for the assay (see Table 3.6). Using a mutation filter threshold of 15%, and when combined with a minimum depth of coverage of 30, the confidence for detection of heterozygous calls has been estimated as 99.995% (De Leeneer et al., 2011). This assumes both the wild type and mutant sequence are amplified equally, however when this is not the case variants may be missed by these thresholds (Rattenberry et al., 2013). For this tNGS assay, all known variants were detected and therefore the sensitivity was 100%. It was noted that many of the same limitations of FS apply to this assay, for example the possibility of allele drop out due to variants in the primer binding sites, inability to detect large-scale deletions, duplications and rearrangements, and the absence of intronic and promoter regions.

Table 3.6: The proportion of reads containing the variant base for all known sequence variants identified in Run 1.

The heterozygous calls ranged from 27.78-50% and shows that all calls would be detected by a mutation filter threshold of less than 25%. The data for the homozygous calls identifies an outlier at 82.98% as the average and median were 98.89% and 100% respectively. This reflects the identification of Variant 1.

Run 1	Heterozygous variants	Homozygous variants
Total number	67	30
Average	47.51%	98.89%
Median	47.83%	100.00%
Minimum	27.78%	82.98%
Maximum	50.00%	100.00%

3.3.3.2.7. False positives

16 potential variants were identified in Runs 1 to 3 in regions of the genes in which FS had not previously been done. Sequencing artefacts had been excluded as being present in more than one sample. The 16 variants were sequenced using FS and the results compared with the tNGS. 9 were not present in the FS results and were false positives (and of these only 6 had tNGS coverage of 30 reads or more) therefore giving a false positive rate of 0.2% (6 false positive results in a total of 5857 ROIs analysed). 7 variants were confirmed as real and were either low frequency polymorphisms or VUS.

3.3.3.3. Data analysis validation

NextGene software was available in the laboratory therefore it was selected as the primary data analysis method for the tNGS assay. Here the overall strategy is described as well as the considerations made to risks identified during the tNGS Optimisation stage.

3.3.3.3.1. File conversion

The .sff file produced by the GS Junior was converted to an .fna file using the NextGene File Conversion tool. These files are the equivalent of .fastq and .fasta, respectively, and the conversion leads to a small number of poor quality reads being discarded. Various parameters of the conversion settings were compared and only small differences to the percentage of bases converted were observed. The settings used for file conversion were median score threshold greater than 20, maximum number of uncalled bases 3 or fewer, called base number of each read of 25 or fewer, trim or reject read when 3 or

more bases had a score of 16 or less. Reads and or bases not meeting these criteria were discarded.

3.3.3.3.2. Contamination checks

In order to check whether contamination from previous runs, or contamination of the negative control, were present, the NextGene Barcode Sort tool was used to sort the .fna file using all possible MID sequences (see 2.1.12.2, Table 2.2). Any MID which had more than a few reads present was investigated to find the cause of contamination. A small number of reads was often present in the negative control, however when these were mapped against the tNGS reference sequences, very few of them matched. This meant that these reads would not interfere with the interpretation of the results and could be ignored.

3.3.3.3.3. Barcode Sort, Alignment and Variant Calling

The .fna file was used to create individual sample projects in which the reads for each MID were mapped against GenBank files for each gene. This was done automatically using the NextGene Autorun tool. The Alignment settings used were a minimum of 100 bases or 85% of bases in a read mapped, ambiguously mapped reads were removed, rigorous alignment and detection of large indels was selected and unmatched ends were hidden. For Variant Calling the mutation filter settings were that variants with less than 15% of reads, less than 3 mutant reads and less than 15 total reads were not called. This threshold was chosen based on calculations from De Leeneer et. al. (De Leeneer et al., 2011).

3.3.3.3.4. Variant Comparison

Variants were called by NextGene using the Variant Comparison tool. To do this modified GenBank files for the six genes were created. Firstly exons with known mutations outside of 5 bases from the intron exon boundaries were noted. Then GenBank files for each gene were annotated with Regions of Interest (ROIs). In general, this was set to + or – 5 bases from the intron exon boundaries, with the exception of those exons in which known mutations existed outside of this. The ROI for those exons were extended accordingly. Then the Variant Comparison tool was used to create a .vcf file listing all the variants called within those regions of interest for all the samples on the run. This allowed information on all samples to be interpreted at the same time, even those in whom a particular variant did not reach the mutation filter cut off (15% of reads). This strategy was useful for the interpretation of sequencing artefacts generated by the GS Junior (see 3.3.3.3.6).

3.3.3.3.5. Variant comparison checking

The Variant Comparison .vcf file for the run was manipulated using Microsoft Excel to annotate and compare the variants. Firstly, a polymorphism list was generated using prior knowledge of the genes and this was used to annotate any variants that were on that list as polymorphisms. Secondly, any variants that were present in only one sample, and not on the polymorphism list, were automatically highlighted for confirmation by FS. Thirdly, other variants were very likely to be sequencing artefacts. The percentage of mutant alleles, and the direction bias were carefully checked to identify outliers from the run as a

whole. Those outliers were highlighted for confirmation by FS. The .vcf file was checked by two independent scorers and the most cautious interpretation of each variant was used.

3.3.3.3.6. False positives

As mentioned in section 3.3.3.2.6, sequencing of homopolymer stretches of DNA are a known problem for 454 sequencing (Loman 2012) and the reads are prone to insertion and deletion errors. This manifested as dozens of false positive variant calls in the validation runs using the mutation filter of 15%. Most of these were insertions or deletions of one base in homopolymer stretches of more than 4 or 5 bases; however artefact base substitutions and errors in other regions of DNA were also routinely seen. One reason for this was that quality scores for variant bases were not taken into account by this analysis pipeline and instead all calls were interpreted manually using the variant comparison method described above. This was in an attempt to minimise the possibility of missing a sequence variant in the same region as an artefact. As described above the mutant allele percentage and direction bias for all samples on the run was evaluated and any outliers were treated like a potential real variant and investigated by FS. Outliers were identified by being outside of two standard deviations from the mean mutation percentage. During the assay validation, potential variants that were identified by these criteria gave normal results using FS (n=76).

3.3.3.3.7. Overlapping fragments

As identified during the assay validation (3.3.3.2.6), sequence variants may be masked if they are present in the same region of DNA as a primer from an overlapping fragment. Ideally primer sequences should be trimmed off, however this was not possible within the NextGene software therefore a manual check of such regions was instigated. There were a total of 8 fragments whose primers overlapped ROIs and these were; *ABCB11* Exon 14A reverse and 14B forward, Exon 20 reverse, Exon 21 forward, *ABCB4* Exon 21A reverse, Exon 21B reverse, Exon 28A reverse, Exon 28B reverse, *ATP8B1* Exon 5 forward, Exon 8 reverse, Exon 21A reverse and Ex21B forward, *NPC1* Exon 8A reverse and Exon 08B forward, Exon 16 reverse and *SLC25A13* Exon 16 reverse and Exon 17 forward.

3.3.3.3.8. Depth of coverage

For the project described in Chapter 4, it was not practical for financial reasons to FS all of the fragments that did not meet the minimum of 30 reads. However the depth of coverage was recorded for all of the ROIs and if one heterozygous sequence variant was identified in any gene, the gaps for that gene were filled by FS. In general, coverage was good and apart from those ROIs removed from the assay during the validation very few failed to meet the minimum threshold. A few patients amplified poorly overall and these were repeated using the tNGS assay.

3.4. Discussion

3.4.1. Microarray sequencing

MS is an attractive alternative to traditional FS for genetic testing in neonatal cholestasis (Liu et al., 2009). The BRUM1 microarray used in this study has a much larger capacity than was utilised in this experiment, as it contains probes for 92 genes associated with inherited disorders (Bruce et al., 2010). Analysis of the ability of this microarray to detect known gene mutations has shown a 97% mutation detection rate for base substitutions (Bruce et al., 2010). The major advantage of MS over FS is increased capacity, allowing sequencing of multiple genes in one experiment as quickly as one gene using FS. The main bottleneck is the requirement to quantify and to pool individual PCR products before hybridisation. This step is time-consuming and prone to error (due to pipetting volume variations and potential sample mix-ups). Therefore, automation of processes to minimise such errors would be useful, if not essential, for adoption of this method into clinical laboratories. Alternatively, use of long range PCR as the preparatory step for the MS protocol could reduce the number of reactions to be pooled, although this is more sensitive to fragmented DNA than standard PCR (Cheng et al., 1995). Finally, microfluidics-based PCR systems, such as the 48.48 Access Array (Fluidigm Corporation, San Francisco, CA), might be combined with MS to avoid the quantification and pooling step altogether. This system allows the simultaneous but separate amplification of up to 48 PCR products for up to 48 samples, in nanolitre-sized reactions, using a semi-automated process. The

end product of the process is a pool of PCR products for each sample, and it is commonly used for target enrichment for benchtop next-generation sequencers.

Of the mutations recorded in the Human Gene Mutation Database in *ATP8B1* and *ABCB11*, 24% and 16% respectively were small insertions or deletions (indels), although their combined frequency in PFIC cases is unknown as most are private mutations. The major disadvantage of MS is it's insensitivity for the detection of indels (Bruce et al., 2010). Whilst known indels can be detected with proper microarray design, novel indels will be missed. Larger insertions and deletions involving whole exons, of the type routinely detected by MLPA (Schouten et al., 2002), are not detected by either MS or FS. Another disadvantage, underscored by the results of this study, is that mis-called base substitutions (both false negative and false positive calls), are frequent and were found in approximately 10% of samples in this study.

In summary, MS allows rapid and cost-effective genetic screening in neonatal cholestasis and yields results relevant for patient management. Many clinical laboratories are experienced in aCGH and thus have access to equipment required for MS, suggesting that MS could be implemented for a relatively small monetary investment. In principle, MS could be applied to many clinical scenarios involving heterogeneous conditions, especially if the mutations tend to be base substitutions.

3.4.2. Targeted next generation sequencing

The custom-designed tNGS strategy used here, incorporating 48.48 Access Arrays, GS Junior sequencing and NextGene analysis software, is a useful and attractive alternative to FS for sequencing genes involved in cholestasis. Our evaluation found a sensitivity of 100% and a specificity of over 99%. The main advantage of tNGS over FS is the capacity and if this is used optimally, the volume of sequencing data generated is greatly increased for the time put in. A few disadvantages were identified, primarily the inaccurate sequencing of homopolymer tracts and the problems of overlapping primer sequences. Both of these potentially could be reduced or eliminated using bioinformatics expertise as described in the results section. The main advantage over MS is that the AA TEM has avoided the PCR bottleneck, and in addition, many samples may be sequenced at the same time vastly reducing both the total preparation time and the hands-on time required compared to FS. As previously mentioned, some of the limitations of FS also apply to the tNGS assay. The tNGS assay described here was chosen as the best method for the sequencing of samples for the research project described in Chapter 4.

3.4.3. Alternative strategies

Since the development of this tNGS assay, Illumina sequencing has become established as the NGS method of choice for the vast majority of research and clinical applications for testing of human DNA. A Pubmed seach for publications using ‘Illumina sequencing’ as the search term, returned 83 results

in 2009 (when this project began) and 1,323 in 2015; compare this with '454 sequencing' which returned 123 publications in 2009 and 358 in 2015.

Illumina now produce a benchtop sequencer, the MiSeq, which has several advantages over the GS Junior. Firstly, the capacity of the MiSeq is much higher therefore the cost per base is reduced (see 1.4.6.2). Secondly, the clonal PCR step of the process is done on the machine instead of on the bench and is fully automated (compared to at least a day of lab work for the GS Junior for the emPCR). Thirdly, Illumina sequencing has generally higher quality scores and less difficulty sequencing homopolymer stretches (Loman et al., 2012). These factors combined mean the MiSeq would be the machine of choice if this assay were being designed now.

Assuming the MiSeq was selected as the sequencer, the next decision would be the TEM. If the same six genes were to be sequenced then a small-scale TEM such as TSCA or Haloplex might have been the best option. However, there are many more genes associated with cholestasis, and in fact, the choice of genes for the tNGS assay was limited by the TEM and sequencer combination that was originally chosen. Instead, a much larger group of genes could be targeted and for this a hybridisation approach might have been selected, for example SureSelect. Alternatively a whole exome or clinical exome approach might have been more suitable, as off-the-shelf products would have required less optimisation than custom-designed products. If so, the genes of interest may be selected at the data analysis stage by applying a ROI filter to the variant call file (.vcf). The disadvantages of targeting a larger

number of genes include that although more mutations might be detected, undoubtedly more VUS be more likely required for larger data analysis projects. Also, computing capabilities for data processing and data storage would become a greater consideration.

Chapter 4 – GENETIC INVESTIGATION OF INFANTS WITH LIVER DISEASE

4.1. Introduction

4.1.1. Liver disease in infants

Neonatal cholestasis occurs at a frequency of approximately 1 in 2,500 births (Dick and Mowat 1985). This diagnosis is considered in neonates presenting with prolonged neonatal jaundice, pale stools or dark urine and urgent investigations to differentiate between unconjugated and conjugated bilirubinaemia are indicated which include a fractionated bilirubin level and also ultrasound scan of the liver and biliary tree (De Bruyne et al., 2011b). Measurement of bilirubin in blood allows detection of conjugated bilirubinaemia; commonly agreed as direct bilirubin of greater than 1mg/dL or greater than 20% of total bilirubin (De Bruyne et al., 2011b). Cholestasis may lead to progressive pruritus, growth retardation, fat-soluble vitamin deficiencies, hepatomegaly, splenomegaly, and cirrhosis and liver failure if left untreated (Roberts, 2003). Treatment for neonatal cholestasis varies dependent upon the underlying defect. For example, early surgical intervention for biliary atresia with the Kasai procedure can successfully re-establish bile flow in some cases (Altman et al., 1997). Also, liver transplantation can be curative in PFIC type 2, however the extra-hepatic features of PFIC type 1 will remain or even worsen following a liver transplant

(Davit-Spraul et al., 2009). Several of the metabolic genetic conditions mentioned in Chapter 1 can be diagnosed using biochemical measurements of metabolites in blood or urine, or by measuring enzyme activity. Others, including NPC, PFIC and NICCD, are difficult to diagnose by methods other than genetic diagnosis (Patterson et al., 2012), (Davit-Spraul et al., 2009).

4.1.2. Genetic testing

Genetic testing for NPC, PFIC and NICCD requires sequencing to identify unknown mutations in the coding regions and splice sites of the six genes involved. Typically this is done using fluorescent Sanger sequencing (FS) however genetic testing is not routine in this patient group and genetic diagnosis may take months or years. As a result, the incidence of NPC, PFIC and NICCD in patients presenting with neonatal liver disease is not known.

4.1.3. Background to chapter

Infants presenting with liver disease were recruited from worldwide centres and sent to the West Midlands Regional Genetics Laboratory where they were tested for mutations in the six genes involved in PFIC, NPC and NICCD. Initially microarray sequencing (MS) was used however this was replaced early in the project by targeted next generation sequencing (tNGS). The development of the tNGS assay has been described in Chapter 3. This study was a collaborative project involving the University of Birmingham, Birmingham Children's Hospital and Birmingham Women's Hospital and was funded by Actelion Pharmaceuticals. Professor Paul Gissen, Professor Deirdre Kelly and Dr Chris Hendrickzs were responsible for the design of the project. Zoe Gray

and Carla Lloyd were responsible for the collection of the clinical data. A manuscript has been prepared for publication and is presented in Appendix II.

4.1.4. Chapter 4 Objectives

- To test infants with liver disease for mutations in the *ATP8B1*, *ABCB11*, *ABCB4*, *NPC1*, *NPC2* and *SLC25A13* genes.
- To determine the incidence of PFIC, NPC and NICCD in infants with liver disease.
- To evaluate where possible the contribution of variation in these genes to infantile liver disease.

4.2. Materials & Methods

4.2.1. Patients

Two hundred and twenty two patients were recruited from twelve centers worldwide; seventy four from Germany, seventy three from the UK, sixteen from Poland, thirteen from Turkey, twelve from India, eleven from Oman, six from Denmark, five from the Netherlands, five from Bulgaria, four from Greece, two from Canada and one from Hungary.

The patients met the recruitment criteria which allowed inclusion for patients presenting under the age of two years with either cholestasis (measured by serum conjugated bilirubin levels greater than 20μmol/L or more than 20% of total bilirubin), acute liver failure (prothrombin time greater than twice the upper limit of normal for age), hepatomegaly or splenomegaly (observed on examination or by ultrasound scan). Patients were excluded if they did not meet the above criteria or if a family member had a diagnosis of a genetic condition known to cause neonatal cholestasis.

Questionnaires were completed by the referring centres detailing clinical features and investigations carried out. Follow-up questionnaires were requested from referring centres after a minimum of one year. In some cases, no follow-up information was obtained from the referring centres, and this is recorded as 'Not available' in tables 4-6.

Ethical approval was obtained for the study prior to commencement, and all patients recruited were done so on this basis with ethical approval obtained

from the individual referral centres. This testing was carried out in parallel to the usual diagnostic procedures used at the individual referral centres. The results were fed back to the referral centres in order to aid with diagnosis where possible.

4.2.2. DNA extraction

DNA was extracted from peripheral blood samples using a variety of DNA extraction methods by the individual referring centres and sent to the West Midlands Regional Genetics laboratory for further processing. Once received the DNA samples were quantified using the Nanodrop Spectrophotometer (See 2.2.2.1).

4.2.3. Sequencing methods

The original sequencing method chosen for this work was MS; validation of this method has been reported previously (Bruce et al., 2010) (McKay et al., 2013). In the early stages of the project, the tNGS method became available, and it was decided to change the sequencing method. The reasoning for the change was that tNGS was less labour intensive and more sensitive than MS. The design and optimisation of this assay is described in Chapter 4. The regions interrogated by the microarray were fully covered by the tNGS assay therefore it was possible to combine the data gathered using both methods. Forty four patients were tested using the MS method and one hundred and eighty seven patients were tested using the tNGS method.

4.2.3.1. BioMix Red PCR

PCR was used to amplify the coding exons of *NPC1, NPC2*, *ATP8B1, ABCB11, ABCB4* and *SLC25A13* including the intron-exon boundaries prior to MS (See 2.2.3.2).

4.2.3.2. Microarray Sequencing

MS and data analysis was performed as described previously (2.2.6.2 and 2.2.8.2).

4.2.3.3. Access Array PCR for tNGS

PCR was to amplify the coding exons of *NPC1, NPC2*, *ATP8B1, ABCB11, ABCB4* and *SLC25A13* including the intron-exon boundaries prior to tNGS. This was done using the Access Array system and is described in the chapter 3 and 2.2.3.4.

4.2.3.4. Targeted next generation sequencing

The tNGS assay was performed using the GS Junior system as described in Chapter 3 and 2.2.6.3.

4.2.3.5. NGS data analysis

Analysis of tNGS data was done using NextGENe™ software as described in the chapter 3.

4.2.3.6. Fluorescent Sanger sequencing

9 amplicons in the tNGS design failed to give adequate sequencing results (see Chapter 3). These were removed and instead tested using FS. In

addition, sequence variants identified by tNGS were confirmed using FS (2.2.6.1).

4.2.3.7. FS data analysis

Analysis of the Sanger sequencing data was done using Mutation Surveyor™ as described in 2.2.8.1.

4.2.4. Mutation nomenclature

GenBank reference files are listed in Chapter 2 (2.2.9).

4.2.5. Variant interpretation

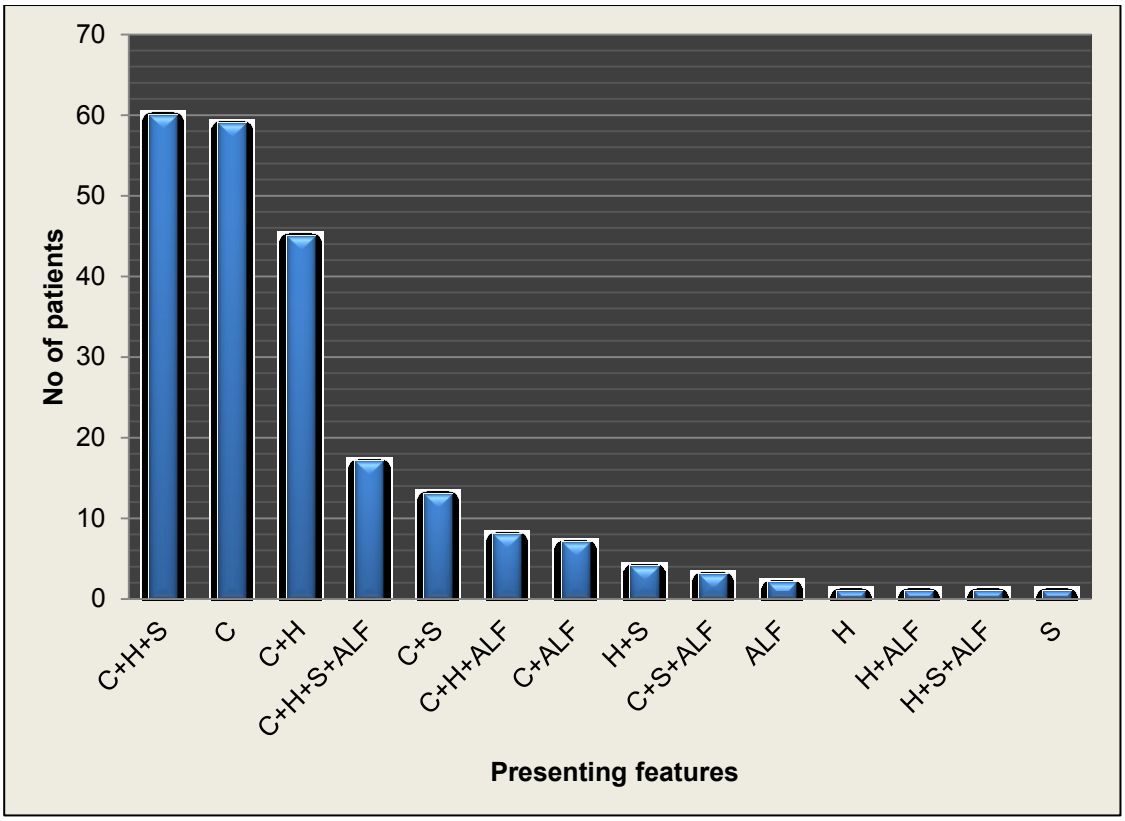
Variant interpretation was done using the methods described in Chapter 2 (2.2.10).

4.3. Results

4.3.1. Presenting features of cohort

Two hundred and twenty two patients were recruited to this study; two hundred and twelve of them had cholestasis, one hundred and thirty seven of them had hepatomegaly, ninety nine had splenomegaly and thirty nine had acute liver failure amongst their presenting features. On the whole cholestasis was the most common feature and acute liver failure was the least common feature, however further grouping of the presenting features of the cohort is shown in Figure 4.1.

Figure 4.1: The presenting features of the cohort at the time of recruitment.
KEY; C=cholestasis, H=hepatomegaly, S=splenomegaly, ALF=acute liver failure.



4.3.2. Sequence variants and interpretation

One hundred and one unique sequence variants were identified by this study. All sequence variants were classified into the following categories; pathogenic, possibly pathogenic, uncertain significance, possibly benign and benign using guidelines previously described (see 2.2.10). The complete results of the analysis of variants are listed in Tables 4.1 to 4.4. Variants classed as benign will not be discussed further.

Nineteen patients were diagnosed with autosomal recessive genetic conditions in this study, and in addition twenty patients were found to have single heterozygous mutations. A further forty four patients were identified who had variants of uncertain clinical significance that did not explain their clinical presentation, and the remainder of the patients (n=139), no sequence variants other than known benign polymorphisms were identified.

Table 4.1: Nonsense and frameshift mutations identified in the cohort. Since these are all predicted to result in nonsense-mediated decay they were all considered pathogenic.

Gene	DNA	Protein	Known/Novel	Conclusion	Patient
ABCB11	c.1081C>T	p.Gln361*	Novel	Pathogenic	180
ABCB11	c.1416T>A	p.Tyr472*	(Knisely et al., 2006)	Pathogenic	156
ABCB11	c.1558A>T	p.Arg520*	(Strautnieks et al., 2008)	Pathogenic	185
ABCB11	c.3904G>T	p.Glu1302*	(Strautnieks et al., 2008)	Pathogenic	11
ABCB11	c.3933C>G	p.Tyr1311*	Novel	Pathogenic	218
ABCB11	c.731_732insA	p.Ile245Thrfs*26	Novel	Pathogenic	194
ATP8B1	c.2788C>T	p.Arg930*	(Klomp et al., 2004)	Pathogenic	36
NPC1	c.2010C>A	p.Cys670*	Novel	Pathogenic	28

Table 4.2: Non-coding changes identified in the study, i.e changes in intronic or untranslated regions of the gene.

Table 4.3: Amino acid changes identified in the study.

Table 4.4 Synonymous changes identified in the study, i.e no change to the protein sequence is predicted, however there is a possibility of altered mRNA splicing patterns.

KEY:

Frequency data includes HET= heterozygosity figure from dbSNP (NCBI) 1KG=1000Genomes data listed as number of alleles observed, ESP=Exome Sequencing Project data listed as number of alleles observed, AA=African American, EA= European American.

Conservation includes nucleotide (NA) conservation and amino acid (AA) conservation as scored by Alamut software.

Protein Predictions include AGVGD=Align GVGD, SIFT and PP=PolyPhen-2 tools.

Splicing predictions were done using SSF=Splice-site finder, MES=MaxEntScan, NNS=NNSplice, GS=Genesplicer and HSF=Human Splicing Finder. Results were 'no change'= no change when compared to wild type sequence, 'donor/acceptor destroyed'= predicted loss of wild type splice site or 'cryptic donor/acceptor'=predicted creation of a novel splice site.

Table 4.2 Non-coding changes.

Gene	DNA	HGMD	SNP reference	Frequency	Splicing prediction tools	Conclusion	Patient
ABCB11	c.1084-2A>G	Novel	No		SSF acceptor destroyed, cryptic acceptor MES acceptor destroyed, cryptic acceptor NNS acceptor destroyed, cryptic acceptor GS acceptor destroyed, cryptic acceptor HSF acceptor destroyed	Pathogenic	211
ABCB4	c.1230+1G>T	Novel	No		SSF donor destroyed MES donor destroyed NNS donor destroyed GS donor destroyed HSF donor destroyed	Pathogenic	200
ATP8B1	c.1918+8C>T	Novel	No		SSF no change MES no change NNS no change GS no change HSF no change	Likely benign	169
ATP8B1	c.555-3T>C	Novel	No		SSF no change MES cryptic donor NNS no change GS no change HSF no change	Likely benign	91
NPC1	c.2131-5_2131-4delTT	Novel	rs201618643	No frequency data	SSF no change MES no change NNS no change GS no change HSF no change	Likely benign	77
NPC2	c.441+1G>A	(Bauer et al., 2013)	rs140130028	dbSNP HET 0.002 1KG 2/2184 alleles ESP AA 7/4406 and EA 76/8600 alleles	SSF donor destroyed MES donor destroyed NNS donor destroyed GS donor destroyed HSF donor destroyed	Uncertain significance	219, 162, 84, 128, 132, 37, 182
SLC25A13	c.-115G>T	Novel	No		SSF no change MES no change NNS no change GS no change HSF cryptic donor	Likely benign	189

Table 4.3 Amino acid changes.

