# Chapter 5 Analysis of mRNA expression profiles of *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{s453}$ mutants

## 5.1 Introduction

The expression of mRNA determines the functional and physiological status of a cell, tissue or organism. The study of transcripts by mRNA expression has proved to be a powerful tool to uncover genes of unknown function and coordinate activities of genes in different biological processes, such as in developmental processes and genetic disorders (Ko, 2001; Smith and Greenfield, 2003). Various platforms have been used to study zebrafish mRNA expression profiles, including the most commonly used DNA microarray (Smith and Greenfield, 2003). This technique can detect thousands of zebrafish transcripts by measuring the intensity of signal produced by hybridisation of cDNA to probes attached to a solid surface. It has successfully identified, amongst others, gene clusters at distinct developmental stages (Mathavan et al., 2005), liver-enriched genes in adult zebrafish (Cheng et al., 2006), genes enriched in the gastrointestinal tract at various developmental stages (Stuckenholz et al., 2009) and genes enriched in *def* mutant (Chen et al., 2005) and in *sbds* knockdown (Provost et al., 2012). A second technology, serial analysis of gene expression (SAGE) (Velculescu et al., 1995), has been used only in a handful of studies in zebrafish, including the analysis of mRNA expression of ovarian follicles (Knoll-Gellida et al., 2006) and livers of male and female zebrafish (Zheng et al., 2013). This method can discover new genes through better quantification of transcription levels compared to microarray analysis but is more expensive (Ibrahim et al., 2005; van Ruissen et al., 2005). RNA-sequencing (RNA-seq) technology has now become the state-of-art method for studying mRNA expression (Wang et al., 2009). The technique involves converting an RNA population into a library containing cDNA fragments that have adapters attached to one or both ends (Wang et al., 2009). The molecules are then subjected to sequencing in a high-throughput manner to produce short sequences from one or both ends and the reads are either assembled *de novo* or aligned to a reference genome or transcripts (Wang et al., 2009). A genome-scale transcription map is

produced and can inform on the gene expression levels and transcriptional structure (Wang et al., 2009). Compared to the DNA microarray, RNA-seq can profile RNA transcript abundance more accurately and within a greater dynamic range, can detect novel transcripts, and can characterise and monitor alternatively spliced transcripts and non-coding RNAs (Marioni et al., 2008). It has been used in various studies including zebrafish mRNA expression at various developmental stages (Vesterlund et al., 2011; Yang et al., 2013), small RNA expression during early development (Wei et al., 2012), mRNA expression profiling of a zebrafish model of Diamond-Blackfan anemia (Jia et al., 2013) and transcriptome analyses of zebrafish mutants genes in *RNA-binding region (RNP1, RRM) containing 3* (*rnpc3*) (Markmiller et al., 2014) and *nom1* mutants (Qin et al., 2014).

Differential Expression Transcript Counting Technique (DeTCT) is a high-throughput mRNA expression profiling technique developed by John Collins (unpublished). It produces count data by tag sequencing the 3' end of transcripts from strand-specific Illumina libraries that have been enriched for short fragments adjacent to the polyA tail. DeTCT is used in the Sanger ZMP to determine the mRNA expression profiles of phenotypic ZMP mutants that are associated with an allele (Kettleborough et al., 2013). The DeTCT pipeline starts with RNA extraction of mutants and wild-type siblings. In brief, the library preparation (described in detail in Section 2.6.1) consists of annealing the poly(T) primer to pull down poly(A) transcripts (Figure 2-1). A first partial Illumina adapter is ligated to the 5' end of the transcripts. Subsequently, an oligonucleotide comprising of a second partial Illumina adapter, a six base index tag, four random bases and 14 T bases is ligated to the 3' end of the transcripts. The RNA ligated to the Illumina adapters is reverse transcribed and is amplified using full-length primers to produce libraries. The DeTCT libraries are subjected to paired-end sequencing on an Illumina HiSeq 2000. The sequencing data produces two sets of reads, the read 1 contains the random bases, the index tag, 14 T bases and the remaining is transcript-specific sequence whereas the read 2 is entirely transcript-specific sequence. The first step of the differential expression analysis is to differentiate between the different libraries by using their unique index tag on read 1 and to remove non-species sequences (Figure 5-1) (Section 2.6.2). Subsequently, Burrow-Wheeler Alignment (BWA) tool is used to align the sequencing reads to the Zv9 genome assembly independently of the known gene annotation (Li and Durbin, 2009). The random bases of read 1 are then used to remove PCR duplicates to reduce the effects of PCR amplification bias introduced during library preparation. The abundance of read 2 is counted where reads accumulate across all libraries

and is attached to the 3' end of regions defined using read 1. DESeq (Anders and Huber, 2010) is then used to determine the genomic regions that show significant differential transcript abundance and the Ensembl annotation is used to match the 3' end of transcripts identified by read 1. Lastly, a gene list of differentially expressed genes is produced filtered on the DESEq adjusted *p*-value and the distance to the nearest Ensembl gene.

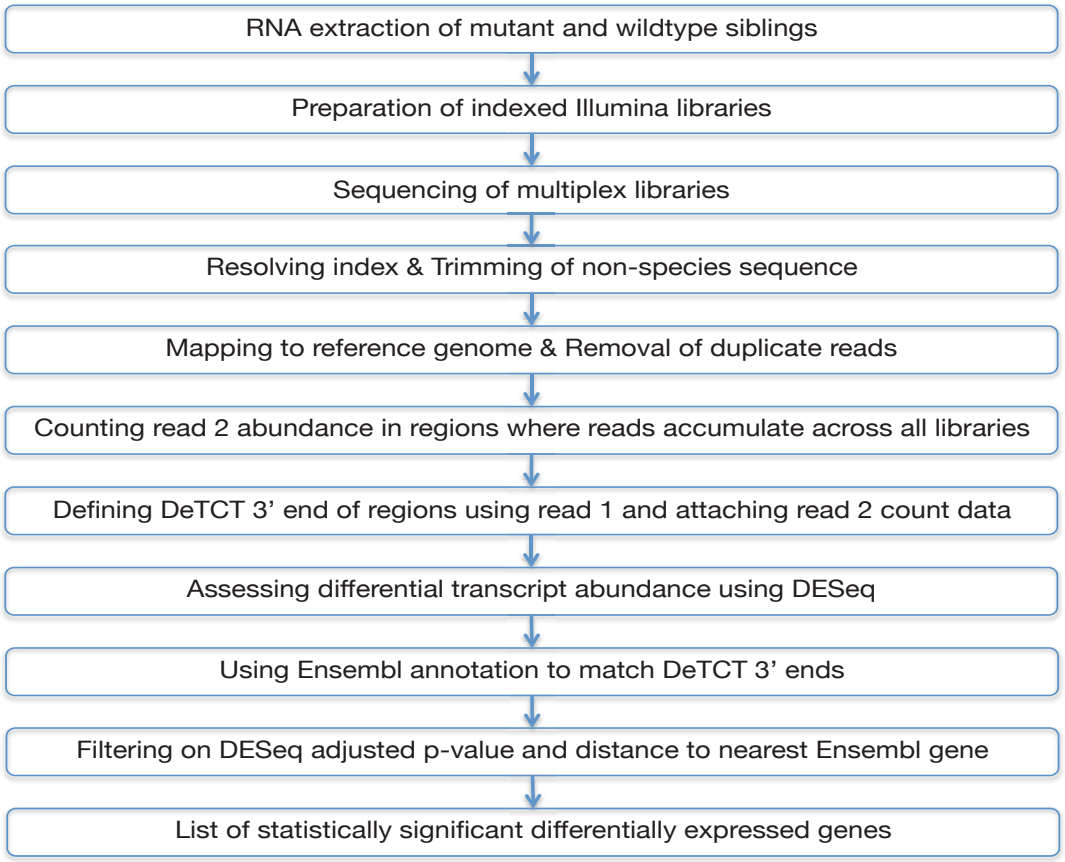


**Figure 5-1 Differential Expression Transcript Counting Technique (DeTCT) pipeline.** RNA of mutants and wild-type siblings is extracted and used for preparation of indexed Illumina libraries that are sequenced by paired-end Illumina HiSeq. Two sets of reads are produced: read 1 consists of random bases, index tag, 14 T bases and the remainder is transcript-specific sequence whilst read 2 is entirely transcript-specific sequence. The index of each library is resolved by using its unique tag and non-species sequences are trimmed. The sequencing reads are mapped to the reference genome by BWA and duplicate reads are removed by using the random bases. The abundance of read 2 is counted where reads accumulate across all libraries. The 3' end of regions is defined using read 1 and the count data is attached. DESeq is used to assess the differential transcript abundance and Ensembl annotation is used to match DeTCT 3' ends. The list is filtered based on the DESeq adjusted *p*-value and distance of transcript to the nearest Ensembl gene to produce a list of statistically significant differentially expressed genes.

Compared to RNA-seq, DeTCT provides a higher dynamic range for the same amount of sequence data since each transcript is sampled only once, and the total number of reads is not dominated by long or highly expressed transcripts (John Collins, personal communication). In addition, the library preparation of DeTCT is faster and less expensive than that of RNA-seq. With DeTCT, it is easier to locate un-annotated regions of the genome that show differential transcript abundance and identify previously unknown transcripts.

In order to provide a better understanding of the function of *nol9* and *las1l* in zebrafish development, we investigated the mRNA expression profiles of *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants using DeTCT with the goal of identifying genes and pathways that, when altered, can contribute to the phenotype of these mutants. In addition, the mRNA expression profiles of two mutants *tti*$^{s450}$ and *set*$^{s453}$ were also examined. Since *tti*$^{s450}$ and *set*$^{s453}$ mutants also have a mutation in rRNA processing genes and display similar phenotypic defects to *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants, comparisons of their expression profiles can help to uncover shared pathways and decipher the molecular basis of developmental defects of zebrafish rRNA processing mutants.

## 5.2 Results

### 5.2.1 The mRNA expression profile of *nol9*$^{sa1022}$ mutants

In this experiment, six pairs of *nol9*$^{sa1022/+}$ zebrafish were incrossed and for each pair, 25 phenotypically mutant and 25 wild-type siblings were collected at 5 d.p.f. for RNA extraction (Section 2.2). Library preparation was carried out on the 12 samples (six mutant and six wild-type pools) and the 12 libraries were sequenced on one single lane of Illumina HiSeq 2000 (Section 2.6.1). Dr Ian Sealy carried out the bioinformatic analysis on the sequence data including resolving the index of each library, trimming of non-zebrafish sequence, mapping to reference genome and removal of duplicate reads. Table 5-1 shows the number and percentage of reads mapped for each pairs of mutant and wild-type siblings and the complexity of each library, i.e. the estimated number of unique molecules in each library. Since the library complexity is similar for each sample, all the six biological replicates were included for the next step in DeTCT analysis.

| Sample | Reads | | | Estimated number of unique molecules |
|---|---|---|---|---|
| | Mapped | Total | % Mapped | |
| Mutant 1 | 22,762,654 | 28,952,219 | 78.62 | 1,323,614,272 |
| Wild-type 1 | 23,844,191 | 28,708,629 | 83.06 | 1,209,790,911 |
| Mutant 2 | 20,371,319 | 26,035,334 | 78.24 | 943,296,898 |
| Wild-type 2 | 23,979,267 | 30,791,253 | 77.88 | 1,231,014,078 |
| Mutant 3 | 23,579,438 | 30,153,578 | 78.20 | 1,418,214,013 |
| Wild-type 3 | 22,151,787 | 27,007,284 | 82.02 | 1,236,748,436 |
| Mutant 4 | 20,806,364 | 26,820,920 | 77.58 | 1,180,581,745 |
| Wild-type 4 | 18,931,237 | 26,063,887 | 72.63 | 1,565,331,425 |
| Mutant 5 | 18,006,129 | 25,116,115 | 71.69 | 1,180,950,054 |
| Wild-type 5 | 18,009,362 | 24,359,753 | 73.93 | 943,381,713 |
| Mutant 6 | 21,268,436 | 26,122,043 | 81.42 | 1,235,860,145 |
| Wild-type 6 | 22,246,054 | 27,863,509 | 79.84 | 1,561,066,036 |
| Average | 21,329,686.5 | 27,332,877 | 77.93 | 1,252,487,477 |
| Stdev | 2,153,066 | 1,990,773 | 3.60 | 197,681,535 |

**Table 5-1 Number and percentage of mapped reads for each sample of *nol9*$^{sa1022}$ mutants and wild-type siblings and the estimated number of unique molecules in each library. (Stdev) Standard deviation.**

The differential transcript abundance analysis was carried out using a DESeq-based pipeline written by Dr James Morris and Dr Ian Sealy (Anders and Huber, 2010) (Section 2.6.2). DESeq allows differential expression analysis by using a method based on negative binomial distribution and uses a shrinkage estimator for variance of the distribution (Anders and Huber, 2010). The list of differentially expressed regions has been filtered to include only regions that are within 100 bases of the nearest 3' end Ensembl gene. At a 5% false discovery

rate (FDR) when Benjamini-Hochberg multiple testing adjustment (Benjamini Y and Hochberg Y, 1995) was used, the number of differentially expressed regions within 100 bases to the nearest 3' end Ensembl gene was 599 corresponding to 566 unique genes. The *nol9*$^{sa1022}$ mutant had significantly fewer transcripts than wild-type siblings in 231 of these genes, with a log2 fold change ranging from -0.69 to -4.59, but had significantly more transcripts than wild-type siblings in the remaining 335 genes, with a log2 fold change ranging from 0.70 to 5.59.

