Chapter 1 Introduction

1.1 Pancreas

The pancreas derives its name from the Greek roots ‘pan’ and ‘creas’ meaning ‘all’ and ‘flesh’ respectively. It is composed of two morphologically and functionally distinct cell populations (Figure 1-1). The endocrine compartment consists of hormone-secreting cells that aggregate to form the islets of Langerhans that are associated with blood vessels, neurons and connective tissue. The endocrine islet is composed of α-, β-, δ-, ε- and PP-cells that secrete glucagon, insulin, somatostatin, ghrelin and pancreatic polypeptide respectively into the bloodstream to regulate glucose homeostasis and nutrient metabolism. The exocrine component comprises of acinar cells that form clusters at the end of pancreatic ducts. Acinar cells secrete enzymes including proteases, lipases and nucleases and duct cells actively secrete bicarbonate ions, mucins and transport digestive enzymes towards the dudodenum (Pan and Wright, 2011; Slack, 1995).

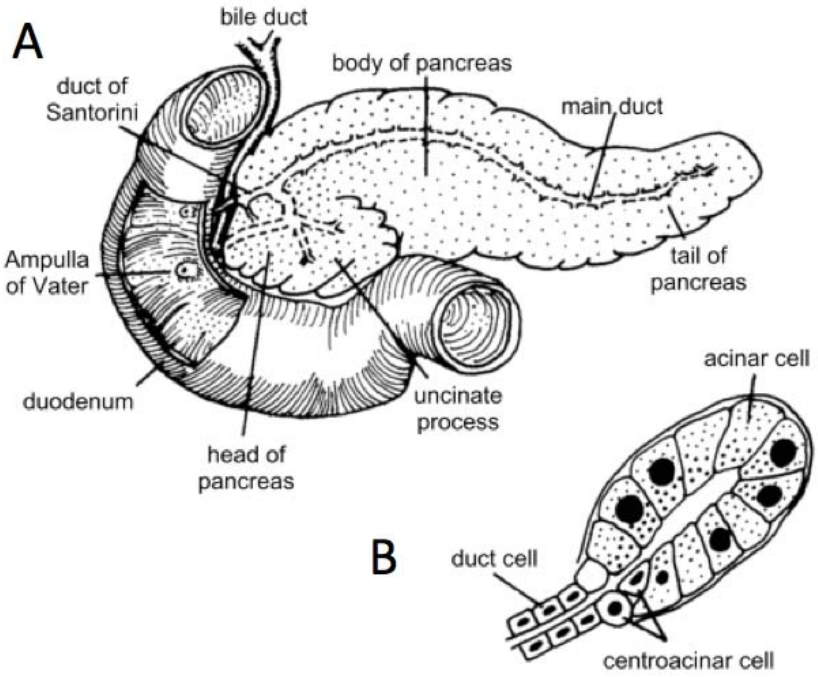

Figure 1-1 (A) Anatomy of the adult human pancreas, (B) Histology of a pancreatic acinus. Taken from Slack, 1995.

1.1.1 Pancreas development in humans

Most of the knowledge of pancreas development in mammals comes from studies in mice. However, human pancreas organogenesis largely resembles the process in the mouse. Pancreas formation starts with the regionalisation of the gut into distinct organ regions by a series of dorsoventral and anteroposterior patterning events. The endoderm evaginates forming the dorsal bud on the 26th day of gestation in humans (G26d) and approximately 9.5 days gestation in mice (E9.5). After 6 days in humans and approximately 12 hours in mice, the ventral bud emerges. The stalk region of the buds undergoes elongation while branching morphogenesis occur in the more apical region. As a consequence of gut rotation and dorsal bud elongation, the pancreatic buds come into contact and fuse at G37d to G42d in humans and E12 to E13 in mice. This causes the ducts of the ventral bud and the distal portion of the dorsal bud to fuse giving rise to the duct of Wirsung. The duct of the proximal portion of the dorsal bud remains as the duct of Santorini. Thereafter, the cellular architecture of the pancreas changes dramatically. A massive differentiation wave occurs and the three main pancreatic lineages are allocated. The number of endocrine cells particularly β-cells increases rapidly while acinar cell differentiation and rapid branching morphogenesis occur in the exocrine pancreas (Gittes, 2009; Pan and Wright, 2011).

1.1.2 Congenital disorders of the pancreas

Several congenital disorders and primary developmental anomalies of the pancreas have been described (Cano et al., 2007). They can be grouped into congenital developmental anomalies and primary malformations of the pancreas, congenital disorders with predominant secretory insufficiency, congenital disorders of pancreatic endocrine function and various other congenital diseases (Table 1-1). Since the zebrafish mutants that are being studied in this thesis show defects in development of the exocrine pancreas, I will describe the congenital disorders with predominant secretory insufficiency in more detail.

Entities	Genes
Developmental anomalies/malformations	
Pancreatic agenesis	*PDX1*/*IPF1*
Pancreatic and cerebellar agenesis	*PTF1A*
Pancreatic hypoplasia including congenital short pancreas	Unknown
Annular pancreas	Unknown
Pancreas divisum	Unknown
Pancreaticobiliary maljunction (common channel syndrome)	Unknown
Ectopic/heterotopic pancreas	Unknown
Congenital pancreatic cysts	
True pancreatic cysts	Unknown
Gastrointestinal duplication cysts	Unknown
Von Hippel-Lindau disease	*VHL*
Autosomal dominant polycystic kidney disease	*PKD1*, *PKD2*
Congenital secretory insufficiency	
Cystic fibrosis	*CFTR*
Shwachman-Diamond syndrome	*SBDS*
Johanson-Blizzard syndrome	*UBR1*
Pearson marrow-pancreas syndrome	mtDNA
Isolated enzyme deficiencies	*PRSS7*
Congenital disorders endocrine function	
Congenital hyperinsulinism and islet cell adenomatosis	*ABCC8*, *KCNJ11*, *GCK*, *HAD*, *INSR*, *GLUD1*
Transient neonatal diabetes mellitus	6q24 alterations, *ABCC8*
Permanent neonatal diabetes mellitus	*GCK*, *ABCC8*, *KCNJ11*
X-linked immunodysregulation, polyendocinopathy, and enteropathy syndrome	*FOXP3*
Miscellaneous	
Jeune syndrome	Unknown
Autosomal-recessive polycystic kidney disease	*PKHD1*
Beckwith-Wiedemann syndrome	11p15 alterations, *CDKN1C*, *H19*, *LIT1*
Renal-hepatic-pancreatic dysplasia	Unknown
Metabolic disorders	
Lipoprotein lipase deficiency	*LPL*
Apolipoprotein C-II deficiency	*APOC2*
Others	
Hereditary pancreatitis	*PRSS1*, (*SPINK1*)

Table 1-1 Congenital Pancreatic Disorders: Congenital developmental anomalies and primary malformations of the pancreas, congenital disorders with predominant secretory insufficiency, congenital disorders of pancreatic endocrine function and other congenital diseases affecting the pancreas and the genes mutated in the human disorder are shown. Taken from Cano et al., 2007.

1.1.2.1 Congenital disorders with predominant secretory insufficiency

Congenital disorders with predominant secretory insufficiency are rare and symptoms include loose and voluminous stools, excess fat in stools, abnormally low level of protein in the blood (hypoproteinemia) and failure to thrive. Exocrine pancreatic failure may manifest only after >90% of the exocrine cells are destroyed due to the high functional capacity of the exocrine pancreas (Stormon and Durie, 2002). Congenital pancreatic secretory insufficiency without diabetes results from either maturation defects or early-onset degeneration of acinar cells rather than an early developmental defect. In some conditions, the disease can progress to combined insufficiency of exocrine and endocrine pancreas.

Cystic fibrosis (CF; OMIM #219700) is the most common inherited cause of exocrine pancreatic insufficiency, with an incidence of approximately 1:2,500 (Roberts, 1990). The causative gene *cystic fibrosis transmembrane conductance regulator* (*CFTR)* was mapped by linkage analysis in the 1980s and encodes an epithelial chloride channel involved in electrolyte transport across epithelial cell membranes (Sheppard and Welsh, 1999). Only about 60% of CF patients have pancreatic insufficiency as newborns with most of the remaining patients losing pancreatic function over time (Waters et al., 1990). The pancreatic acinar tissue in CF is atrophic and pancreatic ducts are obstructed with secretory material. The basic pathogenesis is attributed to a progressive destruction of the pancreas possibly by viscous pancreatic secretions or other mechanisms (Freedman et al., 2000; Imrie et al., 1979).

Shwachman-Diamond syndrome (SDS; OMIM #260400) is the second most common cause of congenital exocrine pancreatic insufficiency, with an approximate incidence of 1:50,000 in the North American population (Cano et al., 2007; Narla and Ebert, 2010). It is an autosomal recessive ribosomopathy (Section 1.4.2), and in addition to the pancreatic defects, it is characterised by haematologic abnormalities most commonly an abnormally low number of neutrophils (neutropenia) and an increased risk of leukaemia. Other clinical features include short stature, skeletal abnormalities, a decrease in the number of red blood cells or a lower level of haemoglobin in blood (anaemia) and a relative decrease in number of platelets in blood (thrombocytopenia). Most SDS infants present signs and symptoms of exocrine pancreatic dysfunction which improves with age, with about 40-60% of patients no longer requiring enzyme supplements due to an improvement in digestive enzyme production (Mack et al., 1996). However, based on quantitative intubation techniques almost all SDS patients display a degree of exocrine pancreatic dysfunction (Ip et al., 2002). The SDS pancreas has

normal islets, ductal architecture and secretory function, but with few or no acinar cells and shows extensive fatty replacement (Cipolli, 2001; Jones et al., 1994). The pathogenesis of the exocrine defects is unknown but it is believed that the pancreatic acini either fail to mature properly or undergo very early degeneration rather than secondary acinar cell atrophy (Cipolli, 2001).

Linkage analysis of families with SDS revealed a disease-associated interval at 7q11 (Goobie et al., 2001; Popovic et al., 2002). Boocock *et al.* identified causal mutations in the gene *Shwachman-Bodian-Diamond syndrome* (*SBDS*) in that candidate region and these explain up to 90% of the cases. In addition, 75% of these mutations result from a gene conversion with *SBDSP,* an adjacent pseudogene that shares 97% homology with *SBDS* (Boocock et al., 2003). The SBDS protein is required for late cytoplasmic maturation of 60S ribosomal subunits and translational activation of ribosomes (Finch et al., 2011; Menne et al., 2007; Wong et al., 2011b). In mammalian cells, SBDS and elongation factor like 1 (EFL1) catalyse the removal of the assembly factor eukaryotic initiation factor 6 (eIF6) from late cytoplasmic pre-60S ribosomal subunits (Finch et al., 2011). Loss of *Sbds* in mice leads to early embryonic lethality (Zhang et al., 2006) whilst knockdown of *sbds* in zebrafish recapitulates the human phenotype: exocrine pancreatic insufficiency, neutropenia and skeletal defects (Provost et al., 2012).

Johanson-Blizzard syndrome (JBS; OMIM #243800) is a rare autosomal recessive disorder with an estimated incidence of 1:250,000 (Zenker et al., 2005). It is characterised by congenital pancreatic exocrine insufficiency and a typical facial appearance distinguished by absence or underdevelopment of the nasal wings. Short stature, scalp defects, deafness, absence of several permanent teeth (oligodontia), hypothyroidism, anorectal malformations and mental retardation may also be present (Al-Dosari et al., 2008). The pancreas of JBS infants has normal islets, ductal architecture and secretory function, but lacks acinar cells and shows replacement by fat and connective tissue (Jones et al., 1994). Early intrauterine destruction of acinar cells similar to prenatal onset pancreatitis could lead to the exocrine defects in JBS (Zenker et al., 2005). With age, the pancreatic disease progresses with a significant proportion of patients eventually developing diabetes.

Zenker *et al.* identified the causative gene *ubiquitin protein ligase E3 component n-recognin 1* (*UBR1*) by performing a genome-wide linkage scan in 7 affected families (Zenker et al., 2005). UBR1 protein is involved in the proteolytic pathway of the ubiquitin system

responsible for degradation of intracellular proteins. It is believed that an excess of non-degraded unidentified proteins leads to the pathogenesis observed in JBS. *Ubr*$^{-/-}$ mice display milder pancreatic abnormalities: impaired secretion of zymogen granules and increased susceptibility to pancreatic injury (Zenker et al., 2005; Zenker et al., 2006).

Pearson marrow-pancreas syndrome (OMIM #557000) is characterised by exocrine pancreatic dysfunction, anaemia with abnormal red blood cell precursors instead of healthy red blood cells (sideroblastic anaemia) and vacuolisation of precursors of the bone marrow. Pancreatic dysfunction manifestations are not always present in early childhood whilst haematologic symptoms often start in infancy. It is similar to Schwachman-Diamond syndrome but can be distinguished by their differing bone marrow morphology and pancreatic abnormalities; the pancreas is fibrotic in Pearson marrow-pancreas syndrome whilst the pancreas shows fatty replacement in SDS (Favareto et al., 1989). Pearson marrow-pancreas syndrome is caused by a single large deletion of several mitochondrial genes involved in oxidative phosphorylation suggesting that the pancreatic dysfunction results from an impaired energy supply in acinar cells (Rotig et al., 1989). The progression of the syndrome to a disease with overt muscle dysfunction, Kearns-Sayre syndrome has been reported (Rotig et al., 1995).

Various deficiencies of pancreatic enzymes have been described although they are very rare and the molecular defects have not yet been proven (Durie, 1996). Congenital pancreatic lipase deficiency results in excess fat in faeces (steatorrhea) (Sheldon, 1964). There have been reports of deficiency of colipase as well as both lipase and colipase (Hildebrand et al., 1982; Sjolund et al., 1991). Deficiency in pancreatic amylase can lead to diarrhoea in a starch-rich diet (Sjolund et al., 1991) whilst trypsinogen deficiency has been reported in children with growth failure, hypoproteinemia, diarrhoea and oedema (Townes, 1965). The few number of cases of isolated enzyme deficiencies suggests that the defect is usually compensated by alternative sources of lipolytic, glycolytic and proteolytic activities (Cano et al., 2007).

1.2 Zebrafish as a model organism

Danio rerio (Zebrafish) are teleosts of the cyprinid family in the Actinopterygii (ray-finned fish) class (Nüsslein-Volhard and Dahm, 2002). They are tropical freshwater fish that originated in the streams of the South-Eastern Himalayan region (Talwar and Jhingran, 1991). Their emergence as a vertebrate model organism that allowed both the application of genetic approaches and embryological methods started with the pioneering work of George Streisinger and his colleagues (Streisinger et al., 1981). Their seminal paper described a method to produce clones of homozygous diploid zebrafish and the first zebrafish mutation, *golden*, that produces a homozygous viable pigment phenotype. Subsequently, the University of Oregon zebrafish group provided a detailed description of zebrafish embryology (Kimmel et al., 1995) and isolated the first two developmental mutants, *cyclops* and *spadetail* that showed abnormal floor plate specification and defective trunk mesoderm respectively (Hatta et al., 1991; Ho and Kane, 1990; Molven et al., 1990). Since then, thousands of mutants have been characterised following the development of highly efficient mutagenesis methods and large-scale screens (Nüsslein-Volhard and Dahm, 2002).

