The role of lin28, an FGF signalling target, in development and miRNA regulation

Fiona Warrander

Thesis submitted for the degree of Doctor of Philosophy (PhD)

The University of York

Department of Biology

September 2012

Abstract

The related genes *lin28a* and *lin28b* code for conserved RNA-binding proteins, which contain two key RNA-binding motifs that determine their functions. Previously, our laboratory identified *lin28a* as a putative downstream target of FGF signalling in the early *Xenopus* embryo. This was found to occur at gastrulation stage, during which key signalling pathways such as the FGF pathway are active in specifying germ layer development.

lin28 is a heterochronic gene in *C. elegans* and controls the timing of developmental events. In vertebrates, the *lin28a* gene shows pluripotent-specific expression, and has come to particular interest as one of four factors used to successfully re-program differentiated somatic cells into induced pluripotent stem cells. Both lin28a and lin28b have a high prevalence of ectopic expression in cancer. A well characterised target of both lin28a and lin28b is the microRNA let-7, inhibiting biogenesis to the mature microRNA form. The lin28 proteins have also been shown to potentiate translation of numerous mRNA targets.

The aim of this project was to identify if lin28 has an important developmental role in vertebrates. Work in the *Xenopus tropicalis* embryo found that both *lin28a* and *lin28b* were targets of FGF signalling at gastrulation, and were required for correct germ layer patterning at this stage. In order to explore conservation in humans, mesenchymal stem cells were used to model the effects of FGF signalling upon the mesodermal germ layer, and whether lin28 played a role in this. Additional pluripotent cell models were investigated for their suitability in which to study lin28 function. Analysis of lin28 targets in *Xenopus* revealed that the miR-17-92 cluster family of miRNA may be positively regulated by the lin28 proteins, with a direct interaction possible with a member of these: pre-mir-363. These miRNAs may have further important developmental roles in response to this regulation by the lin28 proteins.

Contents

List of figures

List of tables

Acknowledgements

I'd like to thank my supervisors Harv and Paul, for getting me through this and helping shape this project over the years. I've been glad to have you both. Additionally, to my first lab supervisor, Anne-Marie Buckle, who helped show me this path I set myself on, but sadly succumbed to cancer a couple of years ago now. Thanks to all members, past and present of the Isaacs, Genever and Pownall labs. There's too many of you to mention, but a special thanks to Betsy and for Laura for working alongside on lots of the lin28 work. Cheers to Dan for helping me understand some of the world of structural biology, and to Oleg in Fred Anston's group and Patty in Mark Cole's group for providing me with samples. Huge thanks to Naveed and Celina in the TF, for helping with the genomics work and being simply brilliant with it.

To the group of Toby, Matt, Sarah Kendall, Rich, Paula, Sally and Sam for countless breaks, cakes, drinks and chats through the years, and particularly Sarah Keenan for being the best housemate for the majority of this time as well. The support from you guys has made all the difference. Likewise to those friends more removed - either by distance or by subject and understanding. Thanks also to everyone that there has been along the way, there's too many to name, for drinks, chats, and anything and everything.

Thanks to my family, to Dad and Mum for unwavering support and plenty of help along the way. To my sister Lynne - maybe soon we can discuss real doctor status. An enormous amount of thanks to John, who's supported me through the highs and lows when I've most needed it, and for being just simply brilliant. Can't believe you put up with me writing this.

I couldn't have done it without any of you, thanks all so absolutely very much.

Declaration

Some of the work presented in this thesis has been included in a paper currently under submission (Faas et al., 2013), which is included at the end of this thesis. Portions of this work have been presented in posters at UKNSCN Fourth Annual Science Meeting (2011), 13th International *Xenopus* Conference (2010), UKNSCN Third Annual Science Meeting (2010), Departmental Cell Biology Symposium (2010), RNA-protein interactions in development and cancer workshop (2009), and the 16th International Society for Developmental Biology Congress (2009). All of the work presented here is that of the author, with the exception of the microarray laboratory work and Agilent microarray analysis, which was carried out by The Biology Department Technology Facility.

Chapter 1. Introduction

1.1 Fibroblast growth factor (FGF) signalling

1.1.1 FGF signalling factors

Fibroblast growth factors (FGFs) are a large family of polypeptide growth factors which are crucial signalling molecules that regulate many cellular functions, including cell proliferation, growth and differentiation. In humans and mice, the FGF family consists of 22 known FGFs (Ornitz and Itoh, 2001); *Xenopus tropicalis* has been discovered to express orthologues to 20 of these, lacking clear orthologues to FGF-17 and FGF-18 (Lea et al., 2009).

FGF ligands share a conserved internal core, but also contain unique domains which result in differing functional roles, along with having different temporal and spatial expression patterns. For example, human FGF-4 is found only in embryonic tissue, whereas FGF-2 is expressed throughout embryonic and adult development (Ornitz and Itoh, 2001). The FGFs can be grouped into subfamilies, based on their different structures and expression (Itoh and Ornitz, 2004) (Table 1.1). Many FGF subfamilies contain N-terminal signal peptides and are secreted, whereas some are able to be secreted without secretion signal peptides (FGF-9 subfamily). FGF-22 contains a secretion signal like other members of its subfamily, but remains adhered to the cell surface rather than being secreted. FGF-1 and FGF-2 also lack a signal peptide, but can be released from cells following damage. Most FGF ligands are typically found in the extracellular matrix and function by binding to extracellular domains of transmembrane FGF receptors (FGFRs), but this is not the case for some FGFs; the FGF-19 subfamily members show very weak activity in activating the FGFRs, and are suggested to act in an endocrine fashion (Zhang et al., 2006). Additionally, the FGF-11 subfamily ligands are not secreted, and may act intracellularly although no pathway for this is currently understood (Ornitz and Itoh, 2001).