The most statistically significant differentially expressed genes in *nol9*$^{sa1022}$ mutants with the minimum *p*-value accepted by the DESeq-based pipeline (an adjusted *p*-value of less than $1x10^{-16}$) are shown in Table 5-2 and Table 5-3. The downregulated genes, with a negative log2 fold change of mutant over wild-type siblings, include several genes that are specifically expressed in the liver (*fabp10a*), intestine (*fabp2*) and pancreas (*try* and *ela3l*) of 5 d.p.f. zebrafish larvae (Table 5-2). The eye-specific gene, *galectin related inter-fiber protein* (*grifin*) showed the biggest fold change in expression in mutants compared to wild-type siblings. Also, the gene *nol9* was downregulated with a log2 fold change of -1.16 at an adjusted *p*-value of 0.0127. This data suggests that that we can detect differential expression of tissue-specific genes and that the number of *nol9* transcripts is reduced in *nol9*$^{sa1022}$ mutants. The upregulated genes, with a positive log2 fold change of mutant over wild-type siblings, include the tumour suppressor *tp53*, indicating that the Tp53 signalling pathway may contribute to the phenotypic defects of *nol9*$^{sa1022}$ mutants. Also, it was found that the expression of *senp3b* that encodes a protein that in humans interact with NOL9, was upregulated with a log2 fold change of 1.14 at an adjusted *p*-value of $1.13x10^{-5}$ whereas no changes were detected in the transcript abundance of the other genes encoding NOL9-interacting proteins including *las1l*, *senp3a*, *tex10* and *wdr18* at an adjusted *p*-value of 0.05 (Refer to Appendix Table A - 1). This suggests that loss of *nol9* is affecting the expression of at least another gene involved in ribosome biogenesis.

| Genome location | Strand | Distance to Ensembl Gene (bp) | Gene name | Gene description | Log2 fold change (*nol9*$^{sa1022}$/wild-type) |
|---|---|---|---|---|---|
| 3:41780644 | -1 | 1 | *grifin* | *galectin related inter-fiber protein* | -4.39 |
| 12:34101458 | -1 | 3 | *ENPP7 (3 of 4)* | *ectonucleotide pyrophosphatase/phosphodiesterase 7* | -4.15 |
| 23:25084905 | 1 | 20 | *olfml3a* | *olfactomedin-like 3a* | -3.60 |
| 21:9817926 | 1 | 62 | *cel.2* | *carboxyl ester lipase, tandem duplicate 2* | -3.50 |
| 9:1035571 | 1 | 85 | *slc15a1a* | *solute carrier family 15 (oligopeptide transporter), member 1a* | -3.46 |
| 20:25564364 | -1 | 84 | *cyp2ad2* | *cytochrome P450, family 2, subfamily AD, polypeptide 2* | -2.88 |
| 17:27320149 | 1 | 7 | *ctsl.1* | *cathepsin L.1* | -2.76 |
| 2:38720820 | -1 | 54 | *rbp2b* | *retinol binding protein 2b, cellular* | -2.59 |
| 24:4835110 | -1 | 1 | *cpb1* | *carboxypeptidase B1 (tissue)* | -2.54 |
| 1:60218572 | 1 | 5 | *cyp3a65* | *cytochrome P450, family 3, subfamily A, polypeptide 65* | -2.41 |
| 2:16581399 | -1 | 3 | *tfa* | *transferrin-a* | -2.40 |
| 5:27756150 | -1 | 0 | *ela3l* | *elastase 3 like* | -2.40 |
| 7:36516813 | -1 | 4 | *zgc:112160* | *zgc:112160* | -2.39 |
| 20:35506189 | 1 | 0 | *mep1a.1* | *meprin A, alpha.1* | -2.28 |
| 16:56122673 | 1 | 1 | *fabp10a* | *fatty acid binding protein 10a, liver basic* | -2.26 |
| 4:21594158 | -1 | 4 | *cd36* | *CD36 antigen* | -2.16 |
| 8:52652107 | 1 | 91 | *msmbl* | *microseminoprotein, beta-like* | -2.12 |
| 22:17813696 | -1 | 1 | *gpx4a* | *glutathione peroxidase 4a* | -1.89 |
| 16:27320046 | -1 | 1 | *try* | *trypsin* | -1.82 |
| 1:24708924 | 1 | 51 | *fabp2* | *fatty acid binding protein 2, intestinal* | -1.76 |

**Table 5-2 List of the most statistically significant genes in *nol9*$^{sa1022}$ mutants with an adjusted *p*-value of less than 1x10$^{-16}$ sorted from the biggest to the smallest negative fold change.**
The genome location, the strand direction, the distance to the nearest Ensembl gene, the gene name and description and the fold change of mutant over wild-type sibling are shown.

| Genome location | Strand | Distance to Ensembl Gene (bp) | Gene name | Gene description | Log2 fold change (*nol9*$^{sa1022}$/wild-type) |
|---|---|---|---|---|---|
| 5:45468027 | -1 | 88 | *BDP1 (2 of 2)* | *B double prime 1, subunit of RNA polymerase III transcription initiation factor IIIB* | 2.75 |
| 3:21453661 | -1 | 6 | *tcap* | *titin-cap (telethonin)* | 2.42 |
| 25:34827000 | -1 | 2 | *RPS27L* | *ribosomal protein S27-like* | 2.42 |
| 2:167259 | -1 | 1 | *igfbp1b* | *insulin-like growth factor binding protein 1b* | 2.19 |
| 20:7033441 | 1 | 1 | *igfbp1a* | *insulin-like growth factor binding protein 1a* | 2.10 |
| 5:25766741 | 1 | 1 | *tp53* | *tumor protein p53* | 2.08 |

**Table 5-3 List of the most statistically significant genes in *nol9*$^{sa1022}$ mutants with an adjusted *p*-value less than 1x10$^{-16}$ sorted from the biggest to the smallest positive fold change.** The genome location, the strand direction, the distance to the nearest Ensembl gene, the gene name and description and the fold change of mutant over wild-type sibling are shown.

### 5.2.2 Enriched Gene Ontology categories in *nol9*$^{sa1022}$ mutants

Given the long list of differentially expressed genes in *nol9*$^{sa1022}$ mutants, it is difficult to infer the biology underlying the *nol9*$^{sa1022}$ phenotype. In order to enhance the interpretation of such list, functional analysis of the differentially expressed genes is an important step. The topGO package provides tools to test for enrichment of Gene Ontology (GO) terms by taking into account the relationships between the different GO terms (Alexa et al., 2006). Dr Ian Sealy carried out the functional analysis using the *elim* method that iteratively removes all the genes that are mapped to a significantly enriched node from more general terms (Alexa et al., 2006) (Section 2.6.4). The Kolmogorov-Smirnov (K-S)-like test statistic (Subramanian et al., 2005) was used to score for the significance of the Gene Ontology term (Alexa et al., 2006). The enrichment analysis for Gene Ontology categories cellular component, molecular function and biological process was produced (Ashburner et al., 2000; Harris et al., 2004). Since a multiple testing adjustment procedure was not carried out, a K-S value of 0.01 was used as an arbitrary threshold, which is a compromise between sensitivity and specificity. The genes displaying altered expression in *nol9*$^{sa1022}$ mutants were markedly enriched for genes assigned to the cellular component categories ‘nucleolus’, ‘cytoplasm’, ‘proteasome core complex’, ‘ribonucleoprotein complex’, ‘proteasome complex’, ‘mitochondrion’ and ‘ribosome’ (Figure 5-2). For the molecular function categories, the enriched terms were ‘RNA binding’, ‘serine-type endopeptidase activity’, ‘translation initiation factor activity’, ‘DNA-directed RNA polymerase activity’ and ‘RNA methyltransferase activity’ (Figure 5-3). The most significantly enriched biological terms included ‘rRNA processing’, ‘tRNA aminoacylation for protein translation’, ‘mRNA processing’, ‘translational initiation’, ‘proteolysis’, ‘ribosome biogenesis’ ‘tRNA processing’, ‘translation’, ‘hematopoiesis’, ‘cell cycle’ and ‘digestive tract development’ (Figure 5-4). However, the other GO terms that described the digestive system namely the ‘exocrine pancreas development’, ‘pancreas development’, endocrine pancreas development’, ‘liver development’,’ digestive system development’ and ‘digestive tract morphogenesis’ were not statistically significantly enriched in *nol9*$^{sa1022}$ mutants using a K-S cut off value of 0.01 (Refer to Appendix Table A - 2).

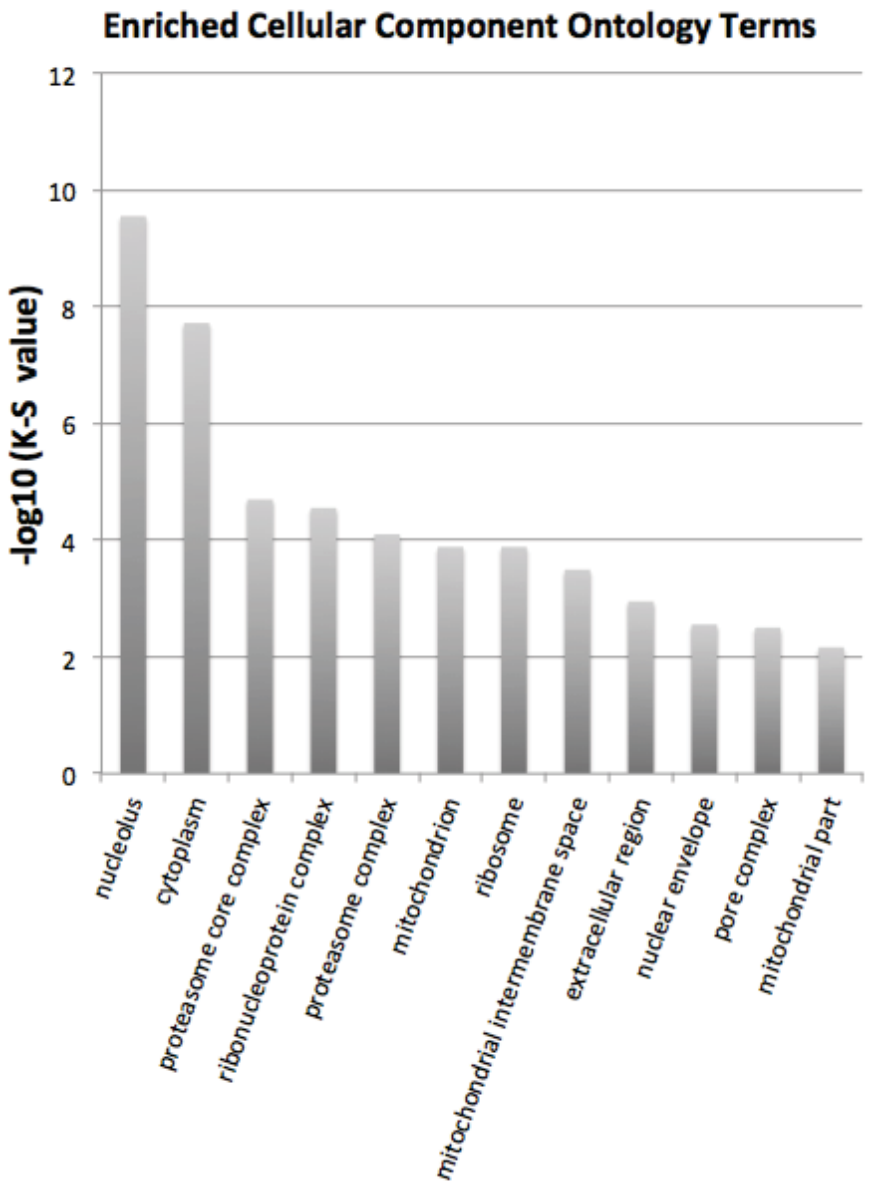


**Figure 5-2 The enriched cellular component ontology terms in *nol9*$^{sa1022}$ mutants at statistical significance of K-S value less than 0.01 (-log10 (K-S value) greater than 2.0).**