Zebrafish is now an indispensable vertebrate model organism for studying embryo development and modelling of human diseases thanks to its numerous attractive features. *Ex utero* development and translucent embryos allow easy visualisation of internal organs. The development of embryos is rapid: all the common vertebrate specific body features including internal organs, compartmentalised brain, ears and eyes are observed after about 2 days post fertilization (d.p.f.). Zebrafish have high fecundity producing a brood size of hundreds of eggs, and have a relatively short generation time of 2-4 months, both characteristics important for genetic analyses. Adults can easily be maintained in a small space, with approximately five adult fish per litre (Nüsslein-Volhard and Dahm, 2002).

1.2.1 Zebrafish genome

A high-quality zebrafish genome, with complete annotation of protein-coding genes is important for the use of zebrafish as a model organism. It enables effective modelling of human diseases since the extent to which zebrafish and human genes are related can be determined, and it has accelerated the identification of mutants by allowing comparisons of mutated and normal sequences (Howe et al., 2013; Schier, 2013). In 2013, Howe *et al.* reported a high-quality zebrafish genome of 1.4 gigabases (Gb) on the Ensembl Zv9 assembly

(Howe et al., 2013). Zebrafish have 26,206 protein-coding genes (Collins et al., 2012), a number greater than any vertebrate previously sequenced. This is probably due to the zebrafish common ancestor undergoing the teleost-specific genome duplication (TSD), a second round of whole-genome duplication (WGD) (Meyer and Schartl, 1999). A comparison between human and zebrafish protein coding genes show that 71% of human genes have at least one zebrafish orthologue, a number which increases to 82% when only Online Mendelian Inheritance in Man (OMIM) genes are considered. This suggests a higher degree of conservation of disease-related genes compared to non-disease-related genes between humans and zebrafish.

1.2.2 Forward genetic approaches

Forward genetic screens involve screening a mutagenised population for a particular phenotype and subsequently identifying the causative gene. These screens have been carried out systematically in plants, flies and worms and have successfully identified numerous genes involved in embryogenesis. A smaller number of mammalian mutants have provided important insights into vertebrate development but identification of a large number is difficult because of intrauterine development and costly supporting research facilities. The remarkable advantages of the zebrafish and the success of the earliest zebrafish genetic screens (Kimmel, 1989) have encouraged large-scale genetic screens to be undertaken (Patton and Zon, 2001).

N-ethyl-N-nitrosurea (ENU) is an alkylating agent that induces random point mutations in the genome, with a preference for A->T transversions and AT->GC transitions (de Bruijn et al., 2009; Nolan et al., 2002). It was found to be the most efficient chemical mutagen in pre-meiotic germ cells in zebrafish making it suitable for large-scale mutagenesis screens (Mullins et al., 1994; Solnica-Krezel et al., 1994). The frequency at which ENU induces mutations is dosage dependent and a significant increase in mutation efficiency is observed when increasing the ENU concentration from 2.0 to 3.0 mM but concentrations significantly higher than 3.0 mM result in high lethality (Solnica-Krezel et al., 1994). An average germ line mutation load of one mutation every 175,000-250,000 base pairs can be achieved consistently by performing six weekly ENU treatments of varying concentrations (Kettleborough et al., 2011).

The groups at Boston (Driever et al., 1996) and Tübingen (Haffter et al., 1996) carried out the first zebrafish screens in which ENU-mutagenised males were outcrossed to wild-type

females to produce F1 offsprings. An F2 generation was raised from F1 sibling matings and subsequently incrossed to produce an F3 generation that contains homozygous induced mutations. The two screens altogether identified ~2000 developmental mutants representing more than 500 genes. These genes are involved in embryogenesis, gastrulation, pigmentation, body shape, haematopoiesis and development of the notochord, muscle, craniofacial skeleton, eye, ear, brain, cardiovascular and digestive organs. Positional cloning of several of these mutants has lead to a better understanding of developmental gene networks (Patton and Zon, 2001).

Insertional mutagenesis is also used in addition to chemical and radiation mutagenesis methods. A group led by Nancy Hopkins injected pseudotyped retroviruses with a genome based on Moloney murine leukemia virus ([M]MLV) and the envelope protein of Vesicular stomatitis Indiania virus (VSIV) into blastula-stage embryos (Gaiano et al., 1996; Lin et al., 1994). They had successfully infected the germ cells and found that insertion events were transmitted to their progeny. Although less efficient than ENU mutagenesis, insertional mutagenesis allows rapid identification of the mutated gene since each insertional event is tagged. A large-scale insertional mutagenesis screen was carried out and had identified 315 genes required for early vertebrate development (Amsterdam et al., 1999; Amsterdam et al., 2004; Golling et al., 2002).

1.2.3 Reverse genetic approaches

Reverse genetics investigate the function of a target gene by modifying its activity and studying the phenotypic consequences. As in forward genetic approaches, random mutagenesis has been the preferred method, since targeted knockout strategies were until recently inadequate in zebrafish. To identify induced mutations in a library of ENU-mutagenised zebrafish, the reverse genetic approach Targeting Induced Local Lesions IN Genomes (TILLING) has been used as an alternative to resequencing which is more laborious and expensive (Wienholds et al., 2003). TILLING, first described in *Caenorhabditis elegans* (Jansen et al., 1997) and *Arabidopsis thaliana* (McCallum et al., 2000), traditionally uses the plant endonuclease CEL1 (Oleykowski et al., 1998) to cleave heteroduplex DNA and analyse fragments to identify induced mutations. Using this approach, Wienholds *et al.* screened a library of 4608 ENU-mutagenised F1 fish for 16 genes and identified 255 mutations (Wienholds et al., 2003).

1.2.3.1 Zebrafish gene knockout resources

The Zebrafish Mutation Project (ZMP), previously known as the Zebrafish Mutation Resource, was launched in 2003 with aims to identify and phenotype disruptive mutations in every protein-coding gene in the zebrafish genome (Kettleborough et al., 2013). To identify mutant alleles, TILLING was used initially but was substituted by direct PCR-based resequencing of target genes (Kettleborough et al., 2011) and subsequently with whole-exome sequencing (Kettleborough et al., 2013). In brief, males are ENU mutagenised and outcrossed to create an F1 generation of carriers that are subjected to exome sequencing (Figure 1-2). Induced mutations are identified and their effect on protein coding is predicted using a modified version of the variant-calling pipeline of the 1000 Genomes Project (Abecasis et al., 2010; Abecasis et al., 2012). To take full advantage of this resource, it is important to screen for phenotypic consequences of disruptive mutations. Phenotypic analysis of homozygous nonsense and essential-splice mutations is carried out in a two-step, multi-allelic approach. The F1 individuals are outcrossed to produce an F2 family. In the first step, crosses of up to 12 pairs of F2 individuals are carried out and phenotypically normal F3 embryos are genotyped for disruptive mutations heterozygous in both parents. Homozygous mutations that are present in the expected Mendelian ratios are deemed to not cause a phenotype whereas if they are present in less than 25% of embryos, they are suspected to cause a phenotype. Secondly, the F2 adults that are heterozygous for any suspected causal mutation are incrossed and the morphological and behavioural phenotypes of the F3 embryos during the first 5 d.p.f. are examined. All phenotypes identified are genotyped for the suspected mutation, the mutation is documented as likely to be causal if over 90% of embryos are homozygous for the specific mutation whereas it is documented as being linked if less than 90% of embryos are homozygous. To date, ZMP has identified nonsense and essential splice site mutations in over 45% of all known protein-coding genes and has examined the phenotypic consequences of more than 1000 alleles. All mutant alleles and data generated by ZMP is available to the scientific community (http://www.sanger.ac.uk/Projects/D_rerio/zmp) (Dooley et al., 2013; Kettleborough et al., 2013). Alleles can be obtained from two international stock centres, the European Zebrafish Resource Center (http://www.itg.kit.edu/ezrc) and the Zebrafish International Resource Center (http://www.zebrafish.org/zirc). ZMP provides a resource to study developmental biology and human diseases and has been used in various studies including pancreas development (Arkhipova et al., 2012; Verbruggen et al., 2010; Wilfinger et al., 2013), heart development (de Pater et al., 2009), osteogenesis (Laue et al., 2008),

muscle development and associated disorders (Hinits et al., 2009; Knight et al., 2011; Lin et al., 2011; Powell and Wright, 2011) and haematopoietic disorders (Cvejic et al., 2008).

In addition to a library of ENU-mutagenised fish, Wang *et al*. reported the establishment of a pseudo-typed retrovirus mediated insertional library (Wang et al., 2007). So far, ~15000 [M]MLV proviral integrations have been mapped to 3054 genes, 0.5% of which have been rescued and characterised. Mutant lines are available to the scientific community through the Zebrafish International Resource Center (http://www.zebrafish.org/zirc) (Amsterdam et al., 2011; Varshney et al., 2013).

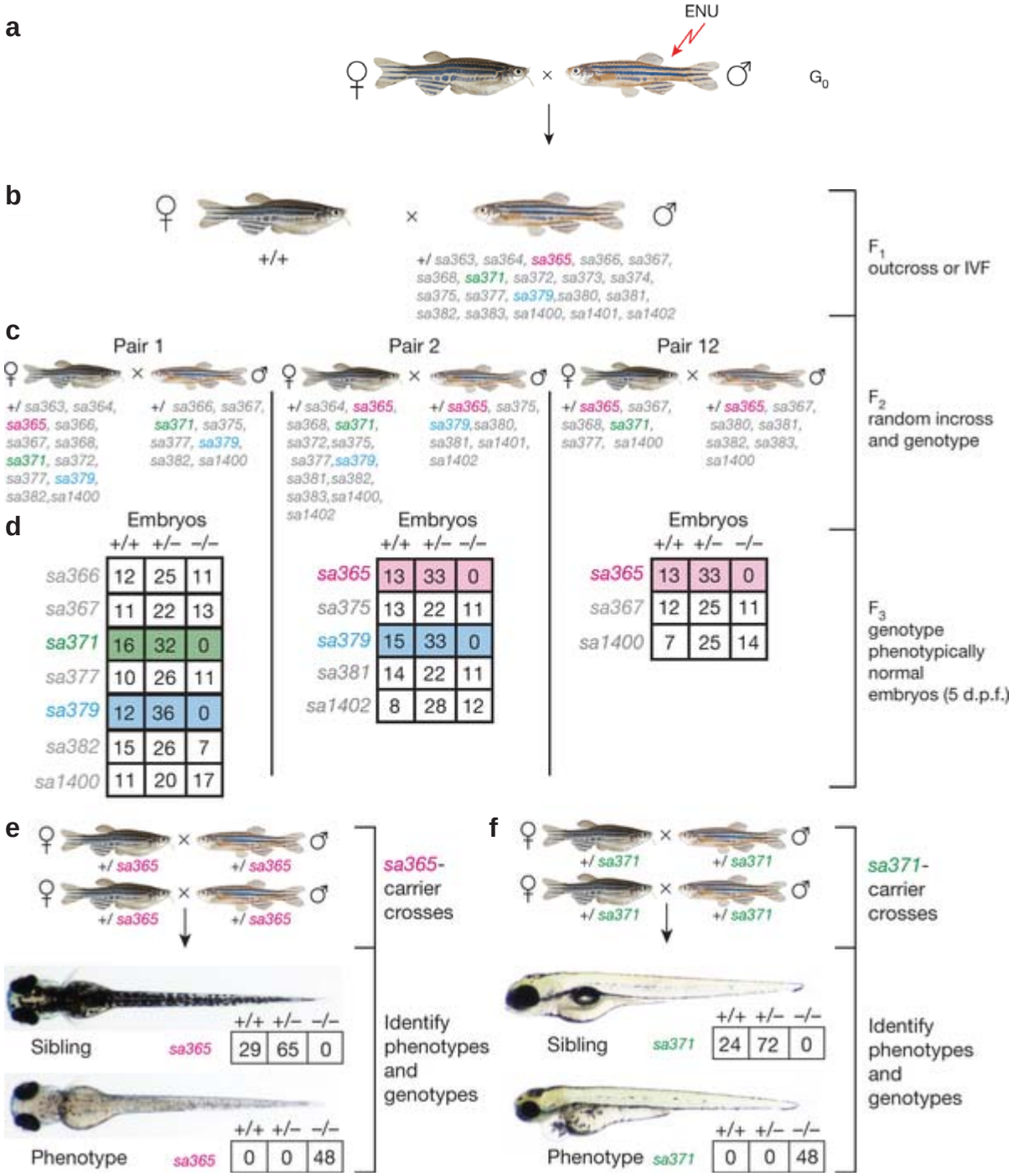

Figure 1-2 Zebrafish Mutation Project (ZMP) Phenotyping Pipeline. (a) G_0 males are ENU-mutagenised and outcrossed to create F_1 individuals and the induced mutations are identified by exome sequencing. (**b)** F_1 individuals are outcrossed or cryopreserved sperms are used in *in vitro* fertilisation (IVF) to produce an F_2 family. The induced disruptive mutations for one family are shown here. (**c**) F_2 individuals are incrossed and genotyped for induced mutations present in F_1 parents. (**d**) First round of phenotyping where 5 d.p.f. embryos of wild-type appearance are collected and genotyped for mutations present in both F_1 parents. Using a chi-squared test and a *p*-value cut-off of 0.05, the number of homozygous mutant embryos is assessed. Mutations that are homozygous in less than 25% of embryos are suspected to cause a phenotype. There are no homozygous embryos for alleles *sa365*, *sa371* and *sa379*. (**e**, **f**), Second round of phenotyping where F_2 adults heterozygous for suspected mutation are incrossed and embryos are phenotyped during the first 5 days and then genotyped. Embryos homozygous for *sa365* and *sa371* are phenotypic. Adapted from Kettleborough et al., 2013.

1.2.4 Transgenic approaches

The different techniques that can be used in zebrafish are summarised in Table 1-2. Transgenesis is an important technology that allows the study of a particular gene. The traditional method of generating transgenic fish involves DNA integration into the genome by injection of linearised or circular DNA plasmids or bacterial artificial chromosomes (BACs) into fertilised eggs. The germline transmission efficiency to the F1 generation is very low, 5-20%, with foreign DNA tending to integrate as concatemers with multiple tandem copies and potentially also causing chromosomal rearrangements. In addition, this repetitive DNA becomes susceptible to methylation over time leading to silencing of the transgene (Culp et al., 1991; Stuart et al., 1988; Thummel et al., 2006). To circumvent these problems, several new techniques have been developed with transposon-based methods being the most popular choice in zebrafish (Davidson et al., 2003; Kawakami et al., 2000).

DNA transposons are genetic elements that move or transpose by a 'cut and paste' mechanism and have been extensively studied in plants and invertebrates. The use of transposons as a genetic tool in vertebrates started with the application of the Sleeping Beauty (SB) transposon system in which the transposase is provided *in trans* and the key DNA cargo is flanked by transposon end sequences (Ivics et al., 1997). This transposon paradigm remains the primary approach employed today with 10 transposons from four different superfamilies (*TcI/Mariner*, *hAT*, *PIF/Harbinger*, *piggyBac*) available for use in vertebrates (Ni et al., 2008). *Tol2* and Sleeping Beauty transposon-based methods show efficient germline transmission in zebrafish (Davidson et al., 2003; Kawakami et al., 2000). Also, *Tol2*-mediated transgenes may not be subject to gene silencing effects since their expression persists through generations (Kawakami, 2007).