Table 1.1 FGF subfamily classifications for humans

FGF subfamily	FGF ligands	Characteristics
FGF-1	FGF-1, FGF-2	Not secreted, but are released from cells
FGF-4	FGF-4, FGF-5, FGF-6	Secreted
FGF-7	FGF-3, FGF-7, FGF-10, FGF-22	Secreted except for FGF-22
FGF-8	FGF-8, FGF-17, FGF-18	Secreted
FGF-9	FGF-9, FGF-16, FGF-20	Secreted, though no N-terminal sequences
FGF-19	FGF-19, FGF-21, FGF-23	Secreted, but only weakly activate FGFRs
FGF-11	FGF-11, FGF-12, FGF-13, FGF-14	Do not signal through FGFRs, remain intracellular

Information on FGFs collected from Itoh and Ornitz (2004), Zhang et al. (2006) and Ornitz and Itoh (2001).

The FGFR family in humans contains four closely related members, FGFR1-4 (Johnson and Williams, 1993), transmembrane receptors which contain intracellular tyrosine kinase domains, and extracellular receptor domains which contain three immunoglobulin (Ig)-like domains. The Ig-like domains are subject to alternative splicing and modulate specificity for FGFs, with splice variants showing preferences for the binding of particular ligands (reviewed in Bottcher and Niehrs, 2005; Zhang et al., 2006). The binding of FGF ligand causes receptor dimerisation, with the formation of both homodimers and heterodimers of the FGFRs 1-4 depending on the ligand and cell type. This results in cross-phosphorylation of the intracellular domains and activation of downstream signalling cascades through the phospho-tyrosine residues. The FGF-FGFR interaction is stabilised by the binding of heparin or heparan sulphate proteoglycans (HSPGs), which bind preferentially to the dimerised FGF-FGFR complex (Ornitz and Itoh, 2001) (Figure 1.1).

1.1.2 FGF signalling pathways

Much of the downstream activity of FGF signalling is activated by proteins containing a Src homology 2 (SH2) domain, which recognises phospho-tyrosine residues (Pawson et al., 1993). Proteins containing these domains can be both adaptor proteins, which transduce signals to downstream pathways, and catalytic proteins capable of carrying out activities once activated via the domain. There are three major pathways activated by FGFRs, which are described below (Figure 1.1) (reviewed in Bottcher and Niehrs, 2005).

One of the main downstream signalling pathways activated by FGF is the mitogen-activated protein kinase (MAPK) pathway. The FGFR substrate 2 (FRS2) protein is associated with FGFRs and contains phosphorylation sites which are targeted by the

activated FGFRs (Kouhara et al., 1997). Once phosphorylated, FRS2 binds the adaptor protein Grb2, which further recruits and forms a complex with son of sevenless (SOS), a nucleotide exchange factor (Hadari et al., 2001). In this complex, SOS then acts on a GTP binding protein, Ras, which interacts with Raf (Dent et al., 1992). Raf is a serine/threonine kinase protein that stimulates the MAPK cascade, phosphorylating MEK which then phosphorylates MAPK/ERK (Gomez and Cohen, 1991). MAPK/ERK then phosphorylates transcription factors, including ETS proteins and c-myc, which relocate to the nucleus and activate gene transcription (Wasylyk et al., 1998). The MAPK pathway is commonly activated by other growth factor pathways too, including epidermal growth factor and nerve growth factor.

A second pathway activated by FGF signalling is the phospholipase Cγ (PLCγ) and calcium release pathway. PLCγ is an enzyme activated following recruitment to the phosphorylated FGFRs (Mohammadi et al., 1991). Once activated, it hydrolyses phosphatidylinositol 4,5-bisphosphate to inositol-1,4,5-triphosphate (IP3) and diacylglycerol (DAG). IP3 is soluble and stimulates the release of intracellular calcium from sites such as the endoplasmic reticulum. DAG activates protein kinase C (PKC), promoting its translocation to the membrane; PKC is additionally activated by calcium and can perform a variety of roles in the cell.

The third major pathway activated downstream of FGF signalling acts via phosphoinositide-3 kinase (PI3K), which is activated via the FGFRs or Grb2 (Hadari et al., 2001). A main component of this pathway which is activated is Akt/protein kinase B, a serine/threonine kinase (Carballada et al., 2001).