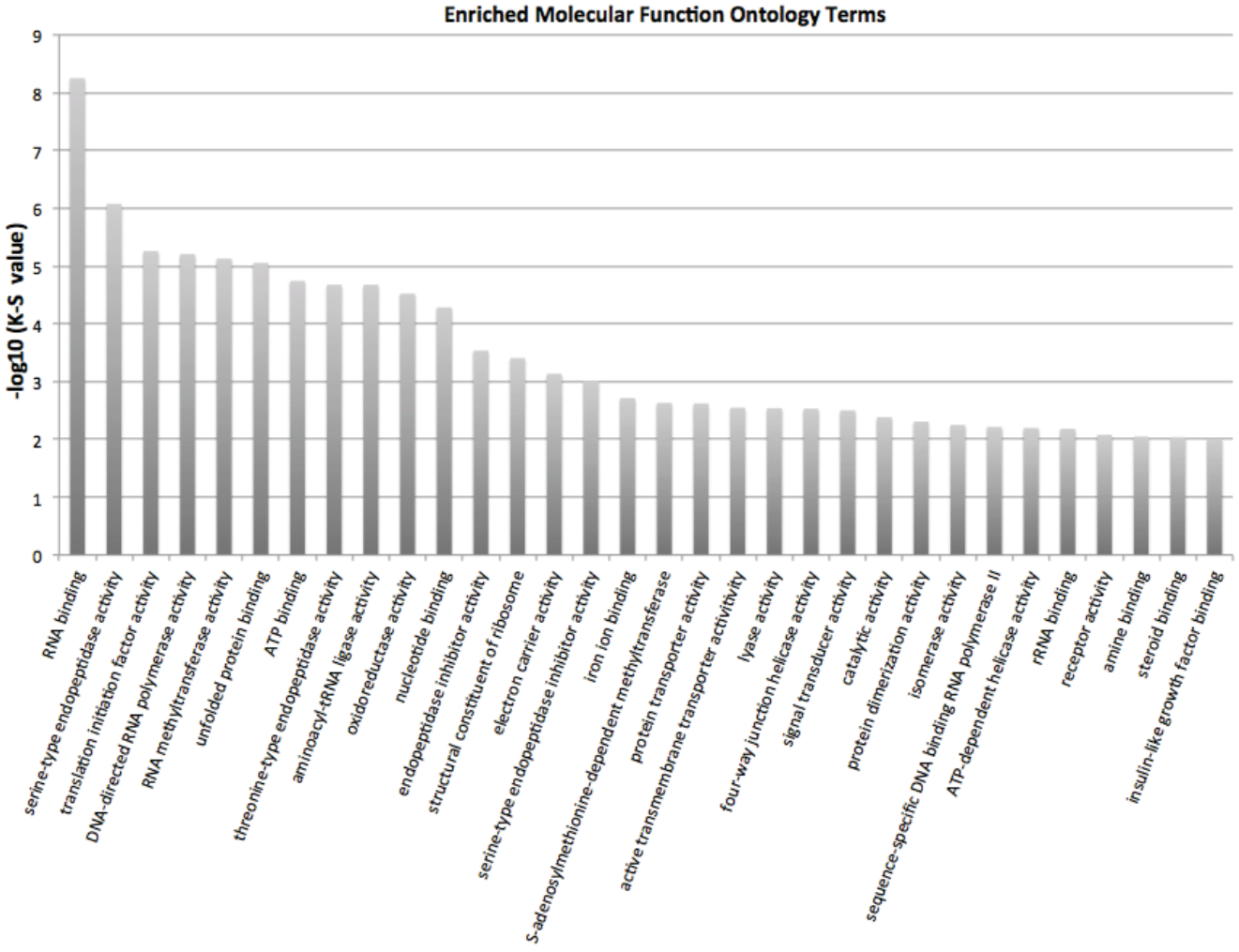


**Figure 5-3 The enriched molecular function ontology terms in *nol9*$^{sa1022}$ mutants at statistical significance of K-S value less than 0.01 (-log10 (K-S value) greater than 2.0).**

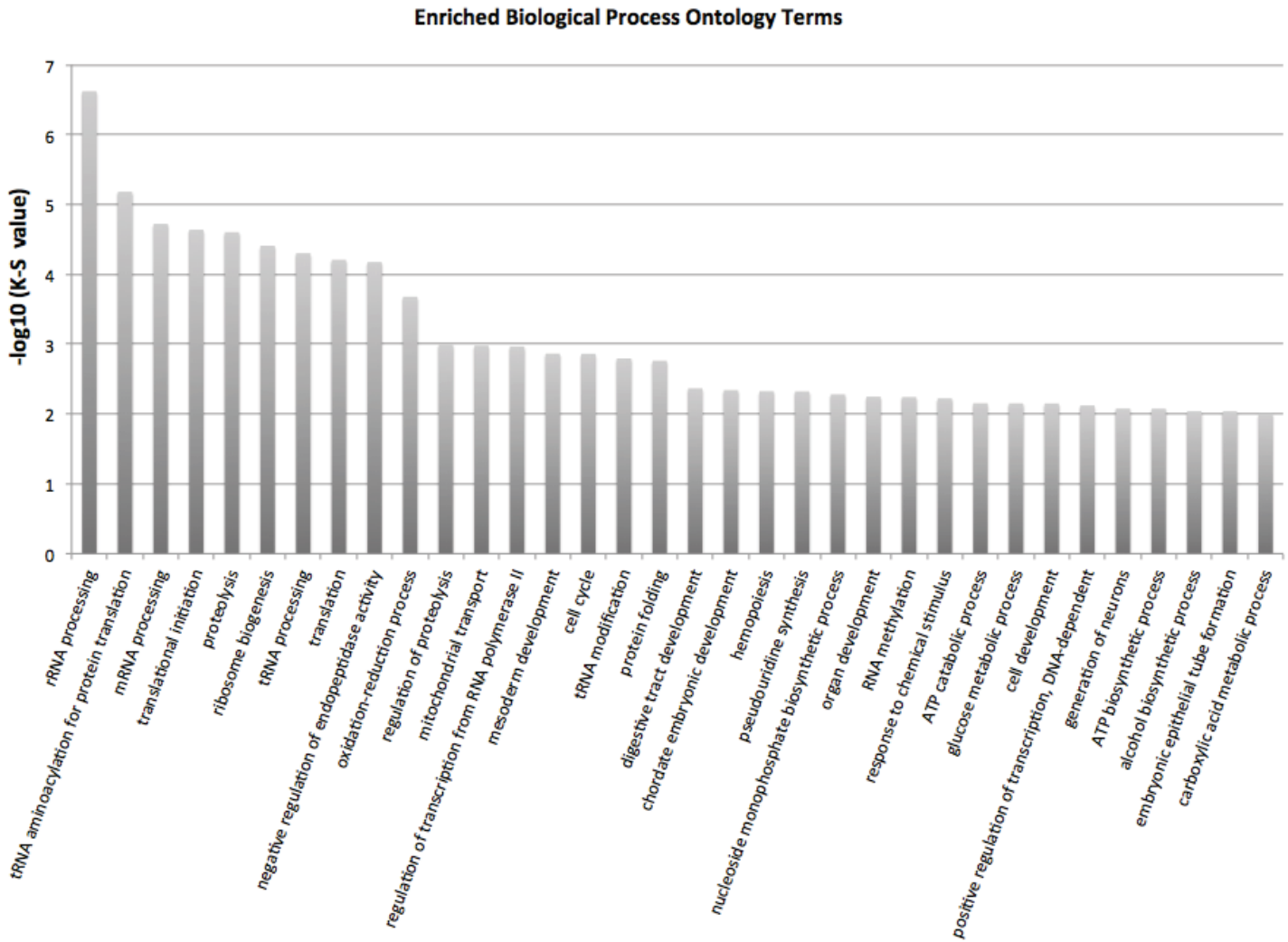


**Figure 5-4 The enriched biological process ontology terms in *nol9*$^{sa1022}$ mutants at statistical significance of K-S value less than 0.01 (log10(K-S value) greater than 2.0).**

### 5.2.3 Enriched KEGG pathways in *nol9*$^{sa022}$ mutants

In order to identify active pathways that differ between *nol9*$^{sa1022}$ mutants and wild-type siblings, I explored the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al., 2003; Huang da et al., 2009b). DAVID provides different data mining tools that combine functionally descriptive information with graphic displays for large lists of genes and proteins (Dennis et al., 2003; Huang da et al., 2009b). *KeggCharts*, one of the analysis modules, ascribes genes to the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic processes (Dennis et al., 2003; Huang da et al., 2009a; Kanehisa and Goto, 2000). The previously described list of 566 unique differentially expressed genes that were statistically significant (at an adjusted *p*-value of less than 0.05, after Benjamini-Hochberg multiple testing adjustment) was compared against a list of 11,170 unique genes that were assayed in this experiment (Section 2.6.5). The DAVID *KeggCharts* was able to provide information for only 131 of the 566 differentially expressed genes tested. At a false discovery rate of 5% after adjustment for multiple testing with the procedure of Benjamini and Hochberg (Benjamini Y and Hochberg Y, 1995), the enriched KEGG terms were 'Aminoacyl-tRNA biosynthesis', 'Steroid biosynthesis', 'Pyrimidine metabolism', 'PPAR signalling pathway' and 'RNA polymerase' (Table 5-4). All of the genes assigned to the 'Aminoacyl-tRNA biosynthesis' and 'RNA polymerase' KEGG pathways, and most of the genes assigned to the 'Pyrimidine metabolism' category, were upregulated in *nol9*$^{sa1022}$ mutants compared to wild-type siblings (Refer to Appendix Table A - 3, Table A - 4, Table A - 5). This data is consistent with the functional analysis results using topGO and reveals an upregulation of genes involved in the transcription and translation in *nol9*$^{sa1022}$ mutants. Interestingly, the KEGG term 'p53 signaling pathway' was also enriched in *nol9*$^{sa1022}$ mutants, albeit only at a 10% false discovery rate (FDR) using Benjamini-Hochberg procedure. All the differentially expressed genes belonging to the 'p53 signalling pathway' are upregulated in *nol9*$^{sa1022}$ mutants compared to wild-type siblings, including *tp53*, *mdm2*, *ccng1*, *cdkn1a*, *gadd45a*, *cycsb*, *casp8* and *baxa* (Table 5-5). These results again highlight the Tp53 signalling pathway as a potential important contributor to the *nol9*$^{sa1022}$ mutant phenotype.