The high throughput nature of zebrafish allows transgenesis to be used as a screening tool to manipulate and probe the genome. Examples of this approach include insertional mutagenesis and gene trapping or enhancer detection (van Ruissen et al., 2005). Gene trapping is a high-throughput approach that introduces insertional mutations across a target genome and is widely used to generate knock-out mice (Mikkola and Orkin, 2005; Misra and Duncan, 2002). *Tol2* and Sleeping Beauty transposon-based methods have demonstrated to be suitable for gene trapping in zebrafish. Gene trap vectors whose principal element is a gene trapping cassette that consists of a reporter gene flanked by a 3' splice acceptor and a 5' transcriptional termination sequence are co-injected with transposase into zebrafish embryos

and result in inactivation of target genes and translation of the reporter gene (Kawakami et al., 2004; Kotani et al., 2006; Song et al., 2012a; Song et al., 2012b). Enhancer trapping uses insertion site context vectors to detect enhancers in the genome and has been extensively used in *Drosophila*. In zebrafish, Sleeping Beauty and *Tol2*-transposon based enhancer trapping methods have successfully identified enhancers (Balciunas et al., 2004; Fisher et al., 2006; Parinov et al., 2004).

1.2.4.1 Spatial and temporal control

Tissue-specific and inducible techniques allow spatial and temporal control of expression of a specific transgene and have been successfully used in various model organisms including mouse and zebrafish (Table 1-2). Many tissue-specific promoter elements have been identified in zebrafish and faithfully mimic the expression patterns of the endogenous gene for expression of a transgene in a specific tissue. To facilitate and provide tight spatial and temporal regulation of transgene expression, several genetic methods have been modified.

The yeast Gal4/UAS system consists of the yeast transcription activator protein Gal4 and the Upstream Activation Sequence (UAS), a Gal4-specific enhancer. Its versatility has been demonstrated in *Drosophila* and has been used to express genes in several model organisms, including zebrafish (Halpern et al., 2008). The Cre/LoxP system is an established genetic method to regulate transgene expression in mouse. Its utility in zebrafish was first demonstrated in 2005 (Langenau et al., 2005) and since then a number of Cre deleter zebrafish lines have been generated to achieve tissue-specific transgene expression (Feng et al., 2007; Langenau et al., 2007; Le et al., 2007; Pan et al., 2005; Thummel et al., 2005). Flp, another site-specific recombinase has been predominantly employed in *Drosophila*. It has also been used in mice and shown to function in zebrafish (Boniface et al., 2009; Wong et al., 2011a).

Several methods have been used to achieve temporal control. One method involves using a modified soldering iron to administer localised heat shock in adult fish with transgene expression regulated by a heat shock protein promoter but the utility of this approach is restricted by its inconsistency and limitation to superficial structures (Hardy et al., 2007). Influenced by the work on inducible Cre/LoxP system in mouse models, zebrafish researchers have used chimeric Cre recombinases to achieve temporal control (Metzger et al., 1995). Cre

fused to the mutated ligand-binding domain of estrogen receptor (Cre-ER^{T2}) show high ligand sensitivity and efficient inducible recombination (Feil et al., 1997; Indra et al., 1999). Site-specific recombination is induced by administration of tamoxifen or 4-hydroxy-tamoxifen, with fast recombination kinetics observed (Hans et al., 2009). The LexPR system is an inducible system developed in zebrafish and temporal control is achieved using the synthetic steroid, mifepristone (RU-486) (Emelyanov and Parinov, 2008; Nguyen et al., 2012). Huang *et al.* developed the tetracycline (Tet)-on system to induce heart-specific expression of GFP in adult fish. Doxycycline, a tetracycline derivative induces rapid and strong gene expression but inactivation of the transgene is very slow upon removal of doxycycline (Campbell et al., 2012; Huang et al., 2005).

1.2.5 Gene inactivation tools

Strategies have been developed to target specific genes by site-specific genome modification instead of relying on targeted genome modification based on large-scale forward genetic screens (Table 1-2). These include sequence-specific chimeric nucleases; i.e. zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-based RNA-guided Endonucleases (RGENs). ZFNs introduce double strand breaks (DSBs) at their target locus leading to insertion/deletion (indel) mutations mostly through error-prone non-homologous end joining (NHEJ) in zebrafish. ZFNs are chimeric proteins made up of a zinc finger DNA binding domain and a DNA cleavage domain derived from bacterial non-specific endonuclease FokI. ZFNs are assembled from multiple three-base recognition motifs comprising of about 30 amino acids. The main disadvantages of ZFNs are their complex manufacture due to nonintuitive ZFN binding rules and their diverse range of targeting efficacy. Despite these issues, ZFNs have been successfully used to generate dozens of zebrafish mutants (Urnov et al., 2010). TALENs are artificial restriction enzymes composed of a DNA binding domain of TALE proteins that are mainly found in the plant pathogen *Xanthomonas* and a DNA cleavage domain derived from FokI. TALENs consist of repeat modules of approximately 34 amino acids that are arranged into TAL effector arrays. Within each repeat module, a two-amino acid repeat variable di-residue (RVD) is responsible for one-to-one base pairing, with the four most commonly used RVDs binding preferentially to bases A, C, G or T. Like ZFNs, TALENs introduce double strand breaks but they are easier to design due to their simpler base recognition code and usually have higher targeting efficiency

and lower off-target effects. In addition, assembly of TALENs is rapid, taking less than a week, and is accessible to even small laboratories since the initiation of the entire platform is possible using clones from a single 96-well plate. Several groups have reported successful somatic and germline gene modifications using TALENs (Clark et al., 2011; Huang et al., 2012). The type II CRISPR/Cas system is derived from eubacteria and archaea and involves a complex of transactivating CRISPR RNA (tracrRNA), CRISPR RNA (crRNA) and Cas9 nuclease. Recently, Jinek *et al*. programmed the type II CRISPR system to target a specific genomic sequence using a guide RNA molecule: a chimera of tracrRNA- and crRNA-derived sequences connected by a four-base loop (Jinek et al., 2012). Using a similar chimeric guide RNA in zebrafish, mutations have been induced with efficiencies similar to ZFNs and TALENs. CRISPR gene editing system is faster and easier to produce than ZFNs or TALENs as it requires only a minimum of two plasmids one coding for *cas9* sequence and another one for a single unique guide RNA sequence. However, the off-target effects and germline efficiency remain to be determined (Blackburn et al., 2013; Hwang et al., 2013).

1.2.6 Gene knockdown tools

Gene knockdown tools have proved indispensable in developmental biology (Table 1-2). The established method to knock down specific target genes in eukaryotes is RNA interference (RNAi). RNAi, first identified in *Caenorhabditis elegans* is an endogenous gene-silencing mechanism involved in various biological processes such as transcriptional gene silencing, post-transcriptional regulation of gene expression and antiviral response (Lee et al., 1993; Plasterk, 2002; Wightman et al., 1993; Zamore, 2002). RNAi-based gene-silencing methods involve triggering the innate mechanism of cells by introducing small interference RNAs (siRNAs) or small hairpin RNAs (shRNAs) that are then processed by the enzyme Drosha to produce precursor microRNA-like molecules. These methods are routinely used in *Drosophila melanogaster*, human and mouse cell lines (Martin and Caplen, 2007). RNAi-mediated gene knockdown has been used in zebrafish but progress has been very slow with varying success rates and nonspecific toxicity observed in developing embryos (Kelly and Hurlstone, 2011). Instead, morpholino is the method of choice to knock down the gene of interest in zebrafish. It is a synthetic oligonucleotide made of a morpholine ring and is usually 25 bases long. It has the advantage of being completely resistant to nucleases, has excellent sequence specificity, generally good targetting predictability and high in-cell efficacy. It can either block translation by binding to the 5'UTR of a specific mRNA or interfere with pre-

mRNA processing steps, thereby modifying splicing and resulting in exon exclusion or intron inclusion (Summerton, 1999; Summerton and Weller, 1997). In addition, varying levels of gene knockdown can be achieved by controlling the dosage of the morpholino and the synergistic and counteractive effects of various genes can be studied by co-injecting several morpholinos. However, morpholino has two key disadvantages: firstly its effects are transient allowing only the study of early development and secondly it is known to produce off-target effects thereby complicating the interpretation of results (Eisen and Smith, 2008).

Techniques	Features
Transgenic methods	
Linearised/Circular DNA and BAC	• Simple procedure • Germline transmission is generally low 5-20% • Integrate as concatemers of many tandem copies and may cause chromosomal rearrangements • Transgene becomes methylated and silenced over time
Transposon-based (*Tol2*, Sleeping Beauty)	• Commonly used in various model organisms including zebrafish • Relative ease of manipulation and high germline transmission efficiency
Tissue-specific and Inducible methods	
Gal4/UAS system	• Highly successful in *Drosophila*, effective in mouse and functional in zebrafish • Variegated or diminished expression over time
Cre/LoxP & Cre-ER^{T2}	• Widely used in mouse and efficient in zebrafish • Efficacy of Cre-ER^{T2} in adult fish remains to be determined
Flp/FRT	• Widely used in *Drosophila* and moderately in mouse • Efficient but not established in zebrafish
LexPR	• First developed in zebrafish • Efficacy in adult fish not determined
Tet-On	• First developed in zebrafish • Rapid and strong gene expression but inactivation very slow
Gene inactivation methods	
ZFN	• Successful in various model organisms including zebrafish • Low germline efficacy with off-target effects observed • Difficult to design and production expensive and time-consuming
TALEN	• Effective in zebrafish • Moderate-high germline efficacy with very low off-target effects observed • Reasonable design and production affordable and rapid (<1 week)
CRISPR/Cas	• Functional in zebrafish • Germline efficacy and off-target effects remain to be determined (Initial results suggest similar germline efficacy to ZFN and TALEN) • Simple design and production inexpensive and very fast
Gene knockdown methods	
RNAi	• Successful and routinely used in mammalian cells • Progress slow in zebrafish • Varying success rates and off-target effects
Morpholino	• Method of choice in zebrafish • Highly specific and efficient • Transient knockdown and off-target effects

Table 1-2 Techniques used in zebrafish. Transgenic, tissue-specific and inducible, gene inactivation and knockdown methods and their different features are shown.

1.3 Zebrafish pancreas development

Zebrafish is an excellent model to study pancreas development for two main reasons. Firstly, the zebrafish and mammalian pancreas share similar anatomy and histology. The adult zebrafish pancreas contains several principal islets that are situated next to the gallbladder and additional smaller accessory islets embedded in the exocrine tissue situated in the intestinal mesentery. In both fish and mammals, the pancreatic islet is composed of a central core of insulin-producing β-cells that is surrounded by glucagon-producing α-cells, somatostatin-producing δ-cells and ghrelin-producing ε-cells (Argenton et al., 1999; Biemar et al., 2001; Devos et al., 2002). Enzyme-producing acinar cells surround the endocrine islets and are arranged into acini. A complex ductular network connects the exocrine cells to the intestine. In fish and mammals, the hepatic and main pancreatic ducts are joined at the site where they enter the intestine (Yee et al., 2005). Secondly, orthologous signalling pathways and transcription factors control pancreas development in zebrafish and mammals in a similar fashion.

Pancreas development in zebrafish, like other vertebrates, is a complex process comprising of three major developmental mechanisms. The first includes induction and pattern formation of the endoderm and defines the location of the pancreatic primordium within the endoderm. The second is the control of cell differentiation of pluripotent endodermal precursor cells into specialised pancreatic cells. The third involves morphogenesis of the pancreas and requires extensive rearrangements of cells and movements of whole organ parts resulting in the location, shape and cellular organisation of the pancreas (Gnugge et al., 2004).

1.3.1 Endoderm induction, patterning and regionalisation

The three primary germ layers, the endoderm, mesoderm and ectoderm form during gastrulation. The endoderm and mesoderm are thought to arise from the mesendoderm, a transient common precursor cell population. An evolutionarily conserved gene regulatory network consisting of Nodal growth factor signalling and downstream transcription factors controls the induction and commitment of mesendoderm to an endodermal lineage (Figure 1-3). After gastrulation, the naïve endoderm transforms into a primitive gut tube by a series of morphogenetic movements. The gut tube regionalises along both the anterior-posterior (A-P)

and dorsal-ventral (D-V) axes into foregut, midgut and hindgut domains. The basic organisation of the gut tube is conserved from fish to mammals (Zorn and Wells, 2009).

The endodermal germ layer derives from blastoderm cells that are mostly located in the dorsal half of the mid-blastula margin (Warga and Nusslein-Volhard, 1999). Both the yolk syncytial layer and the vegetal margin of the blastoderm secrete transforming growth factor beta (TGFβ) molecules (Rodaway et al., 1999). Nodal ligands, members of the TGFβ family, induce the endoderm and mesoderm in a concentration dependent manner; high levels promote endoderm development whilst lower levels promote mesoderm formation (Chen and Schier, 2001; Dougan et al., 2003; Thisse et al., 2000). Mid-blastula-stage marginal cells can give rise to both endoderm and mesoderm whilst late-blastula-stage marginal cells can give rise only to either endoderm or mesoderm. Notch (Kikuchi et al., 2004), Bone morphogenetic protein (BMP) and Fibroblast growth factor (FGF) (Poulain et al., 2006) signalling pathways are involved in segregation of the endoderm and mesoderm.

In the current model of zebrafish endoderm induction, nodal ligands encoded by *ndr1/squint* and *ndr2/cyclops* induce a common endo-mesodermal region (Figure 1-3). These ligands are expressed at the vegetal margin of the blastoderm and are antagonised by members of the Antivin/Lefty family (Agathon et al., 2001). Nodal signals act through Acvr1b/Taram-a receptors (Aoki et al., 2002b) and Oep co-receptors (Gritsman et al., 1999). This leads to intracellular phosphorylation of Smad2 proteins binding to Smad4 to form Smad2-Smad4 complex that translocates to the nucleus (Dick et al., 2000; Muller et al., 1999). Smad2-Smad4 complex interacts with additional cofactors: FoxH1/Fast1 (Pogoda et al., 2000; Sirotkin et al., 2000) and Mixer/Bon (Kikuchi et al., 2000) to activate target genes including three transcription factors: *mixer/bon* (Alexander and Stainier, 1999), *gata5/fau* (Reiter et al., 2001) and *og9x/mezzo* (Poulain and Lepage, 2002) leading to expression of *sox32*/*cas* (Alexander et al., 1999; Alexander and Stainier, 1999; Aoki et al., 2002a). Sox32 then induces the expression of *sox17*, the final determinant of endoderm specification (Alexander and Stainier, 1999). Several forkhead and homeobox transcription factors are also expressed in the endoderm including *foxa2* and *hhex* (Odenthal and Nusslein-Volhard, 1998; Wallace and Pack, 2003; Wallace et al., 2001). In addition, three maternal transcripts contribute to Nodal-dependent endoderm formation. Pou5fl/Spg maintains *sox32* expression in a positive feedback loop and interacts with Sox32 to specify mesoendodermal precursors to an endodermal identity inducing expression of *sox17* and *foxa2* (Lunde et al., 2004; Reim et

al., 2004). Eomesa interacts with both Mixer and Gata5 to induce expression of *sox32* (Bjornson et al., 2005) and Cdx1b regulates *foxa2* and possibly *gata5* expression during endoderm formation and digestive organ development (Cheng et al., 2008).

The next step in pancreas formation is the antero-posterior regionalisation of the digestive tube and is governed by the combined activity of several signalling pathways including Wnt, retinoic acid (RA), BMP, FGF and Hedgehog (Hh) (Figure 1-3) (Tiso et al., 2009).