Gene	DNA	Protein	Known or Novel	SNP reference	Frequency	Conservation	Amino Acid difference	Protein prediction tools	Splicing prediction tools	Conclusion	Patient
ABCB11	c.127G>A	V43I	No	rs183406496	dbSNP not recorded 1KG 2/2196 alleles ESP 4/3246 AA and 24/8214 EA alleles	NA not AA weak	Small	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF Cryptic donor	Likely benign	143, 173
ABCB11	c.779G>A	G260D	Novel	No		NA weak AA moderate	Moderate	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	194
ABCB11	c.1097T>G	V366G	Novel	No		NA high AA moderate	Moderate	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	148
ABCB11	c.1396C>A	Q466K	(Strautnieks et al., 2008)	rs200148505	dbSNP not recorded 1KG not seen ESP 9/3246 AA and 0/8214 EA alleles	NA high AA high	Small	AGVGD C45 SIFT Deleterious PP Probably damaging	SSF Cryptic acceptor MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	101
ABCB11	c.1409G>A	R470Q	(Strautnieks et al., 2008) (Byrne et al., 2009)	No		NA high AA high	Small	AGVGD C35 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Pathogenic	159,
ABCB11	c.1445A>G	D482G	(Strautnieks et al., 2008)	rs72549402	Not frequency data	NA high AA high	Moderate	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF No changes MES Cryptic donor NNS No changes GS No changes HSF Cryptic donor	Pathogenic	181, 180, 179

ABCB11	c.1621A>C	I541L	(Nobili et al., 2006) (Byrne et al., 2009)	No		NA high AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Pathogenic	193
ABCB11	c.1676T>C	M559T	Novel	No		NA high AA high	Moderate	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF Cryptic donor	Possibly pathogenic	218
ABCB11	c.1829T>C	I610T	Novel	No		NA weak AA moderate	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	54
ABCB11	c.2093G>A	R698H	(Anzivino et al., 2013)	rs138642043	dbSNP HET 0.1 1KG 52/2196 ESP 38/3246 AA and 29/8214 EA alleles	NA high AA moderate	Small	AGVGD C0 SIFT Deleterious PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	29
ABCB11	c.2596T>G	S866A	Novel	No		NA moderate AA high	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	139
ABCB11	c.2678C>T	A893V	Novel	No		NA high AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	164
ABCB11	c.2708T>G	V903G	Novel	No		NA high AA high	Moderate	AGVGD C35 SIFT Deleterious PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	14
ABCB11	c.2834G>A	S945N	Novel	rs200857579	dbSNP not recorded 1KG not seen ESP 1/8214 EA alleles, AA not seen	NA weak AA moderate	Small	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	118
ABCB11	c.2918C>T	T973I	Novel	No		NA moderate AA moderate	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	92

ABCB11	c.3517A>G	N1173D	Novel	No		NA high AA moderate	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	216
ABCB11	c.3628A>C	T1210F	(Strautnieks et al., 2008) (Byrne et al., 2009) (Gonzales et al., 2012)	No		NA high AA moderate	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Pathogenic	216
ABCB4	c.217C>G	L73V	(Pauli-Magnus et al., 2004) (Colombo et al., 2011) (Anzivino et al., 2013)	rs8187788	dbSNP HET 0.001 1KG 1/2184 alleles ESP 11/8600 EA alleles, AA not seen	NA weak AA high	Small	AGVGD C15 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	93
ABCB4	c.262A>G	T88A	Novel	No		NA weak AA moderate	Small	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	147
ABCB4	c.523A>G	T175A	10 (Rosmorduc et al., 2001), 11 (Wendum et al., 2012)	rs58238559	dbSNP HET 0.012 1KG 13/2184 alleles ESP 103/8600 EA and 7/4406 AA alleles	NA high AA high	Small	AGVGD C55 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	20
ABCB4	c.524C>T	T175M	Novel	No		NA moderate AA high	Moderate	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF Cryptic acceptor MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	213
ABCB4	c.598G>T	A200S	Novel	No		NA weak AA moderate	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS Cryptic acceptor HSF No changes	Likely benign	30
ABCB4	c.1529A>G	N510S	(Davit-Spraul et al., 2010) (Anzivino et al., 2013)	rs3753315619	dbSNP not recorded 1KG not seen ESP 1/8600 EA alleles, AA not seen	NA weak AA high	Small	AGVGD C0 SIFT Deleterious PP Possibly damaging	SSF Cryptic donor MES No changes NNS No changes GS No changes HSF Cryptic acceptor	Possibly pathogenic	18

ABCB4	c.1584G>C	E528D	(Rosmorduc et al., 2003)	rs8187797	dbSNP HET 0.019 1KG 21/2184 alleles ESP 138/4406 AA alleles, EA not seen	NA not AA moderate	Small	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	23
ABCB4	c.1597C>A	L533M	Novel	No		NA moderate AA high	Small	AGVGD C0 SIFT Tolerated PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	19
ABCB4	c.1624G>C	A542P	Novel	No		NA high AA high	Small	AGVGD C25 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	2
ABCB4	c.1652C>T	P551L	Novel	No		NA high AA high	Moderate	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	148
ABCB4	c.1769G>A	R590Q	(Degiorgio et al., 2007)	rs45575636	dbSNP HET 0.15 1KG 16/2184 alleles ESP 74/8600 EA and 28/4406 AA alleles	NA high AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	33, 4
ABCB4	c.1858_1860delAAG	K620del	Novel	No		AA moderate			SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	203
ABCB4	c.1954A>G	R652G	(Liu et al., 2009) (Davit-Spraul et al., 2010)	rs2230028	dbSNP HET 0.243 1KG ESP 679/8600 EA and 1513 AA alleles	NA not AA weak	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	28, 23, 20, 34
ABCB4	c.2324C>T	T775M	(Degiorgio et al., 2007)	rs148052192	dbSNP HET 0.001 1KG not seen ESP 7/8600 EA and	NA weak AA high	Moderate	AGVGD C0 SIFT Deleterious PP Probably	SSF No changes MES No changes NNS No changes	Uncertain significance	135

					1/4406 AA alleles			damaging	GS No changes HSF No changes		
ABCB4	c.2363G>A	R788Q	(Rosmorduc et al., 2003)	rs8187801	dbSNP HET 0.045 1KG 50/2184 alleles ESP 4/8600 EA and 330/4406 AFR alleles	NA moderate AA high	Small	AGVGD C35 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	202
ABCB4	c.2800G>A	A934T	(Rosmorduc et al., 2003)	rs61730509	dbSNP HET 0.006 1KG 7/2184 alleles ESP 54/4406 AFR alleles, EA not seen	NA high AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	163
ABCB4	c.2809T>C	T937H	Novel	No		NA moderate AA moderate	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	11
ABCB4	c.3317A>G	E1106G	Novel	rs139042803	dbSNP not recorded 1KG not seen ESP 2/4406 AA alleles, EA not seen	NA moderate AA moderate	Moderate	AGVGD C0 SIFT Deleterious PP Possibly damaging	SSF Cryptic acceptor MES Cryptic acceptor NNS No changes GS No changes HSF No changes	Possibly pathogenic	185
ABCB4	c.3403G>A	E1135K	Novel	No		NA high AA high	Small	AGVGD C55 SIFT Deleterious PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	195
ATP8B1	c.134A>C	N45T	(Painter et al., 2005)	rs146599962	dbSNP HET 0.005 1KG 5/2184 alleles ESP 37/8600 EA and 4/4406 AA alleles	NA weak AA moderate	Small	AGVGD C0 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	51
ATP8B1	c.208G>A	D70N	(Klomp et al., 2004) (Mullenbach, 2005) (Folmer et al., 2009)	rs34719006	dbSNP HET 0.007 1KG 8/2184 alleles ESP 26/8600 EA and 5/4406 AA alleles	NA moderate AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	187, 83, 6
ATP8B1	c.287C>G	A96G	Novel	No		NA moderate AA high	Small	AGVGD C0 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	109

ATP8B1	c.607A>G	K203E	(Painter et al., 2005)	rs56355310	dbSNP 2/2184 alleles 1KG 16/8600 EA and 3/4406 AA alleles ESP	NA moderate AA high	Small	AGVGD C0 SIFT Deleterious PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	83
ATP8B1	c.1177A>G	I393V	Novel	rs34315917	dbSNP HET 0.012 1KG 13/2184 alleles ESP 64/8600 EA and 6/4406 AA alleles	NA high AA high	Small	AGVGD C0 SIFT Tolerated PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF Cryptic donor	Likely benign	14
ATP8B1	c.1244A>G	Q415R	Novel	No		NA high AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	199
ATP8B1	c.1286A>C	E429A	(Klomp et al., 2004)	rs34018205	dbSNP HET 0.014 1KG 15/2184 alleles ESP 37/8600 EA and 4/4406 AA alleles	NA weak AA moderate	Moderate	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	160
ATP8B1	c.1367C>T	T456M	(Klomp et al., 2004)	rs121909104	Not frequency data	NA high AA high	Moderate	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	183
ATP8B1	c.2083G>A	E695K	Novel	No		NA high AA high	Small	AGVGD C55 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	183
ATP8B1	c.2425A>C	I809L	Novel	No		NA moderate AA high	Small	AGVGD C0 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	23
ATP8B1	c.2498G>A	R833Q	Novel	No		NA weak AA moderate	Small	AGVGD C0 SIFT Tolerated PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	213

ATP8B1	c.3043T>C	F1015L	Novel	No		NA high AA high	Small	AGVGD C15 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	139
ATP8B1	c.3633C>A	F1211L	Novel	No		NA moderate AA high	Small	AGVGD C15 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	30
ATP8B1	c.3656A>G	D1219G	Novel	No		NA moderate AA high	Moderate	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	33
NPC1	c.467T>C	M156T	Novel	rs147615070	dbSNP not recorded 1KG not seen ESP 1/8600 EA alleles, AA not seen	NA high AA high	Moderate	AGVGD C25 SIFT Deleterious PP benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	116
NPC1	c.873G>T	W291C	Novel	rs138151007	dbSNP HET 0.002 1KG 2/2184 alleles ESP 1/8600 EA and 6/4406 AA alleles	NA high AA high	Large	AGVGD C15 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	153
NPC1	c.885A>C	K295N	Novel	No		NA weak AA moderate	Moderate	AGVGD C0 SIFT Deleterious PP Benign	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	26
NPC1	c.2000C>T	S667L	(Tamura et al., 2006)	No		NA high AA high	Large	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS Cryptic acceptor HSF No changes	Pathogenic	43
NPC1	c.3107C>T	T1036M	(Carstea, 1997)	rs28942104	Not frequency data	NA high AA moderate	Moderate	AGVGD C15 SIFT Deleterious PP Probably damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Pathogenic	115
NPC1	c.3182T>C	I1061T	(Yamamoto et al., 1999) (Pipalia et al., 2011) (Gelsthorpe	rs80358259	dbSNP not recorded 1KG not seen ESP 4/8600 EA and 1/4406 AA alleles	NA high AA high	Moderate	AGVGD C25 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Pathogenic	43

			et al., 2008)								
NPC1	c.3362T>G	L1121R	Novel	No		NA high AA moderate	Moderate	AGVGD C35 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS Cryptic donor HSF No changes	Possibly pathogenic	157
NPC1	c.3614C>T	T1205I	Novel	No		NA high AA high	Small	AGVGD C65 SIFT Deleterious PP Probably damaging	SSF Cryptic acceptor MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	54
NPC2	c.212A>G	K71R	Novel	rs142075589	dbSNP HET 0.001 1KG 1/2184 alleles ESP 2/8600 EA alleles, AA not seen	NA moderate AA moderate	Small	AGVGD C0 SIFT Tolerated PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Uncertain significance	172
SLC25A13	c.639T>G	H213Q	Novel	No		NA weak AA high	Small	AGVGD C0 SIFT Deleterious PP Benign	SSF Cryptic donor MES No changes NNS No changes GS No changes HSF Cryptic acceptor	Uncertain significance	21
SLC25A13	c.1903G>T	D635Y	Novel	No		NA moderate AA high	Large	AGVGD C15 SIFT Deleterious PP Possibly damaging	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Possibly pathogenic	147

Table 4.4 Synonymous changes

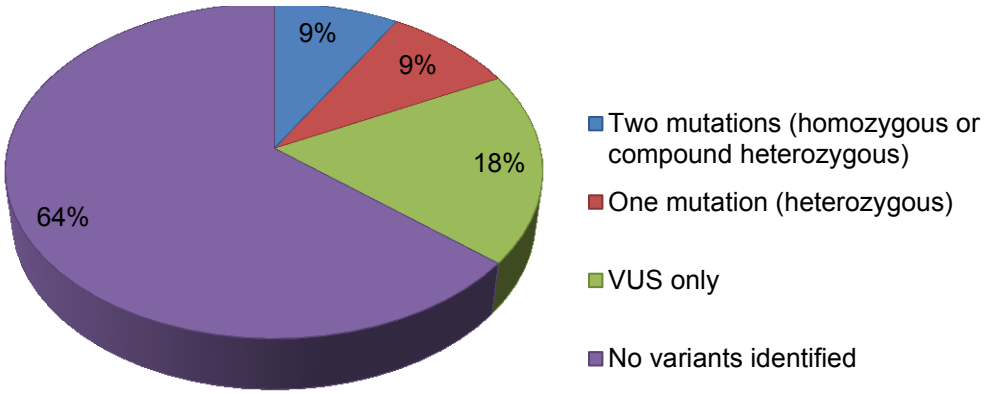
Gene	DNA	Protein	Known or Novel	SNP reference	Frequency summary	Splicing prediction tools	Conclusion	Patient
ABCB11	c.1281C>T	p.=	Novel	rs11568360	dbSNP HET 0.02 1KG 2/2196 alleles ESP 1/3246 AA and 0/8214 EA alleles	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	111
ABCB11	c.1530C>A	p.=	Novel	rs374701109	dbSNP not recorded 1KG not seen ESP 1/8330 AA and 2/3924 EA alleles	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	135
ABCB11	c.3589C>T	p.=	Novel	No		SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	210
ABCB11	c.3846C>T	p.=	Novel	rs369538863	dbSNP not recorded 1KG not seen ESP 2/4240 EA alleles, AA not seen	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	210
ABCB4	c.1143A>G	p.=	Novel	rs142671969	dbSNP not recorded 1KG not seen ESP 1/8600 EA alleles, AA not seen	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	166
ABCB4	c.3037A>C	p.=	Novel	rs2230029	dbSNP HET 0.004 1KG not seen ESP not seen	SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	214, 69
ABCB4	c.696C>T	p.=	Novel	rs8187791	dbSNP HET 0.004 1KG 4/2184 alleles ESP 4/8600 EA and 1/4406 AA alleles	SSF cryptic acceptor MES No changes NNS cryptic acceptor GS No changes HSF No changes	Uncertain significance	155
ATP8B1	c.2937A>C	p.=	Novel	No		SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	148

ATP8B1	c.852A>C	p.=	Novel	No		SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	128
NPC1	c.3519A>G	p.=	Novel	No		SSF No changes MES No changes NNS No changes GS No changes HSF cryptic acceptor	Likely benign	105
NPC1	c.3756G>A	p.=	Novel	No		SSF No changes MES No changes NNS No changes GS No changes HSF No changes	Likely benign	75

A summary of the results is shown by Figure 4.2. Genetic diagnoses of the autosomal recessive conditions were made by identification of two changes in the same gene which were determined to be 'Pathogenic' or 'Possibly pathogenic'; and this was identified in 8.5% (n=19) of the cohort (Table 4.5). Thirteen patients carried homozygous variants, and 6 of these were known to have consanguineous parents. Patient 179 was known to have non-consanguineous parents but is homozygous for a relatively frequent missense mutation in *ABCB11*. Cholestasis was a presenting feature of all of the PFIC 1 and 2 patients as would be expected. Cholestasis was more frequent in the patients with diagnoses than in the whole cohort, presumably because of the frequency of PFIC cases. As a result, patients with cholestasis were more likely to have a genetic diagnosis made than patients without cholestasis.

Single heterozygous variants, classified as pathogenic or possibly pathogenic, were identified in 9% of the patients (n=20, see Table 4.6).

Figure 4.2: Summary of genetic findings in patients (n=222)



No pathogenic variants or variants classified as possibly pathogenic were identified in one hundred and eighty three of the two hundred and twenty two patients tested, however in forty one of these (18%), variants were identified which were classified as of unknown significance (VUS) or as possibly benign (see Table 4.7).

Thirteen sequence variants were classified as VUS because there was evidence for and against pathogenicity. None of these variants were identified in a patient who also had a variant classed as 'pathogenic' or 'possibly pathogenic' in the same gene (see Table 4.7). For this reason it has not been possible to confirm that any of these VUS are causative of autosomal recessive disease.

4.3.3. Clinical and genetic features of patients with *ATP8B1* mutations and variants

Three patients were diagnosed with PFIC-1 (Table 4.5). All three presented with cholestasis, and two of the three also presented with hepatomegaly. One patient had a PEBD aged seven months, one had a liver transplant, and information was not available for the third.

Five patients were identified with single heterozygous mutations in *ATP8B1* (Table 4.6). All five presented with cholestasis, one of whom also had hepatomegaly, and one of whom had hepatosplenomegaly and acute liver failure. Symptoms of progressive liver disease resulted in liver transplantation in three patients, although BA was cited as an alternative diagnosis in two of

these cases. The symptoms of the remaining two cases were reported to have resolved.

Eight patients were identified with VUS in the *ATP8B1* gene (Table 4.7). Seven of these had cholestasis, six had hepatomegaly, five had splenomegaly, and three had acute liver failure. Two of the eight patients also had symptoms of progressive liver disease which resulted in liver transplants, and both had alternative diagnoses of BA and VACTERL association. Another two patients also had alternative diagnoses of biliary atresia, but their condition had stabilised following Kasai portoenterostomy. No follow-up information was available for the remaining four patients with VUS in the *ATP8B1* gene.

4.3.4. Clinical and genetic features of patients with *ABCB11* mutations and variants

Eleven patients were diagnosed with PFIC-2 (Table 4.5), and all presented with cholestasis. Additionally nine patients had hepatomegaly, seven had splenomegaly, and four had acute liver failure. Three patients underwent PEBD procedures, among which cholestasis persisted in two cases. Three patients underwent liver transplants, and all were asymptomatic at follow up. There were no follow-up data available for the remaining five patients.

Four patients were identified with single heterozygous mutations in *ABCB11*, and one of these also had a possibly pathogenic change in *ABCB4* (Table 4.6). Symptoms resolved without intervention in two of these cases, one patient improved after a PEBD procedure, and no information was available for the fourth patient.

Nine patients were identified with only VUS in the *ABCB11* gene (Table 4.7); eight of which presented with cholestasis, seven with splenomegaly, and five with hepatomegaly. None of these patients had acute liver failure. At follow up; one patient was reported to have had progressive liver disease and an alternative diagnosis of AGS. Two patients were found to have BA and were stable following Kasai portoenterostomy. One patient had an alternative diagnosis of Abernathy malformation and was stable. Two were asymptomatic, with no further information available.

4.3.5. Clinical and genetic features of patients with *ABCB4* mutations and variants

Four patients were diagnosed with PFIC-3 (Table 4.5). All four presented with hepatomegaly, three presented with cholestasis, two had acute liver failure, and one had splenomegaly. One patient with cholestasis, hepatomegaly and acute liver failure had a liver transplant and was asymptomatic at follow up. No follow-up information was available for the other three cases.

Single heterozygous *ABCB4* mutations were identified in three patients (excluding the previously mentioned patient 100 who had mutations in *ABCB11* and *ABCB4* (Table 4.6). All three presented with cholestasis: two had splenomegaly, one had hepatomegaly, and one had acute liver failure. Follow-up information was available for two of these three cases. One had BA resulting in a liver transplant, and one died aged fifteen months due to multi-organ failure.

Thirteen patients were identified with only VUS in the *ABCB4* gene (Table 4.7), all of whom presented with cholestasis. Nine also had hepatomegaly, eight had splenomegaly, and three had acute liver failure. Six of the patients were subsequently diagnosed with biliary atresia, three of whom had a liver transplant and three of whom underwent Kasai portoenterostomy. Another patient had symptoms of progressive liver disease and was suspected of having a long-chain fatty acid oxidation defect. One patient was reported to be asymptomatic at follow up, and no follow-up information was available for the remaining four patients.

4.3.6. Clinical and genetic features of patients with *NPC1* mutations and variants

One patient was diagnosed with NPC1 (Table 4.5). This patient presented with cholestasis and splenomegaly but no further information was available at follow-up. A further seven patients were identified with a single heterozygous mutation in the *NPC1* gene (Table 4.6). All seven had cholestasis: five had hepatomegaly, two had splenomegaly, and one had acute liver failure at presentation. In four patients, visceral symptoms resolved with no intervention. One patient with progressive liver disease received an alternative diagnosis of BA and required a liver transplant. However, a further two patients had progressive liver disease and no alternative diagnosis.

Four patients were identified with VUS in the *NPC1* gene (Table 4.7), three of whom presented with cholestasis (two with splenomegaly; and one with acute liver failure). Three patients had symptoms of progressive liver disease that

resulted in liver transplantation. Alternative diagnoses of neonatal hemochromatosis, transient myelodysplastic syndrome associated with Trisomy 21 syndrome, and AGS were reported in three patients.

4.3.7. Clinical and genetic features of patients with *NPC2* mutations and variants

No patients were diagnosed with Niemann Pick disease type C2, and no patients were found with single heterozygous mutations in the *NPC2* gene. VUS in the *NPC2* gene were identified in eight patients (Table 4.7). Patient 172 presented with cholestasis, hepatomegaly and splenomegaly, and had a final diagnosis of extrahepatic biliary atresia (BA). Seven of the eight patients were found to be heterozygous for the same *NPC2* variant; c.441+1G>A (see Figure 4.3). All of these presented with cholestasis, six patients had hepatomegaly, two had splenomegaly, and one had acute liver failure. One patient had progressive liver disease and was subsequently diagnosed with acute lymphoblastic leukaemia. Three more patients had features of progressive liver disease but no alternative diagnosis was found. No follow-up information was available for the remaining three patients. This variant was the most frequent finding in this cohort, occurring in 1.5% of patients.

4.3.8. Clinical and genetic features of patients with *SLC25A13* mutations and variants

No patients were diagnosed with NICCD, but one patient with a single heterozygous mutation in *SLC25A13* was identified (Table 4.6). This patient had cholestasis and hepatosplenomegaly, and these symptoms resolved

without intervention. A further two patients had VUS in the *SLC25A13* gene (Table 4.7). Both presented with cholestasis and were asymptomatic at follow up.

Table 4.5: Genetic findings, presenting features and outcomes in patients diagnosed with autosomal recessive conditions. 12 had PFIC2, 5 had PFIC3, 3 had PFIC1 and 1 had NPC. Parental samples were not tested therefore it has been assumed that patients with two different mutations were compound heterozygotes.

KEY: C=cholestasis, H=hepatomegaly, S=splenomegaly, ALF=acute liver failure, LT=liver transplant, PEBD=partial extrahepatic biliary diversion.

Patient	Mutation 1	Mutation 2	Diagnosis	Presenting features	Features at follow-up
43	NPC1 c.2000C>T p.(S667L)	NPC1 c.3182T>C p.(I1061T)	NPC	C, H	Not available
199	*ATP8B1* c.1244A>G p.(Q415R)	*ATP8B1* c.1244A>G p.(Q415R)	PFIC 1	C, H	Not available
183	*ATP8B1* c.1367C>T p.(T456M)	*ATP8B1* c.2083G>A p.(E695K)	PFIC 1	C, H	PEBD at 7m, C gradually resolved
36	*ATP8B1* c.2788C>T p.(R930*)	*ATP8B1* c.2788C>T p.(R930*)	PFIC 1	C	LT. Current symptoms unknown
194	*ABCB11* c.731_732insA p.(I245Tfs*26)	*ABCB11* c.779G>A p.(G260D)	PFIC 2	C, H	PEBD at 14m Pruritis resolved C persists
180	*ABCB11* c.1081C>T p.(Q361*)	*ABCB11* c.1445A>G p.(D482G)	PFIC 2	C, S	PEBD at 14m. C persists
211	*ABCB11* c.1084-2A>G	*ABCB11* c.1084-2A>G	PFIC 2	C, S, H, ALF	Not available
159	*ABCB11* c.1409G>A p.(R470Q)	*ABCB11* c.1409G>A p.(R470Q)	PFIC 2	C, H	LT. Asymptomatic post-transplant
205	*ABCB11* c.1409G>A p.(Q470R)	*ABCB11* c.1409G>A p.(Q470R)	PFIC 2	C, S, H	Not available
156	*ABCB11* c.1416T>A p.(Y472*)	*ABCB11* c.1416T>A p.(Y472*)	PFIC 2	C, H, ALF	Not available
179	*ABCB11* c.1445A>G p.(D482G)	*ABCB11* c.1445A>G p.(D482G)	PFIC 2	C, S, H	PEBD at 9m. C resolved
218	*ABCB11* c.1676T>C p.(M559T)	*ABCB11* c.3933C>G p.(Y1311*)	PFIC 2	C, S, H, ALF	LT. Asymptomatic post-transplant
14	*ABCB11* c.2708T>G p.(V903G)	*ABCB11* c.2708T>G p.(V903G)	PFIC 2	C, S, H, ALF	Not available
216	*ABCB11* c.3517A>G p.(N1173D)	*ABCB11* c.3628A>C p.(T1210F)	PFIC 2	C	LT. Asymptomatic post-transplant
11	*ABCB11* c.3904G>T p.(E1302*)	*ABCB11* c.3904G>T p.(E1302*)	PFIC 2	C, S, H	Not available
200	*ABCB4* c.1230+1G>T	*ABCB4* c.1230+1G>T	PFIC 3	C, S, H, ALF	Not available
2	*ABCB4* c.1624G>C p.(A542P)	*ABCB4* c.1624G>C p.(A542P)	PFIC 3	C, H	Not available
148	*ABCB4* c.1652C>T p.(P551L)	*ABCB4* c.1652C>T p.(P551L)	PFIC 3	C, H, ALF	LT. Asymptomatic post-transplant
203	*ABCB4* c.1858_1860delAAG p.(K620del)	*ABCB4* c.1858_1860delAAG p.(K620del)	PFIC 3	H	Not available

Table 4.6: Genetic findings, presenting features and outcomes in patients with only single heterozygous mutation. 6 had *NPC1* mutations, 3 had *ATP8B1* mutations, 3 had of *ABCB11* mutations, 2 had *ABCB4* mutations and one had an *ABCB11* mutation plus an *ABCB4* mutation.
KEY: C=cholestasis, H=hepatomegaly, S=splenomegaly, ALF=acute liver failure, LT=liver transplant, PEBD=partial extrahepatic biliary diversion, PN=parenteral nutrition, Dx=diagnosis, BA=biliary atresia.

Gene	Mutation	Patient	Presenting features	Final diagnosis and status at follow-up
NPC1	c.467T>C p.(M156T)	116	C, H, S	Symptoms resolved. No intervention Alternative diagnosis: multisystem juvenile xanthogranuloma
NPC1	c.873G>T p.(W291C)	153	C	Symptoms resolved. No intervention
NPC1	c.2010C>A p.(C670*)	28	C, H	Progressive liver disease Alternative diagnosis: BA LT (22m). Asymptomatic post-transplant
NPC1	c.2010C>A p.(C670*)	22	C, S	Symptoms resolved. No intervention Alternative diagnosis: Hirschsprung's disease necessitating PN.
NPC1	c.3107C>T p.(T1036M)	115	C, H	Symptoms resolved. No intervention
NPC1	c.3614C>T p.(T1205I)	54	C, H	Progressive liver disease with portal hypertension
NPC1	c.3614C>T p.(T1205I)	157	C, H, ALF	Progressive liver disease LT. Asymptomatic post-transplant
ATP8B1	c.287C>G p.(A96G)	109	C, H	Symptoms resolved.
ATP8B1	c.2425A>C p.(I809L)	23	C	Progressive liver disease Alternative diagnosis: BA LT (15m). Asymptomatic post-transplant
ATP8B1	c.3043T>C p.(F1015L)	139	C, H, S, ALF	Progressive Liver disease Alternative diagnosis: BA LT(10m). Asymptomatic post-transplant

ATP8B1	c.3633C>A p.(F1211L)	30	C	Symptoms resolved. No intervention
ATP8B1	c.3656A>G p.(D1219G)	33	C	Progressive liver disease LT(30m). Asymptomatic post-transplant
ABCB11	c.1445A>G p.(D482G)	179	C, ALF	PEBD at 5m, symptoms resolved.
ABCB11 *ABCB4*	c.1558A>T p.(R520*) c.3317A>G p.(E1106G)	185	C, H, S	Symptoms resolved. No intervention
ABCB11	c.1621A>C p.(I541L)	193	C	Symptoms resolved. No intervention
ABCB11	c.2678C>T p.(A893V)	164	C, H, S	Not available
ABCB4	c.524C>T p.(T175M)	213	C, H	Not available
ABCB4	c.1529A>G p.(N510S)	18	C, S	Progressive liver disease Alternative diagnosis: BA LT (8m). Asymptomatic post-transplant
ABCB4	c.3403G>A p.(E1135K)	195	C, S, ALF	Progressive liver disease (multi-organ failure) Died age 15m
SLC25A 13	c.1903G>T p.(D635Y)	134	C, H, S	Symptoms resolved. No intervention

Table 4.7 – Genetic findings, presenting features and outcomes in patients with only VUS or likely benign variants.

KEY: C=cholestasis, H=hepatomegaly, S=splenomegaly, ALF=acute liver failure, EHBA=extrahepatic biliary atresia, LT=liver transplant, PEBD=partial extrahepatic biliary diversion, PN=parenteral nutrition, ICP=intrahepatic cholestasis of pregnancy

Gene	Sequence change	Patient	Presenting features	Final diagnosis and status at follow-up
NPC1	c.885A>C p.(K295N)	26	ALF	Progressive liver disease Alternative diagnosis: Neonatal hemochromatosis LT (1m). Asymptomatic post-transplant
NPC1	c.2131-5_2131-4delTT	77	C, H	Progressive liver disease Alternative diagnosis: Trisomy 21; transient myelodysplastic syndrome LT (3m). Asymptomatic post-transplant
NPC1	c.3519A>G p.(=)	105	C	Suspected Alagille syndrome (phenotypic features); lost to follow-up.
NPC1	c.3756G>A p.(=)	75	C, H	Progressive Liver disease Two LTs; living related LT at 14m and cadaveric LT 4 days later. Asymptomatic post-transplant
NPC2	c.212A>G p.(K71R)	172	C, H, S	Progressive liver disease with portal hypertension Alternative diagnosis: BA. Stable post-Kasai
NPC2	c.441+1G>A	37	C, H	Progressive liver disease
		219	C, H	Progressive liver disease
		155	C, H, S	Progressive liver disease Alternative diagnosis: acute lymphoblastic leukaemia
		182	C, H, S, ALF	Progressive liver disease LT (13m). Asymptomatic post-transplant
		84	C, H	Not available
		128	C, H	Not available
		132	C, ALF	Not available
ATP8B1	c.134A>C p.(N45T)	51	C, S	Stable liver disease Alternative diagnosis: BA and Cat Eye syndrome. Stable post-Kasai
ATP8B1	c.208G>A p.(D70N)	187	H, S, ALF	Progressive liver disease Alternative diagnosis: VACTERL association; required PN support. LT. Asymptomatic post-transplant
		83	C, S, ALF	Not available
		6	C, H, ALF	Not available

ATP8B1	c.555-3T>C	91	C, H, S	Stable liver disease Alternative diagnosis: BA. Stable post-Kasai
ATP8B1	c.852A>C p.(=)	128	C, H	Not available
ATP8B1	c.1918+8C>T (Homozygous)	169	C, H, S	Progressive liver disease Alternative diagnosis: BA LT (10m) Asymptomatic post-transplant
ATP8B1	c.1286A>C p.(E429A)	160	C, H	Not available
ABCB11	c.127G>A p.(V43I)	173	C, S	Asymptomatic
		143	C, H, S	Stable liver disease Alternative diagnosis: BA Stable post-Kasai
ABCB11	c.1281C>T p.(=)	111	C, S	Stable liver disease Alternative diagnosis: Abernethy malformation
ABCB11	c.1396C>A p.(Q466K)	101	C, S	Asymptomatic
ABCB11	c.1530C>A p.(=)	135	C, H, S	Stable liver disease Alternative diagnosis: BA. Stable post-Kasai
ABCB11	c.2093G>A p.(R698H)	29	S	Not available
ABCB11	c.2834G>A p.(S945N)	118	C, H	Stable liver disease Alternative diagnosis: BA Stable post-Kasai
ABCB11	c.2918C>T p.(T973I)	92	C, H, S	Asymptomatic
ABCB11	c.3589C>T p.(=)	210	C, H	Not available
ABCB11	c.3846C>T p.(=)	210	C, H	Not available
ABCB4	c.217C>G (L73V)	93	C	Stable liver disease Alternative diagnosis: BA Type 3 Stable post-Kasai
ABCB4	c.523A>G p.(T175A)	20	C, H, S	Progressive liver disease Alternative diagnosis: BA LT (7m). Asymptomatic post-transplant
ABCB4	c.696C>T p.(=)	155	C, H, S	Not available
ABCB4	c.1143A>G p.(=)	166	C, ALF	Progressive liver disease with recurrent episodes of ALF Alternative diagnosis: Suspected long chain fatty acid oxidation defect
ABCB4	c.1597C>A p.(L533M)	19	C, H	Asymptomatic
ABCB4	c.1769G>A p.(R509Q)	4	C, H, ALF	Not available
ABCB4	c.1954A>G p.(R652G)	20	C, H, S	Progressive liver disease

				Alternative diagnosis: BA LT (7m). Asymptomatic post-transplant
		34	C, H, S	Progressive liver disease Alternative diagnosis: BA LT (8m). Asymptomatic post-transplant
ABCB4	c.2324C>T p.(T775M)	135	C, H, S	Stable liver disease Alternative diagnosis: BA Stable post-Kasai
ABCB4	c.2363G>A p.(R788Q)	202	C, H, S	Stable liver disease Alternative diagnosis: BA Stable post-Kasai
ABCB4	c.2800G>A p.(A934T)	163	C, H, S, ALF	Not available
ABCB4	c.3037A>C p.(=)	214	C, S	No diagnosis.
		69	C, H, S	Asymptomatic
SLC25A13	c.-115G>T	189	C	Asymptomatic
SLC25A13	c.639T>G p.(H213Q)	21	C	Asymptomatic

4.4. Discussion

4.4.1. Incidence of autosomal recessive genetic disorders

A diagnosis of PFIC was made in 8.5% of patients (n=19) and was the most prevalent genetic condition reported in this cohort, with type 2 being the most frequent. PFIC2 is a severe and progressive condition associated with a significant risk of hepatocellular carcinoma at a young age, therefore accurate and early diagnosis of these patients allows appropriate planning for liver transplantation to begin. Identification of only one case of NPC was lower than might have been predicted (Yerushalmi et al., 2002) however it is still a rare cause of infantile liver disease in this cohort. No patients were diagnosed with NPC resulting from *NPC2* mutations or citrin deficiency. This is not unexpected as both are very rare conditions. Twenty four unique variants were identified in this study that have not previously been reported and are highly likely to be disease-causing.

4.4.2. Significance of heterozygous mutations

The combined incidence of the three autosomal recessive conditions is unknown but likely to be in the region of 1 in 100,000. Using the Hardy-Weinberg equilibrium, the expected incidence of mutation carriers in the general population is predicted to be around 1 in 160. Therefore there is an apparent increase in mutation carriers in this cohort (n=20 in 222, or 1 in 11).