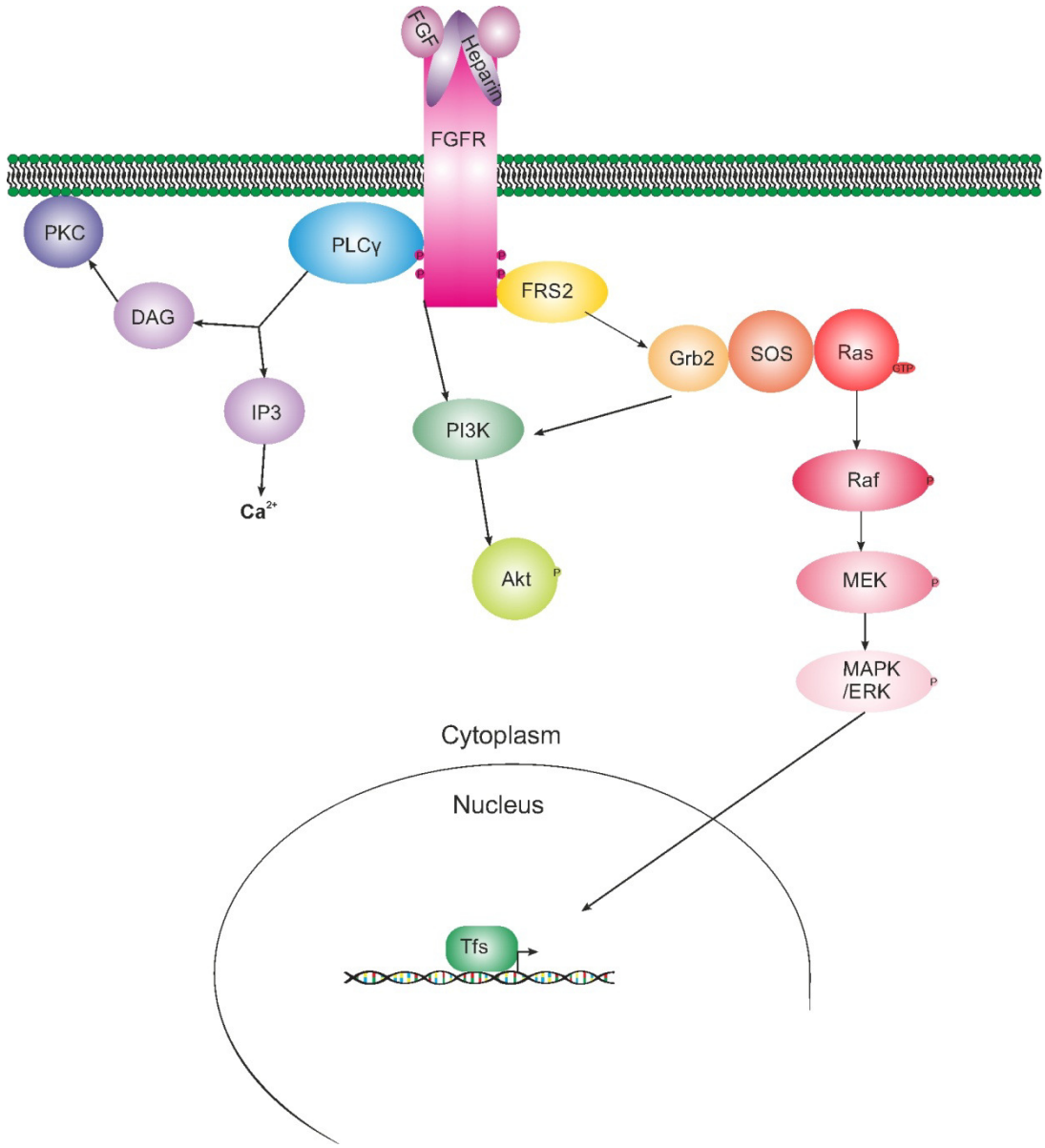


Figure 1.1 FGF signalling pathways within the cell

Intracellular signalling pathways following the activation of FGFRs by ligand binding. The FGF:FGFR:heparin/HSPGs complex activates receptor phosphorylation. FRS2 is phosphorylated and activates the adaptor Grb2. This associates with SOS which activates Ras. Ras activates Raf which sets off the chain of phosphorylation to MEK and MAPK, which activates transcription factors (Tfs). Alternatively, PLCγ binds FGFRs and hydrolyses phosphatidylinositol 4,5-bisphosphate to DAG, which activates PKC, and IP3, causing the release of calcium. FRS2/Grb2 also activate PI3K, which activates Akt.

1.1.3 FGF signalling in patterning the mesoderm

FGF signalling is involved in patterning the vertebrate embryo (Bottcher and Niehrs, 2005), in particular the induction of mesoderm differentiation, which has been well studied in *Xenopus* (Isaacs, 1997). Mesoderm is one of the primary germ layers along with ectoderm and endoderm, and is the precursor for differentiation into mesenchymal cell types, which includes muscle, bone and haematopoietic cells.

Mesoderm formation in embryos is initiated by signals from the vegetal pole. There are two signals, a dorsal signal and a ventral signal, which induce the marginal zone to form dorsal and ventral mesoderm respectively. Explants of cells taken from the animal cap, the top cells of the animal hemisphere, are not capable of developing into mesoderm without the vegetal pole to provide signals, or the signals provided exogenously. Another dorsal signalling centre is established in the marginal zone, known as the Spemann organiser. This releases signals to pattern the dorso-ventral axis of the mesoderm and to induce neural specification in the ectoderm (Figure 1.2). As FGF treatment could induce the development of mesoderm in animal cap cells (Slack et al., 1987), it was suggested that this was one of the vegetal signals. The FGF pathway was found to be required for both the formation and maintenance of the mesoderm (Cornell and Kimelman, 1994), and an inhibition of FGF signalling through the use of a mutant receptor lacking the tyrosine kinase domain which inhibited downstream signalling resulted in the development of much of the mesoderm being lost (Amaya et al., 1991). However, this inhibition was seen to result in disturbed lateral and ventral mesoderm, rather than a complete loss, indicating it is not the primary inducer of mesoderm. FGF was later discovered to be acting as a competence factor, enabling cells to respond to activin (Cornell et al., 1995), and is expressed within the marginal cells and developing mesoderm, rather than the vegetal cells (Isaacs, 1997; Isaacs et al., 1994).

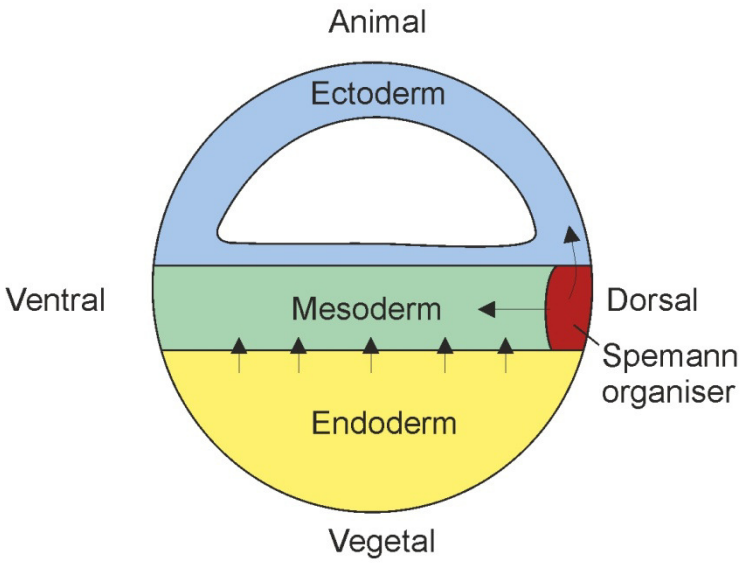


Figure 1.2 Early developmental signals in *Xenopus* development
A model of signals for germ layer induction. Dorsal and ventral signals from the endoderm induce dorsal and ventral mesoderm development from the marginal zone. Spemann's organiser signals are then released; these are further signals to the marginal zone to dorsalise mesoderm, and neural inductive signalling to the ectoderm.

FGF signalling is required not just to induce mesoderm, but also to maintain and pattern it. An early pan-mesodermal marker, *brachyury*, is a target of FGF signalling (Amaya et al., 1993), and brachyury is in turn able to activate expression of FGF, to create an autocatalytic loop (Isaacs et al., 1994). This regulation of *brachyury* is both an early mesoderm-specification event, and also occurs throughout gastrulation to maintain mesoderm. FGF signalling acts to pattern the dorso-ventral axis in the mesoderm, with high FGF levels driving dorsal structures (Furthauer et al., 1997; Isaacs et al., 1994).