Goessling *et al.* showed that during early somitogenesis, endodermal pattern formation requires suppression of Wnt signalling, but later in development Wnt activity alters endodermal fate with increased activity favouring liver growth over pancreas formation (Goessling et al., 2008). Stafford *et al.* studied the *neckless* mutant with disrupted Retinoic Acid (RA) synthetic enzyme (RALDH2) and demonstrated that RA signalling is required for specification of both the pancreas and the liver and that RA synthesised in the mesoderm acts as an instructive signal to induce precursors of the endocrine pancreas (Stafford and Prince, 2002; Stafford et al., 2006). The transcription factors Cdx4 and Cdx1a localise the pancreas by blocking pancreatic identity, possibly through RA inhibition in the posterior endoderm (Kinkel et al., 2008). Bmp2b signalling plays a role in endoderm patterning and controls the fate of bipotential hepatopancreatic progenitors whilst Bmp2a is required to specify the ventral pancreatic bud between 20 and 24 hours post fertilisation (h.p.f.) (Chung et al., 2008; Naye et al., 2012; Tiso et al., 2002). More recently, it was found that in early zebrafish developmental stages prostaglandin E2 (PGE2) synthesis favours liver over pancreatic cell fate by interacting with the bmp2b pathway but later in development PGE2 promotes both liver and pancreas outgrowth in zebrafish (Nissim et al., 2014)

FGF signalling is another pathway critical for pancreatic specification. At early stages, endodermal expression of *fgf24* stimulates the patterning of the pancreatic lateral plate mesoderm and subsequently both Fgf10 and Fgf24 induces ventral pancreas formation and represses the hepatic fate (Manfroid et al., 2007; Naye et al., 2012). Fgf10 refines the boundaries between the hepatopancreatic organs (liver, pancreas and gallbladder) and the ductal epithelium and inhibits the differentiation of the liver and proximal pancreas into hepatic and pancreatic cells respectively (Dong et al., 2007). Song *et al.* showed that both FGF and BMP act genetically upstream of RA signalling and that they direct pancreas formation through the hepactocyte nuclear factor 1 homeobox B (Hnf1b) pathway (Song et

al., 2007). Hnf1b controls the regionalisation and specification of the gut through proper expression of *pdx1* and *shh* (Sun and Hopkins, 2001). The Hnf1b and Wnt pathways were found to synergise in the specification of the ventrally derived hepatopancreas progenitors in studies of conditional loss of Wnt signalling in a hypomorphic *hnf1ba* zebrafish mutant (Lancman et al., 2013).

The Hedgehog (Hh) signalling pathway plays a critical role at two stages during pancreas development (Chung and Stainier, 2008; dilorio et al., 2002; Roy et al., 2001). At the start of gastrulation, Hh signalling is required for migration and differentiation of pancreas progenitors by inhibition of Bmp signalling (Tehrani and Lin, 2011). By the end of gastrulation, Hh acts antagonistically to RA in specification of the endocrine pancreas, although it still promotes differentiation of exocrine progenitor cells (Tehrani and Lin, 2011). The expression of *sonic hedgehog* (*shh*) in the second hedgehog-dependent step is regulated by Meis Homeobox 3 (Meis3) and pre-B-cell leukemia homeobox 4 (Pbx4) (dilorio et al., 2007). Histone deacetylase 1 (Hdac1) restricts foregut fates, promotes specification of the liver and exocrine pancreas and morphogenesis of the endocrine islet (Noel et al., 2008).

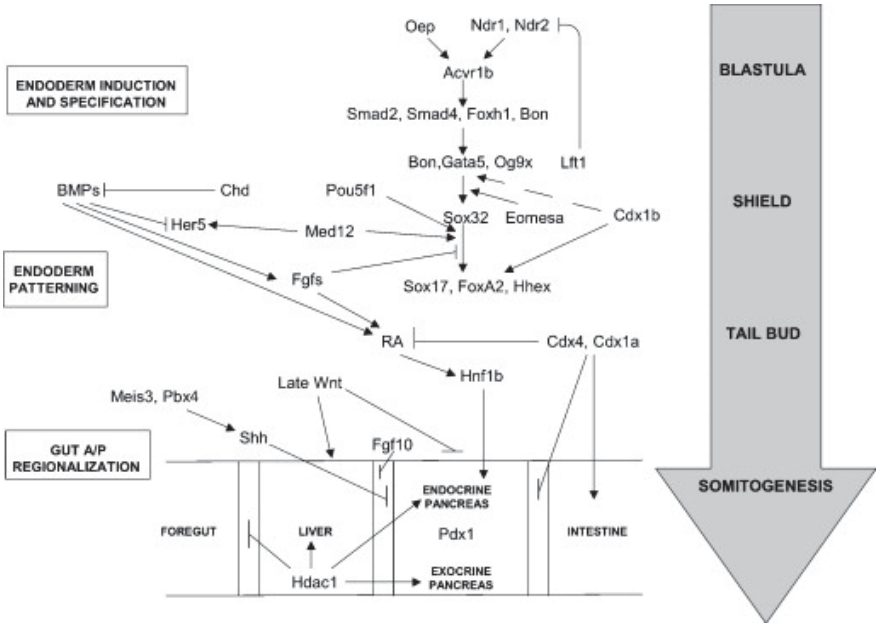

Figure 1-3 Integration of signalling cascades during zebrafish pancreas development. Nodal, Bmp, Fgf, RA, Wnt and Hh signalling pathways are involved in endoderm induction, specification and regionalisation at different embryonic stages (shown on the right) to define a Pdx1-positive pancreatic region within the developing gut tube. Taken from Tiso et al., 2009.

1.3.2 Morphogenesis of pancreas

In zebrafish pancreas morphogenesis, two contiguous areas of the gut bud sequentially and merge to form the definitive pancreas. Expression patterns of specific genes during early development have revealed that pancreatic precursors are already present before budding occurs. At 14 h.p.f., pancreatic precursors are first detected as two bilateral rows of cells adjacent to the midline, expressing *pancreatic and duodenal homeobox 1* (*pdx1*), a bipotential marker of intestinal bulb and pancreatic progenitor cells (Figure 1-4). Expression of endocrine markers *insulin, somatostatin* and *glucagon* are first detected at 15 h.p.f., 17 h.p.f. and 21 h.p.f. respectively. The bilateral group of cells begins to merge at 16 h.p.f. and by 18 h.p.f. is located in the midline, immediately dorsal to the yolk. At 24 h.p.f., the first pancreatic posterodorsal bud protrudes from the intestinal rod to form the primary islet (Argenton et al., 1999) (Biemar et al., 2001). Between 24 h.p.f. and 48 h.p.f., gut looping displaces the posterodorsal bud to right side and the gut to the left side. By 48 h.p.f., the first islet already has a similar organisation to its mammalian counterpart, with a core of β-cells surrounded by α- and δ-cells.

The second pancreatic anteroventral bud forms after 34 h.p.f. within the *pdx1*-expressing region of the ventral region of the gut tube and anteriorly to the main islet. Between 34 h.p.f. and 48 h.p.f., the second bud grows towards the first bud posteriorly and surrounds it, giving rise to exocrine tissue, pancreatic ducts and late-forming endocrine cells (Field et al., 2003a). *Trypsin*-positive cells surround *insulin*-positive cells in the head of the merged pancreas (Field et al., 2003a). *Carboxypeptidase A* and *elastase B*, two other exocrine-specific genes, are detected at 48 h.p.f. and 56 h.p.f. respectively (dilorio et al., 2002; Mudumana et al., 2004). The exocrine tissue continues to grow posteriorly along the intestine in parallel with the establishment of the intrahepatic ductal system (Wan et al., 2006). At 5 d.p.f., the zebrafish pancreas consists of a single islet with about 60 endocrine cells and exocrine acinar tissue with a branching ductal system connecting the pancreas to the gut. During postembryonic growth, the primary islet becomes bigger and smaller secondary islets are formed along the exocrine duct system resulting in several dozen secondary islets in adults (Chen et al., 2007; Pack et al., 1996; Parsons et al., 2009; Pauls et al., 2007; Yee et al., 2005).

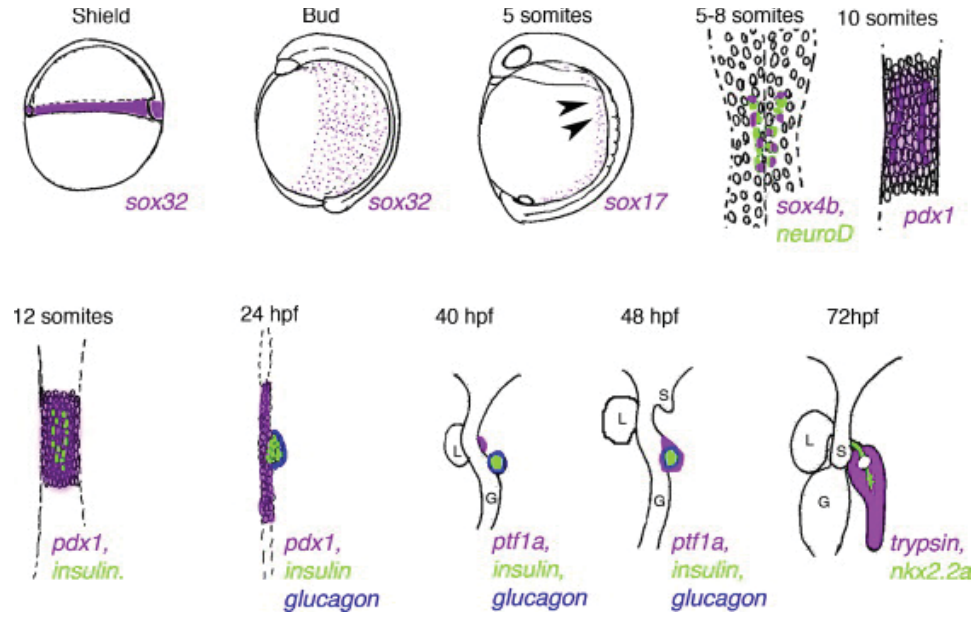

Figure 1-4 Zebrafish pancreas development. At the shield and bud stage, *sox32* is expressed. At the 5-somite stage, arrowheads indicate the *sox17*-positive endodermal area. At the 5-8 somite stages, *sox4b* and *neuroD* are the first genes expressed in the pancreatic primordium. At the 10-somite stage, two bilateral rows of cells express *pdx1*. At the 12-somite stage, *insulin* is detected. At 24 h.p.f., protrusion of the first pancreatic bud occurs with *glucagon* expression also detected. At 40 h.p.f., the second pancreatic bud is formed with *ptf1a* expression observed. At 48 h.p.f., the two buds have merged and at 72 h.p.f. growth of exocrine tissue, liver and gut is apparent. Lateral view dorsal to the right of embryos at shield, bud and 5-somite stages. Dorsal view anterior to the top of embryos at all other stages. L=liver, G=gut, S=swim bladder. Taken from Tiso et al., 2009.

1.3.3 Differentiation of pancreatic progenitor cells

Different signalling pathways and various transcription factors control the differentiation process of pancreatic cells. FGF and Notch signalling are two genetic pathways that are classically involved in cell fate decisions and differentiation from precursor cells. As previously mentioned, FGF appears to be involved in determining the fate towards ducts or organs during patterning and differentiation of the hepatopancreatic ductal system (Dong et al., 2007). Notch signalling is involved in cell fate decisions in both mammals and zebrafish (Appel et al., 1999; Artavanis-Tsakonas et al., 1999). Esni *et al.* showed that Notch signalling genes negatively regulate exocrine pancreatic differentiation in both mouse and zebrafish (Esni et al., 2004). Zebrafish *mindbomb* (*mib*) mutants, where Notch signalling is disrupted, displayed accelerated differentiation of the exocrine pancreas (Esni et al., 2004) (Table 1-3). Zecchin *et al*. determined that different Notch ligands of the Delta and Jagged families control temporally distinct phases of endocrine and exocrine cell type specification (Zecchin et al., 2007). Notch signalling also plays a role in secondary islet formation in zebrafish (Parsons et al., 2009; Wang et al., 2011). Parsons *et al.* showed that Notch inhibition induces loss of pancreatic Notch-responsive cells (PNCs) accompanied by precocious

secondary islet formation (Parsons et al., 2009). PNCs are progenitor cells that are associated with the pancreatic ductal epithelium and differentiate to form secondary islets, ducts and centroacinar cells later in development (Parsons et al., 2009; Wang et al., 2011). By using various transgenic lines the same group found that PNCs are derived from the ventral bud and are distinct from *ptf1a*-expressing pancreatic proegenitor cells (Wang et al., 2011). More recently, Huang *et al.* found that inhibition of RA signalling also induces secondary islet formation and that RA negatively regulates the differentiation of PNCs (Huang et al., 2014). In addition, G protein-coupled receptor (GPCR) signalling plays a role in clustering of endocrine progenitors into islets in both mouse and zebrafish (Serafimidis et al., 2011).

Transcription factors regulate the differentiation process at different stages (Figure 1-5). The pancreatic and duodenal homeobox 1 (Pdx1), homeobox Hb9 (Hlxb9) and pancreas transcription factor 1 subnit alpha (Ptf1a) specify pancreas progenitors at early stages. Subsequently the basic Loop-Helix-Loop factors neurogenin 3 (Ngn3) and neurogenic differentiation (NeuroD) induce the endocrine precursors and the homeodomain factors NKX-homeobox factor 2.2 (Nkx2.2) and 6.1 (Nkx6.1), Paired box 4 (Pax4) and 6 (Pax6) and aristaless related homebox (Arx) control the formation of the endocrine cells.

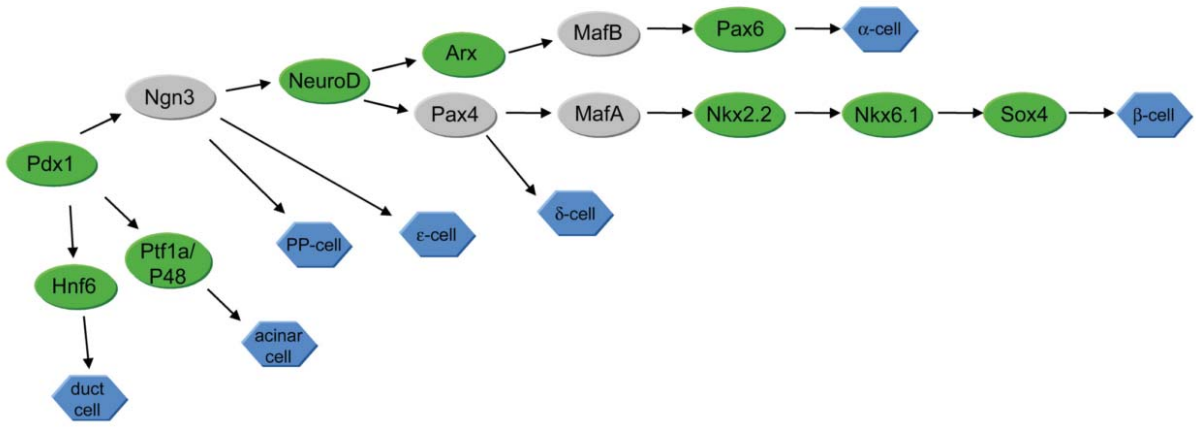

Figure 1-5 Diagram of transcription factors in mouse and zebrafish pancreatic cell differentiation. The lineages are based on data from mouse studies. The function of a transcription factor may not be conserved between mouse and zebrafish. Taken from Kinkel and Prince, 2009.