There are a number of proposed reasons for this finding of increased incidence of heterozygous gene mutations. Firstly, the possibility that a second mutation in these patients has been missed and they are in fact affected by an autosomal recessive condition was considered. The NGS strategy employed has a sensitivity of over

99%, however promoter regions and intronic regions were not sequenced. In addition, large scale deletions, duplications and rearrangements, would not be detected by sequencing. However, the vast majority of mutations reported in the Human Gene Mutation Database for these genes would have been detected by the NGS screen therefore this risk might contribute to but not fully explain the increased rate of mutation carriers in the cohort. For example, patient 179 presented with cholestasis and acute liver failure and the previously reported missense mutation c.1445A>G p.(D482G) was identified in the *ABCB11* gene (Strautnieks et al., 2008). This genotype does not explain this patient's symptoms and therefore a second pathogenic mutation in *ABCB11* cannot be ruled out. In comparison, patient 193 presented with cholestasis which resolved by fifteen months. The previously reported missense mutation c.1621A>C p.(I541L) was identified, however a second pathogenic mutation in this patient would cause more severe progressive symptoms than was described in this patient.

Secondly, the mutations listed in Table 4.7 were classified as either pathogenic or possibly pathogenic (see 2.2.10), and it is possible that some of those listed as possibly pathogenic may in fact be rare benign variation. The features of liver disease in these patients may be caused by mutations in other genes or other factors such as infection or sensitivity to parenteral nutrition. Again this possibility may contribute to the apparent increased rate of mutation carriers in this cohort, but does not fully explain it as seven of the mutations in Table 4.7 did have significant evidence for pathogenicity in the literature or were truncating mutations.

Thirdly, it is possible that single heterozygous mutations in these genes results in an increased risk of liver disease in infancy. The three genes associated with PFIC are also associated with less severe phenotypes. BRIC is traditionally reported to be

autosomal recessive and associated with missense mutations which have some residual protein function (Davit-Spraul et al., 2009). There is evidence that some BRIC patients carry single heterozygous severe mutations (Kubitz et al., 2006). ICP and LPAC are associated with single heterozygous mutations in the same genes (Davit-Spraul et al., 2009). Both BRIC and ICP are known to result in periods of cholestasis in response to some known and some unknown factors, for example pregnancy, oral contraceptives or other drugs. Evidence suggests that patients with mutations in the PFIC genes have a lower threshold of tolerance to these factors than individuals without mutations. Patients with heterozygous mutations produce enough wild type protein under normal conditions, however when the system is stressed internally or by environmental factors there is a susceptibility to cholestasis. A few individuals with heterozygous mutations in the PFIC genes and transient neonatal cholestasis have been reported, however this study provides more evidence that there is a genetic susceptibility to infantile cholestasis similar to ICP. Long term follow up of individuals with single mutations, and retrospective analysis of neonatal phenotypes in BRIC and ICP patients, is required to determine penetrance and prognosis in these patients.

Of the seven patients identified with mutations in *NPC1* (Table 4.6), three had another explanation for their symptoms (patients 116, 28 and 22). Two patients had severe disease (patients 54 and 157) and two patients had symptoms which resolved (patients 153 and 115). Patient 115 carried a known NPC mutation; therefore this patient would be a good candidate for more in-depth analysis of *NPC1* gene expression and protein levels. This type of work was outside the scope of this study. Single mutations in *NPC1* have not previously been reported to be associated with infantile liver disease, however a comparable study of the *NPC1* and *NPC2*

genes in adults with neurological features identified several patients with single mutations (Bauer et al., 2013). The authors speculated about the possibility of a late-onset reduced penetrance form of NPC. It would certainly be interesting to investigate a cohort of *NPC1* mutations carriers to determine whether there is an increased risk of infantile liver disease or adult-onset neurological or psychological problems.

One patient with a novel heterozygous missense in *SLC25A13* was identified (patient 147). Citrin deficiency would be predicted to resolve in infancy and therefore the significance of this finding is not yet clear.

Four of the patients in this group were subsequently diagnosed with biliary atresia (Table 4.7) and all had liver transplants by the age of two years. Two of the patients were reported to be in intensive care and receiving parenteral nutrition (patients 30 and 22) and both were alive and well at follow-up. The possibility of variants in these genes contributing to the risk of biliary atresia and sensitivity to total parenteral nutrition will be discussed later.

4.4.3. Polygenic and epistatic effects

None of the patients tested were found to carry 'pathogenic' mutations in more than one of the genes tested, however two patients (patient 148 and 185) were identified with novel sequence variants that were classified as 'possibly pathogenic' in more than one gene. Patient 148 was found to be homozygous for the missense variant c.1652C>T p.(P551L) in *ABCB4* and heterozygous for the missense variant c.1097T>G p. (V366G) in *ABCB11*. The former was classified as 'possibly pathogenic' was based upon all three *in silico* protein tools predicting a deleterious effect, the latter variant because two of the three tools predicted pathogenicity.

Since it has not been confirmed that either of these variants are definitely pathogenic, no conclusions about the effects of mutations in these two genes can be drawn. Patient 185 was shown to be heterozygous for the nonsense mutation c.1558A>T p.(R520*) in *ABCB11* and heterozygous for the missense variant c.3317A>G p.(E1106G) in *ABCB4*. The missense variant had not previously been reported to be associated with disease; however it had been detected at a low level in the 1000 Genomes project. The nucleotide and codon are moderately conserved and two of the three *in silico* protein tools predicted it was deleterious to the protein. In addition, two of the 5 *in silico* splicing tools suggested it created a cryptic acceptor site. However, without conclusive evidence to support the missense variant as pathogenic, the significance of this finding is unclear. In conclusion, no evidence of polygenic inheritance or epistatic effects was determined by this study.

4.4.4. Variants of uncertain significance (VUS)

A significant proportion of patients without mutations were identified with variants classified as VUS or possibly benign (19%, n=41). This included seven patients who were found to be heterozygous for the *NPC2* variant c.441+1G>A, none of which had another *NPC1* or *NPC2* variant. Most sequence variants at position +1 would be predicted to disrupt the conserved splice donor site, lead to aberrant mRNA splicing and in accordance, all five *in silico* splicing tools predicted splice site disruption for this variant. However, the downstream effect would be determined by whether or not an alternative splice site is used and whether the resulting transcript maintained the reading frame. Since exon 4 is the penultimate exon of *NPC2*, any in-frame transcripts produced may evade nonsense-mediated decay and result in protein production albeit with an altered C-terminus (see Figure 4.3a). In addition, exon 5 of *NPC2* codes for only four amino acids therefore the protein might tolerate such small

changes in amino acid sequence (see Figure 4.3b). Furthermore, this variant has been identified in apparently normal individuals and although the frequency in this cohort is higher than reported in population studies, it is not statistically significant (see Table 4.8). This evidence suggests therefore, that c.441+1G>A is not a loss of function *NPC2* allele; however mRNA or protein studies would be required to confirm this. In relation to the VUS and likely benign variants listed in Table 5.8, this study has not provided any evidence to support these variants as pathogenic and associated with autosomal recessive disease, however the possibility of reduced penetrance or modifier allele status cannot be excluded.

Figure 4.3a: *NPC2* gene structure

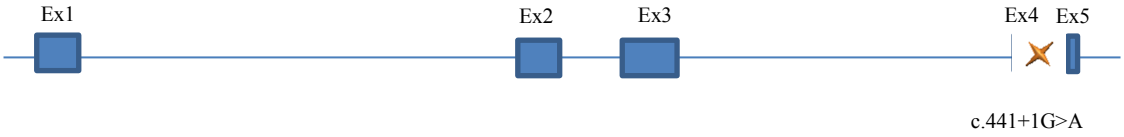


Figure 4.3b: *NPC2* gene sequence

Exons= blue upper case in black box. Intron = black lower case. Sequence variant c.441+1G>A highlighted in orange. Amino acid sequence encoded by exon 5 shown in green bold.

ATAAAACTGGTGGTGGAGTGGCAACTTCAGGATGACAAAAACCAAAGTCTCTTCTGCTGGGAAATCCCAGTACA
GATCgtaagtctatctgggggtgagagggcatgggtggagggaagaaagtggaggagaaatcagactgaaacta
aatcagtgccataagataaaaggaatttcaagactgctgttttatgcctctcaacctctaaaagcatctgagac
cctatttcctgagagctcagaactgttgcctttgtaatatcctttgttaaaaaaaaaaaaggttgaaggaaaga
agagagagtgagactgtagagattgatttaactgtcttcattggtttttttctgtcggagtaaggaaggtccag
ccagacaggacctgaaggaggttgcttaggaatgtctgataacttgccctagggttattgccctgacagtggag
gaaaggttccaaaccacttccactgagctgggacattacatccacagGTTTCTCATCTCTAA
ValSerHisLeu*

Table 4.8: Frequency data for *NPC2* c.441+1G>A

Frequency data and associated p-values (using the Chi-Square statistical test) of *NPC2* c.441+1G>A in various populations including this cohort, the study which first reported the variant (Bauer et al 2013), the 1000Genomes Project Phase 1, the NCBI-GO Exome Sequencing project (ESP) and ClinSeq project. $p<0.05$ was considered significant and $p>0.05$ was not considered significant. In summary, the frequency in this cohort is not significant when compared against the European American ESP population. In addition, the variant was identified in patients with possible adult onset NPC in the Bauer paper, although at a lower frequency. Although ethnicity was not recorded in this study, the country of origin for the majority of patients was within Europe therefore it is likely that the increased frequency reflects the ethnicity of the patients rather than an association with liver disease in infancy.

Cohort	This cohort	Bauer et al 2013	1000 Genomes Project	ESP - African Americans	ESP - European American	ClinSeq
Total number of individuals	222	170	1092	2203	4300	659
Number of heterozygotes	7	3	2	7	76	6
%	3.15	1.76	0.18	0.32	1.77	0.91
p-value		0.39	0	0	0.13	0.02

4.4.5. Patients with no genetic findings

In 62% of patients no genetic findings were identified (n=139) and in 20% of patients only VUS were identified which do not sufficiently explain the clinical presentation (n=44). These patients may have mutations in parts of the genes not tested by this study, or in other genes related to hepatic disease (known or unknown), or may have a non-genetic explanation for their clinical presentation. The contribution of genetic disorders to cholestasis in infancy is not precisely known however biliary atresia is thought to account for around a third of cases, and infectious causes or toxic causes may account for another third. A genetic contribution has been identified in 18% of patients in this study, which is a significant proportion of the possible genetic diagnoses.

4.4.6. Contribution of mutations in patients with other liver conditions

Alagille syndrome (AGS) and alpha-1 anti-trypsin deficiency (AATD) are genetic conditions associated with variable expressivity of liver phenotypes in individuals with the same genotype; both between and within families. The possible contribution of mutations in the six genes studied to the variable expressivity seen in AGS and AATD was evaluated. Four patients were reported to have possible AGS and none had any clearly pathogenic mutations. One had a synonymous change in *NPC1* (patient 105) and one had a missense in *ABCB11* which is not predicted to be damaging to the protein (patient 118, see Table 4.7). Six patients were reported to have AATD and one of those also had a diagnosis of PFIC type 1 (patient 36**)**. No other variants were detected in these patients. The numbers of AGS and AATD cases were too few to draw any meaningful conclusions about the contribution of variants in the six genes to liver disease in these conditions.

Isolated biliary atresia (BA) is described as a progressive obliterative cholangiopathy and is likely to be multifactorial in nature. Variation in the incidence of BA between populations suggest there are susceptibility factors involved however no environmental or genetic link to BA has yet been established. There were forty eight patients in this cohort with biliary atresia (BA) and twelve of these had genetic changes in the genes studied. Two patients were identified who carried single 'possibly pathogenic' mutations in *ATP8B1* (patients 23 and 139). Another ten patients carried single VUS or likely benign variants spread across the six genes.

Seven patients were reported to have been receiving parenteral nutrition (PN). One of these patients was found to carry a single pathogenic mutation in *NPC1* (patient

22), and two patients carried single possibly pathogenic mutations in *ATP8B1* (patients 30 and 33).

4.4.7. Recommendations

Diagnosis of genetic conditions that cause liver disease in infants is important for appropriate management of patients therefore genetic testing (especially for PFIC) should be routinely considered as part of the patient pathway. Next generation sequencing allows multiple genes to be tested simultaneously and helps to reduce genetic testing costs and time to diagnosis. Incorporation of further genes into one test would increase the proportion of diagnoses achievable; including the genes for AGS and AATD.

Genetic testing does not replace other diagnostic methods in most cases and instead should be considered as a complimentary strategy. Correlation of clinical features, biochemical test results and genetic results may be useful, especially for interpretation of ambiguous genetic findings.

Chapter 5 - GENOTYPE-PHENOTYPE CORRELATIONS IN PFIC GENES

5.1. Introduction

5.1.1. Genetic characterisation of PFIC and BRIC

In 1998, Strautnieks et. al. reported that PFIC 2 was caused by mutations in the *ABCB11* gene, which encodes the bile salt export protein (Strautnieks et al., 1998). Around the same time, Bull et. al. reported that mutations in *ATP8B1* cause PFIC1 and BRIC1 (Bull et al., 1998). That year de Vree and colleagues also reported the first human cases of MDR3 deficiency caused by mutations in *ABCB4* and resulting in PFIC3 (De Vree et al., 1998). The genetic cause of BRIC was subsequently expanded to include *ABCB11* (van Mil et al., 2004b). It is generally accepted that PFIC and BRIC are autosomal recessive conditions and are caused by mutations on both alleles of the gene. The difference between them being more severe mutations cause PFIC and milder mutations cause BRIC (Bull et al., 1998), (van Mil et al., 2004b).

In 2003, Mullenbach et al reported a link between *ABCB4* mutations and intrahepatic cholestasis of pregnancy (ICP) and then in 2005 the same group reported screening of the *ATP8B1* gene in British cases of ICP cases (Mullenbach et al., 2003), (Mullenbach, 2005). In 2009, Dixon et al reported that *ABCB11* mutations were also linked to ICP (Dixon et al., 2009). In addition, *ABCB4* mutations have also been linked to transient neonatal cholestasis, adult biliary cirrhosis and LPAC syndrome in small numbers of patients (Jacquemin et al., 2001). The mutation distribution in these genes for the UK population has not been widely reported.

5.1.2. Investigations in subjects suspected of PFIC and related diseases

Mutation detection in subjects suspected of PFIC, and related diseases, was carried out by testing patients referred to the West Midlands Regional Genetics Service between January 2010 and May 2014. Subjects were tested using FS, or tNGS after the assay had been developed in January 2013. Both methods allowed interrogation of the coding regions of the genes plus the intron-exon boundaries allowing the findings to be collated. Sequence variant interpretation was performed on all variants identified except those considered benign polymorphisms. Questionnaires were completed, where possible, by the referring clinicians for patients in whom sequence variants were identified in order to correlate phenotypes with genetic findings.

5.1.3. Chapter 5 Objectives

- To broaden the knowledge of the mutation spectrum of the *ATP8B1*, *ABCB11* and *ABCB4* genes.
- To establish the mutation spectrum of these genes with respect to the UK population.
- To evaluate genotype-phenotype correlations in the PFIC genes.

5.2. Methods

5.2.1. Patient material

Patients suspected of PFIC or related diseases were referred for diagnostic testing; mainly by consultant hepatologists or paediatricians. Consent was taken by the referring clinicians, and this included the anonymous reporting of sequence variants. Samples received were mainly venous blood samples in EDTA or occasionally Lithium Heparin. Some DNA samples were received from outside of the West Midlands region or from other countries. All results were reported back to the patients via the referring clinicians.

5.2.2. DNA extraction

DNA was extracted from blood samples using Gentra Autopure method between January 2010 and February 2013. After that, the lab method switched to the Qiagen Qiasymphony method. Occasionally the phenol chloroform method was used for samples which appeared to be in poor condition, for example if they had a long transit time. Samples of very small volumes or from very young babies were extracted using a manual version of the Autopure method. Quantification of DNA samples was done using the Nanodrop Spectrophotometer. All methods are described in 2.2.1.

5.2.3. PCR primer design

PCR primers were designed for all coding exons of the *ATP8B1*, *ABCB11* and *ABCB4* genes. This was done using the online tools Exon-Primer (UCSC), Primer3 and Primer-Blast (NCBI) using 60^{o}C as the optimal annealing temperature and at least 30 bases of flanking intron each side. All primer pairs were checked using the

NGRL SNPCheck tool to avoid polymorphisms of more than 1% in any population (2.1.14). PCR primer sequences are listed in 2.1.12.1 (Table 2.1).

5.2.4. Analysis methods

All analysis methods are described in Chapter 2. Please see sections 2.2.3.1 for PCR, 2.2.6.1 for FS and 2.2.8.1 for sequencing analysis. FS was used to confirm sequence variants identified using tNGS and to test those exons not covered by the tNGS assay; these were *ATP8B1* exon 18, *ABCB11* exons 3, 6, 22 and 28. Please see sections 2.2.3.4, 2.2.6.3 and 2.2.8.4 for the methods relating to the tNGS assay. These processes were optimised and validated as discussed in Chapter 3. The methods for MLPA are described in section 2.2.7 and GeneMarker analysis in section 2.2.8.2.

5.2.5. Bespoke assay to detect deletion in *ATP8B1*

PCR primers were designed for Long PCR in introns 11 and 14 of *ATP8B1* using Primer3. Long PCR amplification, gel electrophoresis, sequencing and analysis was performed using the methods described Chapter 2 (2.2.3.3).

5.2.6. Variant interpretation

All sequence variants were investigated for potential pathogenicity using the variant interpretation pipeline described in Chapter 2 (2.2.10).

5.2.7. Phenotype questionnaires

Phenotype questionnaires were sent to referring clinicians for all subjects in whom sequence variants were identified in order to gather clinical information. The template is shown in Appendix III.

5.3. Results

5.3.1. Sequence variants detected

One hundred and fifty nine patients with clinical suspicion of an inherited cholestasis disease were tested for mutations in at least one of the *ATP8B1*, *ABCB11* and *ABCB4* genes between 2009 and 2014. At the time of referral, forty five patients were under three years of age, twenty four were between three and twelve years, thirty one were between thirteen and twenty one years and fifty nine were aged twenty two years or over (see Figure 5.1). The majority were referred from centres in the UK, however five were from Ireland, two were from Australia, and single samples were received from Germany, Hong Kong and Jordan. Eighty five of these were tested using the FS method and seventy four using the tNGS method, with FS used to confirm any variants found and to complete any missing exons due to low coverage on the tNGS assay.

Sequence variants were identified in this cohort and classified according to the method described in Chapter 2 (2.2.10). 'Pathogenic' and 'possibly pathogenic' mutations were detected in twenty eight patients (17%) and VUS only were detected in another eighteen patients (11%) meaning a genetic finding was observed in a total of 29% of patients tested. The proportion of patients with genetic variants varied with age (Figure 5.2). This ranged from 16% in those between three and twelve years of age, to 35% in the thirteen to twenty one year olds. Figure 5.3 shows the proportion of such mutations identified per gene screen performed for *ATP8B1*, *ABCB11* and *ABCB4*; this was 10%, 16 and 13%, respectively.

In *ATP8B1*, twelve sequence variants were identified, eighteen were observed in *ABCB11* and sixteen in *ABCB4* (Tables 5.1A-C). The tables include nonsense,

frameshift and splice site (+/- 1 or 2 bases from exon) mutations detected in this cohort, and also missense mutations that had previously been reported in the literature in association with cholestatic conditions. In addition, these tables include sequence variants that were novel and the clinical significance was unclear. In summary, eight nonsense, three frameshift, twenty nine missense, four splice site and two deletions were identified. The tables show the key information gathered in order to classify the sequence variants. Twenty one of these were classified as 'pathogenic', five as 'possibly pathogenic' and twenty as of 'uncertain significance'.

Twenty seven variants were described that had not previously been reported in the literature. Seven novel sequence variants in *ATP8B1* were detected including one nonsense (R271*), one large deletion, two which affect splicing (c.1630+2T>G and c.2707+3T>C), and three missense variants that are of uncertain clinical significance (see Table 5.1C). Eight novel sequence variants in *ABCB11* were also identified; three nonsense mutations (W314*, Q1053*, R1235*) and an in-frame deletion of six amino acids (F383_A389del) as well as four missense variants of uncertain clinical significance. Lastly, twelve novel sequence variants were detected in the *ABCB4* gene. In summary, these were three frameshifts (V336Ffs*5, G362Efs*5 and I980Sfs*19), one nonsense mutation (Q734*), one conserved splice site mutation (c.2211+1G>C) and seven missense variants of uncertain clinical significance.

Figure 5.1: PFIC patients by age at time of referral.

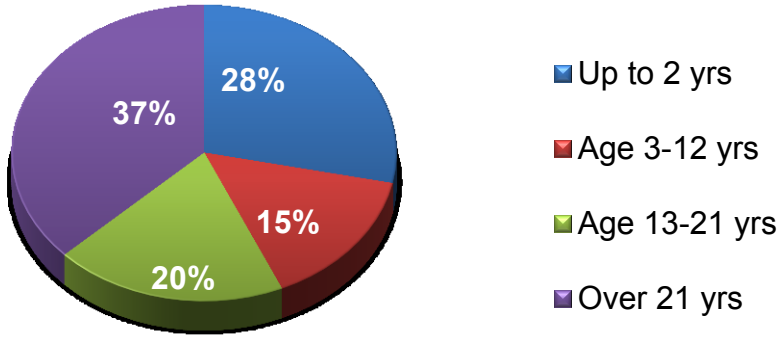


Figure 5.2: Variant detection rate by age group.

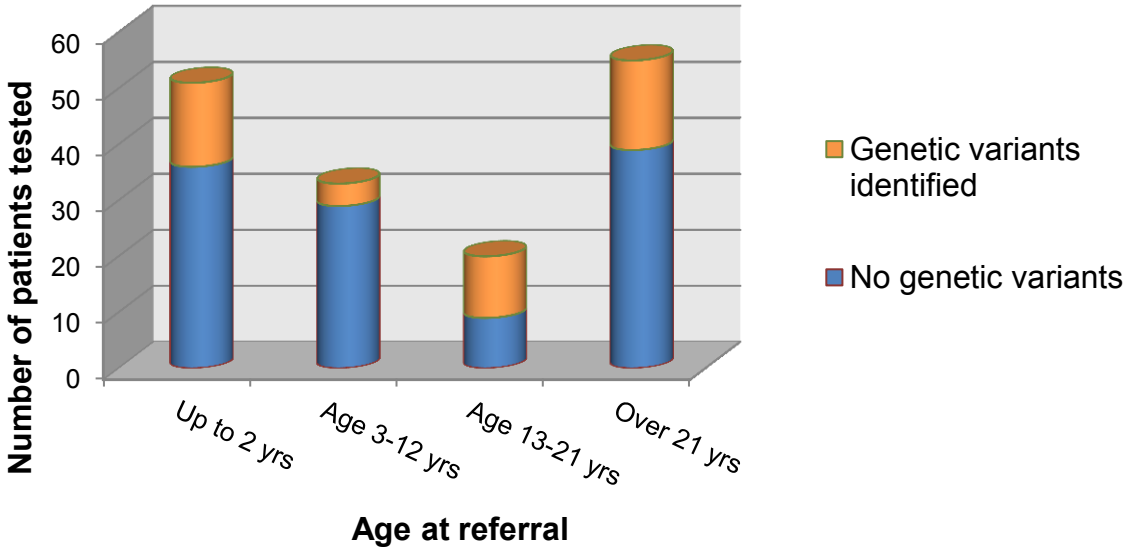


Figure 5.3: Proportion of tests in which genetic variants were found.

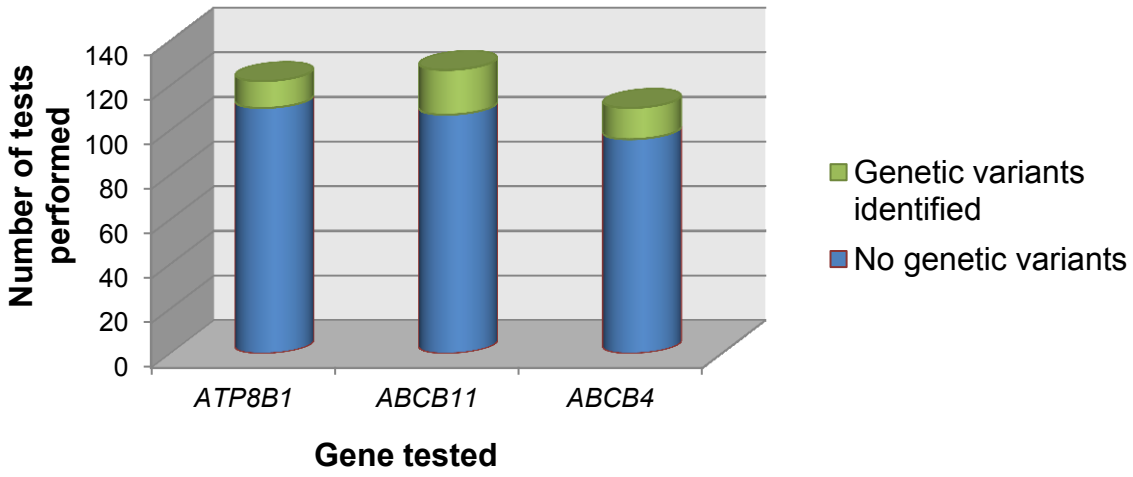


Table 5.1 The sequence variants identified in the cohort, grouped by gene (A=ATP8B1, B=ABCB11 and C=ABCB4).

KEY: DNA change; standard genetic code using single amino acid code and HGVS nomenclature. Type; MS=missense, FS=frameshift, NS=nonsense, SS=splice site, Int=intronic, Del=deletion. Population frequency; 1KG=1000 Genomes project using their three letter code for populations, ESP=exome sequencing project. NA cons=Nucleic acid conservation, AA cons=Amino acid conservation, AA diff=amino acid difference. The terms used were those generated using Alamut software. Protein prediction; AGVGD=Align GVGD, S=SIFT, MT=Mutation Taster, PP=PolyPhen2. Splicing prediction; SSF=Splice site finder, MES=MaxEntScan, NNS=NNSplice, GS=GeneSplicer, HSF=Human Splice Finder.

5.1A *ATP8B1* sequence changes

#	Gene	DNA change	Type	Reference	Population frequency	NA cons AA cons AA diff	Protein prediction	Splicing prediction	Conclusion
P07	*ATP8B1*	c.8C>A p.(T3K)	MS	N	rs202024093 1KG 1 allele	Weak Moderate Moderate	AGVGD C0 S Deleterious MT Polymorphism PP Benign	SSF N MES N NNS N GS N HSF Donor created	Uncertain significance
P42	*ATP8B1*	c.134A>C p.(N45T)	MS	(Painter et al., 2005)	rs146599962 1KG <1% ALL, 1% AMR 8 alleles, ESP 41/13016 alleles	Weak Moderate Small	AGVGD C0 S Deleterious MT Disease-causing PP Possibly damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P43	*ATP8B1*	c.208G>A p.(D70N)	MS	(Klomp et al., 2004), (Mullenbach, 2005), (Folmer et al., 2009), (Strautnieks et al., 2008)	rs34719006 0.7%, 1KG <1% ALL, 1% GBR, 9 alleles, ESP 31/13016 alleles	Moderate High Small	AGVGD C0 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P18	*ATP8B1*	c.601C>A p.(R201S)	MS	N	N	Moderate High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP Probably	SSF N MES N NNS N GS N HSF Acceptor created	Uncertain significance

							damaging		
P44	*ATP8B1*	c.811A/C>T p.(R271*)	NS	N	N	n/a	n/a	n/a	Pathogenic
P18	*ATP8B1*	c.1630+2T>G	SS	N	N	High n/a n/a		SSF Donor destroyed MES Donor destroyed NNS Donor destroyed GS Donor destroyed HSF Donor destroyed	Pathogenic
P09	*ATP8B1*	c.1286A>C p.(E429A)	MS	(Klomp et al., 2004)	rs34018205 1.4%, 1KG 1% of ALL, 6% of LWK, 16 alleles, ESP 25/13016	Weak Moderate Moderate	AGVGD C0 S Tolerated MT Disease-causing PP Benign	SSF N MES N NNS N GS N HSF N	Uncertain significance
P45 P46	*ATP8B1*	c.2498G>A p.(R833Q)	MS	N	N	Weak Moderate Small	AGVGD C0 S Tolerated MT Disease-causing PP Benign	SSF N MES N NNS N GS N HSF Acceptor created	Uncertain significance
P19	*ATP8B1*	c.2707+3G>C	SS	N	N	Moderate n/a n.a		SSF N MES Donor reduced NNS Donor destroyed GS Donor reduced HSF N	Possibly pathogenic
P20	*ATP8B1*	c.2941G>A p.(E981K)	MS	(Numakura et al., 2011), (Stone et al., 2012)	N	High High Small	AGVGD C55 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Pathogenic
P10	*ATP8B1*	c.3040C>T p.(R1014*)	NS	(Klomp et al., 2004)	N	n/a	n/a	n/a	Pathogenic
P08	*ATP8B1*	Del Ex12-14	Del	N	N	n/a	n/a	n/a	Pathogenic

5.1B *ABCB11* sequence changes

#	Gene	DNA change	Type	Reference	Population frequency	NA cons AA cons AA diff	Protein prediction	Splicing prediction	Conclusion
P04	*ABCB11*	c.149T>C p.(L50S)	MS	(Strautnieks et al., 2008), (Byrne et al., 2009)	N	Moderate Moderate Large	AGVGD C0 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Pathogenic
P11	*ABCB11*	c.149T>G p.(L50W)	MS	N	N	Moderate Moderate Small	AGVGD C0 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P01	*ABCB11*	c.673A>C p.(T225P)	MS	(Strautnieks et al., 1998)(Jirsa, 2011)	N	Weak High Small	AGVGD C0 S Deleterious MT Disease-causing PP	SSF N MES N NNS N GS N HSF N	Possibly pathogenic (personal communication)
P04, P02, P21	*ABCB11*	c.890A>G p.(E297G)	MS	(Strautnieks et al., 1998), (Wang et al., 2002)	rs11568372 PharmaGKB, 1/145 alleles	High High Moderate	AGVGD C0 S Deleterious MT Disease-causing PP	SSF N MES N NNS N GS N HSF N	Pathogenic
P03	*ABCB11*	c.908+1G>A	SS	(Strautnieks et al., 2008), (Tibesar et al., 2014)	rs147649016 1KG 2 alleles TSI	High n/a n/a	n/a	SSF Donor destroyed MES Donor destroyed NNS Donor destroyed GS Donor destroyed HSF Donor destroyed	Pathogenic
P22	*ABCB11*	c.942G>A p.(W314*)	NS	N	N	n/a	n/a	n/a	Pathogenic

P23, P24	*ABCB11*	c.1146_1166del21 p.(F383_A389del)	Del	N	N	n/a	n/a	n/a	Possibly pathogenic
P24	*ABCB11*	c.1409G>A p.(R470Q)	MS	(Strautnieks et al., 2008), (Byrne et al., 2009)	N	High High Small	AGVGD C35 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Pathogenic
P01	ABCB11	c.1724G>A p.(R575Q)	MS	N	rs200667815 No freq data	High High Small	AGVGD C0 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P12	ABCB11	c.1739T>C p.(L580P)	MS	N	N	High High Moderate	AGVGD C0 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P26, P27, P13	*ABCB11*	c.2093G>A p.(R698H)	MS	(Anzivino et al., 2013), (El Sherrif et al., 2013)	rs138642043 1%, 1KG <1% of ALL 1% of AFR and GBR, 14 alleles, ESP 67/13016	High Moderate Small	AGVGD C0 S Deleterious MT Disease-causing PP Benign	SSF N MES N NNS N GS N HSF N	Uncertain significance