| KEGG Term | Genes | | Fold Enrichment | *p*-value | Adjusted *p*-value (Benjamini-Hochberg) |
|---|---|---|---|---|---|
| | # | % | | | |
| Aminoacyl-tRNA biosynthesis | 12 | 2.40 | 7.71 | $9.25 \times 10^{-8}$ | $8.33 \times 10^{-6}$ |
| Steroid biosynthesis | 6 | 1.20 | 11.18 | $7.96 \times 10^{-5}$ | 0.003577288 |
| Pyrimidine metabolism | 12 | 2.40 | 3.61 | $3.04 \times 10^{-4}$ | 0.00907344 |
| PPAR signalling pathway | 9 | 1.80 | 4.53 | $5.31 \times 10^{-4}$ | 0.011874735 |
| RNA polymerase | 6 | 1.20 | 6.57 | 0.001446115 | 0.025712569 |
| Drug metabolism | 6 | 1.20 | 5.32 | 0.004006265 | 0.058437616 |
| p53 signalling pathway | 8 | 1.60 | 3.31 | 0.008468176 | 0.103574559 |
| Purine metabolism | 11 | 2.20 | 2.23 | 0.021887645 | 0.220397219 |

**Table 5-4 The enriched KEGG terms of differentially expressed genes of *nol9*$^{sa1022}$ mutants at a statistical significance of *p*-value less than 0.05 using a modified Fisher's exact test.** The KEGG term, the number of genes involved in the terms and the percentage of involved genes over total number of genes, the fold enrichment, the *p*-value and the adjusted *p*-value using Benjamini-Hochberg procedure are shown.

| Gene Name | Gene Description | Adjusted *p*-value (Benjamini–Hochberg) | Log2 fold change (*nol9*$^{sa1022}$/wild-type) |
|---|---|---|---|
| *tp53* | *tumor protein p53* | $<1 \times 10^{-16}$ | 2.08 |
| *mdm2* | *transformed 3T3 cell double minute 2 homolog (mouse)* | $2.72 \times 10^{-10}$ | 1.94 |
| *ccng1* | *cyclin G1* | $5.92 \times 10^{-8}$ | 1.29 |
| *cdkn1a* | *cyclin-dependent kinase inhibitor 1A* | 0.000938222 | 1.32 |
| *gadd45aa* | *growth arrest and DNA-damage-inducible, alpha, a* | 0.001929604 | 1.02 |
| *cycsb* | *cytochrome c, somatic b* | 0.003997489 | 0.84 |
| *casp8* | *caspase 8, apoptosis-related cysteine peptidase* | 0.021165876 | 1.24 |
| *baxa* | *bcl2-associated X protein, a* | 0.028395958 | 0.98 |

**Table 5-5 The differentially expressed genes belonging to the KEGG term 'p53 signalling pathway'.** The gene name and description, the adjusted *p*-value and the fold change of mutant over wild-type siblings are shown.

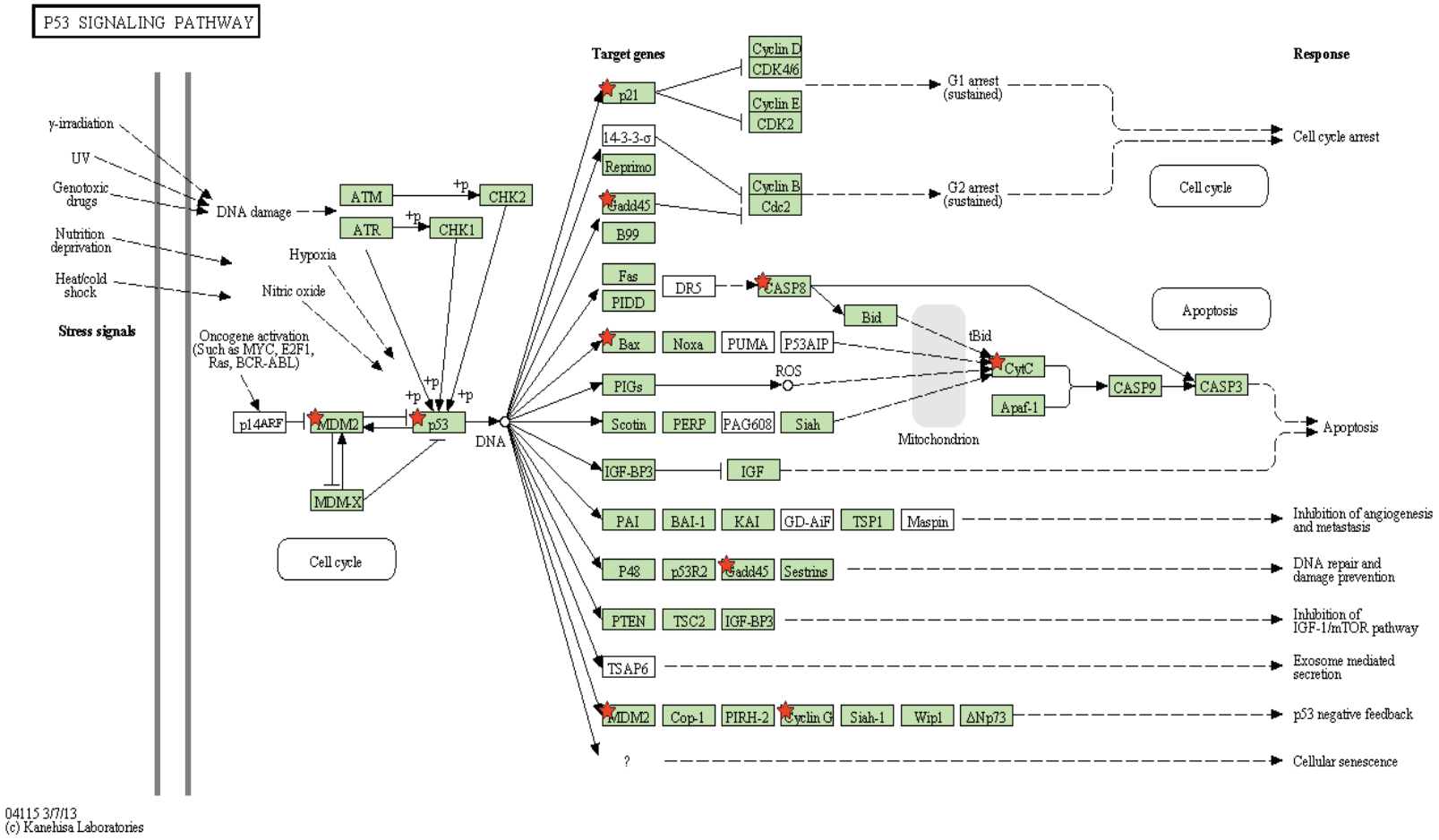


**Figure 5-5 KEGG Tp53 signalling pathway in zebrafish.** Boxes are gene products and red stars indicate genes that are upregulated in *nol9*$^{sa1022}$ mutants compared to wild-type siblings. Taken from KEGG Pathway Database.

### 5.2.4 The small pancreas phenotype of *nol9$^{sa1022}$* mutant is Tp53-independent

The Tp53 pathway has been shown to mediate the response following nucleolar stress and is responsible for many of the clinical features of ribosomopathies by causing perturbations in tissue homeostasis (Fumagalli and Thomas, 2011; Holmberg Olausson, 2012). Moreover, depletion of *NOL9* in HCT116 colon carcinoma cells by RNAi results in Tp53-dependent G1 cell-cycle arrest, stabilisation of Tp53 and increased levels of its transcriptional target *p21* (Castle et al., 2012). Furthermore, we found that many genes involved in the Tp53 signalling pathway are upregulated in *nol9$^{sa1022}$* mutants. In order to investigate whether the phenotype of *nol9$^{sa1022}$* mutant is dependent on Tp53 signalling, the *nol9$^{sa1022/+}$* line was outcrossed to the *tp53$^{zdf1/+}$* line (Berghmans et al., 2005). The *tp53$^{zdf1}$* carries a M214K missense mutation in the DNA-binding domain that ultimately interferes with activation of its target genes. The *nol9$^{sa1022/+}$;tp53$^{zdf1/+}$* zebrafish were incrossed and the 5 d.p.f. larvae were phenotyped, under the dissecting microscope, based on the size of their pancreas and genotyped for both *nol9$^{sa1022}$* and *tp53$^{zdf1}$* alleles (Figure 5-6 A) (Section 2.1.2). Out of the 165 larvae studied, 26% were *nol9$^{-/-}$* and phenotypic, 48% were *nol9$^{+/-}$* and non-phenotypic whilst 27% were *nol9$^{+/+}$* and non-phenotypic. The percentage of phenotypic larvae with genotypes *tp53$^{-/-}$*, *tp53$^{+/-}$* and *tp53$^{+/+}$* were 25%, 52% and 23% respectively. For the 121 non-phenotypic larvae, 19% were *tp53$^{-/-}$*, 55% were *tp53$^{+/-}$* and 26% were *tp53$^{+/+}$*. These results indicate that it was not possible to distinguish among *nol9$^{sa1022}$* mutants with different *tp53* genotypes suggesting that the small pancreas phenotype of *nol9$^{sa1022}$* mutants is Tp53-independent. In order to confirm this, the *nol9$^{sa1022/+}$;tp53$^{zdf1/+}$* zebrafish were outcrossed to the transgenic line *Tg(ins:mCherry)$^{jh2}$;Tg(ptf1a:EGFP)$^{jh1}$* and 5 d.p.f. larvae from an incross of *Tg(ins:mCherry)$^{jh2}$;Tg(ptf1a:EGFP)$^{jh1}$;nol9$^{sa1022/+}$;tp53$^{zdf1/+}$* were genoptyped and the volume of *ptf1a*-expressing region of larvae with genotype *nol9$^{+/+}$;tp53$^{+/+}$*, *nol9$^{+/+}$;tp53$^{-/-}$*, *nol9$^{-/-}$;tp53$^{+/+}$*and *nol9$^{-/-}$;tp53$^{-/-}$* fish was measured (Figure 5-6 B, C) (Sections 2.1.2 and 2.3.4). The *ptf1a*-expressing region of larvae with genotypes *nol9$^{-/-}$;tp53$^{+/+}$* and *nol9$^{-/-}$;tp53$^{-/-}$* appeared similar in size and were both smaller than the *ptf1a*-expressing region of larvae with genotypes *nol9$^{+/+}$;tp53$^{-/-}$* and *nol9$^{-/-}$;tp53$^{+/+}$* which were comparable (Figure 5-6 B). The mean volume of the *ptf1a*-expressing region of *nol9$^{+/+}$;tp53$^{+/+}$* (n=5), *nol9$^{+/+}$;tp53$^{-/-}$* (n=6), *nol9$^{-/-}$;tp53$^{+/+}$* (n=8) and *nol9$^{-/-}$;tp53$^{-/-}$* (n=11) is 0.00157 mm$^3$, 0.00126 mm$^3$, 0.000345 mm$^3$ and 0.000328 mm$^3$ respectively. The difference between the mean pancreatic volume of larvae with genotypes *nol9$^{+/+}$;tp53$^{+/+}$* and *nol9$^{-/-}$;tp53$^{+/+}$*, *nol9$^{+/+}$;tp53$^{+/+}$* and *nol9$^{-/-}$;tp53$^{-/-}$*, *nol9$^{+/+}$;tp53$^{-/-}$* and *nol9$^{-/-}$;tp53$^{+/+}$*, *nol9$^{+/+}$;tp53$^{-/-}$* and *nol9$^{-/-}$;tp53$^{-/-}$* was statistically significant

(Student's *t*-test, $p = 2.76\text{x}10^{-6}$, $p = 2.72\text{x}10^{-8}$, $p = 2.46\text{x}10^{-7}$ and $p = 1.17\text{x}10^{-9}$ respectively) (Figure 5-6 C). This data substantiates the hypothesis that the exocrine pancreas phenotype is independent of the Tp53 signalling pathway.