Pdx1 is one of the earliest specific markers of vertebrate endocrine and exocrine primordia (Milewski et al., 1998; Ohlsson et al., 1993). In zebrafish, *pdx1* expression starts at the 10-somite stage in the cells of the presumptive duodenum and pancreas (Biemar et al.,

2001). Studies of *pdx1* knockdown using morpholinos showed a reduction in exocrine and endocrine tissue with various endocrine cell types still present (Huang et al., 2001; Yee et al., 2001). These results suggest that Pdx1 is not required for specification but for differentiation of pancreatic cell types. Similarly in mice, the initial specification of the pancreatic buds is normal in *Pdx1* knockout embryos but further differentiation and morphogenesis of the pancreas are impaired (Ahlgren et al., 1997; Jonsson et al., 1994; Offield et al., 1996).

In mice, Hlxb9 is required for the development of the pancreas dorsal lobe but not for the initial phases of ventral bud formation (Harrison et al., 1999; Li et al., 1999). However, in zebrafish, morpholino knockdown of *hlxb9* showed that Hlxb9 is not required for early pancreas morphogenesis but is essential for β-cell differentiation (Wendik et al., 2004). Recently, it was found that within the endoderm Hlxb9 functions downstream of RA to promote β- and suppress α-cell fates in the endocrine pancreas progenitor lineage in zebrafish (Dalgin et al., 2011). In addition, Arkhipova *et al*. found that the zebrafish *hlxb9* enhancer is required and sufficient to regulate early expression in β-cells (Arkhipova et al., 2012). The simultaneous knockdown of Pdx1 and Hlxb9 using morpholinos in zebrafish lead to a near complete and persistent β-cell deficiency as opposed to early β-cell reduction and recovery after knockdown of either Pdx1 or Hlxb9 alone suggesting that Pdx1 and Hlxb9 cooperate in β-cell formation (Kimmel et al., 2011). In addition, they found that early-arising endocrine cells are already specified and their differentiation proceeds independently to *pdx1* whilst later-arising PNCs are mostly postmitotic and their differentiation is *pdx1*-dependent (Kimmel et al., 2011).

The transcription factor Ptf1a is required for converting intestinal to pancreatic lineages and for the development of acinar cells in mice (Kawaguchi et al., 2002; Krapp et al., 1998). Morpholino knockdown of *ptf1a* in zebrafish results in loss of exocrine pancreas with normal endocrine gene expression and islet morphology (Zecchin et al., 2004). By studying a *ptf1a* loss of function mutant and using morpholino knockdown, Dong *et al.* showed that high levels of Ptf1a promote exocrine fate but repress endocrine fate whereas low levels of Ptf1a promote endocrine fate (Dong et al., 2008). In addition, Hesselson *et al.* found that Ptf1a is essential in maintaining the pancreatic acinar cell fate by antagonising Ptf1a postembryonically (Hesselson et al., 2011). The *exocrine differentiation and proliferation factor* (*exdpf*) gene is a target of Ptf1a and morpholino knockdown of *expdf* resulted in loss or severe reduction of exocrine cells whilst the endocrine compartment appeared unaffected

(Jiang et al., 2008). The authors also found that *expdf* acts downstream of RA (Jiang et al., 2008). Another gene that was found to be involved in exocrine pancreas development is the Hlxb9-related factor Hlxb9la. Knockdown of *hlxb9la* using morpholino showed that Hlxb9la function in late morphogenesis of the exocrine pancreas (Wendik et al., 2004).

The differentiation of endocrine cells involves numerous transcription factors. Ngn3 was found to be required for specification of a common precursor for pancreatic endocrine cells α, β, δ and PP in mice (Gradwohl et al., 2000). Ngn3 has several downstream targets including NeuroD1 and Nkx2.2. NeuroD1 is a basic helix-loop-helix transcription factor and its loss of function in mice results in a reduction in β-cell number and arrested morphogenesis of the islet (Naya et al., 1997). Anderson *et al*. studied the network of Ngn3, Nkx2.2 and NeuroD1 in both mice and zebrafish and found that the transcriptional activity of *nkx2.2* is necessary to facilitate Ngn3 activation of NeuroD1 within the endocrine progenitor cell and also in maintaining *neurod1* expression in mature β-cells (Anderson et al., 2009a). However, Flasse *et al.* found that endocrine formation appeared normal in zebrafish *ngn3* null mutant embryos, but that endocrine cell differentiation is initiated by the basic helix-loop-helix factor Ascl1b and afterwards controlled by Neurod1 (Flasse et al., 2013). Nkx6.1 has been shown to act downstream of Nkx2.2 and is crucial for differentiation of α- and β-cells in mice and zebrafish (Binot et al., 2010; Henseleit et al., 2005; Sander et al., 2000). Another transcriptional target of Nkx2.2 is the L6 domain tetraspanin family member *tm4sf4* gene that acts in opposition to Nkx2.2 (Anderson et al., 2011).

Paired box 6 (Pax6) is a conserved transcription factor in vertebrates and plays an important role in the development of eye, brain and pancreas (Georgala et al., 2011; Sander et al., 1997; Shaham et al., 2012; St-Onge et al., 1997). Pax6 is involved in differentiation of endocrine cells: in mice, inactivation of *Pax6* lead to a severe reduction in β– and δ-cells, a near complete loss of α-cells and a significant increase of ε-cells (Heller et al., 2005; Heller et al., 2004; Sander et al., 1997; St-Onge et al., 1997) while a *pax6b* null mutant and *pax6b* morpholino injected zebrafish showed complete loss of β–cells, severe reduction of δ-cells, and a significant increase in ε-cell number (Verbruggen et al., 2010). Recently, Arkhipova *et al*. identified a cross-talk between Pax6b and Hlxb9 and showed that in zebrafish Pax6b is critical for maintenance but not induction of pancreatic *hlxb9* expression (Arkhipova et al., 2012). Another member of the Pax family, Pax4 has also been shown to play a role in pancreatic cell differentiation and act antagonistically to the transcription factor Arx. In both

mice and zebrafish, Pax4 acts opposite to Arx to modulate the number of pancreatic α-cells but Pax4 plays a crucial role only in β-cell differentiation in mice and not in zebrafish (Collombat et al., 2005; Collombat et al., 2003; Collombat et al., 2009; Djiotsa et al., 2012; Sosa-Pineda et al., 1997). The transcription factor Sox4b has also been shown to have a crucial function in generation of α-cells in zebrafish (Mavropoulos et al., 2005).

Besides Nkx2.2 and NeuroD1, Ngn3 has additional targets including the recently identified Rfx6 (Soyer et al., 2010). Rfx6 is a winged helix transcription factor that works in parallel with transcription factors NeuroD1, Pax4 and Arx during differentiation of islet cells. In zebrafish, *rfx6* is required for differentiation of cells expressing *glucagon*, *ghrelin* and *somatostatin* and for proper clustering of β-cells.

The LIM-Homeodomain (LIM-HD) transcription factor Isl1 is required for early pancreas morphogenesis and differentiation of the endocrine lineage. In mouse, Isl1 functions independently in exocrine and endocrine tissues (Ahlgren et al., 1997; Du et al., 2009; Liu et al., 2011; Liu et al., 2012). Zebrafish *isl1* mutants display a reduction in expression of endocrine hormones despite normal numbers of endocrine cells and smaller exocrine pancreas. In addition, combined knockdown of two or three *isl1/2* genes results in reduction of exocrine pancreas in a dose-dependent fashion. In zebrafish, *isl1* is important for endocrine cell maturation and *isl1* and *isl2* genes interact for expansion of the exocrine pancreas (Wilfinger et al., 2013).

Recently, Manfroid *et al*. found that the HMG box transcription factor *sox9b* gene plays a critical role in the development of the hepatopancreatic ductal system and the formation of secondary endocrine cells that are derived from pancreatic ducts (Manfroid et al., 2012). The *sox9b* mutants displayed disrupted ductal morphogenesis and differentiation and their β-cell recovery is highly compromised following β-cell ablation (Manfroid et al., 2012).

1.3.4 Role of ribosomal biogenesis genes in exocrine pancreas development

In recent years, several genes involved in ribosome biogenesis have been found to play a role in exocrine pancreas, intestine and liver development in zebrafish (Table 1-3). In 2003, Mayer and Fishman described *nil per os* (*npo*), a zebrafish mutant of *RNA-binding motif protein 19* (*rbm19*) that showed arrested morphogenesis and cytodifferentiation of the intestine and exocrine pancreas in a primordial state (Mayer and Fishman, 2003). The related yeast protein of Rbm19, Mrd1p mediates pre-rRNA processing (Jin et al., 2002). Chen *et al*. described a zebrafish mutant in the gene *digestive expansion factor* (*def*) as having hypoplastic digestive organs: pancreas, liver and intestine (Chen et al., 2005). Def is a member of a novel protein family that is conserved from yeast to human and forms part of the ribosomal small subunit (SSU) processome, a large ribonucleoprotein complex involved in pre-40S ribosome maturation (Charette and Baserga, 2010; Goldfeder and Oliveira, 2010; Harscoet et al., 2010). It was found that the isoform of *tp53*, Δ*113p53*, was selectively upregulated within the digestive organs and triggered expression of Tp53-responsive genes that lead to cell cycle arrest resulting in impaired growth of digestive organs (Chen et al., 2005). It was recently reported that Def is a nucleolar protein whose loss of function results in upregulation of Tp53 protein which accumulates in the nucleolus (Tao et al., 2013a). The authors showed that Def can lead to degradation of Tp53 protein, a process that is dependent on the activity of a cysteine protease, Calpain3 but independent of the proteasome pathway. In addition, adult transgenic fish lines with liver-specific over-expression of Def exhibited abnormal intrahepatic structure indicating that the levels of Def proteins must be controlled to maintain the structural integrity of the liver (Tao et al., 2013b).

Furthermore, Provost *et al*. described a zebrafish model of Shwachman-Diamond syndrome (Provost et al., 2012) in which morpholino knockdown of *sbds* recapitulated the developmental abnormalities characteristic of the human syndrome including pancreatic hypoplasia, skeletal defects and loss of neutrophils. Impaired expansion of *ptf1a*–expressing pancreatic progenitors was observed in *sbds* morphants. Loss of *tp53* by morpholino knockdown or by using *tp53* mutant lines fail to rescue the developmental defects observed in *sbds* knockdown thereby suggesting a Tp53-independent mechanism. Transcriptional profiling by microarray analysis and functional group analysis of differentially expressed genes revealed marked enrichment of genes involved in rRNA processing, ribosome biogenesis and translational initiation in *sbds*-deficient zebrafish. Loss of function studies of

two of these differentially expressed genes *ribosomal protein L3* (*rpl3*) and *pescadillo* (*pes*) also lead to impaired proliferation of pancreatic progenitor cells, a phenotype independent of Tp53 (Provost et al., 2012). The authors went on to study the role of other proteins of the large ribosomal subunits in pancreas development (Provost et al., 2013). They demonstrated that the *rpl* genes have a common expression pattern during development: in early embryos, their expression is initially widespread but afterwards become increasingly restricted to the endoderm. At 48 h.p.f., *rpl* genes were expressed at high levels in *ptf1a*-expressing progenitors within the pancreas. In *rpl23*^{hi2582} and *rpl6*^{hi3655b} mutants, *ptf1a*-expressing pancreatic progenitors failed to expand properly while in heterozygotes, these progenitors showed a recoverable delay in expansion. Knockdown of *tp53* failed to suppress the phenotype observed in *rpl23*^{hi2582} and *rpl6*^{hi3655b} mutant embryos suggesting a Tp53-independent mechanism (Provost et al., 2013).

Similarly, Boglev *et al.* described *titania* (*tti*^{s450}) that exhibited defects in intestine, pancreas, liver and craniofacial development (Boglev et al., 2013). It was identified in the ENU mutagenesis Liver^{plus} screen that was carried out on a transgenic line of zebrafish (*Tg(XlEef1a1:GFP)*^{s854}) expressing GFP specifically in the digestive organs (Field et al., 2003b; Ng et al., 2005; Ober et al., 2006). The *tti*^{s450} mutant harbours a nonsense mutation in *periodic tryptophan protein 2 homolog* (*pwp2h*), a gene encoding a small subunit processome component. Deficiency in *pwp2h* lead to impaired ribosome biogenesis due to a decreased production of mature 18S rRNA and 40S ribosomal subunit. The authors also found that autophagy was upregulated in intestinal epithelial cells of *tti*^{s450} mutants as a survival mechanism and this induction of autophagy was independent of the Tor and Tp53 pathways (Boglev et al., 2013).

More recently, Qin *et al.* identified and characterised a zebrafish mutant of *nucleolar protein with MIF4G domain 1* (*nom1*), a gene encoding a protein with a conserved role in 18S rRNA formation (Qin et al., 2014). The *nom1* mutants displayed defects in exocrine pancreas, intestine, liver and craniofacial development. The authors found that the pancreatic defect was due to impaired proliferation of *ptf1a*-expressing pancreatic progenitor cells and was independent of Tp53 activation. Whole transcriptome analysis by RNA-seq showed that both ribosome biogenesis and pre-mRNA splicing are perturbed in the *nom1* mutants.

Mutant	Gene	Gene Name	Phenotype	Reference
Early development				
heart and mind (*had*)	*atp1a1a.1*	ATPase, Na+/K+ transporting, alpha 1a polypeptide, tandem duplicate 1	Primary lateral *pdx1*-expressing pancreas primordia do not merge medially, perturbed endoderm morphogenesis	Shu et al., 2003
heart and soul (*has*)	*prkci*	protein kinase C, iota	Primary lateral *pdx1*-expressing pancreas de not merge	Field et al., 2003a
hnf1ba	*hnf1ba*	HNF1 homeobox Ba	Patterning of gut endoderm is disrupted, pancreas primordium does not differentiate	Sun and Hopkins, 2001
neckless (nls)	*aldh1a2*	aldehyde dehydrogenase 1 family, member A2	Pancreas primordium does not differentiate	Stafford and Prince, 2002
Endocrine pancreas				
angelina	*agl*	Unidentified	β-cells severely reduced	Kim et al., 2006
cheetah	*chee*	Unidentified	α-, β-, δ- cells scattered	Kim et al., 2006
dalmatian	*dal*	Unidentified	α-, β-, δ- cells scattered	Kim et al., 2006
floating head	*flh*		β-cells reduced	Biemar et al., 2001
knypek (*kny*)	*gpc4*	glypican 4	β-cells bilateral	Biemar et al., 2001
lazarus (*lzr*)	*pbx4*	pre-B-cell leukemia transcription factor 4	Anteriorly shifted endocrine pancreas	Popperl et al., 2000; dilorio et al., 2007
scarlet	*sle*	Unidentified	α-, β-, δ- cells scattered	Kim et al., 2006
schmalspur (*sur*)	*foxh1*	forkhead box H1	β-cells slightly reduced	Biemar et al., 2001
smoothened (*smo*)	*smo*	smoothened homolog (Drosophila)	Endocrine cells fail to develop	Roy et al., 2001; dilorio et al., 2002
sonic you (*syu*)	*shha*	sonic hedgehog a	Endocrine cells fail to develop	Roy et al., 2001; dilorio et al., 2002
spadetail	*tbx16*	T-box gene 16	β-cells bilateral or absent	Biemar et al., 2001
Exocrine Pancreas				
akreas	*ptf1a*	pancreas transcription factor 1 subunit alpha	Arrested growth of exocrine pancreas	Dong et al., 2008
apc	*apc*	adenomatous polyposis coli	Small exocrine pancreas	Goessling et al., 2008
daedalus (*dae*)	*fgf10a*	fibroblast growth factor 10a	Dysmorphic ductal system, ectopic endocrine cells	Manfroid et al., 2007; Dong et al., 2007
digestive expansion factor (*def*)	*def*	digestive expansion factor	Small exocrine pancreas	Chen et al., 2005; Tao et al., 2013a; Tao et al., 2013b
dnmt	*dnmt*	DNA methyltransferase 1	Degeneration of exocrine pancreas	Anderson et al., 2009b
donut	*met*	met proto-oncogene	Small exocrine pancreas	Anderson et al., 2013
ductjam (*dtj*)	*dtj*	unidentified	Small acini and dysmorphic ducts of exocrine pancreas	Yee et al., 2005