P06	*ABCB11*	c.3148C>T p.(R1050C)	MS	(van Mil et al., 2004b), (Lam et al., 2007), (Byrne et al., 2009)	rs72549398 No freq data	High High Large	AGVGD C15 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Pathogenic
P28	*ABCB11*	c.3157C>T p.(Q1053*)	NS	N	N	n/a	n/a	n/a	Pathogenic
P29	*ABCB11*	c.3382C>T p.(R1128C)	MS	(Strautnieks et al., 2008), (Byrne et al., 2009)	N	Weak High Large	AGVGD C25 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Pathogenic
P14	*ABCB11*	c.3634G>T p.(V1212F)	MS	(Evason et al., 2011)	N	High Moderate Small	AGVGD C0 S Deleterious MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Possibly pathogenic
P15, P02	*ABCB11*	c.3669G>C p.(E1223D)	MS	(Byrne et al., 2009)	rs199649780 ESP 5/13016 alleles, AFR	Weak Weak Small	AGVGD C0 S Tolerated MT Disease-causing PP Possibly damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P03	*ABCB11*	c.3703C>T p.(R1235*)	NS	N	N	n/a	n/a	n/a	Pathogenic
P11	*ABCB11*	c.3962G>A p.(S1321N)	MS	N	rs201693189 0.4%, 1KG <1% 2 alleles, ClinSeq 1 allele	High Moderate Small	AGVGD C0 S Tolerated MT Disease-causing PP Probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance

5.1C *ABCB4* sequence changes

#	Gene	DNA change	Type	Literature	Population frequency	NA/AA/AA diff	Protein prediction	Splicing prediction	Classification
P16	*ABCB4*	c.215C>G p.(P72R)	MS	N	N	High High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P30	*ABCB4*	c.551T>C p.(I184T)	MS	N	N	High High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P31	*ABCB4*	c.1006_1009delGTTT p.(V336Ffs*5)	FS	N	N	n/a	n/a	n/a	Pathogenic
P32	*ABCB4*	c.1085delG p.(G362Efs*5)	FS	N	N	n/a	n/a	n/a	Pathogenic
P33	*ABCB4*	c.1505T>C p.(I502T)	MS	N	N	Weak High Moderate	AGVGD C0 S Deleterious MT Disease-causing PP possibly damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P34	*ABCB4*	c.1529A>G p.(N510S)	MS	(Anzivino et al., 2013), (Poupon et al., 2013)	N	Weak High Small	AGVGD C0 S Deleterious MT Disease-causing PP possibly damaging	SSF N MES N NNS N GS N HSF N	Possibly pathogenic
P35	*ABCB4*	c.1546A>G p.(M516V)	MS	N	N	Weak High Small	AGVGD C0 S Deleterious MT Disease-causing PP benign	SSF N MES N NNS N GS N HSF Donor creation	Uncertain significance

P38	*ABCB4*	c.1769G>A p.(R590Q)	MS	(Degiorgio et al., 2007), (Colombo et al., 2011), (Zimmer et al., 2012), (Poupon et al., 2013)	rs45575636 DB 0.9%, 1KG <1% of ALL, 4% LWK, 1% GBR, 22 alleles, ESP 37/4406 alleles	High High Small	AGVGD C0 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P36	*ABCB4*	c.1783C>T p.(R595*)	NS	(Davit-Spraul et al., 2010)	N	Weak High n/a	n/a	n/a	Pathogenic
P37	*ABCB4*	c.1801G>A p.(A601S)	MS	N	N	High High Small	AGVGD C0 S Deleterious MT Disease-causing PP possibly damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P16	*ABCB4*	c.1801G>T p.(A601T)	MS	N	N	High High Moderate	AGVGD C0 S Deleterious MT Disease-causing PP possibly damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance
P17	*ABCB4*	c.2200G>T p.(Q734*)	NS	N	N	Moderate Moderate n/a	n/a	n/a	Pathogenic
P38	*ABCB4*	c.2211+1G>C	SS	N	N	n/a	n/a	SSF Donor destroyed MES Donor destroyed NNS Donor destroyed GS Donor destroyed HSF Donor destroyed	Pathogenic
P39	*ABCB4*	c.2869C>T p.(R957*)	NS	(De Vree et al., 1998)	rs121918440 No freq data	Weak Moderate n/a	n/a	n/a	Pathogenic
P40	*ABCB4*	c.2939_2940delTT p.(I980Sfs*19)	FS	N	N	n/a	n/a	n/a	Pathogenic

P41	*ABCB4*	c.3011C>A p.(A1004E)	MS	N	N	High High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP probably damaging	SSF N MES N NNS N GS N HSF N	Uncertain significance

5.3.2. Multiplex-ligation dependent probe amplification

MDR3 deficiency may be caused by biallelic *ABCB4* mutations, and single heterozygous mutations in *ABCB4* were identified in several patients (see Table 5.2C). Therefore, there was a possibility that a second mutation in *ABCB4* was present in some patients that was not being detected by the sequencing methods employed. In order to reduce the potential proportion of mutations being undetected, patients were tested using MLPA (MRC-Holland). This technique is able to detect homozygous and heterozygous deletions and duplications of whole exons or regions of the gene (Schouten et al., 2002). No additional mutations of this type were identified in *ABCB4* in this cohort. No MLPA kits are currently available for *ATP8B1* or *ABCB11*.

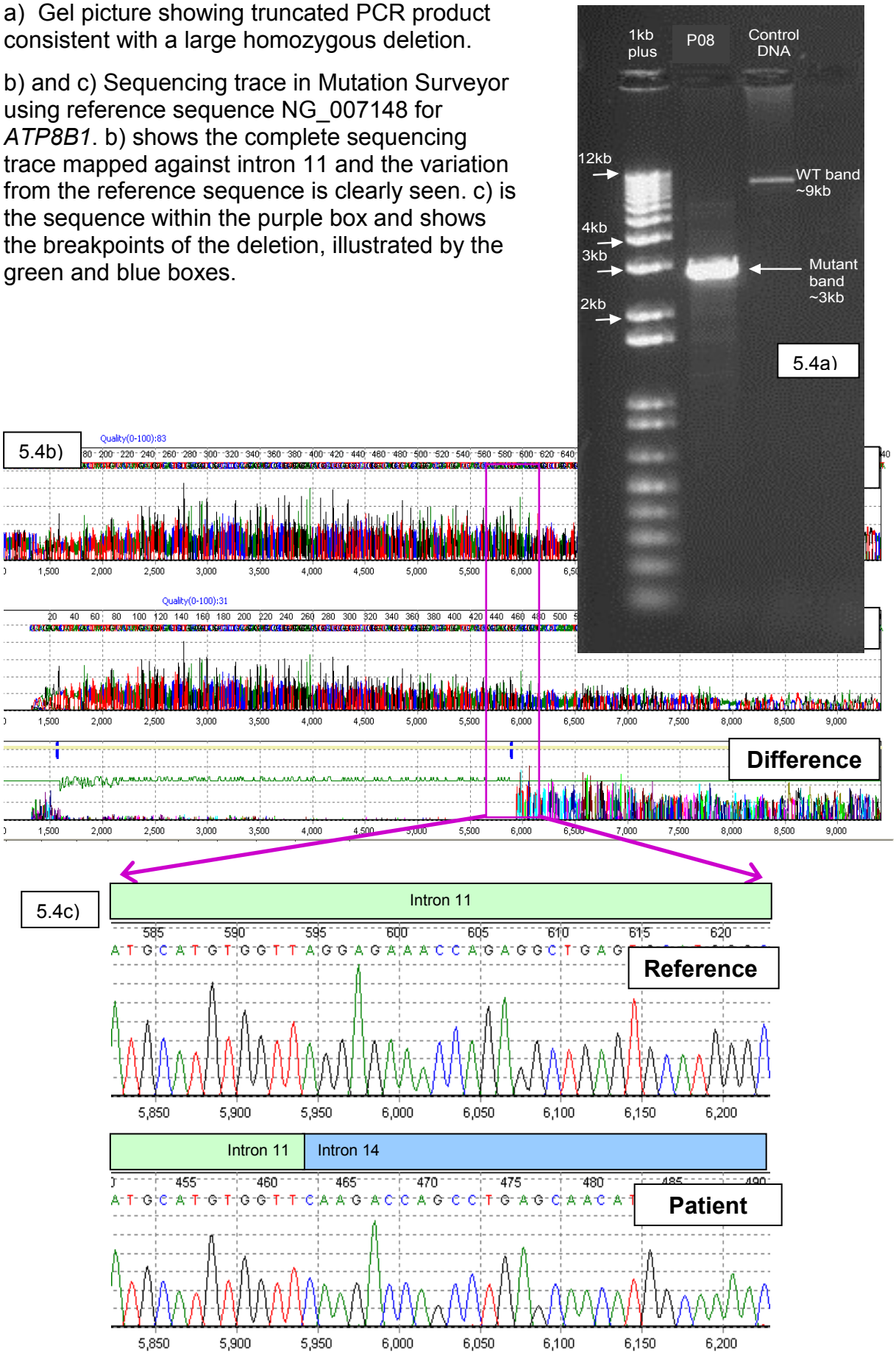
5.3.3. *ATP8B1* deletion case

Three consecutive PCR fragments (exons 12, 13 and 14) of *ATP8B1* failed to amplify in patient P08, therefore the possibility of a large homozygous deletion was investigated. Long range PCR primers were designed in introns 11 and 14 to amplify the region and gel electrophoresis of the PCR products showed a band shift in the patient's sample, consistent with a deletion of at least 5 kilobases (see Figure 5.4a). Sequencing of the PCR products showed the breakpoints of the deletion and allowed the deletion to be sized at 5948 base pairs and characterised as c.1221-490_1630+986del (see Figure 5.4b and 5.4c). Deletion of exons 12-14 is predicted to result in the introduction of a premature stop codon p.(S544Rfs*13) and is therefore predicted to be pathogenic.

Figure 5.4. Characterisation of *ATP8B1* mutation in patient P08.

a) Gel picture showing truncated PCR product consistent with a large homozygous deletion.

b) and c) Sequencing trace in Mutation Surveyor using reference sequence NG_007148 for *ATP8B1*. b) shows the complete sequencing trace mapped against intron 11 and the variation from the reference sequence is clearly seen. c) is the sequence within the purple box and shows the breakpoints of the deletion, illustrated by the green and blue boxes.



5.3.4. Mutation spectrum in UK population

Of the forty six patients with sequence variants identified, forty two of them were from within the UK. Consequently, forty two of the forty six sequence variants identified were present in patients from the UK; including twenty two novel variants (see Tables 5.1 A-C).

5.3.6. Categorisation of patients

The clinical features of the patients in whom pathogenic, possibly pathogenic and unclassified variants were detected are summarised in Tables 5.2A-C. The patients are categorised in sections 5.3.6 to 5.3.11 according to the number and classification of variants detected in order to group patients into useful categories for genotype-phenotype correlations. Section 5.3.6 reviews those patients in which two 'pathogenic' or 'possibly pathogenic' mutations were identified. In contrast section 5.3.7 includes those patients in which two sequence variants were detected, but one was classified as 'uncertain significance'. These patients therefore have a possible diagnosis of PFIC but with a degree of uncertainty. 5.3.8, 5.3.9 and 5.3.10 include those patients in which one pathogenic or possibly pathogenic mutation was identified. Lastly, patients with VUS only are reviewed in section 5.3.11.

5.3.7. PFIC diagnoses

Four patients were diagnosed with PFIC1; P08, P10, P19, P20 and P18 (Table 5.2A). The PFIC1 patients were found, respectively, to have a large homozygous deletion of three exons (section 5.3.3), a homozygous nonsense mutation, a homozygous intronic sequence change that is highly likely to cause aberrant mRNA splicing, and a homozygous missense change that has previously been reported to result in an 80% reduction in enzyme activity (Numakura et al., 2011). The median

age of onset in this group was five months (mean five months, range from one to nine months) and all four patients presented with progressive cholestasis. Two of the patients were reported to have metabolic bone disease which is likely to have resulted from decreased ability to absorb vitamins from the gut (Chen and Chang, 2004). P10 had developmental delay and dysmorphic features, which are unlikely to be related to PFIC1.

Six patients were diagnosed with PFIC2; P03, P04, P05, P24, P28, P29 (Table 5.2B). The median age of onset for these patients was 6 months (range two weeks to 2 years). No further clinical information was available for two of these patients, however progressive cholestasis was noted in the other four.

No patients were diagnosed with PFIC3, which usually accounts for one third of PFIC cases (Davit-Spraul 2009). This finding may reflect the fact that the low serum GGT levels associated with PFIC types 1 and 2 is a more distinct phenotype than the raised serum GGT levels associated with PFIC3, and therefore genetic testing is more often considered by clinicians for these patients.

5.3.8. Possible diagnoses

Two patients were found to carry two sequence variants in *ABCB11* (P01 and P02) but there was uncertainty about the pathological significance of one of them in each case. In addition, both patients were reported to have had an episode of cholestasis rather than progressive cholestasis, further emphasising that the contribution, if any, of the missense variants E1223D and R575Q to BSEP deficiency is unclear. Patient P01 showed presence of MDR3 and BSEP immunostaining, a feature which is not inconsistent with missense mutations in *ABCB11*, but that had directed the clinicians away from a diagnosis of BSEP deficiency. This patient also had the E297G

mutation in ABCB11, a mutation which has previously been associated with PFIC2 and BRIC2 cases (van Mil et al., 2004b).

Patients P16, P35 and P41 had two *ABCB4* variants each. P16 had a more severe and complicated presentation than the other patients, and both variants were in regions highly conserved at the nucleotide and amino acid level, suggesting that they might be contributing to her phenotype. However, no information was known about the health of her relatives and therefore the clinical significance of her genotype is unclear. Reduced MDR3 levels by immunostaining of liver biopsy would support the hypothesis of MDR3 deficiency and UDCA therapy would be indicated. P35 was homozygous for the missense M516V and she presented with progressive cholestasis. Cirrhosis and a relatively early age of onset indicate a more severe presentation than those patients with single heterozygous mutations (section 5.3.10). In contrast, P41 was homozygous for the missense A1004E and had clinical features similar to the patients with the single, severe, heterozygous mutations in *ABCB4*; suggesting that this missense might result in reduced rather than absence of MDR3 protein function. All four of these variants were missense changes with no previous reports in the literature or in normal populations and so have been classified as of uncertain significance (P72R, A601T, M516V and A1004E; see Table 5.1C). If these variants are seen again in other patients it might help to clarify the significance of these variants in future.

5.3.9. Cases with single heterozygous mutations in *ATP8B1*

Patient P18 was an unusual case as he presented at age fifteen with unexplained cholestasis, pruritus and hepatomegaly. He had one pathogenic splice site mutation (c.1630+2T>G) and one novel missense change c.601C>A p.(R201S) of uncertain clinical significance. His parents were tested and the sequence variants were shown

to be on different alleles of *ATP8B1*. It was not clear whether the missense sequence change was contributing to this patient's phenotype; however these results confirmed the clinician's diagnosis of BRIC.

One patient was identified with a single heterozygous *ATP8B1* mutation. P44 was female and presented at age 35 with an unexplained episode of cholestasis, pruritus and jaundice.

5.3.10. Cases with single heterozygous mutations in *ABCB11*

Six patients were identified who carried single heterozygous mutations (pathogenic or possibly pathogenic) in *ABCB11* (P06, P21, P22, P25, P14 and P23). This group all presented with episodic cholestasis at a median age of sixteen years (range two weeks to eighteen years, mean thirteen years). Patient P06 presented much earlier than the other five patients suggesting that the presence of one loss of function mutation in *ABCB11* was not the only contributory factor in this patient's illness, and perhaps additional genetic or environmental factors played a role. The future health of this child, including any subsequent episodes of cholestasis, may help to shed light on whether this is typical BRIC2 which happened to present initially in the newborn period, or whether the child has BSEP deficiency which is more progressive and severe than the *ABCB11* genotype indicates. In three out of five patients with single heterozygous *ABCB11* mutations, use of oral contraceptives was identified as the trigger for the cholestasis episodes.

5.3.11. Cases with single heterozygous mutations in *ABCB4*

Eight patients (P31, P32, P36, P34, P38, P17, P39 and P40) were identified who were heterozygous for a 'pathogenic' or possibly pathogenic' mutation in *ABCB4* (see Table 3.2C). The median age of onset of symptoms in this group was age

twenty five years (mean twenty four years, min ten years, max thirty five years). All had episodes of acute cholestasis, although in two it was recorded that they had signs of underlying progressive disease (P31 and P38). Pregnancy and the oral contraceptive pill were identified as triggers for the episodes in five and one patients respectively; however no trigger was identified in the other three. Gallstones were the most commonly associated feature, seen in three patients, and cholecystectomy had been carried out in another two (whilst the reason for the procedure wasn't recorded, again gallstones is the most likely explanation). A liver biopsy was carried out in seven of the patients and six showed mild biliary changes, two had fibrosis and one ductule proliferation. In one case, ductopenia was the only feature noted (P40). Patient P17 had an unusual presentation due to a significantly earlier age of onset (10 years), and severe fibrosis seen on liver biopsy. It was noted that her mother had ICP which would be consistent with her also being a carrier of the same *ABCB4* mutation. No further mutations were identified in P17 in *ATP8B1*, *ABCB11* or *ABCB4*; however the possibility of other genetic or environmental factors contributing to her phenotype seems likely.

All of the patients with *ABCB4* mutations and VUS were female. This is not surprising since they were mainly were young adults and had come to the attention of liver specialists following episodes of cholestasis during pregnancy or in response to taking oral contraception. All eight patients had a family history of related conditions, and of those with the relationships recorded, all were first degree relatives.

5.3.12. Patients with VUS

Several patients were found to be heterozygous for missense variants of uncertain clinical significance in *ATP8B1* (P07, P09, P42, P43, P45 and P46), *ABCB11* (P12, P13, P15, P26 and P27) and *ABCB4* (P16, P35, P41, P30, P33 and P37). T3K and R833Q in *ATP8B1* have not previously been reported in the literature but *in silico* tools indicate these are weakly conserved nucleotides and are not likely to be damaging to the FIC1 proteins. N45T and D70N have been reported in the literature as causing FIC1-associated disease however they are also present at significant rates in apparently healthy populations (see Table 5.1A). E429A was reported by Klomp et al in 2004 as being of uncertain clinical significance; however since then it has also been identified in normal individuals in the 1000Genomes project. This study has not provided any evidence that these three missenses are capable of leading to severe FIC1 deficiency as they were not detected in any PFIC1 cases; however it remains uncertain whether they can contribute to adult liver disease as observed in P42 and P43.

In *ABCB11*, R698H has been reported as being of uncertain significance (Anzivino et al., 2013) and it is also present in normal individuals in several populations, indicating this is unlikely to be responsible for the phenotypes in patients P13, P26 and P27. Also in *ABCB11*, the E1223D missense was identified in P15. This has been associated with prolonged neonatal hepatitis (Byrne et al., 2009), but has since also been found in normal African American individuals at a low frequency. Clinical details were not reported in P15 therefore it is difficult to draw any conclusions about the effect of this missense. The L580P missense has not previously been reported in the literature or in population frequency databases; *in silico* tools suggest it may have an effect on BSEP however it is not possible to be certain based on those

observations alone. It is worth noting that patients P07 and P09 presented as neonates, and P13 as a young child, and the single VUS identified do not explain their phenotypes. It is likely that another genetic or non-genetic cause is responsible.

Three patients (P30, P33 and P37) were found to be heterozygous for novel missense variants in *ABCB4*; all of which protein prediction tools suggested were damaging to the MDR3 protein. P33 presented age thirty with three episodes of cholestasis triggered by pregnancy. Her liver biopsy showed mild biliary changes and her grandfather had cirrhosis. P37 also presented during pregnancy and had a personal and family history of gallstones. There was no clinical information available for P30. Hopefully, detection of these variants in affected patients or normal populations in future will help to clarify the significance of these variants in liver disease.

Table 5.2 Clinical features of patients with sequence variants in A - *ATP8B1,* B - *ABCB11* and C - *ABCB4.*

A – Patients with *ATP8B1* variants

#	DNA change	Classification	Sex (Country)	Age at presentation	Cholestasis	If Episodic, frequency and triggers	Other features	Liver biopsy	Family history
P08	Del Ex12-15 Hom	Pathogenic	F (UK)	5 months	Progressive		Abnormal liver enzyme levels, metabolic bone disease, liver failure	Inflammation and cholestasis	Not reported
P10	c.3040C>T p.(R1014*) Hom	Pathogenic	M (UK)	9 months	Progressive		Metabolic bone disease, pruritus, developmental delay, dysmorphic appearance, progressive diarrhoea and vomitting, deranged clotting	Cholestasis and fibrosis	Y possibly; a deceased uncle
P19	c.2707+3G>C Hom	Possibly pathogenic	F (UK)	4 weeks	Progressive		None	Non-necrotising granulomas, paucity of bile ducts	N
P20	c.2941G>A p.(E981K) Hom	Pathogenic	M (Ireland)	6 months	Progressive		Migraines	Cholestasis and fibrosis	N
P18	c.601C>A p.(R201S) c.1630+2T>G	Possibly pathogenic Pathogenic	M (UK)	15 years	Episode	1, not identified	Pruritis, hepatomegaly, finger clubbing, mild hypothyroidism	Inflammation and cholestasis	N
P44	c.811A/C>T p.(R271*)	Pathogenic	F (UK)	35 years	Episode	1, not identified	Jaundice, pruritus	Not done	N
P07	c.8C>A p.(T3K)	Uncertain significance	M (UK)	3 weeks	Episode	1, newborn period	None	Fibrosis, ductopenia	N
P09	c.1286A>C p.(E429A)	Uncertain significance	M (UK)	8 weeks	Progressive	1, newborn period	Developed giant cell hepatitis, small bile ducts, extra digit, encephalopathy, liver failure	Neonatal hepatitis, no bile plugging, ductular proliferation	N

P42	c.134A>C p.(N45T)	Uncertain significance	F (UK)	39 years	Episode	6, during all pregnancies and with oral contraceptive pill	N	Inflammation, cholestasis, fibrosis	N
P43	c.208G>A p.(D70N)	Uncertain significance	F (UK)	Adult	Episode	1, pregnancy	Pain	Not done	N
P45	c.2498G>A p.(R833Q)	Uncertain significance	F (UK)	29 years	Slowly progressive		Raised GGT levels, colitis	Mild changes only	N
P46	c.2498G>A p.(R833Q)	Uncertain significance	F (UK)	24 years	Episode	2, pregnancy	Epilepsy, vitamin deficiency, pruritus	Mild changes only	N

B-Patients with *ABCB11* variants

#	DNA change	Classification	Sex	Age at presentation	Cholestasis	If Episodic, frequency and triggers	Other features	Liver biopsy	Family history
P03	c.908+1G>A c.3703C>T p.(R1235*)	Pathogenic Pathogenic	F (UK)	6 months	Not reported	Not reported	Not reported	Not reported	Not reported
P04	c.149T>C p.(L50S) c.890A>G p.(E297G)	Pathogenic Pathogenic	F (UK)	9 months	Progressive		Diarrhoea, ricketts, intermittant pale stools, pruritus	N	N
P05	c.2012-8T>G Hom	Pathogenic	M (UK)	6 months	Progressive		Pruritus, coagulopathy	Fibrosis, giant cell hepatitis, paucity of bile ducts	N
P24	c.1409G>A p.(R470Q) c.1146_1166del21 p.(F383_A389del)	Pathogenic Possibly pathogenic	M (UK)	2 years	Progressive		Liver transplant at 5y	N	N
P28	c.3157C>T p.(Q1053*) Hom	Pathogenic	M (UK)	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
P29	c.3382C>T p.(R1128C) Hom	Pathogenic	M (UK)	13 days	Progressive		Bleeding, pale stools, failure to thrive	N	Yes, 3 siblings
P02	c.890A>G p.(E297G) c.3669G>C p(E1223D)	Pathogenic Uncertain significance	F (Ireland)	6 months	Episode	1, gallstones	None	Non-specific, portal tract inflammation,	N

								rosettes, bile plugs, pericellular cytokeratin, loss of biliary staining	
P01	c.673A>C p.(T225P) c.1724G>A p.(R575Q)	Possibly pathogenic Uncertain significance	F (Australia)	5 months	Episode, but signs of underlying progressive features	1, infection	None	Y, immunostaining showed BSEP and MDR3 present	N
P11	c.149T>G p.(L50W)/c.3962G>A p.(S1321N) Hom	Uncertain significance	F (UK)	13 months	Progressive		Pruritus, hepatosplenomegaly	Distorted architecture, fibrosis, no ductule proliferation, hepatitis, marked cholestasis.	N
P06	c.3148C>T p.(R1050C)	Pathogenic	M (UK)	2 weeks	Episode	1, newborn period	None	Severe cholestasis, resetting. Immunostaining did not confirm PFIC	N
P21	c.890A>G p.(E297G)	Pathogenic	F (UK)	18 years	Episode	Not reported	Not reported	Not reported	Not reported
P22	c.942G>A p.(W314*)	Pathogenic	F (UK)	16 years	Episode	1, not identified	Weight loss, pruritus	Cholestasis	N
P25	c.2012-8T>G	Pathogenic	F (UK)	17 years	Episode	1, oral contraceptive pill	Low serum GGT, jaundice	N	Y, mother
P14	c.3634G>T p.(V1212F)	Possibly pathogenic	F (UK)	14 years	Episode	1, oral contraceptive pill	Pain, pruritus, vomiting, weight loss, lethargy	N	N
P23	c.1146_1166del21 p.(F383_A389del)	Possibly pathogenic	F (UK)	17 years	Episode	1, Norethisterone	Jaundice, pruritus	N	N
P12	c.1739T>C p.(L580P)	Uncertain significance	M (UK)	6 years	Episode	1, gallstones	Gallstones	Fibrosis and cholestasis	N

P13	c.2093G>A p.(R698H)	Uncertain significance	F (UK)	5 years	Episode	1, gallstones	Gallstones, pruritus	N	N
P15	c.3669G>C p.(E1223D)	Uncertain significance	F (UK)	16 years	Not reported	Not reported	Not reported	Not reported	Not reported
P26	c.2093G>A p.(R698H)	Uncertain significance	F (UK)	33 years	Episode but underlying progressive features	1, contraceptive implant following normal pregnancy	Gallstones	Cholestasis	N
P27	c.2093G>A p.(R698H)	Uncertain significance	F (UK)	38 years	Episode	1, not known	Pain	Loss of BSEP polarity and MDR3 staining	N

C-Patients with *ABCB4* variants

#	DNA change	Classification	Sex	Age at presentation	Cholestasis	If Episodic, frequency and triggers	Other features	Liver biopsy	Family history
P31	c.1006_1009delGTTT p.(V336Ffs*6)	Pathogenic	F (UK)	34 years	Acute episode but underlying progressive features	1, pregnancy	Abnormal liver enzyme results, pain, gallstones, cholecystectomy	Biliary disease	Y, father and paternal grandfather
P32	c.1085delG p.(G362Efs*6)	Pathogenic	F (UK)	30 years	Episode	1, oral contraceptive pill	Acute severe hepatitis, gallstone	Not done	Y
P36	c.1783C>T p.(R595*)	Pathogenic	F (UK)	31 years	Episode	2, pregnancy	None	Mild biliary changes	Y, mother
P38	c.2211+1G>C	Pathogenic	F (UK)	30 years	Acute episode but underlying progressive features	1, not identified	Abnormal liver enzyme results, cholecystectomy	Biliary disease, moderate fibrosis	Y, mother
P17	c.2200G>T p.(Q734*)	Pathogenic	F (UK)	10 years	Episode	1, not identified	Pruritus	Severe fibrosis of ductular proliferation	Y, mother

P39	c.2869C>T p.(R957*)	Pathogenic	F (UK)	40 years	Episode	1, pregnancy	Abnormal liver enzyme results, jaundice cholecystectomy	Biliary disease	Y, mother and maternal grandfather
P40	c.2939_2940delTT p.(I980Sfs*20)	Pathogenic	F (UK)	27 years	Episode	1, pregnancy	Abnormal liver enzyme results, gallstones	Ductopenia	Y, mother and sibling
P34	c.1529A>G p.(N510S)	Possibly pathogenic	F (UK)	35 years	Episode	2, pregnancy and post-pregnancy	Pruritus	Mild biliary changes	N
P16	c.215C>G p.(P72R) c.1801G>A p.(A601T) Cmpd Het	Uncertain significance Uncertain significance	F (Australia)	7 years	Slowly progressive	Deterioration following variceal haemorrhage	Fever, lethargy, hepatosplenomegaly, portal hypertension, oesophageal varices, coagulopathy, malnutrition	Chronic liver disease and cirrhosis, copper accumulation.	Not reported
P35	c.1546A>G p.(M516V) Hom	Uncertain significance	F (UK)	19 years	Slowly progressive		Biliary cirrhosis with varices, some pruritus	Moderate biliary fibrosis	N
P41	c.3011C>A p.(A1004E) Hom	Uncertain significance	F (UK)	26 years	Episode	1, pregnancy	Gallstones, cholecystectomy	Chronic inflammation	Y, sibling
P30	c.551T>C p.(I184T)	Uncertain significance	F (UK)	10 years	Not reported	Not reported	Not reported	Not reported	Not reported
P37	c.1801G>T p.(A601S)	Uncertain significance	F (UK)	32 years	Episode	1, pregnancy	Gallstones, cholecystectomy	Not done	Y, aunt
P33	c.1505T>C p.(I502T)	Uncertain significance	F (UK)	30 years	Episode	3, pregnancy	Abnormal liver enzymes	Mild biliary changes	Y, grandfather

5.4. Discussion

5.4.1. General observations

Sequencing of the coding regions of these three genes known to be involved in cholestasis yielded a higher pickup rate in the over age twenty one group. This probably reflects better patient selection in the non-urgent setting. As previously discussed, cholestasis in infancy is much more heterogeneous in origin (Introduction section 1.3).

Routine testing of patients suspected of PFIC and related conditions has detected twenty seven novel mutations, reflecting both that many families harbour discrete mutations and that studies of the UK population are not widely reported in the literature. Although, dozens of mutations in these genes are known about, this study has uncovered several more. In addition, twenty two of these novel sequence changes were found in UK patients. This cohort of patients with sequence variants is larger than any previously reported UK cohort and thus is a significant contribution to the mutation spectrum of these genes.

The remainder of the patients where no mutations were identified may have large exonic deletions or insertions, or deep intronic and promoter mutations, not detectable by the assays employed here. MLPA testing in our *ABCB4* cohort did not yield any further mutations, however the presence of a homozygous deletion in *ATP8B1* was detected in patient P08. Furthermore, it is possible that other genes may be responsible for cholestasis in the remainder of cases (Jacquemin, 2012).

5.4.2. PFIC-related deficiencies

Diagnosis of the four patients with PFIC1 is likely to have been useful for clinicians for prognosis and management; one case was thought to be a lysosomal storage disorder prior to testing. UCDA therapy and PEBD may help reduce symptoms of cholestasis; however it is not usually effective in all cases (Jacquemin, 2012). Liver transplantation is an option, however extrahepatic features such as diarrhoea, may increase in severity following surgery therefore, PFIC1 patients are sometimes considered not suitable for transplant (Jacquemin, 2012). The patient diagnosed with BRIC1 and the patient with a single *ATP8B1* mutation can now be monitored for signs of progressive liver disease, as more becomes known in future about the disease course associated with such genotypes.