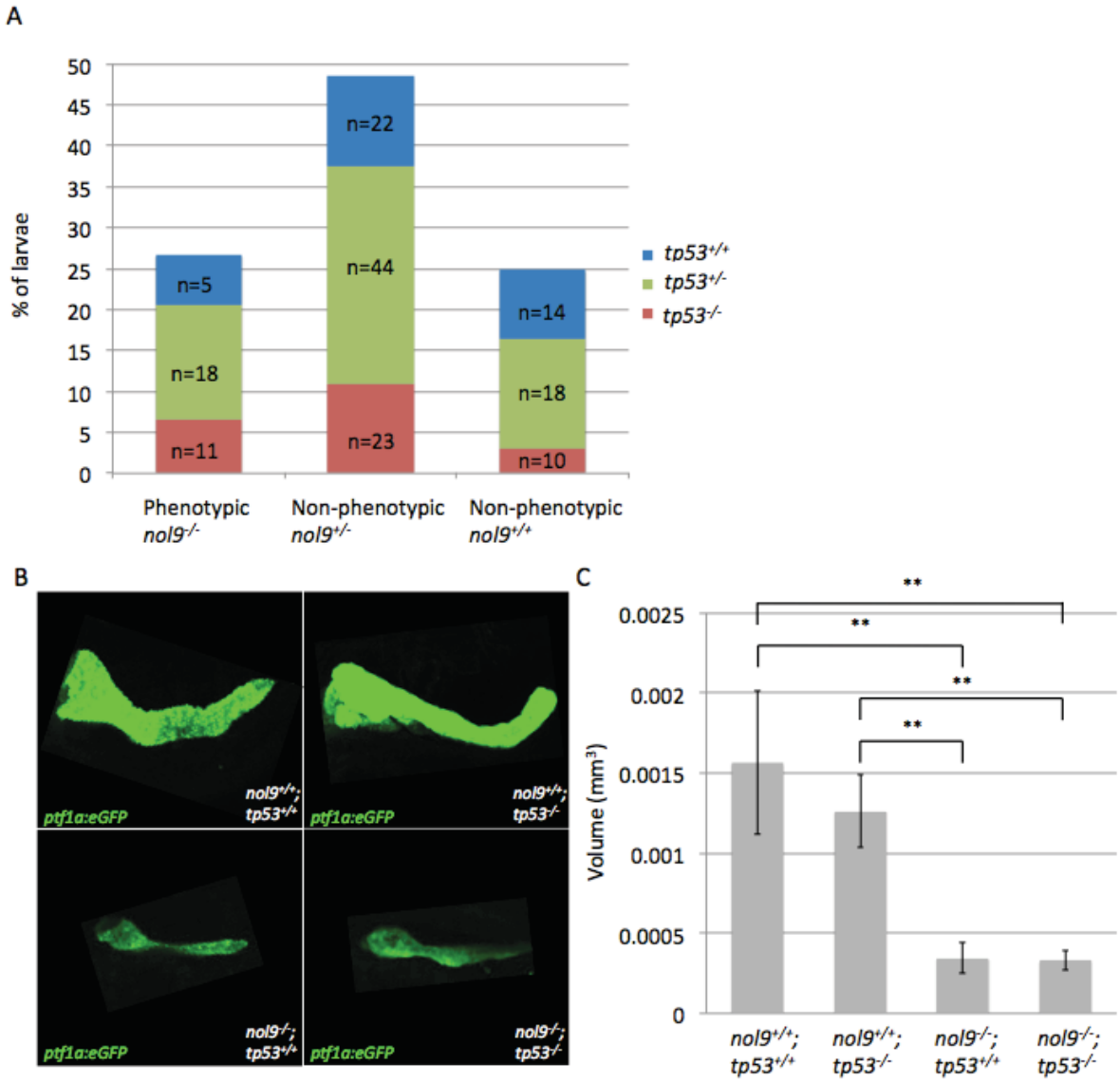


**Figure 5-6 The exocrine pancreas phenotype of *nol9*$^{sa1022}$ mutant is independent of Tp53 signalling pathway. (A)** Graph showing the percentage of 5 d.p.f. phenotypic (*nol9*$^{-/-}$) and non-phenotypic (*nol9*$^{+/-}$ or *nol9*$^{+/+}$) larvae from an incross of *nol9*$^{sa1022/+}$;*tp53*$^{zdf1/+}$ zebrafish that are *tp53*$^{-/-}$ (red), *tp53*$^{+/-}$ (green) or *tp53*$^{+/+}$ (blue). (B) Representative maximum intensity projection images of confocal stacks of 5 d.p.f. larvae from an incross of *Tg(ins:mCherry)*$^{jh2}$;*Tg(ptf1a:EGFP)*$^{jh1}$;*nol9*$^{sa1022/+}$; *tp53*$^{zdf1/+}$ showing that the *ptf1a*-expressing region in *nol9*$^{-/-}$;*tp53*$^{+/+}$ and *nol9*$^{-/-}$;*tp53*$^{-/-}$ is comparable and is smaller than that in *nol9*$^{+/+}$;*tp53*$^{+/+}$ and *nol9*$^{+/+}$;*tp53*$^{-/-}$ which appear similar. (C) Graph showing that the mean volume of the *ptf1a*-expressing region in *nol9*$^{-/-}$;*tp53*$^{+/+}$ (n=8) and *nol9*$^{-/-}$;*tp53*$^{-/-}$ (n=11) is statistically significantly smaller than that in *nol9*$^{+/+}$;*tp53*$^{+/+}$ (n=5) and *nol9*$^{+/+}$;*tp53*$^{-/-}$ (n=6). Data is represented as the mean ± SD, Student's *t*-test ** $p < 0.01$.

### 5.2.5 The mRNA expression profile of *las1l*$^{sa674}$ mutants

We found in Chapter 4 that the *las1l*$^{sa674}$ mutant has similar defects to *nol9*$^{sa1022}$ mutant, and hence we decided to also study the mRNA expression profile of *las1l*$^{sa674}$ by DeTCT. In this experiment, six pairs of *las1l*$^{sa674/+}$ zebrafish were incrossed and for each pair, 25 mutant and 25 wild-type siblings were collected at 5 d.p.f. for RNA extraction (Section 2.2). Only the phenotypic larvae with an inflated swim bladder were collected as their pancreatic size could be assayed with more confidence therefore increasing the accuracy in differentiating between *las1l*$^{sa674}$ mutants and wild-type siblings. Library preparation was carried out on the 12 samples but as one of the libraries failed, the biological replicate pair containing that failed library was excluded from further analysis. Therefore, only 10 libraries were sequenced on one single lane of Illumina HiSeq (Section 2.6.1). The sequence data analysis was carried out by Dr Ian Sealy (Section 2.6.2). Table 5-6 shows the number and percentage of mapped reads for each biological replicate and the library complexity, i.e. the estimated number of unique molecules in each library. All five biological replicates were used for differential transcript abundance using the DESeq-based pipeline written by Dr James Morris and Dr Ian Sealy since the library complexity is similar (Anders and Huber, 2010) (Section 2.6.2).

| Sample | Reads | | | Estimated number of unique molecules |
|---|---|---|---|---|
| | Mapped | Total | % Mapped | |
| Mutant 1 | 26,979,522 | 36,397,692 | 74.12 | 1,846,656,917 |
| Wild-type 1 | 21,611,207 | 27,836,272 | 77.64 | 1,422,215,600 |
| Mutant 2 | 25,531,558 | 33,358,031 | 76.54 | 1,117,281,160 |
| Wild-type 2 | 27,104,447 | 35,712,462 | 75.90 | 1,259,084,206 |
| Mutant 3 | 26,949,361 | 34,128,699 | 78.96 | 1,166,856,015 |
| Wild-type 3 | 24,172,449 | 33,228,696 | 72.75 | 1,597,497,227 |
| Mutant 4 | 25,194,438 | 33,558,009 | 75.08 | 1,591,174,952 |
| Wild-type 4 | 29,615,352 | 38,097,154 | 77.74 | 1,047,899,549 |
| Mutant 5 | 19,475,812 | 24,877,809 | 78.29 | 1,064,422,212 |
| Wild-type 5 | 24,893,796 | 31,197,807 | 79.79 | 1,179,672,647 |
| Average | 25,152,794.2 | 32,839,263.1 | 76.68 | 1,329,276,049 |
| Stdev | 2,910,369.913 | 3,975,929.983 | 2.23 | 271,785,213.7 |

**Table 5-6 Number and percentage of mapped reads for each sample of *las1l*$^{sa674}$ mutants and wild-type siblings and the estimated number of unique molecules in each library. (Stdev) Standard deviation.**

The output of differentially expressed regions was filtered to include only the regions that are within 100 bases to the nearest 3' end Ensembl gene. At a 5% FDR, after adjusting for multiple testing using Benjamini-Hochberg procedure (Benjamini Y and Hochberg Y, 1995), the genes showing differential expression are *grifin*, *RPS27L*, *tp53*, *zgc:153846* and *zgc:136461* (7). The genes that were downregulated in *las1l*$^{sa647}$ mutants compared to wild-type siblings were *zgc:136461*, an orthologue of a human gene *chymotrypsinogen B1* (*CTRB1*) that is expressed specifically in the pancreas, the eye lens genes *grifin* and the zebrafish orthologue of human crystallin, gamma D, *zgc:153846*. These results indicate that the differential expression of pancreatic and eye-specific genes is being detected in *las1l*$^{sa674}$ mutants. Additionally there were no significant changes in transcript abundance of *las1l* and genes encoding proteins that interact with Las1l including *nol9*, *senp3a*, *senp3b*, *tex10* and *wdr18* in *las1l*$^{sa674}$ mutants (Refer to Appendix Table A - 6).

| Genome location | Strand | Adjusted *p*-value (Benjamini-Hochberg) | Distance to Ensembl Gene (bp) | Gene name | Gene description | Log2 fold change (*las1l*$^{sa674}$/wild-type) |
|---|---|---|---|---|---|---|
| 3:41780644 | -1 | $2.27x10^{-9}$ | 1 | *grifin* | *galectin related inter-fiber protein* | -2.235219873 |
| 25:34827000 | -1 | $3.64x10^{-5}$ | 2 | *RPS27L* | *ribosomal protein S27-like* | 1.588163195 |
| 5:25766741 | 1 | 0.000878724 | 1 | *tp53* | *tumor protein p53* | 1.479254501 |
| 9:23051646 | 1 | 0.001022345 | 3 | *zgc:153846* | *zgc:153846* | -3.790242144 |
| 7:36524482 | 1 | 0.007348883 | 5 | *zgc:136461* | *zgc:136461* | -1.528978382 |
| 24:4835110 | -1 | 0.089208372 | 1 | *cpb1* | *carboxypeptidase B1 (tissue)* | -1.103040368 |
| 16:27320046 | -1 | 0.089208372 | 1 | *try* | *trypsin* | -1.086928815 |
| 1:28031527 | 1 | 0.135826525 | 6 | *cryaa* | *crystallin, alpha A* | -1.113336914 |
| 16:58404668 | 1 | 0.229527094 | 1 | *nr0b2a* | *nuclear receptor subfamily 0, group B, member 2a* | 1.279535082 |
| 16:31282459 | 1 | 0.299032485 | 0 | *nes* | *nestin* | 1.165607544 |
| 25:34827000 | -1 | 0.343608453 | 2 | *RPS27L* | *ribosomal protein S27-like* | 1.276394309 |
| 14:23143173 | 1 | 0.344538274 | 5 | *ccng1* | *cyclin G1* | 1.00442056 |
| 4:5407050 | 1 | 0.39003004 | 2 | *si:dkey-14d8.6* | *si:dkey-14d8.6* | -1.506566565 |
| 19:30471007 | 1 | 0.39003004 | 1 | *serpina7* | *serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 7* | 1.444459627 |
| 5:44068461 | -1 | 0.448383904 | 0 | *zgc:123103* | *zgc:123103* | 0.988687546 |
| 7:27756150 | -1 | 0.5250549 | 0 | *ela3l* | *elastase 3 like* | -0.93154133 |
| 7:40027647 | -1 | 0.558443494 | 2 | *c6ast4* | *six-cysteine containing astacin protease 4* | -1.186117809 |
| 5:27756150 | -1 | 0.826714077 | 0 | *ela3l* | *elastase 3 like* | 1.508652156 |
| 25:34827000 | -1 | 0.826714077 | 2 | *RPS27L* | *ribosomal protein S27-like* | -1.038784624 |
| 2:27549277 | 1 | 0.836122625 | 2 | *sepp1b* | *selenoprotein P, plasma, 1b* | 1.202330387 |
| 2:55854933 | 1 | 0.842020919 | 0 | *calrl* | *calreticulin like* | 1.100183091 |
| 1:11229671 | 1 | 0.973752112 | 0 | *pla2g12a* | *phospholipase A2, group XIIA* | 1.005701929 |

**Table 5-7 List of differentially expressed genes in *las1l*$^{sa674}$ mutants at adjusted *p*-value less than 1.0.** The genome location, the strand direction, the adjusted *p*-value, the distance to the nearest Ensembl gene, the gene name and description and the fold change of mutant versus wild-type sibling are shown.