This work has also been confirmed in other species. Active FGF signalling is needed for the proper migration of cells through the primitive streak of both mouse and chick (Chuai et al., 2006; Ciruna et al., 1997), with expression of *brachyury*, or *T* as it is also known, found in the developing mesoderm (Herrmann and Kispert, 1994; Schulte-Merker et al., 1994), and regulated by FGF signalling (Ciruna and Rossant, 2001).

1.1.4 FGF in patterning neural tissue

FGF is also involved as a signalling factor originating from the organiser which induces neural development. Ectoderm will only develop neural fates in the absence of bone morphogenetic protein (BMP) signalling, and key neural inducers expressed by the organiser include the BMP antagonists noggin and chordin (Piccolo et al., 1996; Zimmerman et al., 1996). FGF signalling can antagonise BMP signalling (Streit and Stern, 1999), through phosphorylation of Smad-1, a downstream effector of BMP, by the MAPK

pathway (Pera et al., 2003), and noggin and chordin have additionally been identified as targets of FGF (Branney et al., 2009). However, the inhibition of BMP signalling alone is not sufficient for neural induction; results in the chick embryo indicated that active FGF signalling was needed along with BMP inhibition to drive neural fates (Streit et al., 2000).

1.1.5 Identification of FGF targets during gastrulation

Early specification of the mesoderm occurs during gastrulation, and it is just prior to this stage of development that FGF signalling is first active, signalling through the MAPK pathway (Branney et al., 2009; Christen and Slack, 1999; Lea et al., 2009). At this time of tissue induction, signals are required to drive differentiation in the correct manner, and it is known than FGF signalling has an effect on subsequent expression of numerous genes to alter cell fate. A number of gene targets for FGF signalling at this time have been long known, but previous work in our group was carried out to identify more targets. Branney and colleagues (2009) inhibited FGF signalling in *Xenopus* embryos through the use of dominant negative receptors lacking the tyrosine kinase domains, which therefore could not activate intracellular pathways. Embryos were studied during the important signalling stage of gastrulation, with comparisons between the resultant transcriptomes of control and FGF-inhibited embryos; the study highlighted multiple genes which have been previously validated as FGF targets, but also newly identified other downstream targets. Some validation was carried out, confirming genes such as *ephrin receptor A4* and a *Xenopus* homologue to *DUSP5*, a target in other systems, as positively regulated FGF genes. There were 67 genes found to be significantly down-regulated upon FGF inhibition, and 16 genes were increased in expression, indicating there are more positively regulated targets of FGF than negatively ones regulated at this time. One gene of potential interest was *lin28a*, which showed a 2.7 fold repression, suggestive of being a positively regulated FGF target; *lin28a* is a pluripotency marker expressed in stem cells (Richards et al., 2004), and this gene is discussed further in 1.2.

1.1.6 FGF and embryonic stem cells

One function of FGF that has yet to be fully understood is its role in embryonic stem (ES) cell culture. For human ES cells, FGF has been shown to be required in culture conditions to maintain the cells in a pluripotent, undifferentiated and proliferative state, with inhibition of this pathway resulting in cell differentiation (Amit et al., 2000; Dvorak et al., 2005; Levenstein et al., 2006; Vallier et al., 2005; Xu et al., 2005). This role in ES cells appears despite FGF signalling being previously identified as primarily a differentiation factor. The role in which FGF is acting in ES cells principally remains unknown; the

experimentation using inhibition of FGF signalling results largely in differentiation of the cells (Eiselleova et al., 2009; Vallier et al., 2005) and it is therefore difficult to distinguish between direct effects downstream of FGF signalling and wider indirect effects caused by subsequent differentiation. Active FGF signalling was found to be required for maintained expression of pluripotent genes such as *Oct-4* and *Nanog*, and enhancement of cell survival and adhesion, all contributing to sustained growth and pluripotency (Eiselleova et al., 2009; Greber et al., 2010). ES cells are commonly grown on mouse embryonic fibroblasts (MEFs) as a feeder layer, and increases in FGF dosage can negate the need for this. One theory for how this promotes pluripotency is that active FGF signalling blocks BMP signalling, which would otherwise be driving differentiation, by synergising with the BMP antagonist noggin (Xu et al., 2005). It has further been shown that maintaining pluripotency with FGF culture only occurs when there is Activin or Nodal signalling also active (Vallier et al., 2005), transforming growth factor β (TGFβ) ligands that affect different pathways to BMP, although the downstream effectors of these pathways are also largely unknown in promoting pluripotency.

This requirement for FGF in culturing human ES cells is in stark contrast to mouse ES cells, which do not require FGF supplementation for their continued growth, but instead grow in the presence of leukaemia inhibitory factor (LIF) and BMP (Ying et al., 2003). The culture conditions for human ES cells more accurately match those required for murine stem cells collected from the more restricted epiblast (epiSCs), later in embryonic development, which require FGF to maintain undifferentiated growth (Alberio et al., 2010; Brons et al., 2007; Tesar et al., 2007). It is believed that these epiSCs represent a more primed state of differentiation compared to the more naïve ES cells, with restrictions on their differentiation potentials (Nichols and Smith, 2009). Differences in responses to FGF signalling is believed to be a key factor in these differentiation states, with naïve cells showing less of a propensity to differentiate following inhibition of FGF signalling, although it is not fully understood why (Hanna et al., 2010). Work has also shown that while both human ES cells and murine epiSCs require exogenous FGF for growth, the downstream effects of this treatment do differ, with this signalling required for *Nanog* expression in the ES cells thus being linked to pluripotency maintenance, but not in epiSCs (Greber et al., 2010).