A diagnosis of PFIC2, in contrast to PFIC1, would indicate liver transplantation. UCDA therapy and PEBD surgery might be successful in treating the cholestasis; however there is a high risk of liver cancer in early life in PFIC2 infants, therefore liver transplantation should be recommended (Davit-Spraul et al., 2009). In addition, the extrahepatic complications found in PFIC1 patients following transplant are not observed in PFIC2. A diagnosis of BSEP deficiency in the patient, who presented with an episode of cholestasis in the neonatal period, would be useful to allow close monitoring for signs of progressive liver disease in the future.

The cohort of patients with single *ABCB4* mutations all had a history of related symptoms in close relatives, highlighting that this type of MDR3 deficiency is inherited in an autosomal dominant manner, and suggests these pathogenic, severe *ABCB4* alleles are highly penetrant by adulthood in females. However, since environmental and other factors are involved in the expression of MDR3 deficiency, it is likely that other apparently asymptomatic relatives may also be carriers (i.e.

males). Identification of asymptomatic and symptomatic carrier relatives is important as MDR3 deficiency responds well to UDCA therapy, which reduces gallstones and prevents severe complications such as fibrosis, cirrhosis and the need for cholecystectomy. In addition, UDCA can be taken throughout pregnancy by high-risk women, preventing the onset of severe ICP and the risks to the fetus. Furthermore, a review of patient P41 (5.3.7) suggests that two copies of a 'mild' mutation might lead to a similar clinical presentation as one copy of a 'severe' mutation. This would be supported if the total amount of MDR3 protein produced by the two scenarios was similar, and therefore protein studies would be important to confirm this hypothesis. This hypothesis is also likely to apply to *ABCB11* and *ATP8B1* mutations.

5.4.3. Significance of genetic diagnosis

A genetic diagnosis of PFIC is useful for parents and families for many reasons including; having a name for the condition, allowing access to specific knowledge about a particular condition, allowing more accurate predictions of prognosis, more accurate assessment of genetic risk for siblings, future pregnancies and other relatives and allowing access to prenatal diagnosis or pre-implantation genetic diagnosis. Similarly, diagnosis of PFIC-related disorders in adults can be useful to allow effective treatment planning, for example prior to pregnancy in females with *ABCB4* mutations, and to allow screening of asymptomatic relatives.

5.4.3. Variants of uncertain significance

The cases described above (5.3.11) highlight the need for careful reporting of VUS, as it may be detrimental to patient care if a clinician assumed an incorrect diagnosis due to the misinterpretation of genetic findings. In cases where VUS are identified,

testing of other affected relatives if available would be advised in order to provide evidence that the missense variant segregated with disease, or not. Publication of additional patients with the same genetic variants would lead to greater understanding of their clinical significance in future. It is also possible that other genetic and environmental factors are playing a role in these cases.

Chapter 6 – GENETIC TESTING IN PATIENTS SUSPECTED OF NPC AND CD

6.1. Introduction

6.1.1. Genetic characterisation of NPC

Carstea et. al. mapped the disease-causing gene for Niemann Pick disease type C to a region on chromosome 18 and then later identified the *NPC1* gene (Carstea, 1997). Disease heterogeneity was subsequently exposed and the *NPC2* gene was identified (Naureckiene, 2000b). Since then a large number of mutations have been identified in these genes; although around 95% of NPC cases are linked to mutations in *NPC1* (Vanier et al., 1996). A large cohort of British NPC cases was reported in detail by Imrie et. al. in 2007 (Imrie et al., 2007).

6.1.2. Genetic characterisation of CD

SLC25A13 was identified as the gene responsible for adult-onset citrullinaemia type 2 (CTLN2) in 1999 (Kobayashi et al., 1999). It was first linked to liver disease in infants in 2002 by the same group (Saheki and Kobayashi, 2002). NICCD and CTLN2 are therefore considered allelic conditions, together described as citrin deficiency or citrullinaemia type 2, and neither of them is fully penetrant. This disease has been almost exclusively been described in patients of Japanese and Chinese ethnicity, however recently it has been discovered in other populations including in the UK (Hutchin et al., 2009); summarised by Woo et al., 2014).

6.1.3. Investigations in subjects suspected of NPC and CD

Mutation detection in subjects suspected of NPC and CD was carried out by testing of patients referred to the West Midlands Regional Genetics Service between January 2010 and May 2014. Subjects were tested using FS or tNGS after the assay had been developed in January 2013 (described in Chapter 3). Both methods allowed interrogation of the coding regions of the genes plus the intron-exon boundaries allowing the findings to be collated. Sequence variant interpretation was performed on all variants identified except those considered benign polymorphisms. Questionnaires were completed where possible by the referring clinicians for NPC patients in whom sequence variants were identified, in order to correlate phenotypes with genetic findings.

6.1.4. Chapter 6 Objectives

- To broaden the known mutation spectrum of the *NPC1*, *NPC2* and *SLC25A13* genes.
- To evaluate genotype-phenotype correlations in the *NPC1* gene.

6.2. Methods

6.2.1. Patient material

Patients suspected of NPC or CD were referred for diagnostic testing; mainly by consultant hepatologists or paediatricians, however occasional referrals were made from consultant neurologists, metabolic specialists or geneticists. Consent was taken by the referring clinicians for diagnostic testing, which included the anonymous reporting of sequence variants. Samples received were mainly venous blood samples in EDTA or occasionally Lithium Heparin. Some DNA samples were received from outside of the West Midlands region or from other countries. All results were reported back to the patients via the referring clinicians.

6.2.2. DNA extraction

DNA was extracted from blood samples using Gentra Autopure method between January 2010 and February 2013. At that point, the lab switched to the Qiagen Qiasymphony method. Occasionally the phenol chloroform method was used for samples which appeared to be in poor condition, for example long transit time. Samples of very small volumes or from very young babies were extracted using a manual version of the Autopure method. Quantification of DNA samples was done using the Nanodrop Spectrophotometer. All methods are described in Chapter 2 (2.2.1).

6.2.3. PCR primer design

PCR primers were designed for all coding exons of the *NPC1*, *NPC2* and *SLC25A13* genes. This was done using the online tools Exon-Primer (UCSC), Primer3 and Primer-Blast (NCBI) using 60^{o}C as the optimal annealing temperature and at least 30 bases of flanking intron each side (2.1.14). All primer pairs were checked using

the NGRL SNPCheck tool to avoid polymorphisms of more than 1% in any population. PCR primer sequences are listed in 2.1.12.1, Table 2.1.

6.2.4. Analysis methods

The methods for PCR (2.2.3.1), FS (2.2.6.1), and sequence analysis (2.2.8.1) are described in Chapter 2. FS was also used to confirm sequence variants identified using the tNGS assay and to test those exons not covered by tNGS; *NPC1* exons 1 and 12, *NPC2* exon 1 and *SLC25A13* exon 1. The methods relating to the tNGS assay are described in sections 2.2.3.4, 2.2.8.4 and 2.2.6.3 of Chapter 2. These processes were optimised and validated as discussed in Chapter 3. The methods for MLPA are described section 2.2.7 and for GeneMarker analysis 2.2.8.2.

6.2.5. Variant interpretation

All sequence variants were investigated for potential pathogenicity using the variant interpretation pipeline described in Chapter 2 (2.2.10).

6.2.6. Phenotype questionnaires

Phenotype questionnaires were sent to referring clinicians for NPC referrals in whom sequence variants were identified in order to gather clinical information. The clinical features were based upon the NPC index of suspicion tool (Wijburg et al., 2012). The template is shown in Appendix IV. The NPC suspicion index score was available for some of the patients.

6.3. Results

6.3.1. NPC testing

6.3.1.1. Sequence variants detected in NPC

One hundred and thirty one subjects with clinical suspicion of Niemann Pick disease type C (NPC) were tested for mutations in *NPC1* and *NPC2*. Forty three were two years of age or less, forty eight were between three and fifteen years of age and forty were sixteen years or older at the time of referral. The majority were referred from centres in the UK, however thirty two were from Brazil and single samples were received from Norway, Greece and Canada. Seventy seven of these were tested using the Sanger sequencing method (FS) and fifty four using the next generation sequencing method (tNGS), with Sanger sequencing to confirm any variants found and to complete any missing exons due to low coverage.

The samples were screened for mutations in the *NPC1* and *NPC2* genes using DNA sequencing and MLPA, and Figure 6.1 summarises the method used to detect the mutations detected. Thirty four sequence variants were identified, including fourteen variants which were classified as 'pathogenic', fourteen that were 'possibly pathogenic' and six were classified as of uncertain clinical significance (VUS). Table 6.1 lists the thirty four variants detected and includes nonsense, frameshift, splice site (+/- 1 or 2 bases from exon) mutations detected in this cohort, and also missense mutations that had previously been reported in the literature in association with cholestatic conditions. In addition, this table includes variants that were novel but predicted to be likely pathogenic mutations (using *in silico* methods described 2.2.10). In summary, twenty two were missenses, six were in the conserved splice sites, five altered the reading frame and one was a likely large deletion of an entire

exon. In addition, thirteen of the thirty four variants were novel (38%). These were c.338_339dupGT, c.631+1G>T, P520S, Exon 11 deletion, E606Q, E718D, c.2533delG, c.2972delA, c.3477+2_+3dupTG, c.3477+4A>G, E1166K, F1207S and c.3746-3749delGTTA.

Figure 6.2 summarises the mutation detection rate according to age at referral; 23% in patients under three years of age, 17% in patients between three and fifteen years of age and 25% of patients aged sixteen years or over.

Figure 6.1: NPC variant detection rate by method

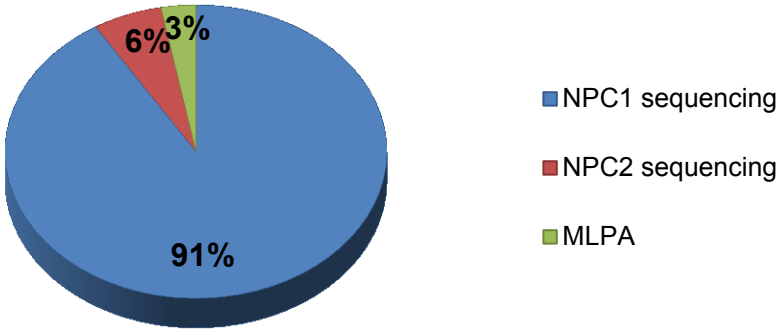


Figure 6.2: Overall genetic outcome by age at referral

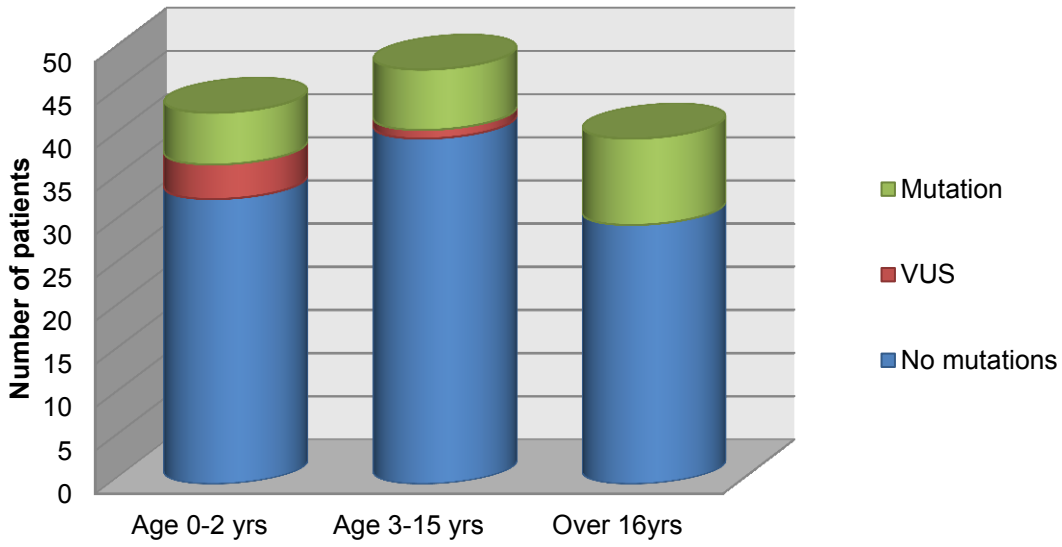


Table 6.1 The sequence variants identified in the NPC cohort.

KEY: DNA change; standard genetic code using single amino acid code and HGVS nomenclature. Type; MS=missense, FS=frameshift, NS=nonsense, SS=splice site, Int=intronic, Del=deletion. Population frequency; 1KG=1000 Genomes project using their three letter code for populations, ESP=exome sequencing project. NA cons=Nucleic acid conservation, AA cons=Amino acid conservation, AA diff=amino acid difference. The terms used were those generated using Alamut software. Protein prediction; AGVGD=Align GVGD, S=SIFT, MT=Mutation Taster, PP=PolyPhen2. Splicing prediction; SSF=Splice site finder, MES=MaxEntScan, NNS=NNSplice, GS=GeneSplicer, HSF=Human Splice Finder.

#	Gene	DNA change	Type	Reference	Population frequency	NA cons AA cons AA diff	Protein prediction	Splicing prediction	Conclusion
N5	*NPC 1*	c.338_339dupGT p.(S114Vfs*9)	FS	Novel	None	n/a	n/a	n/a	Pathogenic
N11	*NPC 1*	c.530G>A p.(C177Y)	MS	(Ribeiro et al., 2001)	rs80358252 No data	High High Large	AGVGD C0 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES Donor creation NNS No change HSF No change	Possibly pathogenic
N12	*NPC 1*	c.631+1G>T	SS	Novel	No data	High	n/a	SS Donor site lost MES Donor site lost NNS Donor site lost HSF Donor site lost	Pathogenic
N18 N19	*NPC 1*	c.1133T>C p.(V378A)	MS	(Millat et al., 2001)	rs120074134 No data	High High Small	AGVGD C25 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Pathogenic

N1	*NPC 1*	c.1408G>C p.(A470P); c.2509A>G p.(I837V)	MS	(Héron et al., 2012)	No data	Moderate High Small	AGVGD C0 S Tolerated MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
					No data	Not Moderate Small	AGVGD C0 S Tolerated MT Polymorphism PP Benign	SS No change MES No change NNS No change HSF Acceptor creation	
N24	*NPC 1*	c.1558C>T p.(P520S)	MS	Novel	No data	High High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Uncertain significance
N13	*NPC 1*	Exon 11 deletion	Ex Del	Novel	No data	High	n/a	n/a	Pathogenic
N1	*NPC 1*	c.1816G>C p.(E606Q)	MS	Novel	No data	High High Small	AGVGD C25 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
N14	*NPC 1*	c.1990G>A p.(V664M)	MS	(Park et al., 2003), (Bauer et al., 2013)	rs376213990 1KG No data ESP 1 EA allele	High High Small	AGVGD C0 S Tolerated MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic

N2	*NPC 1*	c.2154A>C p.(E718D)	MS	Novel	No data	Not Moderate Small	AGVGD C0 S Tolerated MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Uncertain significance
N20 N21	*NPC 1*	c.2292G>A	SS	(Park et al., 2003)	No data	Weak	n/a	SS Acceptor creation MES Acceptor creation NNS Acceptor creation HSF Acceptor creation	Pathogenic
N3	*NPC 1*	c.2533delG p.(V845Ffs*9)	FS	Novel	No data	n/a	n/a	n/a	Pathogenic
N20	*NPC 1*	c.2594C>T p.(S865L)	MS	(Fernandez-Valero et al., 2005)	No data	High High Large	AGVGD C15 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
N10	*NPC 1*	c.2621A>T p.(D874V)	MS	(Millat et al., 2001), (Bauer et al., 2013)	rs372030650 1KG No data ESP 1 EA allele	Moderate Moderate Large	AGVGD C0 S Deleterious MT Disease-causing PP Benign	SS No change MES No change NNS No change HSF Donor creation	Pathogenic
N15	*NPC 1*	c.2660C>T p.(P887L)	MS	(Garver et al., 2010)	No data	High Moderate Moderate	AGVGD C0 S Deleterious MT Disease-causing PP Probably	SS No change MES No change NNS No	Possibly pathogenic

							damaging	change HSF No change	
N15	*NPC 1*	c.2861C>T p.(S954L)	MS	(Greer et al., 1999), (Bauer et al., 2013)	No data	High High Large	AGVGD C35 S Deleterious MT Disease-causing PP Possibly damaging	SS No change MES No change NNS No change HSF No change	Pathogenic
N10	*NPC 1*	c.2882A>G p.(N961S)	MS	(Dvorakova et al., 2006)	rs34084984 dbSNP Het 0.007 1KG 1% of AFR, ESP 53 AA alleles	Moderate Moderate Small	AGVGD C0 S Tolerated MT Disease-causing PP Benign	SS No change MES No change NNS No change HSF Acceptor creation	Uncertain significance
N4	*NPC 1*	c.2972delA p.(Q991Rfs*6)	FS	Novel	No data	n/a	n/a	n/a	Pathogenic
N22 N16	*NPC 1*	c.3019C>G p.(P1007A)	MS	(Greer et al., 1999), (Bauer et al., 2013)	rs80358257 1KG 1 allele ESP 2 EA alleles	High High Small	AGVGD C25 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Pathogenic
N2	*NPC 1*	c.3020C>T (p.P1007L)	MS	(Imrie et al., 2007)	No data	High High Moderate	AGVGD C65 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES Acceptor creation NNS No change HSF No change	Possibly pathogenic
N17	*NPC 1*	c.3056A>G p.(Y1019C)	MS	(Fancello et al., 2009)	No data	High High Large	AGVGD C55 S Deleterious MT Disease-causing	SS No change MES No change	Possibly pathogenic

							PP Probably damaging	NNS No change HSF No change	
N16 N14	*NPC 1*	c.3104C>T p.(A1035V)	MS	(Ribeiro et al., 2001)	rs28942107 No data	High High Small	AGVGD C0 S Tolerated MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
N23	*NPC 1*	c.3107C>T p.(T1036M)	MS	(Fernandez-Valero et al., 2005)	rs28942104 No data	High Moderate Moderate	AGVGD C15 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
N18 N19	*NPC 1*	c.3182T>C p.(I1061T)	MS	(Yamamoto et al., 1999)	rs80358259 ESP 1 AA and 4 AE alleles	High High Moderate	AGVGD C25 S Deleterious MT Disease-causing PP Possibly damaging	SS No change MES No change NNS No change HSF No change	Pathogenic
N9	*NPC 1*	c.3263A>G p.(Y1088C)	MS	(Yamamoto et al., 1999)	rs28942106 No data	High High Large	AGVGD C55 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic
N11	*NPC 1*	c.3477+2_+3dupTG	SS	Novel	No data	High	n/a	SS Donor creation MES Donor creation NNS Donor creation	Possibly pathogenic

								HSF Donor creation	
N24	*NPC 1*	c.3477+4A>G	SS	Novel	rs114073738 dbSNP Het 0.005 1KG 13 alleles ESP 26 AA and 2 EA alleles	High	n/a	SS Donor weakened MES Donor weakened NNS Donor lost HSF Donor weakened	Uncertain significance
N21	*NPC 1*	c.3496G>A p.(E1166K)	MS	Novel	No data	High High Small	AGVGD C55 S Deleterious MT Disease-causing PP Probably damaging	SS No change MES Acceptor creation NNS No change HSF Donor creation	Possibly pathogenic
N6	*NPC 1*	c.3591+4delA	SS	(Park et al., 2003)	No data	High	n/a	SS No change MES Donor site lost/new donor NNS Donor site lost HSF Donor site lost/new donor	Pathogenic
N6	*NPC 1*	c.3620T>C p.(F1207S)	MS	Novel	No data	High High Large	AGVGD C15 S Tolerated MT Disease-causing PP Probably damaging	SS No change MES No change NNS No change HSF No change	Possibly pathogenic

N7 N8 N9	*NPC 1*	c.3662delT p.(F1221Sfs*21)	FS	(Ribeiro et al., 2001)	No data	n/a	n/a	n/a	Pathogenic
N10	*NPC 1*	c.3746-3749delGTTA p.(S1249Tfs*2)	FS	Novel	No data	n/a	n/a	n/a	Pathogenic
N24	*NPC 2*	c.88G>A p.(V30M)	MS	(Park et al., 2003)	rs151220873 dbSNP Het 0.002 1KG 5 alleles ESP 25 EA alleles	Not Weak Small	AGVGD C0 S Tolerated MT Polymorphism PP Possibly damaging	SS No change MES No change NNS No change HSF No change	Uncertain significance
N25 N26 N27 N28	*NPC 2*	C.441+1G>A	SS	(Bauer et al., 2013)	rs140130028 dbSNP Het 0.002 1KG 5 alleles ESP 7 AA and 76 EA alleles	n/a	n/a	SS Donor site lost MES Donor site lost NNS Donor site lost HSF Donor site lost	Uncertain significance

6.3.1.2. Multiplex-ligation dependent probe amplification

DNA sequencing by FS or tNGS will not detect whole exon deletions or duplications therefore MLPA was used to screen for these in *NPC1* and *NPC2* (Schouten et al., 2002). One patient was identified with an apparent homozygous deletion of exon 11 of *NPC1*. In addition, no PCR product could be obtained using flanking primers, excluding the possibility of a homozygous sequence variant in the exon causing MLPA probe dropout. This deletion has not been fully characterised however is likely to extend at least to one or both of the PCR primer sequences (see Figure 6.3). Consequently, this is mostly likely a disease-causing mutation; 'pathogenic'.

Figure 6.3: A diagram showing sequence including *NPC1* exon 11.

The sequence of *NPC1* exon 11 including PCR primer and MLPA probe sequences. Red is forward primer, green is reverse primer, blue is the exonic sequence, purple bold is the MLPA probe. Red text is nucleotide numbering using the A of the first ATG as 1. Blue text is amino acid numbering.

MLPA analysis for this patient showed no amplification of PCR product or MLPA product. Note that the PCR primers and the MLPA probe sequences do not overlap.

```
TATTTCAGCAAACTCTTGTTTAATTCTTAGAGCCCAGAGATACAGTCCATAGCTCCAGTG
AGAAATCTGTGTTGTGATTTTTCCCCTGGTATGTGTCTAATTTTCTGCATGCTTGTAATC
                                                        560
                                                 .  .  .  |  .
TGCTTTTTGATGTCACATTTTTCCTTTCTAGATCAAAACTACAATAACGCCACTGCCCTT
                                      ^1660     ^1670     ^1680
                          570                             580
.  .  .  .  .  .  .  .  .  |  .  .  .  .  .  .  .  .  .  |  .
GTGATTACCTTCCCTGTCAATAATTACTATAATGATACAGAGAAGCTCCAGAGGGCCCAG
      ^1690     ^1700     ^1710     ^1720     ^1730     ^1740

.  .  .  .  .
GCCTGGGAAAAAGAGTGAGTCACTCATGGGTGTGAGCAGACTGTCCTTGCATTGTGGGAA
      ^1750  ^1757
GCTAGACACTTGCGGCAAGCACTTAAACTTCATTTTCTAAATTAACCAATGGCAGATCTG
AGTTACGTAGTTTTGTTCACTTGGTACTAATGATGCTAAGAAGAAACTGTAAAC
```

6.3.1.3. NPC diagnoses

Table 6.2 lists the clinical features of the patients with mutations where these could be obtained. Disease subtype was assigned by the age of onset of symptoms where that information was available, or by the clinical features reported, e.g. neonatal cholestasis was used to assign the subtype 'infantile' and neurological features such as VSGP or gelastic cataplexy indicated 'juvenile'. Where the age at presentation could not be determined 'Not known' was assigned and where the genetic diagnosis of NPC was in doubt 'Not certain' was assigned.

Diagnoses were made in twenty three patients by the identification of at least one pathogenic or possibly pathogenic mutation (18%), and a further four patients had VUS only (3%). There were thirteen infantile cases, seven juvenile cases and one adult case. In addition, there were two cases diagnosed with NPC where the subtype was not clear (not known) and five cases were the diagnosis of NPC was not certain. Four patients diagnosed with NPC were found with only single *NPC1* mutations and two mutant alleles were identified in the remaining nineteen diagnosed cases.

Within the group of thirteen infantile cases, the age of onset was recorded in 8 cases; this ranged from birth to 5 months of age, and averaged 55 days. Clinical features were recorded in eleven cases with visceral features being present in all of these, neurological features in eight cases and psychiatric features in three cases. In the group of seven juvenile cases, the age of onset was recorded in only one. All seven cases had neurological features, four had psychiatric features and three had visceral features. The single adult case presented at age 30 and had neurological and psychiatric features.

6.3.1.4. Infantile NPC patients

The thirteen infantile cases were found to have a total of 18 different mutations, 10 of which were novel, indicating that these mutations are likely to seriously affect the NPC1 protein. Patient N1 had a double missense allele A470P; I837V and a novel missense allele E606Q. This patient presented at 12 weeks of age and had cholestasis, hepatosplenomegaly, hypotonia and delayed developmental milestones, indicating early-onset type NPC. Patient N2 presented at 5 months, had cholestasis and hepatosplenomegaly, and the missense variants E718D and P1007L were identified. Patient N3 was homozygous for the novel frameshift mutation c.2533delG p.(V845Ffs*9) and presented from birth. Similarly, patient N4 was homozygous for the novel frameshift c.2972delA p.(Q991Rfs*6) and presented before the age of 2 years. Patient N5 also had a novel homozygous frameshift mutation c.338_339dupGT p.(S114Vfs*9) and presented at two months of age. Patient N6 was compound heterozygous for the splice site mutation c.3591+4delA and the novel missense F1207S. This patient presented by 9 days with cholestasis and splenomegaly and went on to have acute liver failure at 6 weeks of age. N7 and N8 was both homozygous for a previously reported frameshift mutation c.3662delT p.(F1221Sfs*21), whereas N9 was heterozygous for the same mutation but also carried the previously reported missense c.3263A>G p.(Y1088C). All three had classical filipin staining patterns. N10 presented at 6 weeks with cholestasis and hepatosplenomegaly. Sequencing analysis identified heterozygosity for the D874V missense and a complex allele containing a missense and frameshift; p(N961S; S1249Tfs*2). Patient N11 was compound heterozygous for the missense C177Y and the novel frameshift c.3477+2_+3dupTG. In patient N12, a novel splice site mutation was identified on one allele (c.631+1G>T), however no mutation was

detected on the other allele. This patient presented at 8 weeks of age with cholestasis and hepatomegaly and went on to develop neurological and psychiatric symptoms of NPC. Lastly, N13 had a likely deletion of all or part of exon 11 on both *NPC1* alleles, presented at 6 weeks and had cholestasis, hepatosplenomegaly, hypotonia and spasticity.

6.3.1.5. Juvenile and adult NPC patients

The seven juvenile NPC cases were found to have a total of ten different mutations, only one of which was novel. Ten of these were missense mutations and one was a splicing defect. Patient N14 was found to be heterozygous for two missense mutations V664M and A1035V, had neurological and psychological features and a suspicion index score over 40. Patient N15 presented at age seven years with neurological features of NPC. Heterozygosity for two missense mutations, P887L and S954L, was identified. N16 had neurological features of NPC and splenomegaly and was found to be heterozygous for two missense mutations; P1007A and A1035V. Both N15 and N16 had variant filipin staining patterns. N17 had neurological and psychiatric features of NPC, however only one heterozygous mutation was identified; Y1019C. N20 also had neurological and psychiatric features of NPC, and was found to be heterozygous for the splice site mutation c.2292G>A, and the missense mutation S865L. The splice-site mutation c.2292G>A was also identified in patient N21 along with the novel missense E1166K. This patient had classical neurological symptoms of VSGP, ataxia, dysarthria and dystonia as well as splenomegaly. Lastly, in patient N22 only the missense mutation P1007A was identified. This patient had neurological features of NPC as well as splenomegaly and pre-senile cognitive decline or dementia.

There was one adult case which presented at age 30 years; N23. This patient had neurological and psychiatric features and was found to be heterozygous for one missense mutation; T1036M.

6.3.1.6. Patients with VUS only

There were five patients in which the genetic findings did not fully confirm a diagnosis of NPC, as only VUS were identified. Patient N24 presented at 10 months and had a complex phenotype consisting of cholestasis, hepatosplenomegaly, hypotonia, VSGP and spasticity, as well as coarse facial features and anaemia. In *NPC1* the intronic variant c.3477+4A>G was identified; *in silico* splicing tools predicted this nucleotide is highly conserved and the change is likely to weaken the splice donor site of exon 22, however it has also been identified in normal populations at a low frequency making it less likely to be causative of NPC (rs114073738). In addition, the novel *NPC1* missense P520S was identified. This is a highly conserved nucleotide and amino acid and *in silico* protein prediction tools indicate it is likely to be deleterious to the protein. Lastly the *NPC2* missense V30M was identified in patient N24. This was reportedly identified in the heterozygous state in an NPC patient (Park et al., 2003), although a mutation on the other allele was not identified. However, since then it has also been reported in normal populations (rs151220873) and it is not well conserved at the nucleotide or amino acid level. In conclusion, there is little evidence that V30M is causative however the contribution of the two *NPC1* variants is at this time uncertain. Further characterisation of this patient using filipin staining or complementation studies may be useful to conclusively establish a diagnosis of NPC.

A further four patient (N25 to N28) were found to be heterozygous for the NPC2 variant c.441+1G>A. No other NPC1 or NPC2 variants were identified in these patients.

In two of these patients (N25 and N26) no clinical data was reported, however patients N27 and N28 both presented in the first few weeks of life. N27 had cholestasis and hepatomegaly, and N28 had cholestasis and splenomegaly, as well as VSGP and hypotonia.

Table 6.2 The genotypes and clinical features of patients with NPC sequence variants identified.

KEY: Phase known; Yes or No indicates whether parents have been tested for the variants. SI>40 = NPC suspicion index score (Wijburg et al., 2012) greater than 40, Y/N or numerical value if reported. Age at presentation; d=days, w=weeks, m=months, y=years.

Subtype was assigned by the age of onset of symptoms where that information was available, or by the clinical features reported, eg neonatal cholestasis was used to assign the subtype 'infantile' and neurological features such as VSGP or gelastic cataplexy indicated 'juvenille'. Where the age at presentation could not be determined 'Not known' was assigned and where the genetic diagnosis of NPC was in doubt 'Not certain' was assigned.

Clinical features Visceral; C=prolonged neonatal jaundice or cholestasis, S=Splenomegaly, H=Hepatomegaly, HF=Hydrops fetalis, SA=sibling with fetal ascites.

Neurological; V=vertical supranuclear gaze palsy, G=Gelastic cataplexy, A=Ataxia, clumsiness or frequent falls, DA=dysarthria/dysphagia, DT=dystonia, SP=Acquired and progressive spasticity, HT=hypotonia, DD=delayed developmental milestones, SE=Generalised or partial seizures, M=myoclonus.