ducttrip (*dtp*)	*ahcy*	S-Adenosylhomocysteine hydrolase	Diminished acini and dysmoprhic ducts of exocrine pancreas	Yee et al., 2005; Matthews et al., 2009
earl grey (eal)	*sart3*	squamous cell carcinoma antigen recognised by T cells 3	Exocrine pancreas reduced or absent	Trede et al., 2007
flotte lotte (*flo*)	*ahctf1*	AT hook containing transcription factor 1	Small and degenerated exocrine pancreas	Yee et al., 2005; Davuluri et al., 2008
hdac1	*hdac1*	histone deacetylase 1	Small exocrine pancreas	Davuluri et al., 2008 Noel et al., 2008
ikarus	*fgf24*	fibroblast growth factor 24	Exocrine pancreas reduced	Manfroid et al., 2007
kras	*kras*	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	Arrested growth of exocrine pancreas	Park et al., 2008
mind bomb (*mib*)	*mib*	mind bomb	Premature exocrine differentiation	Esni et al., 2004
mitomess (*mms*)	*mms*	unidentified	Decreased zymogen and hypomorphic ducts of exocrine pancreas	Yee et al., 2005
nil per os (*npo*)	*rbm19*	RNA binding protein 19	Small exocrine pancreas	Mayer and Fishman, 2003
nom1	*nom1*	nucleolar protein with MIF4G domain 1	Small exocrine pancreas	Qin et al., 2014
pescadillo (*pes*)	*pes*	pescadillo	Small exocrine pancreas	Provost et al., 2012
piebald (*pie*)	*pie*	unidentified	Small and degenerated exocrine pancreas	Yee et al., 2005
rpl3	*rpl3*	ribosomal protein L3	Small exocrine pancreas	Provost et al., 2012
rpl6	*rpl6*	ribosomal protein L6	Small exocrine pancreas	Provost et al., 2012
rpl23	*rpl23*	ribosomal protein L23	Small exocrine pancreas	Provost et al., 2012
slimjim (slj)	*polr3b*	RNA polymerase III subunit 2	Small and degenerated exocrine pancreas	Yee et al., 2005; Yee et al., 2007
sweetbread (*swd*)	*swd*	unidentified	Small acini and hypomorphic ducts of exocrine pancreas	Yee et al., 2005
titania (*tti*)	*pwp2h*	periodic tryptophan protein 2 gene homolog (yeast)	Small exocrine pancreas	Boglev et al., 2013
Endocrine and Exocrine Pancreas				
minime	*mnm*	Unidentified	α-, β- cells scattered, β- cells reduced, exocrine pancreas reduced	Kim et al., 2006
peppershaker	*pps*	Unidentified	α-, β-, δ- cells scattered, exocrine pancreas disorganised	Kim et al., 2006
sea dragon	*sdr*	Unidentified	α-, β-, δ- cells reduced, duplicated exocrine pancreas	Kim et al., 2006

Table 1-3 Zebrafish mutations affecting pancreas development. Adapted from Gnugge et al., 2004, Kinkel and Prince, 2009 and Yee, 2010.

1.4 Ribosome

Since ribosomal biogenesis genes play a crucial role in exocrine pancreas development and Nol9 is involved in rRNA processing (Heindl and Martinez, 2010), I will briefly describe ribosomes, their biogenesis and function and ribosomopathies, human disorders where ribosome biogenesis or function is disrupted.

Ribosomes are essential to all organisms as a catalyst for protein synthesis and are required for cell survival, growth and differentiation. They are made up of both ribosomal ribonucleic acid (rRNA) and ribosomal proteins (RPs). The eukaryotic ribosome consists of two subunits: the 40S small subunit (SSU) and the 60S large subunit (LSU). The SSU is composed of a single rRNA (18S) and about 30 RPs, depending on the species whilst the LSU contains three rRNAs (5S, 5.8S, 25S) and about 45 RPs, varying between species (Panse and Johnson, 2010). The process of translation consists of three stages: initiation, elongation and termination, each requiring specific translation factors (Figure 1-6) (Walsh and Mohr, 2011). During the initiation step in eukaryotes, the 40S ribosome subunits bind to eukaryotic translation initiation factors, eIF1, eIF1a, the eIF3 complex, eIF5 and eIF2-GTP to form a 43S pre-initiation complex that is loaded with the initiator-methionine tRNA (Met-tRNA$_i$) in the P site. The multisubunit complex eIF4F, consisting of the cap-binding proteins eIF4E, eIF4G and eIF4A, positions the 43S complex onto the 5' end of a capped polyadenylated mRNA. The poly(A)-binding protein (PABP) recognises the polyadenylated 3' mRNA end and associates with the eIF4G resulting in a 'closed-loop' topology. The MNK kinase binds to eIF4G before phosphorylating eIF4E. After the 48S complex scans the mRNA and recognises the AUG start codon, the 60S subunit joins and triggers the release of the initiation factors (Jackson et al., 2010). During the elongation step, eEF1A•GTP delivers charged tRNAs to the 80S ribosome A site. After peptide-bond formation is catalysed by the 60S subunit, the eukaryotic elongation factor 2 (eEF2) translocates the 80S ribosome resulting in the deacetylated tRNA in the E site, the peptidyl-tRNA in the P site and a re-exposed A site (Herbert and Proud, 2007). During termination, the eukaryotic release factor 1 (eRF1) identifies the stop codon in the A site and stimulates the arrest of 80S and the release of the polypeptide. eRF3 frees eRF1 from the ribosome and the complex is dismantled by a group of initiation factors and nucleotide-hydrolysis by ABCE1, thus promoting the release of mRNA, tRNAs and the recycling of ribosomal subunits (Dinman and Berry, 2007).

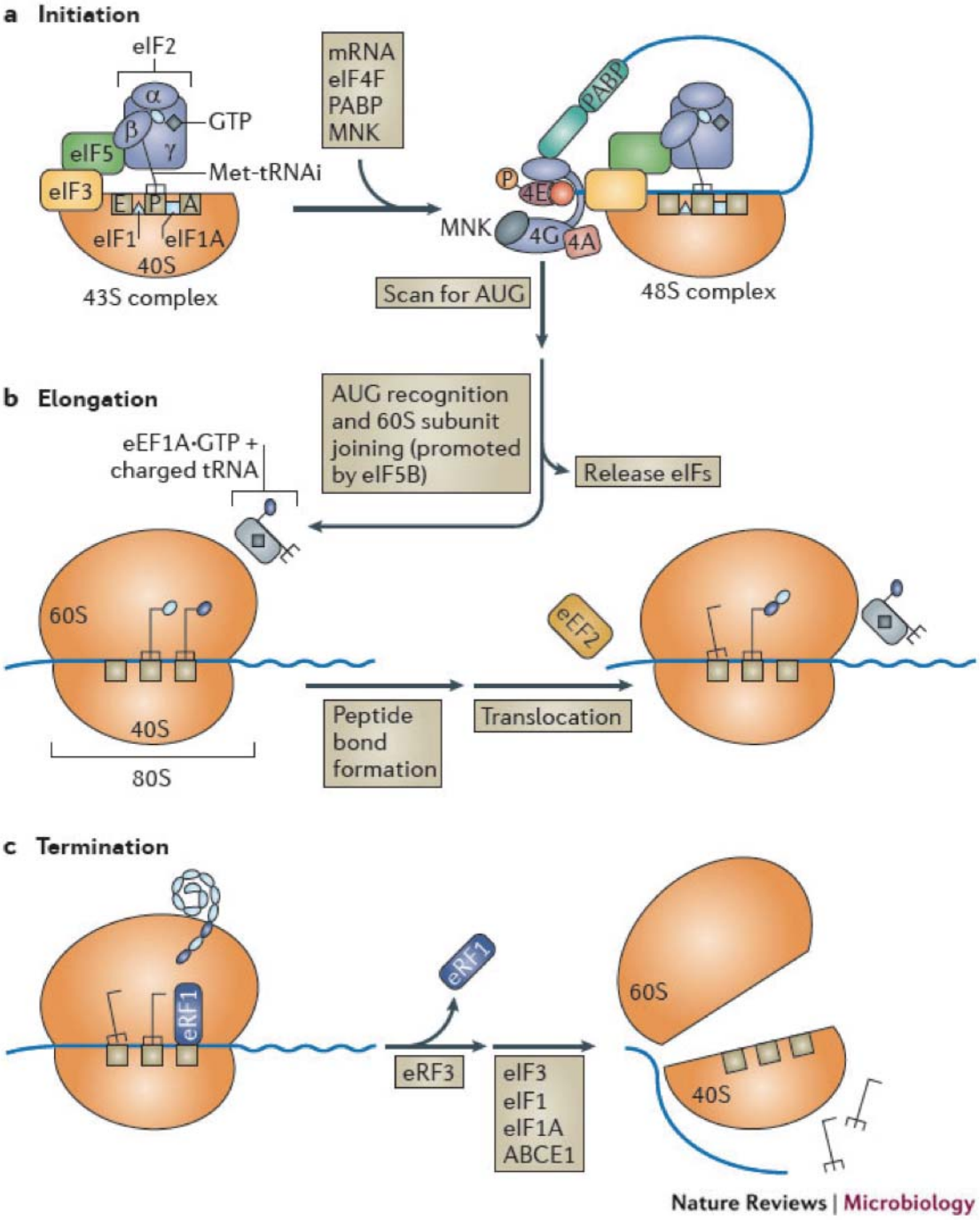

Figure 1-6 Overview of eukaryotic mRNA translation. The process of translation is divided into initiation, elongation and termination. During initiation (a), the 40S ribosome subunits bind to eukaryotic translation initiation factors, the initiator-methionine tRNA (Met-tRNA$_i$), the poly(A)-binding protein (PABP) and MNK to form the 48S complex that scans the mRNA. After AUG start codon recognition and joining of the 60S subunit, the initiation factors are released. During elongation (b), charged tRNAs are delivered by eEF1A•GTP and peptide-bond formation is catalysed by the 60S subunit. The 80S ribosome is then translocated by eukaryotic elongation factor (eEF2). During termination (c), the recognition of a stop codon by the eukaryotic release factor 1 (eRF1) initiates arrest of the 80S and the release of the polypeptide. eRF1 is released by eRF3 and the 80S complex is dismantled by several initiation factors and ABCE1-directed nucleotide hydrolysis resulting in subunit recycling. Taken from Walsh and Mohr, 2011.

1.4.1 Ribosome biogenesis

Ribosome biogenesis is an energy intensive and remarkably complex process that occurs at multiple cell sites. It utilises 60% of total cellular transcription in a growing yeast cell, with 2000 ribosomes synthesised every minute (Warner et al., 2001) whilst 7500 ribosomal subunits are made per minute in a mammalian HeLa cell (Lewis and Tollervey, 2000). It requires the coordinated action of all three RNA polymerase (RNAP I, II, III) and the synthesis of 4 rRNAs, 82 core RPs, more than 200 non-ribosomal proteins and approximately 70 small nucleolar RNAs (snoRNAs) (Panse and Johnson, 2010).

This multi-step process begins in the nucleolus with the RNAP I transcription of the ribosomal deoxyribonucleic acid (rDNA) units that are found in hundreds of copies as tandem repeats across the human genome (Figure 1-7). A single polycistronic 45S pre-rRNA is produced and subjected to several co-transcriptional modifications such as methylation and pseudo-uridylation by small nucleolar RNA-protein complexes (snoRNPs) (Lafontaine and Tollervey, 2006). Ribosomal Proteins (RPs) are synthesised in the cytoplasm and imported into the nucleus/nucleolus where they assemble on the nascent pre-rRNA in a hierarchical fashion. The resulting new structures recruit non-ribosomal *trans*-acting factors that are important in liberating mature RNAs from pre-rRNA precursors. Additional pre-rRNA processing involves removal of the flanking and internal spacer regions by a series of endo- and exo-nucleolytic cleavage events. Cleavage at site A2, the spacer region between the 18S and 5.8S rRNA results in the formation of pre-40S and pre-60S particles. The 5S rRNA is synthesised concurrently in the nucleoplasm by RNA pol III and is incorporated into the pre-60S subunit. Figure 1-8 and Figure 1-9 show the pre-ribosomal particles and non-ribosomal proteins including the ones mentioned in this thesis along the 40S (Figure 1-8) and the 60S assembly pathways (Figure 1-9) in *Saccharomyces cerevisiae* (Kressler et al., 2010).

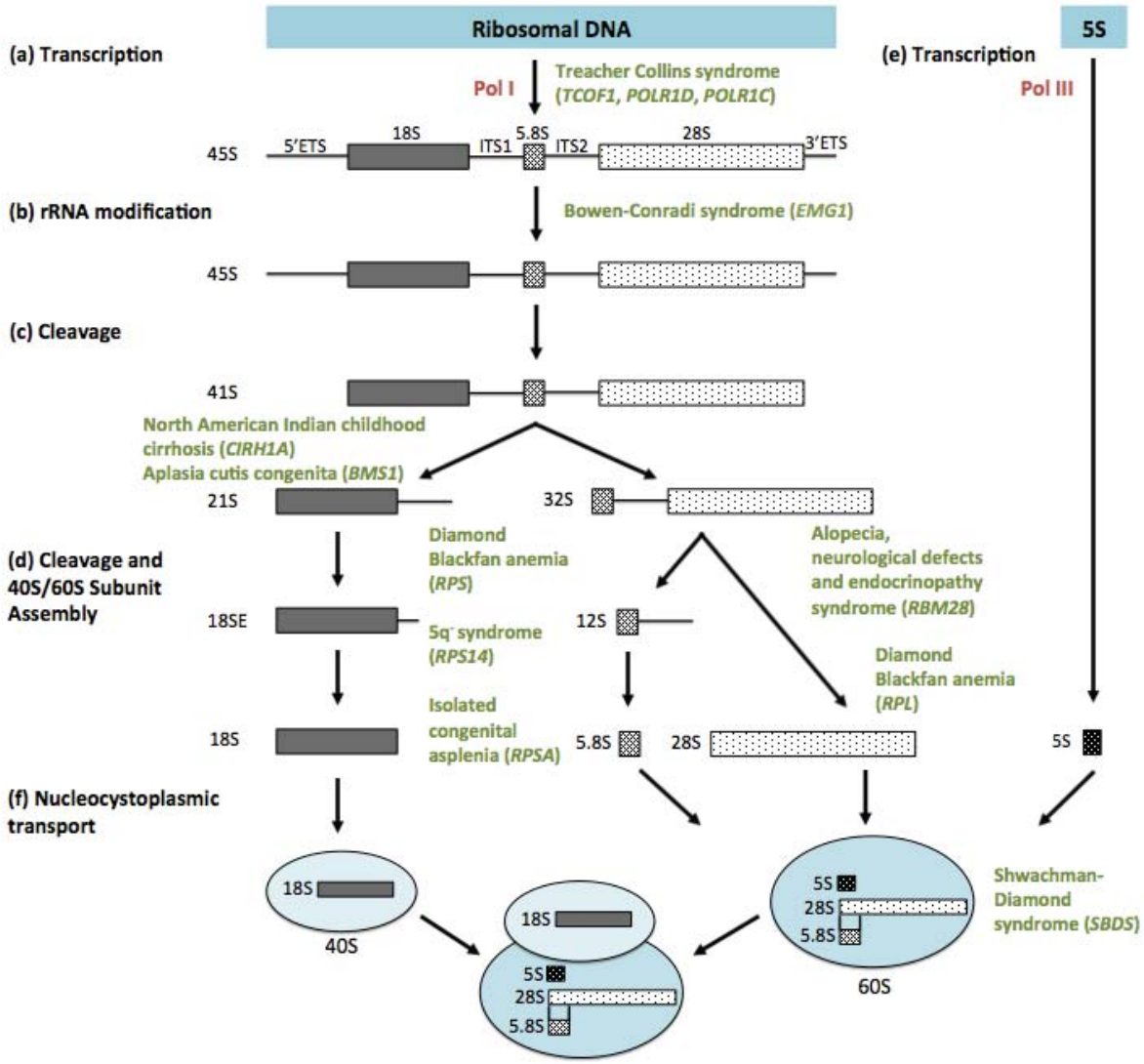

Figure 1-7 Simplified schematic of ribosome biogenesis and the stages which is affected in different ribosomopathies. Ribosome biogenesis starts with RNAP I transcription of rDNA (a) to produce a 45S pre-rRNA that is modified (b) and cleaved by exo- and endo-nucleases (c, d) to produce pre-40S rRNA (18S) and pre-60S rRNA (5.8S and 28S) particles. Ribosomal proteins assemble on the nascent pre-rRNA particles to form pre-40S and pre-60S particles (d). Meanwhile the 5S rRNA is transcribed by RNAP III and is incorporated into the pre-60S particles (e). The pre-ribosomal particles are exported to the cytoplasm (f) where final steps of maturation occur to produce translationally competent ribosomes. The names of the various ribosomopathies and their causative genes in brackets are shown. Adapted from Narla and Ebert, 2010 and Liu et al., 2013.