### 5.2.6 Enriched Gene Ontology categories in *las1l*$^{sa674}$ mutants

There were only a small number of differentially expressed genes in *las1l*$^{sa674}$ mutants that reached statistical significance at 5% FDR using Benjamini-Hochberg adjustment for multiple testing. To enhance the interpretation of such list, the topGO analysis was carried out on the differentially expressed genes of *las1l*$^{sa674}$ mutants by Dr Ian Sealy using the *elim* method and the Kolmogorov-Smirnov (K-S)-like test statistic to score for GO enrichment (Alexa et al., 2006) (Section 2.6.4). At an arbitrary threshold of K-S value 0.01, the enriched GO cellular component terms were 'cytoplasm', 'cytoplasmic part', 'intracellular organelle', 'ribonucleoprotein complex' and nuclear terms 'nuclear lumen', 'nuclear membrane' and 'nucleus' (Figure 5-7). For the molecular function ontology categories, the most significant enriched terms were 'ATP binding', 'protein tyrosine kinase activity', 'nucleotide binding', 'structural molecule activity' and 'RNA binding' (Figure 5-8). The genes displaying altered expression were markedly enriched for the biological process categories 'regulation of transcription from RNA polymerase II promoter', 'mRNA processing', 'cell cycle', 'ATP biosynthetic process', 'regulation of transcription, DNA-dependent' and also 'pancreas development' and 'digestive tract development' (Figure 5-9). However, the additional GO terms describing the digestive system namely the 'liver development', 'endocrine pancreas development', 'digestive tract morphogenesis' and 'digestive system development' were not statistically significantly enriched in *nol9*$^{sa1022}$ mutants using a K-S cut off value of 0.01 (Refer to Appendix Table A - 7).

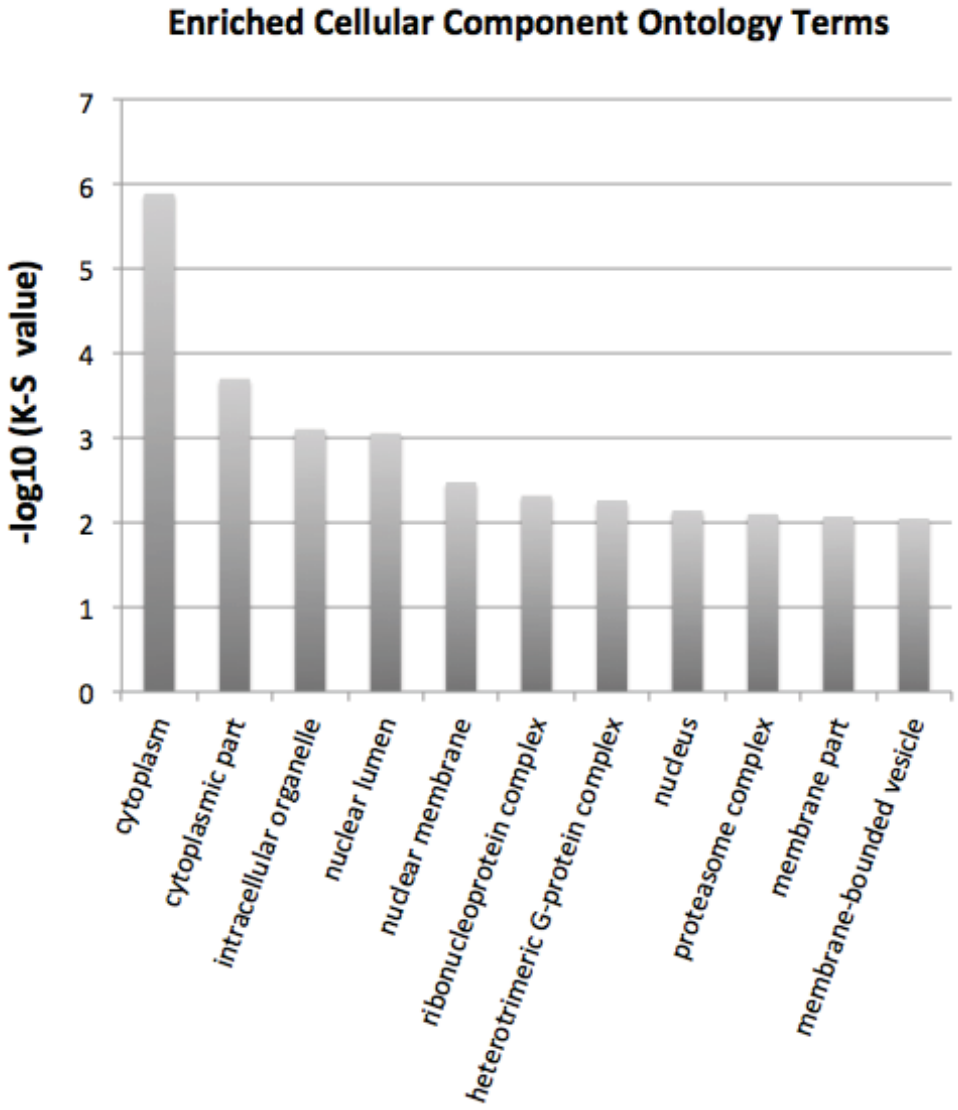


**Figure 5-7 The enriched cellular component ontology terms in *las1l*$^{sa674}$ mutants at statistical significance of K-S value less than 0.01 (-log10 (K-S value) greater than 2.0).**

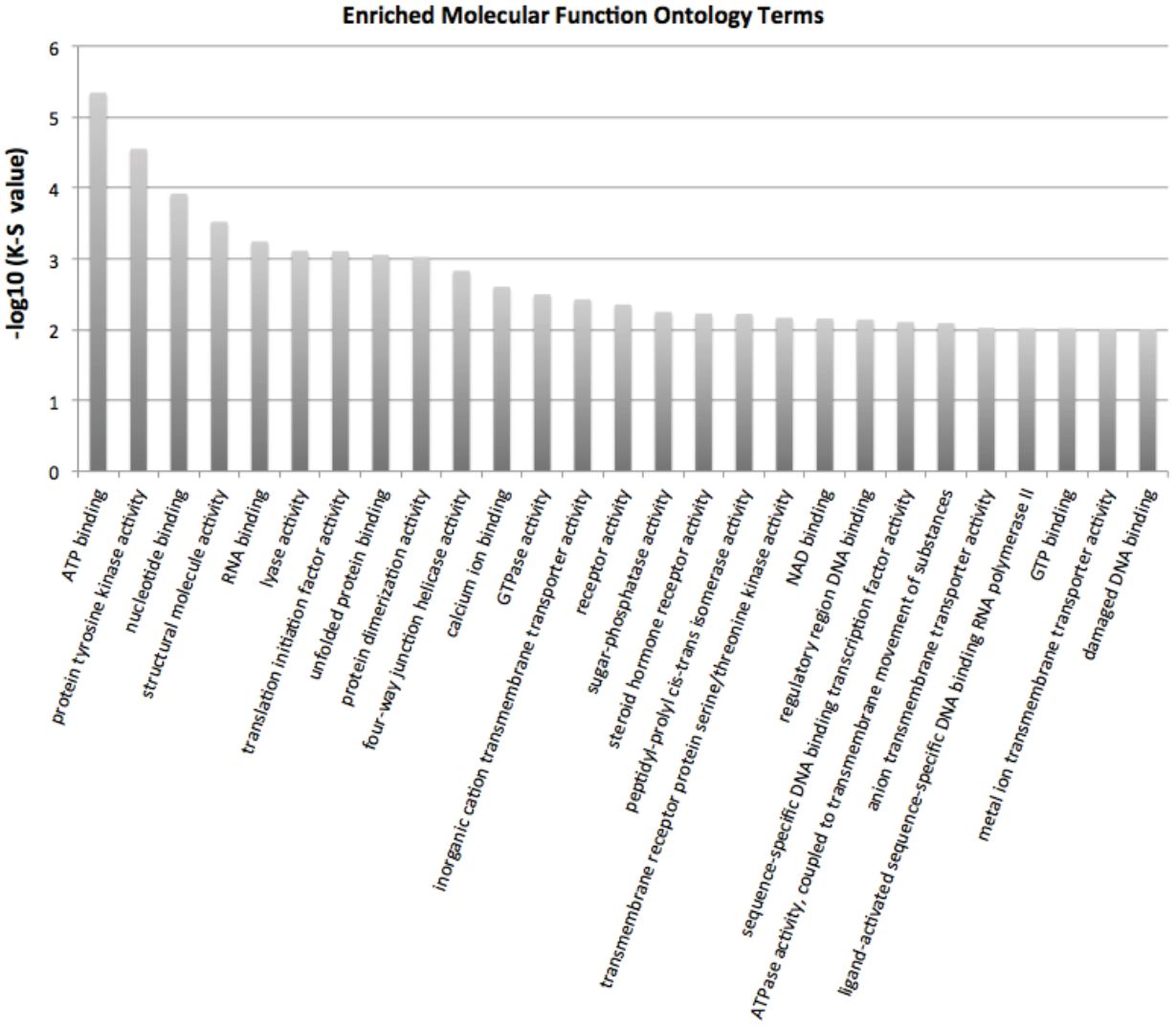


**Figure 5-8 The enriched molecular function ontology terms in *las1l*$^{sa674}$ mutants at statistical significance of K-S value less than 0.01 (-log10 (K-S value) greater than 2.0).**

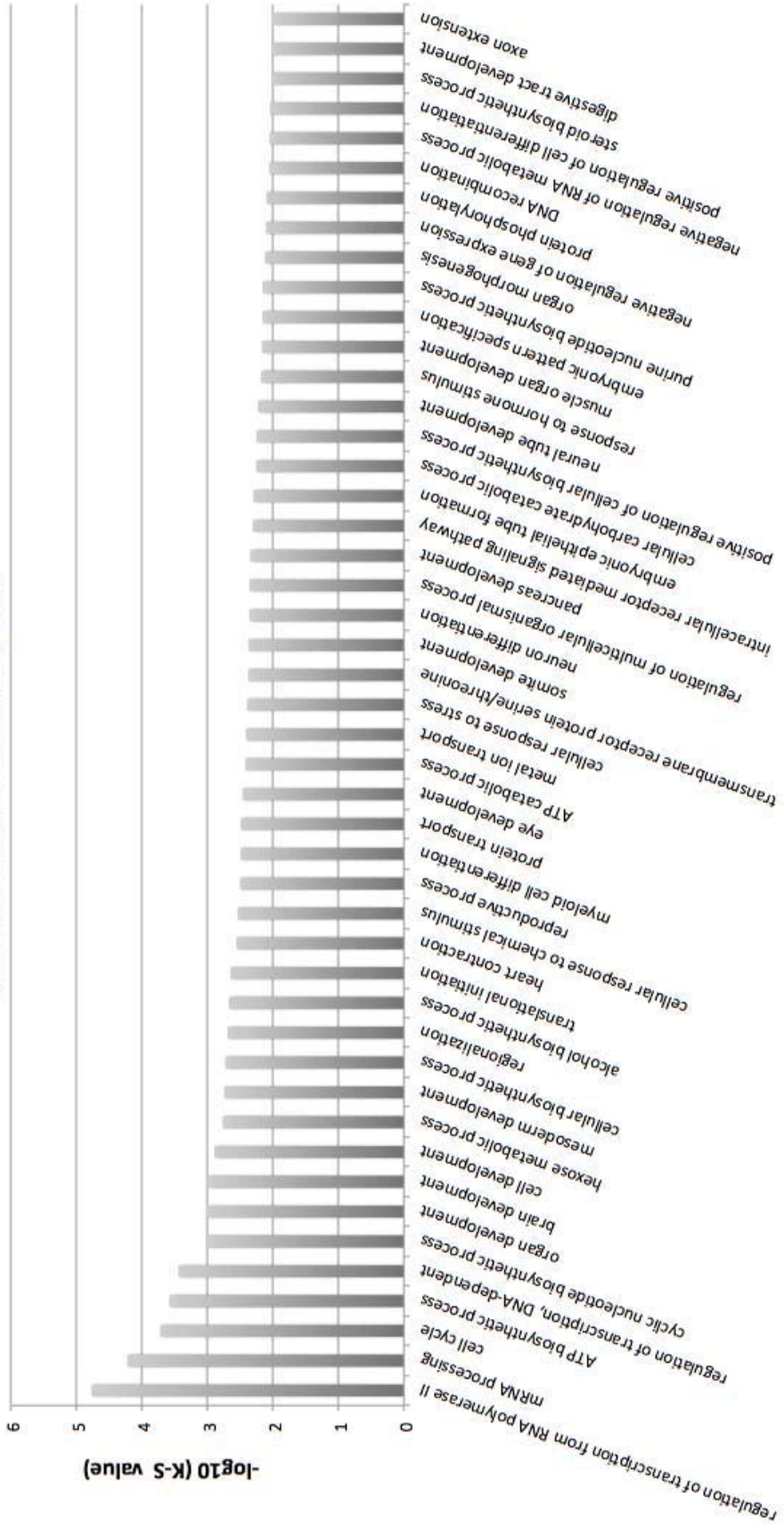


**Figure 5-9 The enriched biological process ontology terms in *las11*$^{sa674}$ mutants at statistical significance of K-S value less than 0.01 (-log10 (K-S value) greater than 2.0).**