Psychatric; D=Pre-senile cognitive decline or dementia, P=Psychotic symptoms, TR= treatment-resistant psychiatric symptoms, PS=other psychiatric symptoms, B=aggressive or disruptive behaviour

#	Country	Allele 1	Allele 2	Phase known	SI>40	Subtype	Age at presentation	Clinical features		
								Visceral	Neurological	Psychiatric
NPC1										
N1	UK	c.1408G>C;2509A>G p.(A470P; I837V)	c.1816G>C p.(E606Q)	Yes	Not known	Infantile	12 w	C, S, H,	HT, DD	-
N2	UK	c.2154A>C p.(E718D)	c.3020C>T (p.P1007L)	No	Not known	Infantile	5 m	C, H	-	-
N3	UK	c.2533delG p.(V845Ffs*9)	c.2533delG p.(V845Ffs*9)	Yes	Not known	Infantile	From birth	Not known		
N4	UK	c.2972delA p.(Q991Rfs*6)	c.2972delA p.(Q991Rfs*6)	Yes	Not known	Infantile	<2 y	Not known		

N5	UK	c.338_339dupGT p.(S114Vfs*9)	c.338_339dupGT p.(S114Vfs*9)	Yes	Not known	Infantile	2 m	C, S, H,	HT, DT	-
N6	UK	c.3591+4delA	c.3620T>C p.(F1207S)	Yes	Not known	Infantile	9 d	C, S	-	-
N7	Brazil	c.3662delT p.(F1221Sfs*21)	c.3662delT p.(F1221Sfs*21)	Yes	148	Infantile	-	C, S, SA,	HY, DD, V, DA, DT, SP	-
N8	Brazil	c.3662delT p.(F1221Sfs*21)	c.3662delT p.(F1221Sfs*21)	No	231	Infantile	-	S	DD, V, A, DA, DT, SP	D
N9	Brazil	c.3662delT p.(F1221Sfs*21)	c.3263A>G p.(Y1088C)	Yes	147	Infantile	-	C, S, SA,	DD, V, A, DT, SP	-
N10	UK	c. 2882A>G; 3746_3749delGTTA p.(N961S; S1249Tfs*2)	c.2621A>T p.(D874V)	Yes	Not known	Infantile	6 w	C, H	-	-
N11	Brazil	c.530G>A p.(C177Y)	c.3477+2_+3dupTG	Yes	276	Infantile	-	C, S,	V, G, A, DA, DT, SP, SE	D
N12	UK	c.631+1G>T	Not identified	n/a	Not known	Infantile	8 w	C, H,	HY, DD, SE, M	D, P, TR, PS, B
N13	UK	Exon 11 deletion	Exon 11 deletion	No	Not known	Infantile	6 w	C, S, H,	HT, SP	-
N14	Brazil	c.1990G>A p.(V664M)	c.3104C>T p.(A1035V)	No	116	Juvenille	-	-	V, A , DA, DT, SP, SE	D
N15	Brazil	c.2660C>T p.(P887L)	c.2861C>T p.(S954L)	No	Yes	Juvenille	7 y	-	V, A, DT, SE	-
N16	Brazil	c.3019C>G p.(P1007A)	c.3104C>T p.(A1035V)	No	117	Juvenille	-	S	V, A, DA, DT, SE, SP	-

N17	Brazil	c.3056A>G p.(Y1019C)	Not identified	n/a	141	Juvenille	-	-	V, G, A, DA, SE, SP	D, PS
N18	UK	c.1133T>C p.(V378A)	c.3182T>C p.(I1061T)	No	Not known	Not known	-	Not known		
N19	UK	c.1133T>C p.(V378A)	c.3182T>C p.(I1061T)	No	Not known	Not known	-	Not known		
N20	Brazil	c.2292G>A	c.2594C>T p.(S865L)	No	106	Juvenille	-	-	V, A, DA, SP	D
N21	Brazil	c.2292G>A	c.3496G>A p.(E1166K)	Yes	130	Juvenille	-	S	V, A, DA, DT	-
N22	Brazil	c.3019C>G p.(P1007A)	Not identified	n/a	256	Juvenille	-	S	V, A, DA, DT, SP, SE	D
N23	UK	c.3107C>T p.(T1036M)	Not identified	n/a	Not known	Adult	30 y	-	G, A, DD	D
N24	UK	c.1558C>T p.(P520S)	c.3477+4A>G	No	Not known	Not certain	10 m	C, S, H	V, SP, HT	-
		NPC2 c.88G>A p.(V30M)	Not identified							
NPC2										
N25	UK	c.441+1G>A	Not identified	n/a	Not known	Not certain	-	Not known		
N26	Canada	c.441+1G>A	Not identified	n/a	Not known	Not certain	-	Not known		
N27	UK	c.441+1G>A	Not identified	n/a	Not known	Not certain	4 w	C, H	-	-
N28	UK	c.441+1G>A	Not identified	n/a	Not known	Not certain	8 w	C, S	V, HT	-

6.3.2. Mutation detection in subjects suspected of CD

51 patients with clinical suspicion of citrin deficiency (CD) *SLC25A13* gene between 2009 and 2015; 48 were from the UK and single cases were received from Germany, France and Brazil. 20 were under 2 years of age and 18 were between 2 and 17 years of age at the time of referral. 45 of these were tested using the FS method and 6 using the tNGS method, with FS used to sequence exon 1 of *SLC25A13* and to confirm any variants found.

9 mutations in the *SLC25A13* gene were identified in 11 patients (21%) and the mutations identified are listed in Table 6.3; 4 missense, 3 nonsense and 2 frameshift mutations. Five of the nine mutations identified in the *SLC25A13* gene were previously reported, however four of them were novel and these are listed in Table 6.4. One was a nonsense mutation and the other 3 were missense variants that were not seen in normal populations, were conserved at the nucleotide and amino acid levels and were predicted to be deleterious to the protein.

Table 6.3: The *SLC25A13* sequence variants found in the cohort.
No parental testing has been done in the patients marked with *, therefore it has been assumed that the mutations are present on opposite alleles of *SLC25A13*.

#	Origin (ethnicity)	Allele 1	Reference/Novel	Allele 2	Reference/Novel
C1*	UK (Indian)	c.550C>T p.(R184*)	(Kobayashi, 2003)	c.550C>T p.(R184*)	(Kobayashi, 2003)
C2	UK	c.848G>T p.(G283V)	Novel	Not detected	n/a
C3*	UK	c.852_855delTATG p.(M285Pfs*2)	(Kobayashi et al., 1999)	c.852_855dalTATG p.(M285Pfs*2)	(Kobayashi et al., 1999)
C4*	UK	c.852_855delTATG p.(M285Pfs*2)	(Kobayashi et al., 1999)	c.1095delT p.(F365Lfs*43)	(Fu et al., 2011)
C5	UK	c.852_855delTATG p.(M285Pfs*2)	(Kobayashi et al., 1999)	Not detected	n/a
C6	UK	c.1173T>G p.(Y391*)	Novel	c.1173T>G p.(Y391*)	Novel
C7	UK	c.1763G>A p.(R588Q)	(Tabata et al., 2008)	c.1763G>A p.(R588Q)	(Tabata et al., 2008)
C8	UK	c.1763G>A p.(R588Q)	(Tabata et al., 2008)	c.1763G>A p.(R588Q)	(Tabata et al., 2008)
C9	UK	c.1766C>T p.(S589F)	Novel	c.1766C>T p.(S589F)	Novel
C10	UK	c.1781G>A p.(G594D)	Novel	c.1781G>A p.(G594D)	Novel
C11	Denmark	c.1813C>T p.(R605*)	(Yasuda et al., 2000)	c.1813C>T p.(R605*)	(Yasuda et al., 2000)

Table 6.4: The novel mutations identified in the CD cohort.

None were reported on dbSNP, on HGMD or in the literature. None showed any effect on mRNA splicing according to 4 in silico tools (data not shown). KEY: N=nonsense, M=missense, NA = nucleic acid, AA = amino acid, cons = conservation, diff = difference. Protein prediction tools AGVGD = Align GVGD, S = SIFT, MT = Mutation Taster, PP = PolyPhen2. Data from Alamut software (see 2.1.13).

Exon	DNA change	Type	NA cons	AA cons	AA diff	Protein prediction tools	Conclusion
8	c.848G>T p.(G283V)	M	High	High	Moderate	AGVGD - C0 S - Deleterious MT - Disease-causing PP - Probably damaging	Possibly pathogenic
11	c.1173T>G p.(Y391*)	N	n/a	n/a	n/a	n/a	Pathogenic
17	c.1766C>T p.(S589F)	M	High	High	Large	AGVGD – C0 S – Tolerated MT - Disease-causing PP – Possibly damaging	Possibly pathogenic
17	c.1781G>A p.(G594D)	M	Moderate	High	Moderate	AGVGD – C25 S – Deleterious MT - Disease-causing PP – Probably damaging	Possibly pathogenic

6.4. Discussion

6.4.1. Observations from testing in patients suspected of NPC

Mutations in *NPC1* and *NPC2* which were detectable using MLPA accounted for only 3% of those detected in this cohort. Such mutations are rare in NPC patients, and HGMD lists ten such mutations, compared to three hundred and seventy five mutations detectable by sequencing methods (Patterson 1993 GeneReviews, Human Gene Mutation Database). 6% of variants detected were present in *NPC2* however none of these were clearly pathogenic. As a result, fewer mutations in *NPC2* were detected than is reported for NPC worldwide (GeneReviews lists *NPC2* mutations as accounting for 4% cases); however it is similar to the genetic distribution for NPC previously reported in the UK population (Imrie et al., 2007).

Thirteen novel sequence variants in *NPC1* were identified; six of which were classified as pathogenic, four were possibly pathogenic and the rest were VUS. This study highlights that despite there being dozens of *NPC1* mutations reported in the literature, genetic testing of NPC patients is still likely to uncover novel mutations. This supports the evidence that many of the mutations in *NPC1* are family-specific (Vanier and Millat, 2003).

A detailed description of a large cohort of UK NPC patients previously found the common *NPC1* allele I1061T accounted for 25.3% of allele (Imrie et al., 2007). In this study, I1061T was found in only 3.6% of alleles in total and in 6.2% of alleles from UK patients (N18 and N19). A smaller sample number and possibly the contribution of different ethnic groups to the two cohorts may account for this difference. V378A was identified in combination with I1061T in two UK patients; however the rest of the UK patients had different *NPC1* alleles.

Four patients (14%) were identified with only single heterozygous mutations in *NPC1*. This observation has previously been reported in other NPC cohorts (Park et al., 2003) (Bauer et al., 2013).

There was a higher pickup rate in the patients referred from Brazil than those from the UK (34% versus 13%). This was expected as all the Brazilian patients had NPC suspicion index scores greater than 40 (Wijburg et al., 2012), increasing the likelihood of NPC in those referred. Most of the mutations found in the Brazilian patients had previously been reported in the literature.

6.4.2. Genotype-phenotype correlations in NPC

It can be assumed that mutations which lead to absence of NPC1 protein would be associated with a severe form of NPC, whereas missense mutations could have more variable effects on phenotype (Park et al., 2003). The patients with mutations in this cohort were grouped by their phenotypes into Infantile, Juvenille and Adult groups and the mutations identified in these groups seem to correlate well with this hypothesis.

In the thirteen infantile cases (those cases presenting under the age of three years where known or with cholestasis) the mutation was identified in twenty five of the twenty six alleles. Twelve of the alleles were small deletions or insertions which would result in a frameshift. Two alleles were a large deletion of an entire exon. Two alleles disrupted conserved splice-sites and another was a VUS but close to a splice-site. The remaining eight mutations were missense changes, three of which were novel and five had previously been reported in the literature.

Ribeiro et. al first identified C177Y and found it to be associated with a decreased level of NPC1 protein in cultured fibroblasts and unusually variant filipin staining

(Ribeiro 2001). They reported one homozygous and one compound heterozygous patient with this mutation. The homozygous patient had a moderate hepatosplenomegaly and early onset of neurological features with cataplexy noted at age three years. In contrast, Fernandez-Valero et. al found C177Y in four compound heterozygous patients and one homozygous patient and reported that all five cases had classical filipin staining. The homozygous patient had hepatosplenomegaly and hypotonia at birth. Patient N11 in this cohort had no filipin staining result, however was known to have had cholestasis and splenomegaly and a very high suspicion index score supporting the hypothesis that it associated with infantile-onset NPC.

Homozygosity for P1007L has previously been identified in a UK patient presenting by age two years with hepatosplenomegaly and then developing neurological NPC features (Imrie 2007). Y1088C has been reported in two patients with juvenile-onset NPC however on both occasions the mutation was compound heterozygous and so genotype-phenotype conclusions are difficult to ascertain (Yamamoto et al., 1999), (Chiba et al., 2014). In contrast, Imrie et. al. reported a patient who presented at age 8 years, however who was also noted to have had prolonged jaundice and hepatosplenomegaly at birth, indicating this was infantile-onset with hindsight (Imrie et al., 2007). Héron et. al found the same complex allele A470P;I837V as found in patient N1 in this cohort, in the homozygous form in a patient with late-infantile onset of neurological features and a classical filipin staining pattern (Héron et al., 2012). Millat et. al. observed the homozygous D874V mutation in a NPC patient with a classical filipin staining pattern but whom had not developed neurological features by age seven years (Millat et al., 2001). The observation of A470P;I837V and D874V associated with infantile-onset NPC in this cohort (patients N1 and N10 respectively) may represent variability in expression of these alleles or differences in phenotypic

classification between the different studies. Lastly, the identification of the three novel missenses in infantile NPC cases (E606Q, E718D and F1207S) can therefore allow the assumption that they result in severely reduced or absent NPC1 protein function.

Twelve of the fourteen mutant alleles were detected in the juvenile cases, as well as one of the two mutant alleles in the adult case. In contrast to the infantile cases, all of the juvenile and adult mutation alleles detected in this cohort were missense changes. There were nine different mutations detected in this group; one was novel and eight had previously been reported in the literature.

V664M was first reported by Park et. al. in their cohort of clinically and biochemically diagnosed NPC patients, but they made no conclusion about the severity of the missense (Park et al., 2003). Bauer et. al. also observed this mutation in an adult with psychiatric features of NPC, as compound heterozygous with another missense of uncertain clinical significance (Bauer et al., 2013). Finally, Greenberg et. al. described a unusual adult case with old mild visceral symptoms and no neurological NPC features aged forty eight (Greenberg et al., 2015). This patient had an abnormal filipin staining pattern and was compound heterozygous for V664M and V378A, indicating that both missenses are likely to result in reduced rather than absent NPC1 protein function.

In 2003 Park et. al. reported the variant c.2292G>A in their cohort (Park et al., 2003). They made no conclusion about the severity of this mutation but did indicate that the synonymous change created a new splice site. The *in silico* analysis performed in this study gave the same prediction using four different splice site tools. Fernandez-Valero et. al. reported a patient who was compound heterozygous for S865L and a

nonsense presumed severe mutation (Fernandez-Valero et al., 2005). This patient presented at age two years with hepatosplenomegaly and psychomotor retardation and had classical filipin staining. Xiong et. al also reported this mutation in a Chinese patient with cognitive decline at age seven years (Xiong et al., 2012). This patient was homozygous for S865L indicating that it is associated with reduced rather than absent NPC1 protein function. Two patients in this cohort were compound heterozygous for c.2292G>A; N20 with S865L on the other allele and N21 with the novel missense E1166K on the other allele. As yet, the severity of the three mutations is not clear and will require additional case studies or functional studies.

P887L was identified by Garver et. al. in a patient who was compound heterozygous with a frameshift on the other allele (Garver et al., 2010). Unfortunately the clinical features of the case were not described and therefore no conclusions can be draw about the severity P887L. A single S954L allele was described by Greer et. al. in a patient with an abnormal filipin staining pattern (Greer et al., 1999). Jahnova et. al. identified S954L eight times in their cohort of Czech patients (Jahnova et al., 2014). In three juvenile-onset and one adult-onset case S954L was found to be heterozygous with a severe frameshift allele, suggesting that it is associated with reduced rather than absent NPC1 protein function. P1007A is a commonly detected NPC1 allele and has been reported as having a moderate effect on NPC1 protein due to its identification in four patients with a variant filipin staining pattern, and either a nonsense mutation or the known severe I1601T mutation on the other allele (Millat et al., 2001). Y1019C has been reported in the homozygous state in one patient however no clinical or biochemical features were described (Fancello et al., 2009). Lastly, A1035V was observed several times in the cohort described by Ribeiro et. al.(Ribeiro et al., 2001). A1035V in the homozygous state was associated with the

classical filipin staiing pattern and decreased NPC1 protein. The patient presented had early onset neurological features supporting its association with a severe effect on NPC1 protein. The two patients in this cohort with A1035V were compound heterozygotes; N14 with V664M and N16 with P1007A, both assigned as mild/moderate *NPC1* alleles.

6.4.3. Observations from testing in patients suspected of CD

GeneReviews estimates this testing strategy should detect over 95% of mutations causing CD. The lower pickup rate in this cohort (21%) reflects a large number of referrals of infants who may have a non-specific clinical presentation. Two patients were identified with only single heterozygous mutations in *SLC25A13* (C2 and C5). It is possible another mutation not detectable by sequencing of the coding regions may be present in these patients, especially as these are likely from ethnic groups other than the extensively studied East Asian populations.

The vast majority of the mutations reported in *SLC25A13* are in patients of Chinese or Japanese ethnicity, therefore the testing of patients from the UK will expand the mutation spectrum known in this gene; as illustrated by the identification of four novel mutations found in patients referred from UK centres. The UK population is increasingly ethnically mixed; therefore the mutations spectrum in all patients referred from UK centres is relevant. Ethnic groups that favour consanguineous unions are likely to be over-represented in cohorts of patients with autosomal recessive genetic conditions (Bundey and Alam, 1993).

Although a rare condition, NICCD is a clinically important one to diagnose because of the high risk of the life-threatening features of CTLN2 in later life (Saheki and Kobayashi, 2002). In addition, siblings of an NICCD infant may be affected but were

not symptomatic in the neonatal period and are therefore also at risk of CTLN2. Once mutations have been identified in a family therefore, screening of siblings and other relatives is useful in order to pick up early signs of CTLN2. This has been done in three of these families so far (C7, C8 and C10).

Chapter 7 – DISCUSSION

7.1. Summary of thesis

Patients presenting with neonatal and infantile cholestasis have often indistinct phenotypes and a wide variety of underlying aetiologies. The known contribution of genetic disorders has increased in recent years with increased availability of diagnostic genetic testing. New sequencing methods have the potential to increase the numbers of genes that can be sequenced in one test and thus reduce the cost of genetic testing.

In Chapter 3, new sequencing methods (microarray sequencing and targeted next generation sequencing) have been assessed for their suitability in testing for genetic causes of cholestasis. Targeted next generation sequencing has proven to be the method with most potential (in terms of both mutation detection capabilities and potential for increased capacity) and in the years since this work was begun has become the strategy of choice for many genetically heterogeneous conditions. Genetic testing for cholestasis is now available in diagnostic laboratories using this broad strategy; although the target enrichment method, sequencer and analysis methods have all been overtaken by new and improved methods.

In Chapter 4, the assays designed and assessed in Chapter 3, were used to test over two hundred infants presenting with liver disease for mutations in the *ATP8B1*, *ABCB11*, *ABCB4*, *NPC1*, *NPC2* and *SLC25A13* genes. Diagnoses were made in 8.5% of patients, with PFIC type 2 being the most common diagnosis. In addition, single heterozygous mutations were identified in 9% of patients, supporting the idea

that individuals with these genotypes are predisposed to liver disease in the infantile period.

In Chapter 5, patients suspected of PFIC-related diseases were tested for mutations in the *ATP8B1*, *ABCB11* and *ABCB4* genes. Significant genetic findings were observed in 17% of the patients tested and twenty seven novel genetic variants were identified. Similarly in Chapter 6, patients suspected of Niemann Pick disease type C and patients suspected of citrin deficiency were tested for mutations in *NPC1* and *NPC2*, and *SLC25A13*, respectively. Thirteen novel mutations were identified in *NPC1* and four novel findings were observed in *SLC25A13*. In Chapters 5 and 6, the clinical phenotypes are presented along with the genotypes in order for conclusions about the contributions of novel or uncertain sequence variants to be drawn.

7.2. Current methods in DNA sequencing

Developments in DNA sequencing methods in the last 10 years have completely transformed the field of genetics research by allowing the conception of projects and research aims that were simply not feasible before. The original human genome project used cloning and Sanger sequencing. It involved several collaborating groups, took 13 years and cost around 2 billion US dollars to complete ("Human Genome Project Information," 2014). Nowadays a full human genome can be sequenced using one NGS machine in a week at a cost approaching $1000. It is now possible to identify the gene for a rare genetic condition by sequencing the genomes of a few affected individuals and identifying which mutated gene or genes are in common. This strategy means researchers do not require the previously essential large pedigrees or consanguineous families as study subjects. Also time

consuming linkage analysis and candidate gene analysis strategies can be circumvented. Obviously, whole genome or exome sequencing produces vast amounts of information and it can be difficult to identify the DNA variant responsible for a disease, especially if it is a missense change, therefore in some cases a combination of DNA sequencing and more traditional strategies, like linkage analysis, has the greatest power (Thomas et al., 2010). However, NGS methods have led to a surge in the discovery of genes responsible for rare disease in recent years (Shen et al., 2015), and are also being applied to the fields of cancer, neurodevelopmental disorders and epidemiology amongst others (Deng et al., 2016; Schulze et al., 2016; Sener et al., 2016).

7.3. Future developments in sequencing methodology

Despite the rapid advancements in recent years, the field of sequencing technologies shows no sign of slowing down. There are several current challenges that, if overcome, will again revolutionise sequencing in research and then subsequently in clinical settings. These challenges are; the detection of copy number variants, the detection of changes in methylation, and the requirement for target enrichment (which always includes an element of selection bias).

7.4. DNA sequencing in clinical practice

Translation of NGS methods from the research setting into clinical laboratories requires additional considerations; sensitivity, cost, quality control, unsolicited findings, bioinformatics requirements, computing requirements etc. As a result, targeted sequencing analysis is still regarded as the most appropriate testing strategy for the vast majority of patients suspected of heterogeneous genetic

conditions. The targeting can be either at the technical stage prior to sequencing (using target enrichment methods/TEMs), at the bioinformatics stage after sequencing (filtering), or both (eg targeting groups of genes within an exome kit). It is likely that as the costs of the sequencing component of the process continues to reduce, the trend towards bioinformatics-driven targeting will continue; to avoid the necessity for time-consuming and costly TEMs. This trend will require increased sequencing capacity, bioinformatics expertise and computing capability.

Even before the problem of fast, accurate, cheap DNA sequencing of the human genome has been completely solved, interpretation of the data is rapidly becoming the new bottleneck in the process. Analysis of human exome data typically yields around 70,000 sequence variants per patient, at least 100 of which are likely to be loss of function; nonsense, frameshift, conserved splice site (Lindor et al., 2015). This list of variants needs to be filtered down to allow the precise disease-causing mutation or mutations to be identified. This process is made more difficult when the DNA change results in a missense or synonymous change, when the DNA change is present in a gene not known to be associated with any aspect of the phenotype of the patient, and when the DNA change is present in a region of the exome with low depth of coverage in terms of sequencing reads. The danger of filtering variant data is that the real disease-causing mutation is missed for one of these reasons whilst other seemingly important variants act as red herrings in the search. Successive rounds of filtering using different parameters may prove to be a useful strategy. In addition, the analysis of parent-patient trios can significantly reduce the number of variants to be investigated.

One area where analysis of the entire genome is probably warranted is the diagnosis of developmental disorders and intellectual disability, where the DDD project and

other studies have shown there are a wide variety of causative lesions, ranging from copy number variants of several megabases down to single base changes (Deciphering Developmental Disorders Study, 2015).

The advantages and limitations of different sequencing strategies, bioinformatics tools and variant filtering methods need to be fully appreciated to enable translation into diagnostic testing.

7.5. Genetic testing for neonatal cholestasis

Herbst and colleagues described a 93-gene panel for infantile cholestasis which has been developed for diagnostic use in Germany (Herbst et al., 2015). Diagnoses were made in four of six patients; including one NPC case, one PFIC1 case, one PFIC2 case and one autosomal recessive polycystic kidney disease case. Therefore, this appears to be a small and highly selected group of patients, and it will be interesting to read about their diagnostic outcomes in a larger dataset in a future publication. They should be able to shed light on the disorders most commonly involved in infantile cholestasis and the possibility of mutations in more than one gene contributing to the phenotypes.

Goldschmidt and colleagues investigated the possibility of mutations in more than one gene contributing to phenotype using a microarray sequencing method to look for mutations in *SERPINA1*, *JAG1*, *ATP8B1*, *ABCB11* and *ABCB4* in 717 infants with idiopathic cholestasis (Goldschmidt et al., 2016). Despite the inclusion of genes for Alagille syndrome (AGS; *JAG1* gene) and alpha-1 anti-trypsin deficiency (AATD; *SERPINA1* gene) their diagnosis rate was 7.5%, lower than found in the study described in Chapter 4. In addition, they identified a statistically significant increased

rate of double and triple heterozygotes. There were differences between the two studies, including different gene selection, different strategies for variant classification, and a broader patient selection for the Goldschmidt study compared to study described in Chapter 4; patient age ranged from six weeks to seventeen years compared to less than two years, respectively. In this study, only a few patients were identified in Chapter 4 with more than one mutation; less stringent classification of variants in the Goldschmidt study may have increased the numbers in their cohort. Although the 1000 Genomes dataset is an excellent source of 'normal' variation, the frequencies of variants can vary widely in different populations. Therefore, the lack of ethnically-matched controls may have contributed to this finding. In addition, *SERPINA1* mutations, other than that which results in the Z allele, are not known to be associated with cholestasis in alpha-1 anti-trypsin deficiency (AATD) patients. Therefore the significance of the identification of several patients with heterozygous variants in this gene is unclear. It will certainly be interesting to see if this increase in double and triple heterozygotes is replicated by other studies, as since the PFIC proteins in particular contribute to common pathways; it seems possible that deficiencies from more than one gene might contribute to the phenotype.

One particular advantage of gene-panel approaches compared to sequential sequencing approaches is that it avoids clinical mis-diagnosis, especially in an often clinically indistinct population of patients such as those with infantile cholestasis. Grochowski and colleagues describe the identification of compound heterozygous mutations in the *ATP8B1* gene in a patient clinically diagnosed with Alagille syndrome (Grochowski et al., 2015). Despite AGS being one of the more distinct clinical syndromes associated with cholestasis, and this case presenting with four of the five features suggestive of AGS, no mutations were identified in the *JAG1* or

NOTCH2 genes. Whole exome sequencing subsequently identified two clearly deleterious mutations in *ATP8B1*, and the inheritance of these was consistent with the mutations being present on opposite alleles of the gene. This finding nicely illustrates the advantages of hypothesis-free testing strategies; those being to allow more accurate diagnosis, to allow more accurate genetic counselling for the family, to expand the recognised clinical presentations of established disorders.

The use of genome-wide analysis methods like WES or WGS has led to the identification of new genes involved in cholestasis. The *NR1H4* gene which encodes the FXR protein has been a candidate gene for cholestasis since it was established at a regulator of both BSEP and MDR3 proteins (involved in PFIC2 and PFIC3, respectively). *N1HR4* mutations have been identified in ICP cohorts (Davit-Spraul et al., 2012; van Mil et al., 2007) and in one case of idiopathic infantile cholestasis (Chen et al., 2012). More recently, Gomez-Ospina and colleagues described homozygous loss of function mutations in *NR1H4* in four patients (two families) with neonatal cholestasis (Gomez-Ospina et al., 2016). The group used whole exome sequencing and SNP microarrays to test two unrelated patients with neonatal cholestasis. The same genotype was then confirmed in an additional affected family member in each case. One family had a homozygous stop codon and the other was compound heterozygous for an in-frame insertion of three nucleotides and 31kb deletion encompassing the same region; both alleles resulting in undetectable levels of *NR1H4* mRNA. This gene looks likely to be a good candidate for PFIC-like patients in whom mutations in *ATP8B1*, *ABCB11* and *ABCB4* have been excluded.

Another protein which may be important in the PFIC cohort is CDC50A (gene name *TMEM30A*). This protein is known to be essential for the subcellular transport of FIC1 protein and its phospholipid flippase function (Paulusma et al., 2009). At the

moment there is no known link to cholestasis but it might be an interesting gene to include on a large gene panel.

7.6. Further areas of study

There are many avenues of further study that would be possible using the work presented in this thesis as a basis. Firstly, other sequencing methods could be evaluated and compared against those already studied. In particular, use of larger capacity target enrichment methods and sequencers would allow the design of a much larger panel of relevant genes to be sequenced that the one described in Chapter 3. The performance, workflow, cost and mutation detection rate of such a panel could be compared with this tNGS assay. Such technologies could also allow the inclusion of intronic and promoter regions of genes; potentially a source of unidentified disease-causing mutations (Cebecauerová et al., 2012). Investigation of copy number variation is another area where additional mutations may be discovered, and therefore methods for detecting these would be useful to investigate. MLPA is the most convenient method for individual genes; however for large gene panels this method becomes impractical. New methods which combine sequencing information and copy number information are sought after and are likely to be studied by others in the near future. Whole genome sequencing (WGS) should allow the detection of sequence and copy number variants simultaneously, and targeted bioinformatics analysis could mean that virtual panels could be created using WGS as a basis. Cost is the main barrier to this strategy at the moment, but is following a downward trend; therefore it may not be too far-fetched for an area of study soon. WGS is likely to have advantages over WES, including better coverage

of GC-rich regions and less interference from pseudogenes or repetitive DNA sequences (Meienberg et al., 2016).

Secondly, further investigation of novel variants and variants of uncertain significance would be an interesting area of study, and would help to reduce the number of patients in whom uncertainty over their diagnosis remains. Mini-gene assays to investigate potential splicing defects and *in vitro* functional studies to investigate uncertain missense changes, could allow more important information to be gathered about these mutations and the proteins. Assays such as fluorescent staining for the PFIC proteins on liver biopsy and the filipin staining test for NPC, are already available, but the results are not often correlated with the genetic results and the clinical phenotype. Perhaps a multi-disciplinary team meeting to discuss patients with equivocal results would be useful in the clinical setting.

A third area of potential future study, would be to continue to discover more genes associated with cholestasis, and then to determine the incidence of these in the neonatal cholestasis, infantile cholestasis and adult cholestasis populations. Large-scale sequencing projects that use genome-wide screens and a hypothesis-free testing strategy have the most potential to discover these new genes, although candidate gene approaches have also been successful. The 100,000 Genomes Project, in particular, has gene discovery for rare disease in its remit. It is probable that each newly discovered gene for cholestasis will account for only very few cases, however if a several rare causes of cholestasis are added to diagnostic testing panels, this could significantly increase the diagnosis rate in the near future. It is also possible that small contributions from a couple or a few genes might contribute to cholestasis, especially when other environmental factors are also present (for

example, mutations or polymorphisms in more than one gene). Elucidating these complexities would be another large area of work for the future.

Lastly, the discovery of more of the genetic basis of cholestasis will also lead to a fuller understanding of the processes involved in bile formation, liver disease and response to environmental factors. This knowledge should help towards developing strategies for new treatments for liver disease in future, for both the specific cases of cholestasis and the wider population of patients with liver disease.

Appendix III – PFIC QUESTIONNAIRE

Patient details		Referrer details	
Patient identifier		Name	
Date of Birth		Centre	
Age at presentation		Completed by	
Age at diagnosis		Date	

Did the patient present with cholestasis? If not, describe presenting features.

If patient had cholestasis, was it progressive or episodic?

If episodic, please give frequency and duration of episodes?

If episodic, was a trigger identified, eg pregnancy, contraceptive pill, another?

Did the patient have or develop other clinical features?

Was liver biopsy done and what were the results? Include immunostaining if performed.

Did the genetic findings alter the management and/or prognosis of the patient?

Is there a family history of cholestasis or liver disease?

Was the patient (or family) referred to Clinical Genetics?

Appendix IV – NPC QUESTIONNAIRE

Patient details		Referrer details	
Patient identifier		Name	
Date of Birth		Centre	
Age at presentation		Completed by	
Age at diagnosis		Date	

Please indicate whether the above patient has (or has had) any of the following symptoms.