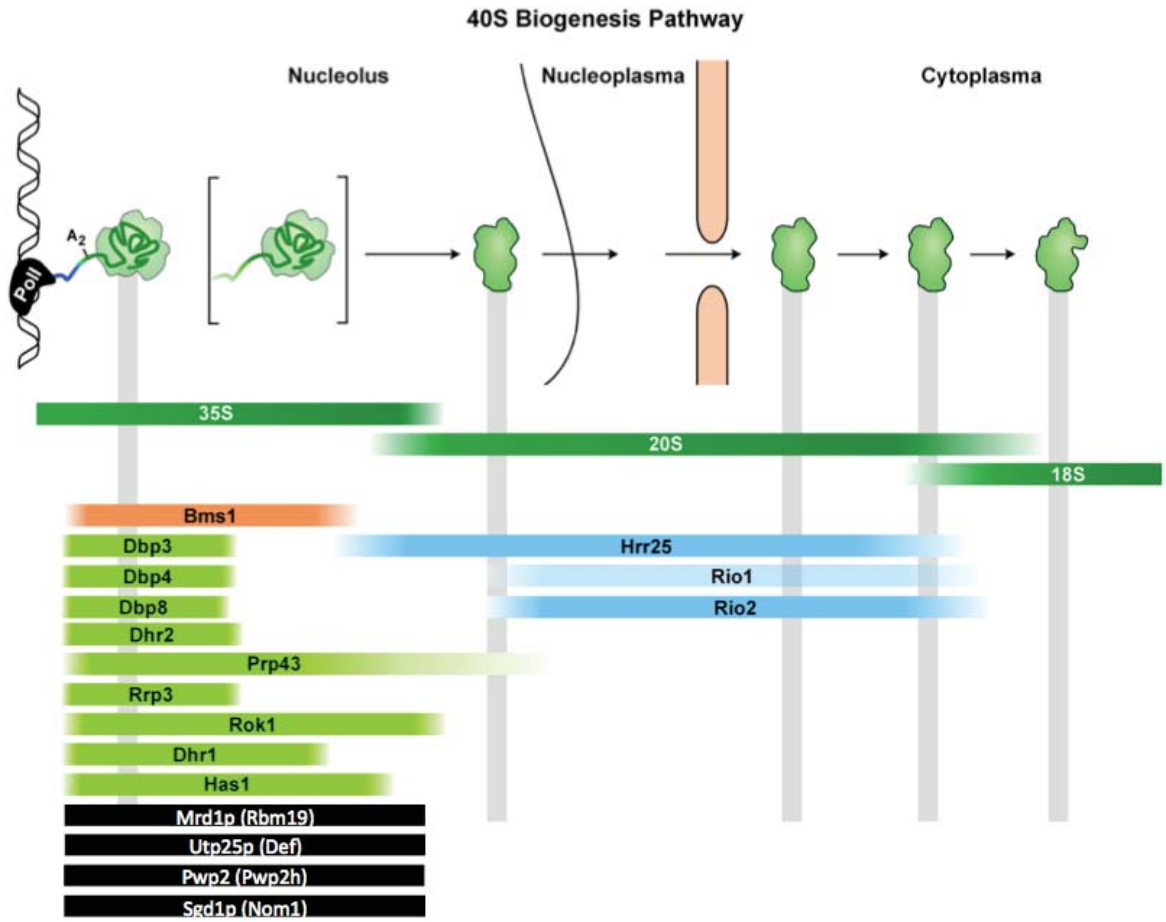

Figure 1-8 The 40S assembly pathway in *Saccharomyces cerevisiae*. The major 40S pre-ribosomal intermediates, their rRNAs (dark green), DExD/H-box ATPases (green), kinases (light blue) and GTPase (orange) are shown. The four proteins Mrd1p, Utp25p, Pwp2 and Sgd1p and their zebrafish orthologues in brackets (black) are mentioned in this thesis. The 35S pre-rRNA is transcribed by RNA Pol I and modified by snoRNPs. RPs and non-ribosomal factors assemble co-transcriptionally with the pre-rRNA and cleavage at site A2 generates the 20S pre-rRNA. The final rRNA processing occurs in the cytoplasm. Adapted from Kressler et al., 2010.

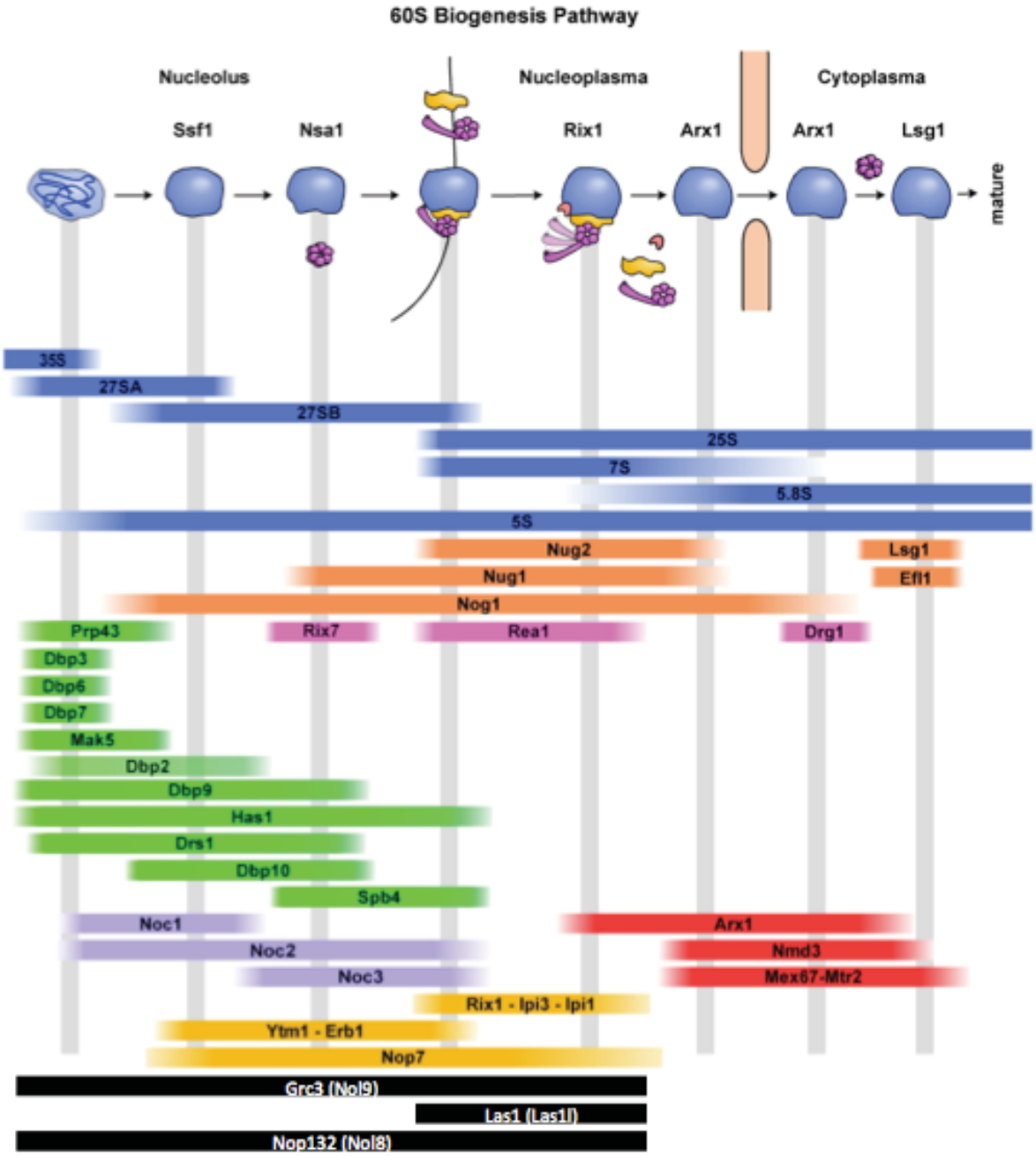

Figure 1-9 The 60S assembly pathway in *Saccharomyces cerevisiae*. The 60S pre-ribosomal intermediates, their rRNAs (blue), GTPases (orange, DexD/H-box, ATPases (green), AAA-type ATPases (pink), subcomplexes (purple/yellow) and export factors are shown. The proteins Grc3, Las1 and Nop132 and their zebrafish orthologues in brackets (black) are mentioned in this thesis. Adapted from Kressler et al., 2010.

The subsequent biogenesis and export pathways of pre-40S and pre-60S particles are independent. Pre-ribosomal particles are transported through the nuclear pore complex (NPC) to be released into the cytoplasm, where the final steps of maturation occur. Previously associated *trans*-acting and export factors are released from the pre-ribosomal particles and the final rRNA processing steps occur before the subunits can achieve translational competence (Liu et al., 2013; Panse and Johnson, 2010). Figure 1-10 shows the export of the pre-60S particles and the release of non-ribosomal factors in *Saccharomyces cerevisiae* (Kressler et al., 2010).

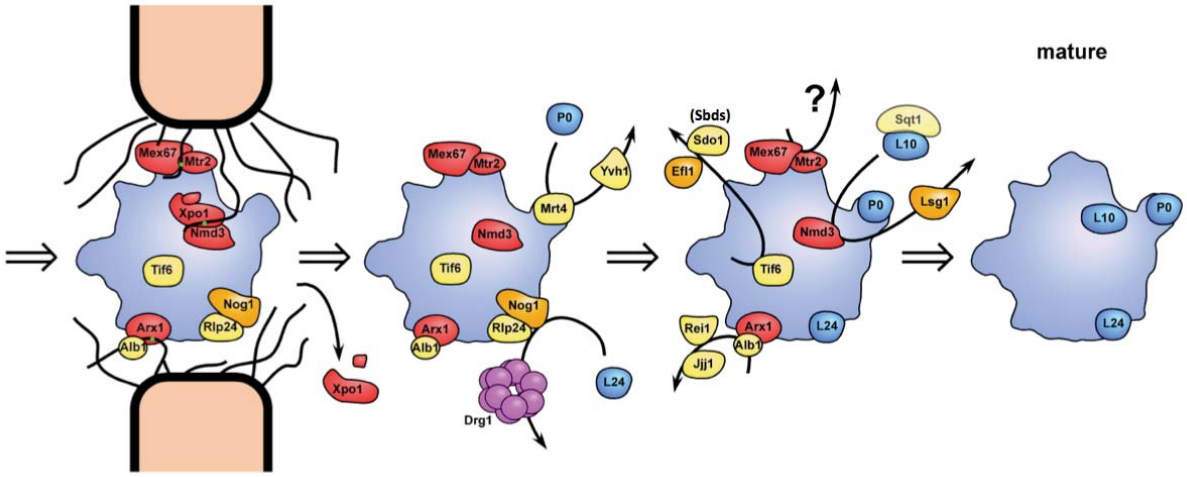

Figure 1-10 The export and maturation of pre-60S particles in *Saccharomyces cerevisiae*. The export factors (red), GTPases (orange), Drg1 AAA-ATPase (pink), other non-ribosomal factors (yellow) and ribosomal proteins (blue) are shown. Sdo1 and its zebrafish orthologue in bracket are mentioned in this thesis. Adapted from Kressler et al., 2010.

1.4.2 Ribosomopathies

Ribosomopathies consist of a group of disorders with defects in genes involved in ribosome biogenesis and function (Figure 1-7, Table 1-4). They show a wide range of symptoms, modes of inheritance and impaired molecular processes but there are some shared patterns between them. Firstly, most of the diseases are caused by haploinsufficiency or partial loss of protein expression suggesting that homozygous null mutations of genes involved in ribosome biogenesis are embryonic lethal. Secondly, even though there is no one symptom present in all diseases, shared clinical features between diseases are observed including defects in growth and development, haematological phenotypes such as anaemia and congenital anomalies such as thumb and craniofacial defects and susceptibility to cancer. The affected cell compartments show hypoplasia with decreased proliferation and increased apoptosis. Thirdly, a common cell death mechanism is probable since mutations in genes involved in ribosome biogenesis can activate the Tp53 pathway (Armistead and Triggs-Raine, 2014; Freed et al., 2010; Liu et al., 2013; Narla and Ebert, 2010).

The ribosomopathies that are inherited in an autosomal dominant fashion include Treacher Collins syndrome (TCS), Diamond Blackfan anemia (DBA), isolated congenital asplenia (ICAS) and aplasia cutis congenita (ACC). TCS (OMIM #606847 and #613717) and its autosomal recessive form (OMIM #248390) is characterised by craniofacial defects including abnormalities of the eyes, ears and facial bones including those of the cheek and lower jaw. In 1996, the Treacher Collins Syndrome Collaborative Group had identified five mutations in the Treacle gene (*TCOF1*) by positional cloning (The Treacher Collins Syndrome Collaborative Group, 1996) and currently the *TCOF1* database contains over 200 mutations including mostly nonsense, insertion, deletion and alternative splicing mutations (www.genoma.ib.usp.br/TCOF1_database). The protein Treacle has been shown to be essential for transcription of ribosomal DNA and is a component of the snoRNP complexes that plays a role in methylation of pre-rRNA transcripts (Gonzales et al., 2005; Valdez et al., 2004). Recently, mutations in genes *POLR1C* and *POLR1D* that encode shared subunits of RNA polymerase I and III have been identified as causing TCS (Dauwerse et al., 2011). By performing genome-wide copy number analysis, Dauwerse *et al*. identified a de novo deletion in *POLR1D* gene in an individual and subsequently detected 20 additional heterozygous mutations. They also sequenced the candidate gene *POLR1C* since it interacts strongly with

POLR1D and discovered compound heterozygous mutations in three patients (Dauwerse et al., 2011).