### 5.2.7 Comparison of the mRNA expression profiles of *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{s453}$ mutants

Similarly to the *titania* (*tti*) mutant (described in Section 1.3.4), the *setebos* (*set*) mutant was also identified in the ENU mutagenesis Liver$^{plus}$ screen that was carried out on a transgenic line that expresses GFP in the digestive organs (*Tg(XlEef1a1:GFP)*$^{s854}$) (Field et al., 2003b; Ng et al., 2005; Ober et al., 2006). The *setebos*$^{s453}$ mutant is impaired in the development of many tissues including the intestine, exocrine pancreas, craniofacial cartilages, eye and brain (Dr Joan Heath, Personal Communication). It is mutated in the *nucleolar protein 8* (*nol8*) and exhibit defects in 28S rRNA processing and 60S ribosomal subunit biogenesis (unpublished). The *titania*$^{s450}$ and *setebos*$^{s453}$ mutants have similar phenotypes to *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants and the causative genes of those mutants play a similar role in rRNA processing. In order to help infer the underlying biology of the zebrafish rRNA processing mutants, I compared the mRNA expression profiles of these mutants to identify genes and pathways that may play an important role in the hypoplastic digestive organs of all four mutants.

Dr Joan Heath kindly provided the total RNA for mRNA expression profiling of *titania* and *setebos* mutants. For *titania* and *setebos*, library preparation was carried out on RNA of 12 biological replicates, i.e. 6 pairs of 25 mutants and 25 wild-type siblings (Section 2.6.1). For both *titania* and *setebos*, libraries of 3 biological replicates failed therefore only six libraries of each mutant were sequenced on a single Illumina HiSeq lane. Dr Ian Sealy carried out the analysis on the sequence data, which included resolving the index of each library, removal of non-species sequence, reference genome mapping, and removal of duplicate reads (Section 2.6.2). In order to compare the four mutants *tti*$^{s450}$, *set*$^{453}$, *nol9*$^{sa1022}$ and *las1l*$^{sa674}$, three biological replicates were obtained for each mutant. As the total number of reads for the different libraries varied, each sample had to be normalised to a total read count of 26 M reads (the smallest number of reads across all samples from the different libraries) (Section 2.6.3). Subsequently, the duplicate reads were removed and differential transcript abundance was carried out for all four mutants using the DESeq-based pipeline (Sections 2.6.2 and 2.6.3).

For each mutant, only the differentially expressed regions within 100 bases to the nearest 3' end Ensembl gene were examined. The number of regions and genes that were significant at 5% FDR using a Benjamini-Hochberg adjustment (Benjamini Y and Hochberg Y, 1995) for each mutant is shown in Table 5-8.

| Mutant | # differentially expressed regions | # differentially expressed genes | # differentially expressed genes upregulated in mutants vs wild-type siblings | # differentially expressed genes downregulated in mutants vs wild-type siblings |
|---|---|---|---|---|
| *titania*$^{s450}$ | 774 | 728 | 457 | 271 |
| *setebos*$^{s453}$ | 651 | 607 | 363 | 244 |
| *nol9*$^{sa1022}$ | 230 | 218 | 107 | 111 |
| *las1l*$^{sa674}$ | 4 | 4 | 2 | 2 |

**Table 5-8 The number of differentially expressed regions, genes and their direction of effect in *tti*$^{s450}$, *set*$^{s453}$, *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants.**

The number of differentially expressed genes that are shared between *nol9*$^{sa1022}$, *tti*$^{s450}$, *set*$^{453}$ and *las1l*$^{sa674}$ mutants is shown in Figure 5-10. The two differentially expressed genes that were significant at a 5% FDR using the Benjamini-Hochberg procedure in all four mutants were *grifin* and *rps27.2*. RPS27L, the homologue of *rps27.2*, is a direct Tp53 inducible target and modulates the DNA damage response (He and Sun, 2007; Li et al., 2007). The *tti*$^{s450}$, *set*$^{453}$ and *nol9*$^{sa1022}$ mutants shared an additional 145 differentially expressed genes. Another 263 differentially expressed genes were shared between only the *tti*$^{s450}$ and *set*$^{453}$ mutants adding to a total of 401 differentially expressed genes shared between *tti*$^{s450}$ and *set*$^{453}$ mutants.


**Figure 5-10 The number of differentially expressed genes that are significant at a 5% FDR using Benjamini-Hochberg procedure that are shared between *nol9*$^{sa1022}$, *tti*$^{s450}$, *set*$^{s453}$, and *las1l*$^{sa674}$ mutants.**

To identify the active pathways that are enriched in *tti*$^{s450}$ and *set*$^{453}$ mutants, the lists of 728 and 607 unique differentially expressed genes that were statistically significantly differentially expressed (an adjusted *p*-value of less than 0.05, after Benjamini-Hochberg multiple testing adjustment) in *tti*$^{s450}$ and *set*$^{453}$ mutants respectively were compared against a

list of unique genes that were assayed in this experiment (10527 and 10535 genes for *tti*$^{s450}$ and *set*$^{453}$ mutants respectively) using DAVID (Section 2.6.5). The DAVID *KeggCharts* provided information for 167 of the 730 differentially expressed genes and 146 of the 607 differentially expressed genes for *tti*$^{s450}$ and *set*$^{453}$ mutants respectively. In the Appendix Table A - 8 and Table A - 9 show the enriched KEGG terms of differentially expressed genes of *tti*$^{s450}$ and *set*$^{453}$ mutants at a statistical significance of *p*-value less than 0.05 using a modified Fisher's exact test. In all three mutants the enriched KEGG pathways were "Aminoacyl-tRNA biosynthesis", "RNA polymerase", "Steroid biosynthesis" and "PPAR signalling pathway".

The lists of differentially expressed genes for *tti*$^{s450}$, *set*$^{453}$, *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants were investigated for enrichment for GO terms using topGO and this analysis was carried out by Dr Ian Sealy (Alexa et al., 2006) (Section 2.6.4). The enriched GO terms at an arbitrary K-S value of 0.01 for categories cellular component, molecular function biological process of differentially expressed genes in *tti*$^{s450}$ and *set*$^{453}$ mutants are shown in the Appendix (Figure A - 1, Figure A - 2, Figure A - 3, Figure A - 4, Figure A - 5, Figure A - 6).

The enriched GO terms at an arbitrary K-S value of 0.01 of all four mutants were compared (Section 2.6.3). The GO cellular component categories that were enriched in all four mutants were 'cytoplasm' and 'proteasome complex'. The *tti*$^{s450}$, *set*$^{453}$ and *nol9*$^{sa1022}$ mutants also shared enrichment for the categories 'nucleolus', 'ribonucleoprotein complex' and 'mitochondrial intermembrane space'. For the molecular function ontology categories, the enriched terms in all four mutants were 'RNA binding', 'translation initiation factor activity', 'nucleotide binding' and 'unfolded protein binding'. The *tti*$^{s450}$, *set*$^{453}$ and *nol9*$^{sa1022}$ mutants additionally shared the categories 'aminoacyl-tRNA ligase activity', 'serine-type endopeptidase activity', 'endopeptidase inhibitor activity', 'DNA-directed RNA polymerase activity' and 'ATP-dependent helicase activity'. The biological process term enriched in all four mutants was 'translational initiation'. In addition, *tti*$^{s450}$, *set*$^{453}$ and *nol9*$^{sa1022}$ mutants also showed enrichment for terms 'rRNA processing', 'tRNA aminoacylation for protein translation' and 'ribosome biogenesis'.

Similarly to the *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants, the development of digestive organs of *tti*$^{s450}$ and *set*$^{453}$ mutants is impaired (Boglev et al., 2013). However, the GO terms associated with the digestive organs namely 'pancreas development', 'exocrine pancreas development' endocrine pancreas development', 'liver development',' digestive system

development', 'digestive tract development' and 'digestive tract morphogenesis' all had a K-S value greater than 0.01 in both *tti*$^{s450}$ and *set*$^{453}$ mutants (Refer to Appendix Table A - 10, Table A - 11). In addition to the digestive organ defects, the *tti*$^{s450}$ and *set*$^{453}$ mutants displayed other phenotypes including smaller eyes, head and craniofacial defects (Boglev et al., 2013). The differentially expressed genes in *tti*$^{s450}$ and *set*$^{453}$ mutants were not enriched in GO terms associated with the eye, brain, cartilage and skeletal muscle except for 'skeletal muscle tissue development' in *set*$^{453}$ mutants (Refer to Appendix Table A - 10, Table A - 11). Moreover, autophagy was found to be upregulated in the intestinal epithelium of *tti*$^{s450}$ and *set*$^{453}$ mutants (Boglev et al., 2013). However, the GO term 'autophagy' was not enriched in *tti*$^{s450}$, *set*$^{453}$, *nol9*$^{sa1022}$ or *las1l*$^{sa674}$ mutants with K-S values of 0.80, 0.65, 0.15 and 0.11 respectively and none of the *autophagy-related genes* (*atg*) that were within 100 bases of the nearest 3' end Ensembl gene was statistically significantly differentially expressed in any of the four mutants (Refer to Appendix Table A - 12).

## 5.3 Discussion

In this chapter, the mRNA expression profiles of four different rRNA processing mutants *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ were investigated using Differential Expression Transcript Counting Technique (DeTCT) in order to provide insight into the molecular basis of the phenotypes in these mutants.

The mRNA expression profiling of *nol9*$^{sa1022}$ mutants revealed an increased expression of genes involved in ribosome biogenesis, rRNA processing, tRNA aminoacylation for protein translation and translational initiation. In comparison, the set of genes differentially expressed in the *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ mutants was enriched in ribosome- and translation-related functions. This agrees with the previously reported zebrafish loss of function model of SBDS in which differentially expressed genes showed enrichment in the GO categories 'ribosome biogenesis', 'rRNA processing' and 'translational initiation', and these trends were particularly robust amongst genes that showed upregulation in *sbds*-deficient embryos (Provost et al., 2012). In the *nom1* mutants also, differentially expressed genes were enriched in ribosome-related Gene Ontology categories (Qin et al., 2014). Provost *et al.* hypothesised that the increased expression of genes related to ribosome biogenesis and translation represents a mechanism by which cells compensate for the disruptions in ribosome assembly. Interestingly, there is a marked enrichment of upregulated genes belonging to the 'proteasome complex' in *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ mutants, it is possible that this is due to cells trying to degrade unrequired or damaged proteins following nucleolar stress induced by impairments in rRNA processing.