	Yes	No
Prolonged unexplained neonatal jaundice or cholestasis		
Isolated unexplained splenomegaly		
Hepatomegaly		
Hydrops fetalis		
Siblings with fetal ascites		
Vertical supranuclear gaze palsy		
Gelastic cataplexy		
Ataxia, clumsiness or frequent falls		
Dysarthria and/or dysphagia		
Dystonia		
Acquired and progressive spasticity		
Hypotonia		
Delayed developmental milestones		
Seizure (partial or generalised)		
Myoclonus		
Pre-senile cognitive ecline or dementia		
Psychotic symptoms (hallucinations, delusions and/or thought disorders)		
Treatment-resistant psychiatric symptoms		
Other psychiatric symptoms		
Disruptive or aggressive behaviour in childhood or adolescence		
Other, please expand.		

Was filipin staining performed in cells from this patient?
If so, what were the results?

Was an oxysterol measurement performed in a sample from this patient?
If so, what were the results?

Does this patient have any family members who also have symptoms of NPC (any of those listed above)?

Appendix X – List of relevant publications

These are publications which I have contributed to during the time of my post-graduate research.

Blackmore, L., Knisely, A.S., Hartley, J.L., McKay, K., Gissen, P., Marcus, R., and Shawcross, D.L. (2013). Polymorphisms in ABCB11 and ATP8B1 Associated with Development of Severe Intrahepatic Cholestasis in Hodgkin's Lymphoma. J Clin Exp Hepatol *3*, 159–161.

McKay Bounford, K., and Gissen, P. (2014). Genetic and laboratory diagnostic approach in Niemann Pick disease type C. J Neurol *261*, 569–575.

Nahorski, M.S., Lim, D.H.K., Martin, L., Gille, J.J.P., McKay, K., Rehal, P.K., Ploeger, H.M., van Steensel, M., Tomlinson, I.P., Latif, F., et al. (2010). Investigation of the Birt-Hogg-Dube tumour suppressor gene (FLCN) in familial and sporadic colorectal cancer. J. Med. Genet. *47*, 385–390.

Rattenberry, E., Vialard, L., Yeung, A., Bair, H., McKay, K., Jafri, M., Canham, N., Cole, T.R., Denes, J., Hodgson, S.V., et al. (2013). A Comprehensive Next Generation Sequencing–Based Genetic Testing Strategy To Improve Diagnosis of Inherited Pheochromocytoma and Paraganglioma. The Journal of Clinical Endocrinology & Metabolism *98*, E1248–E1256.

Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T., et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet *385*, 1305–1314.

List of references

Adeva, M., El-Youssef, M., Rossetti, S., Kamath, P.S., Kubly, V., Consugar, M.B., Milliner, D.M., King, B.F., Torres, V.E., Harris, P.C., 2006. Clinical and Molecular Characterization Defines a Broadened Spectrum of Autosomal Recessive Polycystic Kidney Disease (ARPKD): Medicine (Baltimore) 85, 1–21. doi:10.1097/01.md.0000200165.90373.9a

Ali, M., Rellos, P., Cox, T.M., 1998. Hereditary fructose intolerance. J. Med. Genet. 35, 353–365.

Altman, R.P., Lilly, J.R., Greenfeld, J., Weinberg, A., van Leeuwen, K., Flanigan, L., 1997. A multivariable risk factor analysis of the portoenterostomy (Kasai) procedure for biliary atresia: twenty-five years of experience from two centers. Ann. Surg. 226, 348–355.

American Thoracic Society/European Respiratory Society, 2003. American Thoracic Society/European Respiratory Society Statement: Standards for the Diagnosis and Management of Individuals with Alpha-1 Antitrypsin Deficiency. Am. J. Respir. Crit. Care Med. 168, 818–900. doi:10.1164/rccm.168.7.818

Anzivino, C., Odoardi, M.R., Meschiari, E., Baldelli, E., Facchinetti, F., Neri, I., Ruggiero, G., Zampino, R., Bertolotti, M., Loria, P., Carulli, L., 2013. ABCB4 and ABCB11 mutations in intrahepatic cholestasis of pregnancy in an Italian population. Dig. Liver Dis. 45, 226–232. doi:10.1016/j.dld.2012.08.011

Aslanidis, C., Ries, S., Fehringer, P., Büchler, C., Klima, H., Schmitz, G., 1996. Genetic and biochemical evidence that CESD and Wolman disease are distinguished by residual lysosomal acid lipase activity. Genomics 33, 85–93.

Baerlocher, K., Gitzelmann, R., Steinmann, B., Gitzelmann-Cumarasamy, N., 1978. Hereditary fructose intolerance in early childhood: a major diagnostic challenge. Survey of 20 symptomatic cases. Helv. Paediatr. Acta 33, 465–487.

Bartlett, J.R., 2009. Genetic Modifiers of Liver Disease in Cystic Fibrosis. JAMA 302, 1076. doi:10.1001/jama.2009.1295

Bauer, P., Balding, D.J., Klünemann, H.H., Linden, D.E.J., Ory, D.S., Pineda, M., Priller, J., Sedel, F., Muller, A., Chadha-Boreham, H., Welford, R.W.D., Strasser, D.S., Patterson, M.C., 2013. Genetic screening for Niemann-Pick disease type C in adults with neurological and psychiatric symptoms: findings from the ZOOM study. Hum. Mol. Genet. 22, 4349–4356. doi:10.1093/hmg/ddt284

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara E. Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna

Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., vandeVondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59. doi:10.1038/nature07517

Bosch, A.M., 2006. Classical galactosaemia revisited. J. Inherit. Metab. Dis. 29, 516–525. doi:10.1007/s10545-006-0382-0

Bruce, C.K., Smith, M., Rahman, F., Liu, Z., McMullan, D.J., Ball, S., Hartley, J., Kroos, M.A., Heptinstall, L., Reuser, A.J.J., Rolfs, A., Hendriksz, C., Kelly, D.A., Barrett, T.G., MacDonald, F., Maher, E.R., Gissen, P., 2010. Design and validation of a metabolic disorder resequencing microarray (BRUM1). Hum. Mutat. 31, 858–865. doi:10.1002/humu.21261

Bull, L.N., Roche, E., Song, E.J., Pedersen, J., Knisely, A.S., van der Hagen, C.B., Eiklid, K., Aagenaes, Ø., Freimer, N.B., 2000. Mapping of the Locus for Cholestasis-Lymphedema Syndrome (Aagenaes Syndrome) to a 6.6-cM Interval on Chromosome 15q. Am. J. Hum. Genet. 67, 994–999.

Bull, L.N., van Eijk, M.J., Pawlikowska, L., DeYoung, J.A., Juijn, J.A., Liao, M., Klomp, L.W., Lomri, N., Berger, R., Scharschmidt, B.F., Knisely, A.S., Houwen, R.H., Freimer, N.B., 1998. A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. Nat. Genet. 18, 219–224. doi:10.1038/ng0398-219

Bundey, S., Alam, H., 1993. A five-year prospective study of the health of children in different ethnic groups, with particular reference to the effect of inbreeding. Eur. J. Hum. Genet. EJHG 1, 206–219.

Byrne, J.A., Strautnieks, S.S., Ihrke, G., Pagani, F., Knisely, A.S., Linton, K.J., Mieli-Vergani, G., Thompson, R.J., 2009. Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology 49, 553–567. doi:10.1002/hep.22683

Carlson, J.A., Rogers, B.B., Sifers, R.N., Hawkins, H.K., Finegold, M.J., Woo, S.L., 1988. Multiple tissues express alpha 1-antitrypsin in transgenic mice and man. J. Clin. Invest. 82, 26–36. doi:10.1172/JCI113580

Carlton, V.E.H., Harris, B.Z., Puffenberger, E.G., Batta, A.K., Knisely, A.S., Robinson, D.L., Strauss, K.A., Shneider, B.L., Lim, W.A., Salen, G., Morton, D.H., Bull, L.N., 2003. Complex inheritance of familial hypercholanemia with associated mutations in TJP2 and BAAT. Nat. Genet. 34, 91–96. doi:10.1038/ng1147

Carstea, E.D., 1997. Niemann-Pick C1 Disease Gene: Homology to Mediators of Cholesterol Homeostasis. Science 277, 228–231. doi:10.1126/science.277.5323.228

Cebecauerová, D., Strautnieks, S.S., Byrne, J.A., Jirsa, M., Thompson, R.J., 2012. ATP8B1 Gene Expression Is Driven by a Housekeeping-Like Promoter Independent of Bile Acids and Farnesoid X Receptor. PLoS ONE 7, e51650. doi:10.1371/journal.pone.0051650

Chagnon, P., Michaud, J., Mitchell, G., Mercier, J., Marion, J.-F., Drouin, E., Rasquin-Weber, A., Hudson, T.J., Richter, A., 2002. A Missense Mutation (R565W) in Cirhin (FLJ14728) in North American Indian Childhood Cirrhosis. Am. J. Hum. Genet. 71, 1443–1449.

Chen, H.-L., Chang, M.-H., 2004. Growth failure and metabolic bone disease in progressive familial intrahepatic cholestasis. J. Pediatr. Gastroenterol. Nutr. 39, 328–330.

Chen, X.-Q., Wang, L.-L., Shan, Q.-W., Tang, Q., Deng, Y.-N., Lian, S.-J., Yun, X., 2012. A novel heterozygous NR1H4 termination codon mutation in idiopathic infantile cholestasis. World J. Pediatr. 8, 67–71. doi:10.1007/s12519-011-0299-z

Cheng, J.B., Jacquemin, E., Gerhardt, M., Nazer, H., Cresteil, D., Heubi, J.E., Setchell, K.D.R., Russell, D.W., 2003. Molecular genetics of 3beta-hydroxy-Delta5-C27-steroid oxidoreductase deficiency in 16 patients with loss of bile acid synthesis and liver disease. J. Clin. Endocrinol. Metab. 88, 1833–1841. doi:10.1210/jc.2002-021580

Cheng, S., Chen, Y., Monforte, J.A., Higuchi, R., Van Houten, B., 1995. Template integrity is essential for PCR amplification of 20-to 30-kb sequences from genomic DNA. Genome Res. 4, 294–298.

Chiba, Y., Komori, H., Takei, S., Hasegawa-Ishii, S., Kawamura, N., Adachi, K., Nanba, E., Hosokawa, M., Enokido, Y., Kouchi, Z., Yoshida, F., Shimada, A., 2014. Niemann-Pick disease type C1 predominantly involving the frontotemporal region, with cortical and brainstem Lewy bodies: An autopsy case: Frontotemporal-predominant NPC with LBs. Neuropathology 34, 49–57. doi:10.1111/neup.12047

Colombo, C., Vajro, P., Degiorgio, D., Coviello, D.A., Costantino, L., Tornillo, L., Motta, V., Consonni, D., Maggiore, G., 2011. Clinical Features and Genotype-Phenotype Correlations in Children With Progressive Familial Intrahepatic Cholestasis Type 3 Related to ABCB4 Mutations: J. Pediatr. Gastroenterol. Nutr. 52, 73–83. doi:10.1097/MPG.0b013e3181f50363
Cross, N.C., Tolan, D.R., Cox, T.M., 1988. Catalytic deficiency of human aldolase B in hereditary fructose intolerance caused by a common missense mutation. Cell 53, 881–885.
Cullinane, A.R., Straatman-Iwanowska, A., Seo, J.K., Ko, J.S., Song, K.S., Gizewska, M., Gruszfeld, D., Gliwicz, D., Tuysuz, B., Erdemir, G., Sougrat, R., Wakabayashi, Y., Hinds, R., Barnicoat, A., Mandel, H., Chitayat, D., Fischler, B., Garcia-Cazorla, A., Knisely, A.S., Kelly, D.A., Maher, E.R., Gissen, P., 2009. Molecular investigations to improve diagnostic accuracy in patients with ARC syndrome. Hum. Mutat. 30, E330–E337. doi:10.1002/humu.20900
Davit-Spraul, A., Fabre, M., Branchereau, S., Baussan, C., Gonzales, E., Stieger, B., Bernard, O., Jacquemin, E., 2010. ATP8B1 and ABCB11 analysis in 62 children with normal gamma-glutamyl transferase progressive familial intrahepatic cholestasis (PFIC): Phenotypic differences between PFIC1 and PFIC2 and natural history. Hepatology 51, 1645–1655. doi:10.1002/hep.23539
Davit-Spraul, A., Gonzales, E., Baussan, C., Jacquemin, E., 2009. Progressive familial intrahepatic cholestasis. Orphanet J. Rare Dis. 4, 1. doi:10.1186/1750-1172-4-1
Davit-Spraul, A., Gonzales, E., Jacquemin, E., 2012. NR1H4 analysis in patients with progressive familial intrahepatic cholestasis, drug-induced cholestasis or intrahepatic cholestasis of pregnancy unrelated to ATP8B1, ABCB11 and ABCB4 mutations. Clin. Res. Hepatol. Gastroenterol. 36, 569–573. doi:10.1016/j.clinre.2012.08.008
De Bruyne, R., Van Biervliet, S., Vande Velde, S., Van Winckel, M., 2011a. Clinical practice: Neonatal cholestasis. Eur. J. Pediatr. 170, 279–284. doi:10.1007/s00431-010-1363-8
De Bruyne, R., Van Biervliet, S., Vande Velde, S., Van Winckel, M., 2011b. Clinical practice: neonatal cholestasis. Eur. J. Pediatr. 170, 279–284. doi:10.1007/s00431-010-1363-8
De Leeneer, K., De Schrijver, J., Clement, L., Baetens, M., Lefever, S., De Keulenaer, S., Van Criekinge, W., Deforce, D., Van Nieuwerburgh, F., Bekaert, S., Pattyn, F., De Wilde, B., Coucke, P., Vandesompele, J., Claes, K., Hellemans, J., 2011. Practical Tools to Implement Massive Parallel Pyrosequencing of PCR Products in Next Generation Molecular Diagnostics. PLoS ONE 6, e25531. doi:10.1371/journal.pone.0025531
Deans, Z., Watson, C.M., Charlton, R., Ellard, S., Wallis, Y., Mattocks, C., Abbs, S., 2015. Association of Clinical Genetic Science - Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation. [WWW Document]. URL http://www.acgs.uk.com/media/983872/bpg_for_targeted_next_generation_sequencing_-_approved_dec_2015.pdf (accessed 6.3.16).
De Vree, J.M.L., Jacquemin, E., Sturm, E., Cresteil, D., Bosma, P.J., Aten, J., Deleuze, J.-F., Desrochers, M., Burdelski, M., Bernard, O., others, 1998. Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. Proc. Natl. Acad. Sci. 95, 282–287.
Deciphering Developmental Disorders Study, 2015. Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223–228. doi:10.1038/nature14135
Degiorgio, D., Colombo, C., Seia, M., Porcaro, L., Costantino, L., Zazzeron, L., Bordo, D., Coviello, D.A., 2007. Molecular characterization and structural implications of 25 new ABCB4 mutations in progressive familial intrahepatic cholestasis type 3 (PFIC3). Eur. J. Hum. Genet. 15, 1230–1238.
Deng, X., Bakker, H.C. den, Hendriksen, R.S., 2016. Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. Annu. Rev. Food Sci. Technol. 7, 353–374. doi:10.1146/annurev-food-041715-033259
Dixon, P.H., van Mil, S.W.C., Chambers, J., Strautnieks, S., Thompson, R.J., Lammert, F., Kubitz, R., Keitel, V., Glantz, A., Mattsson, L.-A., Marschall, H.-U., Molokhia, M., Moore, G.E., Linton, K.J., Williamson, C., 2009. Contribution of variant alleles of ABCB11 to susceptibility to intrahepatic cholestasis of pregnancy. Gut 58, 537–544. doi:10.1136/gut.2008.159541
Dooley, J., Sherlock, D.S., 2011. Sherlock's diseases of the liver and biliary system / edited by James S. Dooley ... [et al.]., 12th ed. ed. Wiley-Blackwell.
Dvorakova, L., Sikora, J., Hrebicek, M., Hulkova, H., Bouckova, M., Stolnaja, L., Elleder, M., 2006. Subclinical course of adult visceral Niemann–Pick type C1 disease. A rare or underdiagnosed disorder? J. Inherit. Metab. Dis. 29, 591–591. doi:10.1007/s10545-006-0330-z
El Sherrif, Y., Potts, J.R., Howard, M.R., Barnardo, A., Cairns, S., Knisely, A.S., Verma, S., 2013. Hepatotoxicity from anabolic androgenic steroids marketed as dietary supplements:

contribution from *ATP8B1/ABCB11* mutations? Liver Int. 33, 1266–1270. doi:10.1111/liv.12216
Emerick, K.M., Rand, E.B., Goldmuntz, E., Krantz, I.D., Spinner, N.B., Piccoli, D.A., 1999. Features of Alagille syndrome in 92 patients: frequency and relation to prognosis. Hepatology 29, 822–829.
Evason, K., Bove, K.E., Finegold, M.J., Knisely, A.S., Rhee, S., Rosenthal, P., Miethke, A.G., Karpen, S.J., Ferrell, L.D., Kim, G.E., 2011. Morphologic Findings in Progressive Familial Intrahepatic Cholestasis 2 (PFIC2): Correlation With Genetic and Immunohistochemical Studies. Am. J. Surg. Pathol. 35, 687–696. doi:10.1097/PAS.0b013e318212ec87
Fairbanks, K.D., Tavill, A.S., 2008. Liver disease in alpha 1-antitrypsin deficiency: a review. Am. J. Gastroenterol. 103, 2136–2141; quiz 2142. doi:10.1111/j.1572-0241.2008.01955.x
Fancello, T., Dardis, A., Rosano, C., Tarugi, P., Tappino, B., Zampieri, S., Pinotti, E., Corsolini, F., Fecarotta, S., D'Amico, A., Di Rocco, M., Uziel, G., Calandra, S., Bembi, B., Filocamo, M., 2009. Molecular analysis of NPC1 and NPC2 gene in 34 Niemann–Pick C Italian Patients: identification and structural modeling of novel mutations. neurogenetics 10, 229–239. doi:10.1007/s10048-009-0175-3
Ferdinandusse, S., Denis, S., IJlst, L., Dacremont, G., Waterham, H.R., Wanders, R.J.A., 2000. Subcellular localization and physiological role of α-methylacyl-CoA racemase. J. Lipid Res. 41, 1890–1896.
Fernandez-Valero, E., Ballart, A., Iturriaga, C., Lluch, M., Macias, J., Vanier, M., Pineda, M., Coll, M., 2005. Identification of 25 new mutations in 40 unrelated Spanish Niemann-Pick type C patients: genotype-phenotype correlations: Identification of 25 new mutations. Clin. Genet. 68, 245–254. doi:10.1111/j.1399-0004.2005.00490.x
Ferrarotti, I., Carroll, T.P., Ottaviani, S., Fra, A.M., O'Brien, G., Molloy, K., Corda, L., Medicina, D., Curran, D.R., McElvaney, N.G., Luisetti, M., 2014. Identification and characterisation of eight novel SERPINA1 Null mutations. Orphanet J. Rare Dis. 9. doi:10.1186/s13023-014-0172-y
Folmer, D.E., van der Mark, V.A., Ho-Mok, K.S., Oude Elferink, R.P.J., Paulusma, C.C., 2009. Differential effects of progressive familial intrahepatic cholestasis type 1 and benign recurrent intrahepatic cholestasis type 1 mutations on canalicular localization of ATP8B1. Hepatology 50, 1597–1605. doi:10.1002/hep.23158
Fregonese, L., Stolk, J., 2008. Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. Orphanet J. Rare Dis. 3, 16. doi:10.1186/1750-1172-3-16
Fu, H.-Y., Zhang, S.-R., Wang, X.-H., Saheki, T., Kobayashi, K., Wang, J.-S., 2011. The mutation spectrum of the SLC25A13 gene in Chinese infants with intrahepatic cholestasis and aminoacidemia. J. Gastroenterol. 46, 510–518. doi:10.1007/s00535-010-0329-y
Gabolde, M., Hubert, D., Guilloud-Bataille, M., Lenaerts, C., Feingold, J., Besmond, C., 2001. The mannose binding lectin gene influences the severity of chronic liver disease in cystic fibrosis. J. Med. Genet. 38, 310–311.
Garver, W.S., Jelinek, D., Meaney, F.J., Flynn, J., Pettit, K.M., Shepherd, G., Heidenreich, R.A., Vockley, C.M.W., Castro, G., Francis, G.A., 2010. The National Niemann-Pick Type C1 Disease Database: correlation of lipid profiles, mutations, and biochemical phenotypes. J. Lipid Res. 51, 406–415. doi:10.1194/jlr.P000331
Gelsthorpe, M.E., Baumann, N., Millard, E., Gale, S.E., Langmade, S.J., Schaffer, J.E., Ory, D.S., 2008. Niemann-Pick type C1 I1061T mutant encodes a functional protein that is selected for endoplasmic reticulum-associated degradation due to protein misfolding. J. Biol. Chem. 283, 8229–8236. doi:10.1074/jbc.M708735200
Gissen, P., Johnson, C.A., Morgan, N.V., Stapelbroek, J.M., Forshew, T., Cooper, W.N., McKiernan, P.J., Klomp, L.W.J., Morris, A.A.M., Wraith, J.E., McClean, P., Lynch, S.A., Thompson, R.J., Lo, B., Quarrell, O.W., Di Rocco, M., Trembath, R.C., Mandel, H., Wali, S., Karet, F.E., Knisely, A.S., Houwen, R.H.J., Kelly, D.A., Maher, E.R., 2004. Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis–renal dysfunction–cholestasis (ARC) syndrome. Nat. Genet. 36, 400–404. doi:10.1038/ng1325
Gissen, P., Tee, L., Johnson, C.A., Genin, E., Caliebe, A., Chitayat, D., Clericuzio, C., Denecke, J., Di Rocco, M., Fischler, B., FitzPatrick, D., García-Cazorla, A., Guyot, D., Jacquemont, S., Koletzko, S., Leheup, B., Mandel, H., Sanseverino, M.T.V., Houwen, R.H.J., McKiernan, P.J., Kelly, D.A., Maher, E.R., 2006. Clinical and molecular genetic features of ARC syndrome. Hum. Genet. 120, 396–409. doi:10.1007/s00439-006-0232-z
Goldschmidt, M.L., Mourya, R., Connor, J., Dexheimer, P., Karns, R., Miethke, A., Sheridan, R., Zhang, K., Bezerra, J.A., 2016. Increased frequency of double and triple heterozygous gene

variants in children with intrahepatic cholestasis. Hepatol. Res. Off. J. Jpn. Soc. Hepatol. 46, 306–311. doi:10.1111/hepr.12545
Gomez-Ospina, N., Potter, C.J., Xiao, R., Manickam, K., Kim, M.-S., Kim, K.H., Shneider, B.L., Picarsic, J.L., Jacobson, T.A., Zhang, J., He, W., Liu, P., Knisely, A.S., Finegold, M.J., Muzny, D.M., Boerwinkle, E., Lupski, J.R., Plon, S.E., Gibbs, R.A., Eng, C.M., Yang, Y., Washington, G.C., Porteus, M.H., Berquist, W.E., Kambham, N., Singh, R.J., Xia, F., Enns, G.M., Moore, D.D., 2016. Mutations in the nuclear bile acid receptor FXR cause progressive familial intrahepatic cholestasis. Nat. Commun. 7, 10713. doi:10.1038/ncomms10713
Gonzales, E., Grosse, B., Cassio, D., Davit-Spraul, A., Fabre, M., Jacquemin, E., 2012. Successful mutation-specific chaperone therapy with 4-phenylbutyrate in a child with progressive familial intrahepatic cholestasis type 2. J. Hepatol. 57, 695–698. doi:10.1016/j.jhep.2012.04.017
Greenberg, C.R., Barnes, J.G., Kogan, S., Seargeant, L.E., 2015. A rare case of Niemann–Pick disease type C without neurological involvement in a 66-year-old patient. Mol. Genet. Metab. Rep. 3, 18–20. doi:10.1016/j.ymgmr.2015.02.004
Greer, W.L., Dobson, M.J., Girouard, G.S., Byers, D.M., Riddell, D.C., Neumann, P.E., 1999. Mutations in NPC1 highlight a conserved NPC1-specific cysteine-rich domain. Am. J. Hum. Genet. 65, 1252–1260.
Grochowski, C.M., Rajagopalan, R., Falsey, A.M., Loomes, K.M., Piccoli, D.A., Krantz, I.D., Devoto, M., Spinner, N.B., 2015. Exome sequencing reveals compound heterozygous mutations in ATP8B1 in a JAG1/NOTCH2 mutation-negative patient with clinically diagnosed Alagille syndrome. Am. J. Med. Genet. A. 167, 891–893. doi:10.1002/ajmg.a.36946
Grosse, S.D., Khoury, M.J., 2006. What is the clinical utility of genetic testing? Genet. Med. 8, 448–450. doi:10.1097/01.gim.0000227935.26763.c6
Hacia, J.G., 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nat. Genet. 21, 42–47. doi:10.1038/4469
Hadj-Rabia, S., Baala, L., Vabres, P., Hamel-Teillac, D., Jacquemin, E., Fabre, M., Lyonnet, S., de Prost, Y., Munnich, A., Hadchouel, M., others, 2004. Claudin-1 gene mutations in neonatal sclerosing cholangitis associated with ichthyosis: a tight junction disease. Gastroenterology 127, 1386–1390.
Hao, X., Liu, S., Dong, Q., Zhang, H., Zhao, J., Su, L., 2014. Whole Exome Sequencing Identifies Recessive PKHD1 Mutations in a Chinese Twin Family with Caroli Disease. PLoS ONE 9, e92661. doi:10.1371/journal.pone.0092661
Hartley, J.L., Gissen, P., Kelly, D.A., 2013. Alagille Syndrome and Other Hereditary Causes of Cholestasis. Clin. Liver Dis. 17, 279–300. doi:10.1016/j.cld.2012.12.004
Hellen, B., 2009. Splice Site Tools A Comparative Analysis Report. Natl. Genet. Ref. Lab. Manch.
Herbst, S.M., Schirmer, S., Posovszky, C., Jochum, F., Rödl, T., Schroeder, J.A., Barth, T.F., Hehr, U., Melter, M., Vermehren, J., 2015. Taking the next step forward – Diagnosing inherited infantile cholestatic disorders with next generation sequencing. Mol. Cell. Probes 29, 291–298. doi:10.1016/j.mcp.2015.03.001
Héron, B., Valayannopoulos, V., Baruteau, J., Chabrol, B., Ogier, H., Latour, P., Dobbelaere, D., Eyer, D., Labarthe, F., Maurey, H., others, 2012. Miglustat therapy in the French cohort of paediatric patients with Niemann-Pick disease type C. Orphanet J Rare Dis 7, 36.
Hertz, J.M., 2009. Alport syndrome. Molecular genetic aspects. Dan. Med. Bull. 56, 105–152.
Human Genome Project Information [WWW Document], 2014. URL http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml (accessed 6.10.16).
Hutchin, T., Preece, M.A., Hendriksz, C., Chakrapani, A., McClelland, V., Okumura, F., Song, Y.-Z., Iijima, M., Kobayashi, K., Saheki, T., McKiernan, P., Baumann, U., 2009. Neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD) as a cause of liver disease in infants in the UK. J. Inherit. Metab. Dis. 32, 151–155. doi:10.1007/s10545-009-1116-x
Imrie, J., Dasgupta, S., Besley, G.T.N., Harris, C., Heptinstall, L., Knight, S., Vanier, M.T., Fensom, A.H., Ward, C., Jacklin, E., Whitehouse, C., Wraith, J.E., 2007. The natural history of Niemann–Pick disease type C in the UK. J. Inherit. Metab. Dis. 30, 51–59. doi:10.1007/s10545-006-0384-7
Ioannou, Y.A., 2000. The Structure and Function of the Niemann–Pick C1 Protein. Mol. Genet. Metab. 71, 175–181. doi:10.1006/mgme.2000.3061
Jacquemin, E., 2012. Progressive familial intrahepatic cholestasis. Clin. Res. Hepatol. Gastroenterol. 36, S26–S35.
Jacquemin, E., De Vree, J.M., Cresteil, D., Sokal, E.M., Sturm, E., Dumont, M., Scheffer, G.L., Paul, M., Burdelski, M., Bosma, P.J., Bernard, O., Hadchouel, M., Elferink, R.P., 2001. The wide

spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. Gastroenterology 120, 1448–1458.
Jahnova, H., Dvorakova, L., Vlaskova, H., Hulkova, H., Poupetova, H., Hrebicek, M., Jesina, P., 2014. Observational, retrospective study of a large cohort of patients with Niemann-Pick disease type C in the Czech Republic: a surprisingly stable diagnostic rate spanning almost 40 years. Orphanet J. Rare Dis. 9, 1.
Jeppsson, J.-O., 1976. Amino acid substitution Glu→Lys in α1-antitrypsin PiZ. FEBS Lett. 65, 195–197. doi:10.1016/0014-5793(76)80478-4
Jirsa, M., 2011. [Molecular diagnosis of hereditary canalicular cholestasis and familial hyperbilirubinemias]. Casopís Lékařů Českých 150, 7–13.
Jmoudiak, M., Futerman, A.H., 2005. Gaucher disease: pathological mechanisms and modern management. Br. J. Haematol. 129, 178–188. doi:10.1111/j.1365-2141.2004.05351.x
Kamath, B., Bason, L., Piccoli, D., Krantz, I., Spinner, N., 2003. Consequences of JAG1 mutations. J. Med. Genet. 40, 891–895. doi:10.1136/jmg.40.12.891
Kelly, D.A., 2006. Intestinal Failure–Associated Liver Disease: What Do We Know Today? Gastroenterology 130, S70–S77. doi:10.1053/j.gastro.2005.10.066
Kelly, D.A., Portmann, B., Mowat, A.P., Sherlock, S., Lake, B.D., 1993. Niemann-Pick disease type C: diagnosis and outcome in children, with particular reference to liver disease. J. Pediatr. 123, 242–247.
Kimura, N., Kubo, N., Narumi, S., Toyoki, Y., Ishido, K., Kudo, D., Umehara, M., Yakoshi, Y., Hakamada, K., 2013. Liver transplantation versus conservative treatment for adult-onset type II citrullinemia: our experience and a review of the literature. Transplant. Proc. 45, 3432–3437. doi:10.1016/j.transproceed.2013.06.016
Klomp, L., Bull, L., Knisely, A., Vanderdoelen, M., Juijn, J., Berger, R., Forget, S., Nielsen, I., Eiberg, H., Houwen, R., 2000. A missense mutation in is associated with greenland familial cholestasis. Hepatology 32, 1337–1341. doi:10.1053/jhep.2000.20520
Klomp, L.W.J., Vargas, J.C., van Mil, S.W.C., Pawlikowska, L., Strautnieks, S.S., van Eijk, M.J.T., Juijn, J.A., Pabón-Peña, C., Smith, L.B., DeYoung, J.A., Byrne, J.A., Gombert, J., van der Brugge, G., Berger, R., Jankowska, I., Pawlowska, J., Villa, E., Knisely, A.S., Thompson, R.J., Freimer, N.B., Houwen, R.H.J., Bull, L.N., 2004. Characterization of mutations in *ATP8B1* associated with hereditary cholestasis. Hepatology 40, 27–38. doi:10.1002/hep.20285
Knisely, A.S., Strautnieks, S.S., Meier, Y., Stieger, B., Byrne, J.A., Portmann, B.C., Bull, L.N., Pawlikowska, L., Bilezikçi, B., Ozçay, F., László, A., Tiszlavicz, L., Moore, L., Raftos, J., Arnell, H., Fischler, B., Németh, A., Papadogiannakis, N., Cielecka-Kuszyk, J., Jankowska, I., Pawłowska, J., Melín-Aldana, H., Emerick, K.M., Whitington, P.F., Mieli-Vergani, G., Thompson, R.J., 2006. Hepatocellular carcinoma in ten children under five years of age with bile salt export pump deficiency. Hepatol. Baltim. Md 44, 478–486. doi:10.1002/hep.21287
Ko, D.C., Binkley, J., Sidow, A., Scott, M.P., 2003. The integrity of a cholesterol-binding pocket in Niemann–Pick C2 protein is necessary to control lysosome cholesterol levels. Proc. Natl. Acad. Sci. U. S. A. 100, 2518–2525. doi:10.1073/pnas.0530027100
Kobayashi, K., 2003. Screening of nine SLC25A13 mutations: their frequency in patients with citrin deficiency and high carrier rates in Asian populations. Mol. Genet. Metab. 80, 356–359. doi:10.1016/S1096-7192(03)00140-9
Kobayashi, K., Sinasac, D.S., Iijima, M., Boright, A.P., Begum, L., Lee, J.R., Yasuda, T., Ikeda, S., Hirano, R., Terazono, H., Crackower, M.A., Kondo, I., Tsui, L.C., Scherer, S.W., Saheki, T., 1999. The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. Nat. Genet. 22, 159–163. doi:10.1038/9667
Krantz, I.D., Piccoli, D.A., Spinner, N.B., 1997. Alagille syndrome. J. Med. Genet. 34, 152–157.
Kubitz, R., Keitel, V., Scheuring, S., Köhrer, K., Häussinger, D., 2006. Benign recurrent intrahepatic cholestasis associated with mutations of the bile salt export pump. J. Clin. Gastroenterol. 40, 171–175.
Laberge, C., 1969. Hereditary tyrosinemia in a French Canadian isolate. Am. J. Hum. Genet. 21, 36–45.
Lam, P., Pearson, C.L., Soroka, C.J., Xu, S., Mennone, A., Boyer, J.L., 2007. Levels of plasma membrane expression in progressive and benign mutations of the bile salt export pump (Bsep/Abcb11) correlate with severity of cholestatic diseases. AJP Cell Physiol. 293, C1709–C1716. doi:10.1152/ajpcell.00327.2007
Li, L., Krantz, I.D., Deng, Y., Genin, A., Banta, A.B., Collins, C.C., Qi, M., Trask, B.J., Kuo, W.L., Cochran, J., Costa, T., Pierpont, M.E., Rand, E.B., Piccoli, D.A., Hood, L., Spinner, N.B.,