There is low-level autocrine or paracrine FGF signalling in human ES cells (Eiselleova et al., 2009) but there is a different innate self-renewal programme in mouse ES cells, which causes them to respond differently to exogenous FGF. It is suggested that mouse ES cells produce higher levels of FGF, and require culturing with LIF and BMP to repress the resultant ERK signalling (Ying et al., 2008). FGF signalling and ERK activation have been

shown to promote differentiation in mouse ES cells (reviewed in Villegas et al., 2010), into lineages including trophectoderm (Schenke-Layland et al., 2007), mesoderm (Willems and Leyns, 2008), neural (Kunath et al., 2007) and endoderm (Funa et al., 2008). This all indicates that further understanding of the nature of ES cells is required. Much of the downstream effects of FGF in inducing differentiation appear to function via the ERK pathway (Kunath et al., 2007; Stavridis et al., 2007; Yoshida-Koide et al., 2004).

Due to the multiple FGF ligands found in species, particular effects may be attributed to specific ligands. FGF-4 is an important embryonic FGF ligand, expressed within the inner cell mass (ICM) of the embryo (Niswander and Martin, 1992), as well as in ES and embryonal carcinoma (EC) cells. Expression of FGF-4 is driven by the pluripotency factors Sox2 and Oct-4 (Yuan et al., 1995). *In vivo*, FGF-4 was shown to be required for proliferation of the ICM in the mouse embryo, from which ES cells are derived (Feldman et al., 1995); however, FGF-4 null ES cells are able to self-replicate and differentiate (Wilder et al., 1997), indicating that the function of FGF may differ depending upon cell state. FGF-4 in murine ES cells was shown to be required for driving neural and mesodermal lineage differentiation (Kunath et al., 2007), and also to be involved in specifying trophectoderm (Schenke-Layland et al., 2007).

1.1.7 FGF and other stem cells

The culture of other multipotent stem cells does not show a requirement for FGF supplementation; however supplementing media with FGF-2 has been shown to boost culture in some cases.

Multipotent mesenchymal stromal/stem cells (MSCs) showed an increased rate of proliferation following exposure to exogenous FGF-2, inhibiting cellular senescence; they also showed the ability to be cultured for longer periods of time and maintained an elongated morphology with FGF treatment (Coutu et al., 2011; Martin et al., 1997; Solchaga et al., 2005; Tsutsumi et al., 2001). Exposure to FGF-2 also increased the potency of the MSCs; cells showed enhanced osteogenic potential when expanded in FGF-2 before being treated to differentiate, producing higher levels of mineral deposits and expressing increased levels of bone markers (Martin et al., 1997; Pri-Chen et al., 1998; Tsutsumi et al., 2001). Chondrogenesis differentiation was also elevated with FGF-2 treatment (Solchaga et al., 2005; Tsutsumi et al., 2001), but adiopogenic differentiation was not improved. These effects of both promoting undifferentiated growth and increasing osteogenic differentiation are believed to reflect differing roles for FGF signalling based upon the maturity of the cell (Lai et al., 2011). FGF-2 treatment in MSCs activates both the MAPK

and PI3K pathways, which are involved in the promotion of proliferation and prevention of senescence (Choi et al., 2008; Coutu et al., 2011). Research using the overexpression of FGFR2 in MSCs found phosphorylation of ERK and activation of PKC signalling, both of which were demonstrated to be involved in promoting osteogenesis (Miraoui et al., 2009). MSCs treated with FGF-4 have also been found to exhibit increased proliferation rates (Choi et al., 2008; Farre et al., 2007).

Haematopoietic stem cells (HSCs) are another type of multipotent stem cell derived from the bone marrow, and display a limited potential in cell culture, with population renewal potential lost rapidly. It was found that culturing these cells with the addition of FGF-2 or FGF-1 increased the number of cells with stem cell activity, thus supporting HSC growth (Gabbianelli et al., 1990; Yeoh et al., 2006). However, alternative work indicated that FGF was not working directly on the haematopoietic progenitors, and must be having the effect by an indirect method (Berardi et al., 1995). Indeed, the HSCs that showed boosted growth with FGF treatment by Yeoh and colleagues (2006) were not purified HSCs, with the growth of purified HSCs not being supported by FGF treatment. More recent work has indicated that it is MSCs present in the bone marrow that are responsible for supporting HSC growth in a response to FGF treatment, with MSCs also being expanded in population size (Itkin et al., 2012).

Clonal growth was also induced in neural stem cells, isolated from the adult mouse brain and capable of forming astrocytes, oligodendrocytes and neurons, following FGF-2 treatment; the cells formed proliferating spheres from individual precursor cells (Gritti et al., 1996). In addition, FGF-2 was found to promote proliferation of neural precursors, particularly cells restricted to the neuronal fate, in which it also increased differentiation (Ray et al., 1993; Reynolds and Weiss, 1992; Richards et al., 1992). Both the MAPK and PI3K pathways were again found to contribute to the increased proliferation of these neural progenitor cells (Jin et al., 2005; Learish et al., 2000).

1.2 lin28

1.2.1 The *lin28* genes

The discovery of *lin-28* identified it as a heterochronic gene in *Caenorhabditis elegans*, where mutations causing a loss of function resulted in precocious development (Ambros and Horvitz, 1984). lin-28 was identified as a cytoplasmic protein containing both a cold-shock domain (CSD) and zinc finger motifs, two known RNA-binding domains (Ambros and Horvitz, 1984; Moss et al., 1997) (Figure 1.3).

In humans and many other vertebrates, there are two related *lin28* genes: *lin28a* and *lin28b*, which code for the only proteins known to contain both these CSD and zinc finger motif RNA-binding domains. Sequence identity conservation in humans between these two proteins is 52.3%, and is highest within the RNA-binding domains and lower in the disordered N- and C-termini. lin28b is a much larger protein than lin28a; this difference is mainly contributed to by a long C-terminus region (Figure 1.3, Figure 1.4A). Splice variants for both genes exist for a small number of species, and one of these is *Xenopus* where there are two isoforms of lin28a, these share 94.9% identity and differ only by a few amino acids at the N-terminus (Figure 1.3, Figure 1.4B). As for *C. elegans*, *Drosophila melanogaster* only contains one copy of the *lin-28* gene, and consequently only a solo lin-28 protein. Homology between species for both lin28a and lin28b proteins is apparent within the RNA-binding domains, with less sequence conservation within the C- and N-termini (Figure 1.4C, D); this has directed much of the research focus on these proteins to concentrate on functions of the conserved RNA-binding domains, thought to impart the biological activity of these proteins. The lin28 proteins show high homology between mammalian species, with 96.7% conservation found between the human and mouse lin28a protein. Sequence similarity percentages for the lin28a and lin28b proteins are given below for a selection of species that do not contain splice variants for either gene, with the exception of *X. tropicalis* (Table 1.2, Table 1.3).