As the mRNA expression analysis was carried out on RNA from pooled larvae of mutants or wild-type siblings, one concern of using this approach is its sensitivity in detecting tissue-specific changes. The *rps19*-deficient embryos are characterised by haematopoietic and developmental abnormalities similar to Diamond-Blackfan anaemia (Danilova et al., 2008; Uechi et al., 2008). RNA-seq on pooled embryos injected with *rps19* morpholino, *rps19* and *tp53* morpholino or control morpholino revealed that genes enriched in functions in haematological systems, skeletal and muscular disorders and nervous system development, showed significant differential expression between *rps19*-morphant and control embryos (Jia et al., 2013). This suggests that tissue-specific effects are still detectable when studying gene expression in pools of whole embryo mRNA. Our analysis of the differentially expressed

genes in *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants revealed an enrichment of genes associated with digestive system indicating that tissue-specific effects were also detectable in our system. In addition, genes that are specifically expressed by the liver, pancreas and intestine are amongst those that showed the most significant decrease in transcript abundance in *nol9*$^{sa1022}$ mutants. This observation is consistent with the *nol9*$^{sa1022}$ mutants having fewer transcripts specific to digestive organs due to the smaller size of the digestive organs in *nol9*$^{sa1022}$ mutants compared to those in wild-type siblings.

The two rRNA processing mutants *titania* and *setebos* exhibit digestive organ defects, smaller eyes and head, craniofacial defects and autophagy was found to be increased in their intestinal epithelium (Boglev et al., 2013) (Dr Joan Heath, Personal Communication). The eye-specific gene *grifin* is amongst the genes that showed the most statistically significant decrease in transcript abundance in *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ mutants. It would be interesting to confirm this result by examining the expression of *grifin* by *in situ* hybridisation and investigate whether rRNA processing genes play a role in eye development. The analysis of differentially expressed genes in *tti*$^{s450}$ and *set*$^{453}$ mutants did not detect any significant enrichment of GO terms associated with the digestive organs, eye, brain, cartilage, skeletal system and autophagy. There are a number of reasons to explain this: firstly, those particular tissue-specific changes cannot be distinguished in the RNA of whole larvae. Secondly, the mRNA expression analysis is being carried out on larvae at a stage when changes in the expression of genes associated with those organs and the autophagic process cannot be detected. It will be worth analysing the gene expression profiles of these mutants at a stage when the morphological defects begin to manifest, i.e. 3 d.p.f. for *tti*$^{s450}$ and *set*$^{453}$ mutants and 4 d.p.f. for *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants. Moreover, the examination of changes in the level and cellular localisation of proteins is critical to complement the gene expression analysis as changes in the latter do not always correlate with changes at the protein level. Protein levels can be ascertained through western blot analysis while antibody staining can reveal the localisation of proteins of interest. These will potentially provide insight into the activity of proteins and pathways and help understand the mechanisms underlying the defects in rRNA processing mutants.

Haematological defects are prevalent in ribosomopathies (Narla and Ebert, 2010). The knockdown of *sbds* in zebrafish causes loss of neutrophils and in a zebrafish model of Diamond Blackfan anemia, haematological abnormalities were present and differentially

expressed genes between *rps19*-morphant and control embryos were enriched in functions in haematological systems (Jia et al., 2013). The functional analysis of the mRNA expression profiles of *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants revealed that differentially expressed genes were enriched in the Gene Ontology term "hemopoiesis" in *nol9*$^{sa1022}$ mutants and "myeloid cell differentiation" in *las1l*$^{sa674}$ mutants. These observations suggest that *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants, similar to *sbds* and *rps19* morphants, may also have haematological defects and thus encourage a detailed examination of haematopoiesis in *nol9*$^{sa1022}$ and *las1l*$^{sa674}$ mutants.

The mRNA expression profiling of *las1l*$^{sa674}$ mutants revealed only a handful of differentially expressed genes that reached statistical significance compared to *nol9*$^{sa1022}$, *tti*$^{s450}$ and *set*$^{453}$ mutants. This could be due to a less severe phenotype and therefore fewer differentially expressed genes compared to wild-type siblings. Similarly, the number of differentially expressed genes in *tti*$^{s450}$ and *set*$^{453}$ mutants was about three times more than the number of differentially expressed genes in *nol9*$^{sa1022}$ mutants and could be due to additional morphological defects in *tti*$^{s450}$ and *set*$^{453}$ mutants. Several differentially expressed genes in *nol9*$^{sa1022}$, *tti*$^{s450}$ and *set*$^{453}$ mutants have not been previously characterised and future work will investigate their role in the rRNA processing mutants. Furthermore, in this experiment, the differentially expressed regions that are more than 100 bases from the nearest 3' end of an Ensembl annotated gene have been excluded from the analysis, but could potentially be of interest and explored in the future.

The tumor suppressor protein Tp53 is the principal guardian of the genome and of cell integrity (Lane, 1992). The gene expression analysis in *nol9*$^{sa1022}$ mutants showed an increased expression of *tp53* and Tp53 target genes *mdm2*, *ccng1*, *cdkn1a*, *gadd45aa* and *baxa* (Tokino and Nakamura, 2000). Both Gadd45a and Cyclin G1 can induce cell cycle arrest at the G2/M phase whilst Cdkn1a is a cell cycle regulator at the G1/S phase and inhibits the activity of cyclin-CDK2, cyclin-CDK1 and cyclin-CDK4/6 complexes (Gartel and Radhakrishnan, 2005; Kimura et al., 2001; Rosemary Siafakas and Richardson, 2009). This increased expression of genes involved in cell cycle control could explain the cell cycle defects and impairments in cell proliferation that contribute to the exocrine pancreas phenotype in the *nol9*$^{sa1022}$ mutants. Additionally, an increase in expression of genes involved in apoptosis including *baxa*, *casp8* and *cycsb* was observed (Kruidering and Evan, 2000; Liu et al., 1996; Oltvai et al., 1993) although an increase in apoptosis was not detected in *nol9*$^{sa1022}$ mutants. Interestingly, *rps27.2*, the zebrafish orthologue of *RPS27L* was found to be

significantly upregulated in *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ mutants. The *ribosomal protein S27-like* (*RPS27L*) gene is a Tp53 transcriptional target that determines cell fate in response to DNA damage (He and Sun, 2007; Li et al., 2007). *RPS27L* and *RPS27* form a complex with *MDM2* and inhibits Mdm2-mediated Tp53 ubiquitination thereby increasing the half-life and levels of the Tp53 protein (Xiong et al., 2011). It will be interesting to investigate whether Rps27.2 is involved in the mechanism responsible for the hypoplastic digestive organs of rRNA processing mutants. This can be achieved by studying the effects of knocking down *rps27.2* expression by morpholino and/or combined loss of function of *rps27.2* and *nol9.*

The disruption of ribosome biogenesis and/or nucleolar structure can activate a Tp53-dependent stress response (Holmberg Olausson, 2012). The two main models to explain the mechanism of activation of Tp53 are the stabilisation of Tp53 protein by default following disruption of nucleolar structure and/or function (Rubbi and Milner, 2003; Zauberman et al., 1995) and the inhibition of MDM2, a repressor of Tp53, by ribosomal proteins redistributed to the nucleoplasm after nucleolar disruption (Holmberg Olausson, 2012). Tp53 has been shown to be involved in the pathogenesis of ribosomopathies and ribosome biogenesis mutants. In the Treacher-Collins syndrome mouse model, reduction in expression of *Tp53* rescues craniofacial dysmorphology and in the zebrafish *wdr43* mutant that displays defects in early development in neural tube, eye, heart and pharyngeal arches and then in craniofacial cartilages, knockdown of *tp53* partially rescues the craniofacial defects (Jones et al., 2008; Zhao et al., 2014). In zebrafish *rpl11* and *rps29* mutants, inhibition of Tp53 partially rescues haematopoietic phenotypes in zebrafish *rpl11* and *rps29* mutants (Danilova et al., 2011; Taylor et al., 2012). In addition in the *def*$^{hi429}$ mutant, where upregulation of Δ*113p53*, a shortened isoform and target gene of Tp53, induces the expression of Tp53 response genes, knockdown of *tp53* and Δ*113p53* partially rescues the mutant phenotype (Chen et al., 2009; Chen et al., 2005). In our study, the mean volume of 5 d.p.f. *nol9*$^{-/-}$*tp53*$^{+/+}$ and *nol9*$^{-/-}$*tp53*$^{-/-}$ larvae were not statistically significantly different meaning that the small exocrine pancreas phenotype of *nol9*$^{sa1022}$ mutant was not rescued by Tp53 loss of function and suggests the involvement of a Tp53-independent mechanism. This result is comparable to the pancreatic defects of *sbds* deficient embryos and *pes*$^{hi2Tg}$, *rpl3*$^{hi2437}$, *rpl23a*$^{hi2582}$, and *rpl6*$^{hi3655b}$ mutants also not being rescued by genetic loss of Tp53 (Provost et al., 2012; Provost et al., 2013). In addition, Tp53 loss of function does not abolish the induction of autophagy in the intestine of

*tti*$^{s450}$ mutants even though the *tti*$^{s450}$ larvae displayed increased transcription of the Tp53 target genes Δ*113p53*, *cdkn1a*, *ccng1* and *mdm2* (Boglev et al., 2013).

Tp53-independent mechanisms are thought to be involved in the pathogenesis of ribosmopathies and loss of Tp53 does not rescue the phenotypes completely in zebrafish *rps29, rpl11, wdr43* and *def*$^{hi429}$ mutants (Chen et al., 2005; Danilova et al., 2011; Donati et al., 2012; Holmberg Olausson, 2012; Taylor et al., 2012; Zhao et al., 2014). Several Tp53-independent pathways that mediate the nucleolar stress response have been described and include an RPL11-dependent mechanism whereby inhibition of rRNA synthesis by POLR1A depletion impaired cell-cycle progression in Tp53-inactivated cancer cell lines by downregulation of the transcription factor E2F-1. RPL11 released after nucleolar stress, binds to MDM2 inactivating the E2F-1 stabilising function of MDM2 leading to a downregulation of E2F-1 protein and decreasing the expression of E2Fs target genes that are required for entry and progression at the S-phase (Donati et al., 2011). A second Tp53-independent mechanism is the downregulation of c-Myc in response to inhibition of ribosome biogenesis due to RPL11 binding to c-Myc and reducing its transcriptional activity. This results in reduced cell proliferation even in *tp53*-null cells (Dai et al., 2007). A third Tp53-independent mechanism involves the downregulation of PIM1 expression, caused by deficiency of ribosomal proteins or other ribosomal stressors, which results in stabilisation and activation of CDKN1B leading to cell-cycle arrest (Iadevaia et al., 2010). Other potential p53-independent mechanisms that have been put forward involve the Tp53-independent effects of MDM2 or p63 and p73, two members of the p53 superfamily although there are no studies to demonstrate that these mechanisms are nucleolar stress responses (Holmberg Olausson, 2012). Future work will include investigating whether any of these p53-independent mechanisms contribute to the hypoplastic digestive organs of *nol9*$^{sa1022}$ mutants. Interestingly the expression of *mdm2*, *ccng1* and *cdkn1a* were not completely abolished in *tti*$^{-/-}$;*tp53*$^{-/-}$ mutants and those genes were in fact significantly upregulated in *tti*$^{-/-}$;*tp53*$^{-/-}$ mutants compared to *tti*$^{+/+}$;*tp53*$^{-/-}$ larvae. Furthermore, studying the mRNA expression profiles of *nol9*$^{-/-}$;*tp53*$^{-/-}$ mutants and *nol9*$^{-/-}$;*tp53*$^{+/+}$ mutants will help uncover genes that contribute to the phenotype of *nol9*$^{sa1022}$ mutants in a Tp53-independent mechanism.

The mRNA expression analyses of *nol9*$^{sa1022}$, *las1l*$^{sa674}$, *tti*$^{s450}$ and *set*$^{453}$ mutants served as a discovery tool to identify alterations in genes and pathways in zebrafish rRNA processing mutants. Elucidating the mechanisms of nucleolar stress in the context of a whole

organism is critical to understanding the pathogenic mechanisms operating in ribosomopathies. In particular, determining the Tp53-independent stress pathways will benefit the development of treatments for ribosomopathies and cancers with defective Tp53.