DBA (OMIM #105650, #610629, #612527, #612528, #612561, #612562, #612563, #613308, #613309, #614900, #615550 and #615909) is a congenital bone marrow failure syndrome and is characterised by large red blood cells due to low haemoglobin levels (macrocytic anaemia), abnormal decrease in immature red blood cells (reticulocytopenia), absence or selective decrease of erythroid precursors and increased risk of malignancy. In 40-62% of patients, additional physical abnormalities are present including cardiac defects, short stature and thumb abnormalities (Lipton and Ellis, 2009). The most commonly affected gene in DBA is *RPS19* (Draptchinskaia et al., 1999) but mutations in other ribosomal protein genes have also been found and include *RPS24* (Gazda et al., 2006), *RPS17* (Cmejla et al., 2007), *RPL35A* (Farrar et al., 2008), *RPL5* , *RPL11*, *RPS7* (Gazda et al., 2008), *RPS26*, *RPS10* (Doherty et al., 2010), *RPL26* (Gazda et al., 2012), *RPL15* (Landowski et al., 2013) and *RPS29* (Mirabello et al., 2014). In addition, causative mutations in the haematopoietic transcription factor gene *GATA1* have been found in patients with DBA (Ludwig et al., 2014; Parrella et al., 2014; Sankaran et al., 2012). More recently, Gripp *et al.* reported mutations in *RPS26*, *TSR2* and *RPS28* in patients that present with combining features of TCS and DBA (Gripp et al., 2014).

ICAS (OMIM #71400) is characterised by the absence of spleen with no other developmental defects at birth. Most affected individuals are susceptible to life-threatening bacterial infections (Mahlaoui et al., 2011). Recently, Bolze *et al.* identified heterozygous mutations in *ribosomal protein SA* (*RPSA*) in 18 patients from 8 families (Bolze et al., 2013). *RPSA* encodes a protein that is involved in pre-rRNA processing (O'Donohue et al., 2010) and is also a component of the small subunit (Ben-Shem et al., 2011).

ACC (OMIM #107600) is a non-syndromic disorder characterised by a congenital absence of skin usually on the scalp vertex (Nousbeck et al., 2008). Marneros identified the causative mutation of the autosomal dominant form of ACC in the ribosomal GTPase *BMS1* by genome-wide linkage analysis and exome sequencing (Marneros, 2013). The yeast orthologue of *BMS1*, *BMS1P* is required for the synthesis of the 18S rRNA precursors of the 40S ribosomal subunits (Gelperin et al., 2001).

Shwachman-Diamond syndrome (SDS), Bowen-Conradi syndrome (BCS), North American Indian childhood cirrhosis (NAIC) and alopecia, neurological defects and endocrinopathy (ANE) syndrome are amongst the autosomal recessive ribosomopathies. BCS (OMIM #211180) has been described mostly in the Hutterite population. It is a lethal and autosomal recessive syndrome characterised by pre- and post-natal growth retardation, smaller head circumference (microcephaly), slowed mental and physical processes (psychomotor delay) and multiple joint abnormalities (Bowen and Conradi, 1976). Armistead *et al.* (Armistead et al., 2009) recently reported a missense mutation in *EMG1*, one of 35 candidate genes in the interval mapped by linkage and haplotype analysis (Lamont et al., 2005). Emg1 is a putative methyltransferase and is involved in maturation of 18S rRNA and biogenesis of the 40S ribosomal subunit (Eschrich et al., 2002; Leulliot et al., 2008; Liu and Thiele, 2001).

NAIC (OMIM #604901) is found in the Ojibway-Cree population in Canada and manifests as neonatal jaundice, eventually leading to progressive destruction of the small bile ducts of the liver (biliary cirrhosis) (Betard et al., 2000). By genome-wide scan (Betard et al., 2000) and sequencing of candidate genes in the haplotype shared by all patients, a homozygous missense mutation (R565W) in the gene *CIRH1A* was reported to cause NAIC (Chagnon et al., 2002). Cirhin is a component of the small ribosomal subunit (SSU) processome t-UTP subcomplex and is required for processing of 18S rRNA and optimal rDNA transcription in both yeast and human cells (Gallagher et al., 2004; Prieto and McStay, 2007). Recently, Nol11 was found to interact with Cirhin and this interaction is partially disrupted by the R565W mutation implicating a role for *NOL11* in the pathogenesis of NAIC. Nol11 is also a component of the SSU processome and has similar function to Cirhin (Freed et al., 2012). *Cirh1a* knockout in mice was reported to be embryonic lethal whilst heterozygotes appeared to develop normally (Yu et al., 2009). Recently, a zebrafish model of NAIC was generated by morpholino knockdown of *cirh1a* (Wilkins et al., 2013). Cirhin-deficient 5 d.p.f. zebrafish larvae showed defects in biliary and canalicular morphology and hepatobiliary function (Wilkins et al., 2013).

ANE syndrome (OMIM #612079) is characterised by hair loss, microcephaly, mental retardation, progressive motor retardation and adrenal insufficiency. Nousbeck *et al.* identified a loss-of-function mutation in *RBM28* in a family with ANE syndrome by homozygosity mapping and candidate gene analysis (Nousbeck et al., 2008). *NOP4P*, the

yeast orthologue of *RBM28* is required for production of mature rRNA of the 60S ribosomal subunits (Sun and Woolford, 1994, 1997)

In contrast to the ribosomopathies described so far, a region that contains 40 genes in the long arm of chromosome 5 is deleted in the $5q^-$ syndrome (OMIM #153550) (Van den Berghe et al., 1974). $5q^-$ syndrome is a myelodysplastic syndrome (MDS) that shows ineffective production of non-lymphocyte blood cells and is characterised by macrocytic anaemia. *RPS14*, one of the genes deleted in the $5q^-$ region, was shown to be associated with the syndrome in an RNA interference-based functional screen (Ebert et al., 2008). Inactivating mutations in ribosomal protein genes in DBA is analogous to acquired haploinsufficiency for *RPS14* in $5q^-$ syndrome and contribute to the erythroid defects in both disorders.

Cartilage hair hypoplasia (CHH; OMIM #250460) and its variants anauxetic dysplasia (OMIM #607095) and metaphyseal dysplasia without hypotrichosis (OMIM #250460) (Bonafe et al., 2002; Ridanpaa et al., 2001; Thiel et al., 2005) and X-linked dyskeratosis congenita (X-DC; OMIM #305000) (Heiss et al., 1998) have initially been described as ribosomopathies. However the role of ribosome pathology in these two disorders is currently being challenged and therefore they will not be addressed in this chapter.

In general, the most surprising feature of ribosomopathies is that specific phenotypes in particular cell types are observed even though ribosomes are universally required for protein synthesis and ribosomal biogenesis proteins are ubiquitously expressed. The underlying mechanisms are not fully understood but the main hypothesis is that that highly proliferating tissues have a greater need for ribosomes and therefore are more severely affected than tissues with lower proliferating activity.

Ribosomopathies	Gene	Genetic Defect	Impaired Molecular Function	Clinical Features	References
Treacher Collins syndrome (TCS)	*TCOF1, POLR1D, POLR1C*	*TCOF1* and *POLR1D*: Autosomal Dominant *POLR1C:* Autosomal Recessive	Transcription of rDNA; *TCOF1* also involved in methylation of 18S rRNA	Cranofacial abnormalities	Shu et al., 2003, The Treacher Collins Syndrome Collaborative Group, 1996, Dauwerse et al., 2011.
Diamond Blackfan anemia (DBA)	*RPS19, RPS24, RPS17, RPL35A, RPL5, RPL11, RPS7, RPS10, RPS26, RPL26, RPL15, RPS29, GATA1*	Sporadic, autosomal dominant (40-45%)	Processing of pre-rRNA of 40S or 60S ribosomal subunit and proliferation/differentiation of haemopoietic progenitors	Macrocytic anaemia, cranofacial abnormalities,thumb abnormalities, short stature	Draptchinskaia et al., 1999, Gazda et al., 2006, Cmejla et al., 2007, Farrar et al., 2008, Gazda et al., 2008, Doherty et al., 2010, Gazda et al., 2012, Sankaran et al., 2012, Landowski et al., 2013, Ludwig et al., 2014, Mirabello et al., 2014, Parrella et al., 2014.
Treacher Collins syndrome and Diamond Blackfan anemia	*RPS26, TSR2, RPS28*	Autosomal Dominant	Processing of pre-rRNA of 40S ribosomal subunit	Macrocytic anaemia, cranofacial abnormalities	Gripp et al. 2014
Isolated congenital asplenia (ICAS)	*RPSA*	Autosomal Dominant	Component of 40S ribosomal subunit	Absence of spleen	Bolze et al., 2013
Aplasia cutis congenita (ACC)	*BMS1*	Autosomal Dominant	Ribosomal GTPase	Agenesis of skin, usually on scalp vertex	Marneros et al., 2013
Shwachman-Diamond syndrome (SDS)	*SBDS*	Autosomal Recessive	Maturation and export of 60S ribosomal subunit	Neutropenia/infections, pancreatic insufficiency, short stature	Boocock et al., 2003
Bowen-Conradi syndrome (BCS)	*EMG1*	Autosomal Recessive	Maturation of 40S ribosomal subunit	Growth retardation, psychomotor delay	Armistead et al., 2009
North American Indian childhood cirrhosis (NAIC)	*CIRH1A*	Autosomal Recessive	Maturation of 18S rRNA and rDNA transcription	Cirrhosis	Chagnon et al., 2002
Alopecia, neurological defects and endocrinopathy (ANE) syndrome	*RBM28*	Autosomal Recessive	Processing of pre-rRNA of 60S ribosomal subunit	Hair loss, microcephaly mental retardation, progressive motor retardation, adrenal insufficiency	Nousbeck et al., 2008
$5q^-$ syndrome	*RPS14*	Unknown	Maturation and export of 60S ribosomal subunit	Myelodysplastic syndrome, macrocytic anaemia	Van den Berghe et al., 1974

Table 1-4 Summary of ribosomopathies with details about the causative gene(s), genetic defect, impaired molecular function and clinical features. Adapted from Freed et al., 2010, Narla and Ebert 2010, Liu et al., 2013 and Armistead et al., 2014.

1.5 Thesis aims and objectives

The aim of this thesis is to characterise *nol9*^{sa1022} mutants in order to determine the function of Nol9 in zebrafish pancreas development and to provide insight into the mechanisms involved in rRNA processing mutants. To achieve this, the following objectives will need to be attained:

1. Confirm that *nol9* is the affected gene in *nol9*^{sa1022} mutant (Chapter 3)

Zebrafish mutants generated by ENU mutagenesis contain several disruptive mutations. To provide support that *nol9* is responsible for the pancreas phenotype in *nol9*^{sa1022} mutants, genotype-phenotype correlation and *nol9* morpholino knockdown was carried out in transgenic line *Tg(ins:mCherry)*^{jh2}*;Tg(ptf1a:EGFP)*^{jh1}.

2. Examine *nol9*^{sa1022} mutant for additional non-pancreatic defects (Chapter 3)

Zebrafish mutants of ribosomal biogenesis genes have defects in other organs including the intestine, liver and jaw. The *nol9*^{sa1022} mutant was analysed for those defects by staining and labelling using tissue-specific probes.

3. Identify the stage at which pancreas development is affected in *nol9*^{sa1022} mutant (Chapter 3)

It is important to determine the developmental stage at which pancreatic organogenesis is affected, i.e. during endoderm induction, specification of pancreas progenitor cells or proliferation of differentiated pancreatic cells. The pancreas development was studied using pancreatic-stage specific probes and the transgenic line *Tg(ins:mCherry)*^{jh2}*;Tg(ptf1a:EGFP)*^{jh1}.

4. Examine the formation of the endocrine pancreas namely the pancreatic islet and the secondary islets in *nol9*^{sa1022} mutant (Chapter 3)

The study of the endocrine pancreas is valuable since formation of the exocrine pancreas is impaired in *nol9*^{sa1022} mutant. The formation and differentiation of endocrine cells of the pancreatic islet was investigated using various antibodies and the formation of the secondary islets was examined using the Notch-inhibitor DAPT and the transgenic line *Tg(ins:mCherry)*^{jh2}*;Tg(ptf1a:EGFP)*^{jh1}.

5. Assess cell proliferation and cell death in *nol9*^{sa1022} mutant (Chapter 3)

Impaired cell proliferation and increased cell death can contribute to the defects in digestive organ development in *nol9*^{sa1022} mutant. Therefore the cell cycle of *nol9*^{sa1022} mutant was examined using flow cytometry analysis, cell proliferation rate and cell death of the exocrine pancreas of *nol9*^{sa1022} mutant were studied using Bromodeoxyuridine (BrdU) incorporation assay and terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay respectively.

6. Determine the expression pattern of *nol9* during zebrafish development (Chapter 3)

The expression pattern of *nol9* may help explain the tissue-specific defects observed in *nol9*^{sa1022} mutants and was examined using *in situ* hybridisation with a probe targeting the *nol9* gene.

7. Determine the function of Nol9 in rRNA processing and ribosome biogenesis in zebrafish (Chapter 3)

The human NOL9 protein has been shown to be involved in the processing of 28S rRNA of the large ribosomal subunit. The function of Nol9 in rRNA processing and ribosome biogenesis in zebrafish was determined by Northern blot analysis and polysome fractionation respectively.

8. Characterise *las1l*^{sa674}, a mutant in a Nol9-interacting protein (Chapter 4)

Assessment of the literature identified proteins that interact with Nol9 including LAS1-like (LAS1). The *las1l*^{sa674}, a mutant in the zebrafish orthologue of *LAS1L* was characterised using staining and the transgenic line *Tg(ins:mCherry)*^{jh2}*;Tg(ptf1a:EGFP)*^{jh1}.

9. Identify changes in gene expression in *nol9*^{sa1022} and *las1l*^{sa674} mutants (Chapter 5)

The mRNA expression profiles of *nol9*^{sa1022} and *las1l*^{sa674} mutants can reveal insight into the functions of *nol9* and *las1l* during development and was conducted using Differential Expression Transcript Counting Technique (DeTCT). The enrichment for gene ontology terms and pathways was carried out using the R topGO package and the Database for Annotation, Visualization and Integrated Discovery (DAVID).

10. Determine whether the mechanism involved in the pancreatic defects of *nol9*^{sa1022} mutants is Tp53-dependent or –independent (Chapter 5)

The Tp53 signalling pathway is known to function as a surveillance mechanism in response to defective ribosome biogenesis. It was determined whether the mechanism operating in *nol9*^{sa1022} and is Tp53-dependent or –independent by using a *tp53* loss of function mutant line and the transgenic line *Tg(ins:mCherry)*^{jh2}*;Tg(ptf1a:EGFP)*^{jh1}.

11. Compare the mRNA expression profiles of *nol9*^{sa1022} and *las1l*^{sa674} mutants with those of other rRNA processing mutants, *tti*^{s450} and *set*^{s453} (Chapter 5)

The mRNA expression profiles of *nol9*^{sa1022}, *las1l*^{sa674}, *tti*^{s450} and *set*^{s435} mutants were compared to identify shared genes and signalling pathways so as to help in deciphering the mechanism involved in the defective development of digestive organs of all four mutants.