1997. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat. Genet. 16, 243–251. doi:10.1038/ng0797-243
Lindblad, A., Glaumann, H., Strandvik, B., 1999. Natural history of liver disease in cystic fibrosis. Hepatology 30, 1151–1158.
Lindor, N.M., Schahl, K.A., Johnson, K.J., Hunt, K.S., Mensink, K.A., Wieben, E.D., Klee, E., Black, J.L., Highsmith, W.E., Thibodeau, S.N., Ferber, M.J., Aypar, U., Ji, Y., Graham, R.P., Fiksdal, A.S., Sarangi, V., Ormond, K.E., Riegert-Johnson, D.L., McAllister, T.M., Farrugia, G., McCormick, J.B., 2015. Whole-Exome Sequencing of 10 Scientists: Evaluation of the Process and Outcomes. Mayo Clin. Proc. 90, 1327–1337. doi:10.1016/j.mayocp.2015.05.021
Liu, L.-Y., Wang, Z.-L., Wang, X.-H., Zhu, Q.-R., Wang, J.-S., 2009. ABCB11 gene mutations in Chinese children with progressive intrahepatic cholestasis and low γ glutamyltransferase: ABCB11 mutations in Chinese children. Liver Int. 30, 809–815. doi:10.1111/j.1478-3231.2009.02112.x
Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. 30, 434–439. doi:10.1038/nbt.2198
Lorent, K., 2004. Inhibition of Jagged-mediated Notch signaling disrupts zebrafish biliary development and generates multi-organ defects compatible with an Alagille syndrome phenocopy. Development 131, 5753–5766. doi:10.1242/dev.01411
Lyseng-Williamson, K.A., 2014. Miglustat: a review of its use in Niemann-Pick disease type C. Drugs 74, 61–74. doi:10.1007/s40265-013-0164-6
Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods 7, 111–118. doi:10.1038/nmeth.1419
McDaniell, R., Warthen, D.M., Sanchez-Lara, P.A., Pai, A., Krantz, I.D., Piccoli, D.A., Spinner, N.B., 2006. NOTCH2 Mutations Cause Alagille Syndrome, a Heterogeneous Disorder of the Notch Signaling Pathway. Am. J. Hum. Genet. 79, 169–173.
McDonald-McGinn, D.M., Emanuel, B.S., Zackai, E.H., 1999. 22q11.2 Deletion Syndrome, in: Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J., Bird, T.D., Fong, C.-T., Mefford, H.C., Smith, R.J., Stephens, K. (Eds.), GeneReviews(®). University of Washington, Seattle, Seattle (WA).
McKay, K.E., Bruce, C.K., Hartley, J.L., Knisely, A.S., Baumann, U., Bockisch, S.-S., Sturm, E., Hendriksz, C.J., Kelly, D.A., Macdonald, F., Gissen, P., 2013. Mutation detection in cholestatic patients using microarray resequencing of ATP8B1 and ABCB11. F1000Research 2. doi:10.12688/f1000research.2-32.v2
Mckiernan, P.J., 2002. Neonatal cholestasis. Semin. Neonatol. 7, 153–165. doi:10.1053/siny.2002.0103
Meienberg, J., Bruggmann, R., Oexle, K., Matyas, G., 2016. Clinical sequencing: is WGS the better WES? Hum. Genet. 135, 359–362. doi:10.1007/s00439-015-1631-9
Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46. doi:10.1038/nrg2626
Mieli-Vergani, G., Howard, E.R., Portman, B., Mowat, A.P., 1989. Late referral for biliary atresia--missed opportunities for effective surgery. Lancet Lond. Engl. 1, 421–423.
Millat, G., Marçais, C., Tomasetto, C., Chikh, K., Fensom, A.H., Harzer, K., Wenger, D.A., Ohno, K., Vanier, M.T., 2001. Niemann-Pick C1 disease: correlations between NPC1 mutations, levels of NPC1 protein, and phenotypes emphasize the functional significance of the putative sterol-sensing domain and of the cysteine-rich luminal loop. Am. J. Hum. Genet. 68, 1373–1385.
Miller, J.N., Pearce, D.A., 2014. Nonsense-Mediated Decay in Genetic Disease: Friend or Foe? Mutat. Res. Rev. Mutat. Res. 0, 52–64. doi:10.1016/j.mrrev.2014.05.001
Moreira, R.K., Cabral, R., Cowles, R.A., Lobritto, S.J., 2012. Biliary atresia: a multidisciplinary approach to diagnosis and management. Arch. Pathol. Lab. Med. 136, 746–760.
Moskowitz, S.M., Chmiel, J.F., Sternen, D.L., Cheng, E., Cutting, G.R., 1993. CFTR-Related Disorders, in: Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J., Bird, T.D., Fong, C.-T., Mefford, H.C., Smith, R.J., Stephens, K. (Eds.), GeneReviews(®). University of Washington, Seattle, Seattle (WA).
Mullenbach, R., 2005. ATP8B1 mutations in British cases with intrahepatic cholestasis of pregnancy. Gut 54, 829–834. doi:10.1136/gut.2004.058115
Mullenbach, R., Linton, K., Wiltshire, S., Weerasekera, N., Chambers, J., Elias, E., Higgins, C., Johnston, D., McCarthy, M., Williamson, C., 2003. ABCB4 gene sequence variation in women with intrahepatic cholestasis of pregnancy. J. Med. Genet. 40, e70. doi:10.1136/jmg.40.5.e70

Nakamura, K., Tanoue, A., 2013. Etiology of biliary atresia as a developmental anomaly: recent advances. J. Hepato-Biliary-Pancreat. Sci. 20, 459–464. doi:10.1007/s00534-013-0604-4
Naureckiene, S., 2000a. Identification of HE1 as the Second Gene of Niemann-Pick C Disease. Science 290, 2298–2301. doi:10.1126/science.290.5500.2298
Naureckiene, S., 2000b. Identification of HE1 as the Second Gene of Niemann-Pick C Disease. Science 290, 2298–2301. doi:10.1126/science.290.5500.2298
NICE Guidelines, 2010. Jaundice in newborn babies under 28 days | Guidance and guidelines | NICE [WWW Document]. URL https://www.nice.org.uk/guidance/cg98 (accessed 5.31.16).
Nobili, V., Di Giandomenico, S., Francalanci, P., Callea, F., Marcellini, M., Santorelli, F.M., 2006. A new ABCB11 mutation in two Italian children with familial intrahepatic cholestasis. J. Gastroenterol. 41, 598–603. doi:10.1007/s00535-006-1816-z
Nogueira, C., Almeida, L.S., Nesti, C., Pezzini, I., Videira, A., Vilarinho, L., Santorelli, F.M., 2014. Syndromes associated with mitochondrial DNA depletion. Ital. J. Pediatr. 40, 34. doi:10.1186/1824-7288-40-34
Oda, T., Elkahloun, A.G., Pike, B.L., Okajima, K., Krantz, I.D., Genin, A., Piccoli, D.A., Meltzer, P.S., Spinner, N.B., Collins, F.S., Chandrasekharappa, S.C., 1997. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. Nat. Genet. 16, 235–242. doi:10.1038/ng0797-235
Oliver, G.R., Hart, S.N., Klee, E.W., 2015. Bioinformatics for Clinical Next Generation Sequencing. Clin. Chem. 61, 124–135. doi:10.1373/clinchem.2014.224360
Painter, J.N., Savander, M., Ropponen, A., Nupponen, N., Riikonen, S., Ylikorkala, O., Lehesjoki, A.-E., Aittomäki, K., 2005. Sequence variation in the ATP8B1 gene and intrahepatic cholestasis of pregnancy. Eur. J. Hum. Genet. 13, 435–439. doi:10.1038/sj.ejhg.5201355
Park, W.D., O'Brien, J.F., Lundquist, P.A., Kraft, D.L., Vockley, C.W., Karnes, P.S., Patterson, M.C., Snow, K., 2003. Identification of 58 novel mutations in Niemann-Pick disease type C: Correlation with biochemical phenotype and importance of *PTC1* -like domains in *NPC1*. Hum. Mutat. 22, 313–325. doi:10.1002/humu.10255
Patterson, M., 2000. Niemann-Pick Disease Type C, in: Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Amemiya, A., Bean, L.J., Bird, T.D., Fong, C.-T., Mefford, H.C., Smith, R.J., Stephens, K. (Eds.), GeneReviews(®). University of Washington, Seattle, Seattle (WA).
Patterson, M.C., Hendriksz, C.J., Walterfang, M., Sedel, F., Vanier, M.T., Wijburg, F., NP-C Guidelines Working Group, 2012. Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. Mol. Genet. Metab. 106, 330–344. doi:10.1016/j.ymgme.2012.03.012
Pauli-Magnus, C., Kerb, R., Fattinger, K., Lang, T., Anwald, B., Kullak-Ublick, G.A., Beuers, U., Meier, P.J., 2004. BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. Hepatol. Baltim. Md 39, 779–791. doi:10.1002/hep.20159
Paulusma, C.C., de Waart, D.R., Kunne, C., Mok, K.S., Elferink, R.P.J.O., 2009. Activity of the Bile Salt Export Pump (ABCB11) Is Critically Dependent on Canalicular Membrane Cholesterol Content. J. Biol. Chem. 284, 9947–9954. doi:10.1074/jbc.M808667200
Pawlikowska, L., Strautnieks, S., Jankowska, I., Czubkowski, P., Emerick, K., Antoniou, A., Wanty, C., Fischler, B., Jacquemin, E., Wali, S., Blanchard, S., Nielsen, I.-M., Bourke, B., McQuaid, S., Lacaille, F., Byrne, J.A., van Eerde, A.M., Kolho, K.-L., Klomp, L., Houwen, R., Bacchetti, P., Lobritto, S., Hupertz, V., McClean, P., Mieli-Vergani, G., Shneider, B., Nemeth, A., Sokal, E., Freimer, N.B., Knisely, A.S., Rosenthal, P., Whitington, P.F., Pawlowska, J., Thompson, R.J., Bull, L.N., 2010. Differences in presentation and progression between severe FIC1 and BSEP deficiencies. J. Hepatol. 53, 170–178. doi:10.1016/j.jhep.2010.01.034
Phaneuf, D., Labelle, Y., Bérubé, D., Arden, K., Cavenee, W., Gagné, R., Tanguay, R.M., 1991. Cloning and expression of the cDNA encoding human fumarylacetoacetate hydrolase, the enzyme deficient in hereditary tyrosinemia: assignment of the gene to chromosome 15. Am. J. Hum. Genet. 48, 525–535.
Pipalia, N.H., Cosner, C.C., Huang, A., Chatterjee, A., Bourbon, P., Farley, N., Helquist, P., Wiest, O., Maxfield, F.R., 2011. Histone deacetylase inhibitor treatment dramatically reduces cholesterol accumulation in Niemann-Pick type C1 mutant human fibroblasts. Proc. Natl. Acad. Sci. U. S. A. 108, 5620–5625. doi:10.1073/pnas.1014890108
Poupon, R., Rosmorduc, O., Boëlle, P.Y., Chrétien, Y., Corpechot, C., Chazouillères, O., Housset, C., Barbu, V., 2013. Genotype-phenotype relationships in the low-phospholipid-associated cholelithiasis syndrome: a study of 156 consecutive patients. Hepatol. Baltim. Md 58, 1105–1110. doi:10.1002/hep.26424

Reichardt, J.K., Woo, S.L., 1991. Molecular basis of galactosemia: mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase. Proc. Natl. Acad. Sci. U. S. A. 88, 2633–2637.
Ribeiro, I., Marcão, A., Amaral, O., Sá Miranda, M., Vanier, M.T., Millat, G., 2001. Niemann-Pick type C disease: NPC1 mutations associated with severe and mild cellular cholesterol trafficking alterations. Hum. Genet. 109, 24–32. doi:10.1007/s004390100531
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., 2015. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. Off. J. Am. Coll. Med. Genet. 17, 405–424. doi:10.1038/gim.2015.30
Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245, 1066–1073.
Roberts, E.A., 2003. Neonatal hepatitis syndrome. Semin. Neonatol. SN 8, 357–374. doi:10.1016/S1084-2756(03)00093-9
Rosmorduc, O., Hermelin, B., Boelle, P.-Y., Parc, R., Taboury, J., Poupon, R., 2003. ABCB4 gene mutation-associated cholelithiasis in adults. Gastroenterology 125, 452–459.
Rosmorduc, O., Hermelin, B., Poupon, R., 2001. MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis. Gastroenterology 120, 1459–1467.
Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T., Bustillo, J., 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348–352. doi:10.1038/nature10242
Saheki, T., Kobayashi, K., 2002. Mitochondrial aspartate glutamate carrier (citrin) deficiency as the cause of adult-onset type II citrullinemia (CTLN2) and idiopathic neonatal hepatitis (NICCD). J. Hum. Genet. 47, 333–341.
Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74, 5463–5467.
Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G., 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 30, e57.
Schulze, K., Nault, J.-C., Villanueva, A., 2016. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J. Hepatol. doi:10.1016/j.jhep.2016.05.035
Schwarz, M., Wright, A.C., Davis, D.L., Nazer, H., Björkhem, I., Russell, D.W., 2000. The bile acid synthetic gene 3β-hydroxy-Δ5-C27-steroid oxidoreductase is mutated in progressive intrahepatic cholestasis. J. Clin. Invest. 106, 1175–1184.
Sener, E.F., Canatan, H., Ozkul, Y., 2016. Recent Advances in Autism Spectrum Disorders: Applications of Whole Exome Sequencing Technology. Psychiatry Investig. 13, 255–264. doi:10.4306/pi.2016.13.3.255
Setchell, K.D., Schwarz, M., O'Connell, N.C., Lund, E.G., Davis, D.L., Lathe, R., Thompson, H.R., Weslie Tyson, R., Sokol, R.J., Russell, D.W., 1998. Identification of a new inborn error in bile acid synthesis: mutation of the oxysterol 7alpha-hydroxylase gene causes severe neonatal liver disease. J. Clin. Invest. 102, 1690–1703.
Setchell, K.D., Suchy, F.J., Welsh, M.B., Zimmer-Nechemias, L., Heubi, J., Balistreri, W.F., 1988. Delta 4-3-oxosteroid 5 beta-reductase deficiency described in identical twins with neonatal hepatitis. A new inborn error in bile acid synthesis. J. Clin. Invest. 82, 2148–2157.
Sevin, M., Lesca, G., Baumann, N., Millat, G., Lyon-Caen, O., Vanier, M.T., Sedel, F., 2006. The adult form of Niemann-Pick disease type C. Brain 130, 120–133. doi:10.1093/brain/awl260
Shen, T., Lee, A., Shen, C., Lin, C.J., 2015. The long tail and rare disease research: the impact of next-generation sequencing for rare Mendelian disorders. Genet. Res. 97. doi:10.1017/S0016672315000166
Song, Y.-Z., Deng, M., Chen, F.-P., Wen, F., Guo, L., Cao, S.-L., Gong, J., Xu, H., Jiang, G.-Y., Zhong, L., Kobayashi, K., Saheki, T., Wang, Z.-N., 2011. Genotypic and phenotypic features of citrin deficiency: five-year experience in a Chinese pediatric center. Int. J. Mol. Med. 28, 33–40. doi:10.3892/ijmm.2011.653

Spiegel, R., Raas-Rothschild, A., Reish, O., Regev, M., Meiner, V., Bargal, R., Sury, V., Meir, K., Nadjari, M., Hermann, G., Iancu, T.C., Shalev, S.A., Zeigler, M., 2009. The clinical spectrum of fetal Niemann-Pick type C. Am. J. Med. Genet. A. 149A, 446–450. doi:10.1002/ajmg.a.32642

Stone, A., Chau, C., Eaton, C., Foran, E., Kapur, M., Prevatt, E., Belkin, N., Kerr, D., Kohlin, T., Williamson, P., 2012. Biochemical Characterization of P4-ATPase Mutations Identified in Patients with Progressive Familial Intrahepatic Cholestasis. J. Biol. Chem. 287, 41139–41151. doi:10.1074/jbc.M112.413039

Strautnieks, S.S., Bull, L.N., Knisely, A.S., Kocoshis, S.A., Dahl, N., Arnell, H., Sokal, E., Dahan, K., Childs, S., Ling, V., Tanner, M.S., Kagalwalla, A.F., Németh, A., Pawlowska, J., Baker, A., Mieli-Vergani, G., Freimer, N.B., Gardiner, R.M., Thompson, R.J., 1998. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat. Genet. 20, 233–238. doi:10.1038/3034

Strautnieks, S.S., Byrne, J.A., Pawlikowska, L., Cebecauerová, D., Rayner, A., Dutton, L., Meier, Y., Antoniou, A., Stieger, B., Arnell, H., Özçay, F., Al–Hussaini, H.F., Bassas, A.F., Verkade, H.J., Fischler, B., Németh, A., Kotalová, R., Shneider, B.L., Cielecka–Kuszyk, J., McClean, P., Whitington, P.F., Sokal, É., Jirsa, M., Wali, S.H., Jankowska, I., Pawłowska, J., Mieli–Vergani, G., Knisely, A.S., Bull, L.N., Thompson, R.J., 2008. Severe Bile Salt Export Pump Deficiency: 82 Different ABCB11 Mutations in 109 Families. Gastroenterology 134, 1203–1214.e8. doi:10.1053/j.gastro.2008.01.038

Sveger, T., 1976. Liver disease in alpha1-antitrypsin deficiency detected by screening of 200,000 infants. N. Engl. J. Med. 294, 1316–1321. doi:10.1056/NEJM197606102942404

Sveger, T., Eriksson, S., 1995. The liver in adolescents with alpha 1-antitrypsin deficiency. Hepatol. Baltim. Md 22, 514–517.

Svensson, J., Makin, E., 2011. Gallstone disease in children. Semin. Pediatr. Surg. 21, 255–265. doi:10.1053/j.sempedsurg.2012.05.008

Tabata, A., Sheng, J.-S., Ushikai, M., Song, Y.-Z., Gao, H.-Z., Lu, Y.-B., Okumura, F., Iijima, M., Mutoh, K., Kishida, S., Saheki, T., Kobayashi, K., 2008. Identification of 13 novel mutations including a retrotransposal insertion in SLC25A13 gene and frequency of 30 mutations found in patients with citrin deficiency. J. Hum. Genet. 53, 534–545. doi:10.1007/s10038-008-0282-2

Tamamori, A., Okano, Y., Ozaki, H., Fujimoto, A., Kajiwara, M., Fukuda, K., Kobayashi, K., Saheki, T., Tagami, Y., Yamano, T., 2002. Neonatal intrahepatic cholestasis caused by citrin deficiency: severe hepatic dysfunction in an infant requiring liver transplantation. Eur. J. Pediatr. 161, 609–613. doi:10.1007/s00431-002-1045-2

Tamura, H., Takahashi, T., Ban, N., Torisu, H., Ninomiya, H., Takada, G., Inagaki, N., 2006. Niemann-Pick type C disease: novel NPC1 mutations and characterization of the concomitant acid sphingomyelinase deficiency. Mol. Genet. Metab. 87, 113–121. doi:10.1016/j.ymgme.2005.07.025

T. Strachan, 2011. Human molecular genetics / Tom Strachan and Andrew Read., 4th ed. ed. Garland Science.

Thomas, S., Encha-Razavi, F., Devisme, L., Etchevers, H., Bessieres-Grattagliano, B., Goudefroye, G., Elkhartoufi, N., Pateau, E., Ichkou, A., Bonnière, M., Marcorelle, P., Parent, P., Manouvrier, S., Holder, M., Laquerrière, A., Loeuillet, L., Roume, J., Martinovic, J., Mougou-Zerelli, S., Gonzales, M., Meyer, V., Wessner, M., Feysot, C.B., Nitschke, P., Leticee, N., Munnich, A., Lyonnet, S., Wookey, P., Gyapay, G., Foliguet, B., Vekemans, M., Attié-Bitach, T., 2010. High-throughput sequencing of a 4.1 Mb linkage interval reveals FLVCR2 deletions and mutations in lethal cerebral vasculopathy. Hum. Mutat. 31, 1134–1141. doi:10.1002/humu.21329

Tibesar, E., Karwowski, C., Hertel, P., Scheimann, A., Karnsakul, W., 2014. Two Cases of Progressive Familial Intrahepatic Cholestasis Type 2 Presenting with Severe Coagulopathy without Jaundice. Case Rep. Pediatr. 2014. doi:10.1155/2014/185923

Tsuji, S., Choudary, P.V., Martin, B.M., Stubblefield, B.K., Mayor, J.A., Barranger, J.A., Ginns, E.I., 1987. A Mutation in the Human Glucocerebrosidase Gene in Neuronopathic Gaucher's Disease. N. Engl. J. Med. 316, 570–575. doi:10.1056/NEJM198703053161002

Turnpenny, P.D., Ellard, S., 2012. Alagille syndrome: pathogenesis, diagnosis and management. Eur. J. Hum. Genet. 20, 251–257.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., Sidow, A., Fire, A., Johnson, S.M., 2008. A high-resolution,

nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 18, 1051–1063. doi:10.1101/gr.076463.108
van de Steeg, E., Stránecký, V., Hartmannová, H., Nosková, L., Hřebíček, M., Wagenaar, E., van Esch, A., de Waart, D.R., Oude Elferink, R.P.J., Kenworthy, K.E., Sticová, E., al-Edreesi, M., Knisely, A.S., Kmoch, S., Jirsa, M., Schinkel, A.H., 2012. Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. J. Clin. Invest. 122, 519–528. doi:10.1172/JCI59526
van Mil, S.W.C., Milona, A., Dixon, P.H., Mullenbach, R., Geenes, V.L., Chambers, J., Shevchuk, V., Moore, G.E., Lammert, F., Glantz, A.G., Mattsson, L., Whittaker, J., Parker, M.G., White, R., Williamson, C., 2007. Functional Variants of the Central Bile Acid Sensor FXR Identified in Intrahepatic Cholestasis of Pregnancy. Gastroenterology 133, 507–516. doi:10.1053/j.gastro.2007.05.015
van Mil, S.W.C., van der Woerd, W.L., van der Brugge, G., Sturm, E., Jansen, P.L.M., Bull, L.N., van den Berg, I.E.T., Berger, R., Houwen, R.H.J., Klomp, L.W.J., 2004a. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 127, 379–384. doi:10.1053/j.gastro.2004.04.065
van Mil, S.W.C., van der Woerd, W.L., van der Brugge, G., Sturm, E., Jansen, P.L.M., Bull, L.N., van den Berg, I.E.T., Berger, R., Houwen, R.H.J., Klomp, L.W.J., 2004b. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 127, 379–384. doi:10.1053/j.gastro.2004.04.065
Vanier, M.T., 2015. Complex lipid trafficking in Niemann-Pick disease type C. J. Inherit. Metab. Dis. 38, 187–199. doi:10.1007/s10545-014-9794-4
Vanier, M.T., 2010. Orphanet Journal of Rare Diseases. Orphanet J. Rare Dis. 5, 16.
Vanier, M.T., Duthel, S., Rodriguez-Lafrasse, C., Pentchev, P., Carstea, E.D., 1996. Genetic heterogeneity in Niemann-Pick C disease: a study using somatic cell hybridization and linkage analysis. Am. J. Hum. Genet. 58, 118.
Vanier, M.T., Millat, G., 2003. Niemann-Pick disease type C. Clin. Genet. 64, 269–281.
Vanier, M.T., Wenger, D.A., Comly, M.E., Rousson, R., Brady, R.O., Pentchev, P.G., 1988. Niemann-Pick disease group C: clinical variability and diagnosis based on defective cholesterol esterification. A collaborative study on 70 patients. Clin. Genet. 33, 331–348.
Wada, M., Toh, S., Taniguchi, K., Nakamura, T., Uchiumi, T., Kohno, K., Yoshida, I., Kimura, A., Sakisaka, S., Adachi, Y., others, 1998. Mutations in the canalicular multispecific organic anion transporter (cMOAT) gene, a novel ABC transporter, in patients with hyperbilirubinemia II/Dubin-Johnson syndrome. Hum. Mol. Genet. 7, 203–207.
Wang, D.-W., Yin, Y.-M., Yao, Y.-M., 2013. Advances in the management of acute liver failure. World J. Gastroenterol. WJG 19, 7069–7077. doi:10.3748/wjg.v19.i41.7069
Wang, L., Soroka, C.J., Boyer, J.L., 2002. The role of bile salt export pump mutations in progressive familial intrahepatic cholestasis type II. J. Clin. Invest. 110, 965–972. doi:10.1172/JCI200215968
Waterham, H.R., Ebberink, M.S., 2012. Genetics and molecular basis of human peroxisome biogenesis disorders. Biochim. Biophys. Acta BBA - Mol. Basis Dis. 1822, 1430–1441. doi:10.1016/j.bbadis.2012.04.006
Wendum, D., Barbu, V., Rosmorduc, O., Arrivé, L., Fléjou, J.-F., Poupon, R., 2012. Aspects of liver pathology in adult patients with MDR3/ABCB4 gene mutations. Virchows Arch. Int. J. Pathol. 460, 291–298. doi:10.1007/s00428-012-1202-6
Whitington, P., 2007. Neonatal Hemochromatosis: A Congenital Alloimmune Hepatitis. Semin. Liver Dis. 27, 243–250. doi:10.1055/s-2007-985069
Whitington, P.F., Kelly, S., 2008. Outcome of Pregnancies at Risk for Neonatal Hemochromatosis Is Improved by Treatment With High-Dose Intravenous Immunoglobulin. PEDIATRICS 121, e1615–e1621. doi:10.1542/peds.2007-3107
Wijburg, F.A., Sedel, F., Pineda, M., Hendriksz, C.J., Fahey, M., Walterfang, M., Patterson, M.C., Wraith, J.E., Kolb, S.A., 2012. Development of a suspicion index to aid diagnosis of Niemann-Pick disease type C. Neurology 78, 1560–1567. doi:10.1212/WNL.0b013e3182563b82
Williams, S., 2012. Analysis of in silico tools for evaluating missense variants. Natl. Genet. Ref. Lab. Manch.
Wilson, R.K., Chen, C., Avdalovic, N., Burns, J., Hood, L., 1990. Development of an automated procedure for fluorescent DNA sequencing. Genomics 6, 626–634.
Woo, H.I., Park, H.-D., Lee, Y.-W., 2014. Molecular genetics of citrullinemia types I and II. Clin. Chim. Acta 431, 1–8. doi:10.1016/j.cca.2014.01.032

Xiong, H., Higaki, K., Wei, C.-J., Bao, X.-H., Zhang, Y.-H., Fu, N., Qin, J., Adachi, K., Kumura, Y., Ninomiya, H., Nanba, E., Wu, X.-R., 2012. Genotype/phenotype of 6 Chinese cases with Niemann-Pick disease type C. Gene 498, 332–335. doi:10.1016/j.gene.2012.01.026

Yamamoto, T., Nanba, E., Ninomiya, H., Higaki, K., Taniguchi, M., Zhang, H., Akaboshi, S., Watanabe, Y., Takeshima, T., Inui, K., Okada, S., Tanaka, A., Sakuragawa, N., Millat, G., Vanier, M.T., Morris, J.A., Pentchev, P.G., Ohno, K., 1999. NPC1 gene mutations in Japanese patients with Niemann-Pick disease type C. Hum. Genet. 105, 10–16. doi:10.1007/s004399900059

Yamamoto, Y., Sawa, R., Okamoto, N., Matsui, A., Yanagisawa, M., Ikemoto, S., 1986. Deletion 14q(q24.3 to q32.1) syndrome: significance of peculiar facial appearance in its diagnosis, and deletion mapping of Pi(alpha 1-antitrypsin). Hum. Genet. 74, 190–192.

Yasuda, T., Yamaguchi, N., Kobayashi, K., Nishi, I., Horinouchi, H., Jalil, M.A., Li, M.X., Ushikai, M., Iijima, M., Kondo, I., Saheki, T., 2000. Identification of two novel mutations in the SLC25A13 gene and detection of seven mutations in 102 patients with adult-onset type II citrullinemia. Hum. Genet. 107, 537–545.

Yerushalmi, B., Sokol, R.J., Narkewicz, M.R., Smith, D., Ashmead, J.W., Wenger, D.A., 2002. Niemann-pick disease type C in neonatal cholestasis at a North American Center. J. Pediatr. Gastroenterol. Nutr. 35, 44–50.

Zimmer, V., Krawczyk, M., Mahler, M., Weber, S.N., Müllenbach, R., Lammert, F., 2012. Severe hepatocellular dysfunction in obstetric cholestasis related to combined genetic variation in hepatobiliary transporters. Clin. Exp. Obstet. Gynecol. 39, 32–35.

List of websites and webtools

Roche 454 sequencing http://www.454.com/

Illumina http://www.illumina.com/

Applied Biosystems http://www.thermofisher.com/uk/en/home/brands/applied-biosystems.html

SIFT http://sift.jcvi.org/

PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/

AlignGVGD http://agvgd.iarc.fr/

MutationTaster http://www.mutationtaster.org/

Primer3 http://primer3.ut.ee/

ExonPrimer https://genome.ucsc.edu/

PrimerBlast http://www.ncbi.nlm.nih.gov/tools/primer-blast/

NRGL SNPCheck https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm

HGVS http://www.hgvs.org/mutnomen/

HGMD http://www.hgmd.cf.ac.uk/ac/index.php

Exome variant server http://evs.gs.washington.edu/EVS/

1000 genomes project http://browser.1000genomes.org/index.html