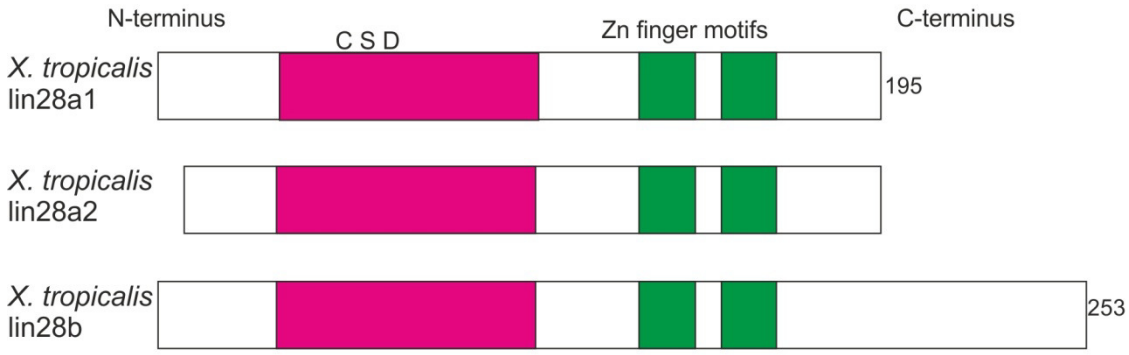


Figure 1.3 Structure of lin28 proteins

Pictorial representation of lin28 protein structures in *X. tropicalis*, indicating locations of the RNA-binding domains CSD and zinc-finger motifs.

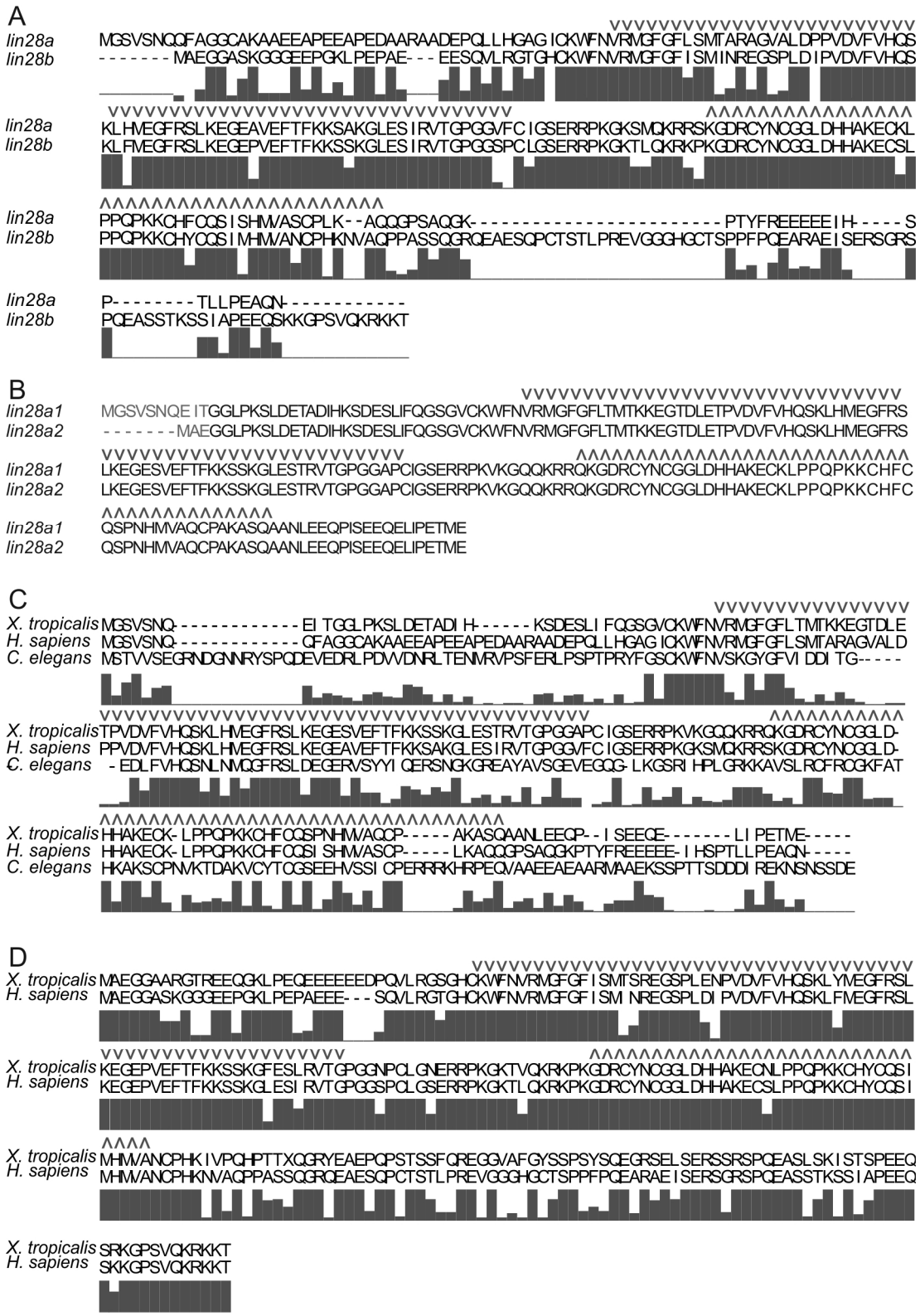


Figure 1.4 Alignments of lin28 protein sequences

Alignments of lin28 protein sequences, arranged in Clustal. A) Human lin28a and lin28b proteins, B) *Xenopus* lin28a1 and lin28a2, grey sequence shows non-conserved region, C) *X. tropicalis* lin28a1, human lin28a, and *C. elegans* lin-28, D) *X. tropicalis* and human lin28b. vvvv = cold shock domain, ^^^^ = zinc finger motifs. Underlying graph shows level of conservation between sequences (A, C, D).

Table 1.2 Percentage conservation between lin28a proteins

lin28a	*Caenorhabditis elegans* lin-28	*Drosophila melanogaster* lin-28	*Xenopus tropicalis* lin28a2	*Xenopus tropicalis* lin28a1	*Danio rerio*	*Gallus gallus*	*Bos taurus*	*Mus musculus*
Human	19.1	34.3	61.2	64.6	64.9	79.4	93.8	96.7
M. musculus	26.5	34.8	61.2	64.6	64	80.4	95.7	
B. taurus	26.5	35.2	62.9	66.3	65.2	82.4		
G. gallus	27.6	37.1	64.9	68.3	66.5			
D. rerio	27.1	33.7	62	61.5				
X. tropicalis a1	27.3	36.4	94.9					
X. tropicalis a2	26.5	36.4						
D. melanogaster lin-28	28.2							

Percentage of sequence conservation between lin28a proteins from varying species, with the exceptions of *C. elegans* and *D. melanogaster*, which only have one lin-28 protein. Species selected for comparison all had confirmed protein sequences and, with the exception of *X. tropicalis*, were without multiple isoforms. Percentage conservation calculated using Vector NTI (Invitrogen).

Table 1.3 Percentage conservation between lin28b proteins

lin28b	*C. elegans* lin-28	*D. melanogaster* lin-28	*X. tropicalis*	*G. gallus*	*B. taurus*	*M. musculus*
Human	25.5	32.1	78.7	82.5	76.7	78.1
M. musculus	22.6	29.2	65	69	67.6	
B. taurus	24.3	29.5	66.7	69.5		
G. gallus	26.6	29.2	83.8			
X. tropicalis	25.1	30.3				
D. melanogaster lin-28	28.2					

Percentage of sequence conservation between lin28b proteins from varying species, with the exceptions of *C. elegans* and *D. melanogaster*, which only have one lin-28 protein. Species selected for comparison all had confirmed sequences and were without multiple isoforms. Percentage conservation calculated using Vector NTI (Invitrogen).

1.2.2 lin28 structure

Both of the characterised domains of lin28 proteins are RNA-binding domains. The CSD of the lin28 proteins is homologous to regions in bacterial cold shock proteins and eukaryotic Y-box proteins, and contains the RNA-binding motifs RNP-1 and RNP-2 (Ermolenko and Makhatadze, 2002; Moss et al., 1997). CSDs commonly fold into five-stranded β barrels; the CSD in lin28a and lin28b contain key differences to vertebrate Y-box proteins, with differences in areas of β strands that are known to contribute to binding capabilities (Ermolenko and Makhatadze, 2002; Moss and Tang, 2003). The zinc finger motifs contain the characteristic CCHC cysteine and histidine motifs and have been found to show homology to retroviral nucleocapsid proteins which package viral genomic RNA, particularly the HIV-1 nucleocapsid NCp7 (Desjardins et al., 2011; Moss et al., 1997). The zinc fingers are flanked by basic amino acids, with a lysine/arginine-rich domain similar to NCp7, and these basic amino acids influence the binding of RNA (De Rocquigny et al., 1993; Desjardins et al., 2011). A flexible linker is present between the RNA-binding domains and lacks defined structure; this may allow the protein to bend around RNA targets to bind multiple target sites within one RNA sequence, with the linker allowing varying differences in the distances between the sites (Nam et al., 2011). This linker region is believed to contain a nucleolar localisation signal for lin28b (Piskounova et al., 2011).

There has not yet been any structural analysis of the N- and C-termini of the lin28 proteins; however, with potential functions being mediated by these domains (Jin et al., 2011), it is clear that further characterisation of them is needed to be certain of the function of the proteins. With the extension seen in the C-terminus of lin28b compared to lin28a, it is likely that this does impart some different physiological functions upon the proteins. One of these differences is the inclusion in this region of a nuclear localisation signal (Piskounova et al., 2011), but it is likely that there are further key differences in these domains that may control functions.

The Sliz group were the first to crystallise part of the lin28a protein, a truncated form of the protein lacking the N- and C-termini (Nam et al., 2011). This was a landmark development as difficulties had previously been reported with purification due to aggregations using the full length protein (Balzer and Moss, 2007; Desjardins et al., 2011); the crystallisation was only possible when the protein was bound to target RNA, which allowed an in-depth analysis of the way in which the protein interacted with the RNA, and this is discussed further in 1.3.7.

1.2.3 lin28 and pluripotency

lin28a was found to be expressed in pluripotent cells and downregulated upon differentiation (Darr and Benvenisty, 2009; Moss and Tang, 2003; Polesskaya et al., 2007; Richards et al., 2004), although expression in mice was found to persist in the adult in particular tissues including cardiac cells, gut epithelial cells, and some kidney structures, all rapidly regenerating cell types (Yang and Moss, 2003).

The *lin28a* gene emerged to wider interest when it was used as one of four factors, along with *Oct-4, Sox2* and *Nanog* to successfully re-program somatic cells into cells resembling ES cells, known as induced pluripotent stem (iPS) cells (Yu et al., 2007); however, another group found that cells could be re-programmed with an alternative set of four genes, with the substitution of *Klf4* and *c-myc* in place of *lin28* and *Nanog*, although they did not examine *lin28* as a candidate gene (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). *Nanog* and *lin28* were both shown to improve the efficiency of re-programming (Yu et al., 2007), with *Oct-4* and *Sox2* the critical components for re-programming, having been selected for use by both the Yamanaka and Thomson groups. Oct4 and Sox2 are both transcription factors that target numerous pluripotency genes, with expression of both required to maintain pluripotency (Masui et al., 2007), and Nanog expression is required for stem cells to have true 'ground-state pluripotency' (Silva et al., 2009). It is not fully understood what lin28a contributes to promoting re-programming, and why its omission is

not detrimental to the process, but numerous targets of lin28a have been identified that are involved in the maintenance of pluripotency and may be key for re-programming.

1.2.4 lin28 expression in the cell

The expression of lin28a in the cell is most commonly reported in the cytoplasm (Moss et al., 1997; Piskounova et al., 2011; Qiu et al., 2010; Sakurai et al., 2012), however in some cell types it has also been reported in the nucleus or nucleolus (Balzer and Moss, 2007; Polesskaya et al., 2007). Expression of lin28a in the cytoplasm has been found in localised foci (Balzer and Moss, 2007), which is discussed more in relation to its function in section 1.2.5. Levels of nuclear lin28a were increased when RNA-binding domains were mutated, leading to the suggestion that lin28a is present in the nucleus to associate with mRNA, which it then shuttles to the cytoplasm (Balzer and Moss, 2007). Expression of lin28b was first reported in hepatocarcinoma cells predominantly in the cytoplasm, although in a very small proportion of cells there was also nuclear expression (Guo et al., 2006). Nuclear expression was again seen in a different carcinoma cell line, where expression was further detected in the nucleolus; the group identified both nuclear and nucleolar localisation sequences within the protein (Piskounova et al., 2011).

1.2.5 Functions of lin28

In *C. elegans*, lin-28 translation is repressed by lin-4, the first microRNA (miRNA) to be discovered, and let-7, another miRNA (Moss et al., 1997; Reinhart et al., 2000). It continues to be repressed in mammalian cells by the family of let-7 miRNAs, and the lin-4 homolog miR-125 (Moss and Tang, 2003; Wang et al., 2012). These target the 3' untranslated region (UTR), and prevent protein synthesis.

The lin28 proteins interact with both mRNA and miRNA in the cell. One target of lin28, which has been highly characterised, is the miRNA let-7; this will be discussed further in sections 1.3.6 and 1.3.7. This regulation results in a negative feedback loop. lin28 proteins additionally target multiple mRNAs and positively regulate translation of these.

Within the cytoplasm lin28a associates in mRNA and protein complexes with other proteins including poly(A) binding protein and the initiation factor eIF3β, and with actively translating polysomes; the association with components of these complexes is generally dependent on the presence of mRNA, with the exception of eIF3β, which was seen to remain bound to lin28 in the absence of mRNA, although the importance of this interaction is not yet understood (Balzer and Moss, 2007; Polesskaya et al., 2007). It has

been found that with these interactions with the polysomes, lin28a promotes the association of target mRNAs, thus aiding their translation (Jin et al., 2011; Polesskaya et al., 2007). However, the mere association of mRNA with polysomes via lin28 interaction is not sufficient to improve translation, and the recruitment of RNA helicase A (RHA), a ubiquitously expressed RNA helicase that is involved in translation, is proposed as a mechanism for this; lin28 binds to both the C- and N-termini of RHA and recruits it to polysomes with target mRNA, which then functions to increase translation (Jin et al., 2011), and a loss of RHA diminishes the ability of lin28s to promote translation (Qiu et al., 2010). Interestingly, the unstructured C-terminus of lin28 is important for this interaction with RHA, allowing concurrent interaction with target mRNA via the RNA-binding domains (Jin et al., 2011).

Identification of mRNA targets of lin28 are helping to identify what role lin28 is playing during pluripotency. One target which shows enhanced translation with lin28 is Oct-4, a key pluripotency marker (Qiu et al., 2010), with further regulation being exhibited over cell cycle and proliferation genes; inhibition of lin28a expression in stem cells prevented cells from entering mitosis, with a stimulated growth rate following lin28 overexpression (Xu and Huang, 2009; Xu et al., 2009). Immunoprecipitation experiments carried out by the Huang group have generated a large list of potential mRNA targets for lin28, many of which have functions in metabolism and ribosomal function, which can affect the rate of cell cycle (Peng et al., 2011). Particular targets have been confirmed as including Histone H2A, cyclins A and B, and cdk4 (Xu and Huang, 2009; Xu et al., 2009). The cyclins and cdk genes are important genes in controlling the rapid cell cycle in ES cells, which is important in the process of self-renewal (reviewed in White and Dalton, 2005). This effect upon the cell cycle has also been seen in cancer cells, where a loss of lin28 reduced proliferation and the number of cells entering S-phase; there was also a loss in expression of *CDC25a*, which is involved in S-phase transition (Pan et al., 2011).

Although primarily a pluripotency marker, lin28a expression and regulation of particular mRNA targets is important during the differentiation of particular tissue types. Insulin-like growth factor-2 (IGF-2) is a target of lin28, with translation enhanced by lin28 during skeletal muscle development (Polesskaya et al., 2007). lin28 also promotes neurogenesis, at the expense of glial development; this process was found to be let-7 independent, and although IGF-2 was showing regulation by lin28 at this time, this was not confirmed as the primary target that was mediating the neuron-inducing effects (Balzer et al., 2010). lin28 is further required for development of primordial germ cells, through the regulation of let-7 (West et al., 2009), and expression has been found of *lin28b*, but not *lin28a*, in foetal HSCs, with an overexpression of lin28a in adult HSCs generating increased B-cells and T-cells

similar to those developing from foetal HSCs, indicative of a role in lymphopoiesis (Yuan et al., 2012).

Early work by the Huang group could not find a common consensus for lin28 binding in all of their identified mRNA targets, and as such they proposed that lin28 recognises RNA structural features more than a specific sequence (Peng et al., 2011). This has further been identified to be partially true, with lin28 recognising an 'A' bulge flanked by GC pairs forming part of a stem structure; this was not a purely structural recognition, however, as mutations of the 'A' to a 'U' did not alter the structure of the RNA, but did reduce binding affinity to lin28 (Lei et al., 2012). Work has suggested that it is the zinc finger motifs in lin28 that are predominantly responsible for binding to the mRNA and promoting translation, with mutations in these regions causing a reduction in association with both mRNA and polysomes (Balzer and Moss, 2007; Lei et al., 2012). This is different to the manner in which it recognises and binds miRNA, which is discussed more extensively in section 1.3.7.

Separate cytoplasmic expression of lin28 has been found in foci, which were characterised as P-bodies (Balzer and Moss, 2007); these are RNA granules in which mRNA can be degraded and decapped, repressed by miRNAs, or merely stored pre-translation. Expression in P-bodies does not co-localise with poly(A) binding protein or polysomes, indicating multiple functions for the protein, although it is not yet clear what role lin28 has in the P-bodies.

1.2.6 The involvement of lin28 in cancer

The role of lin28 in maintaining undifferentiated cells is reflected in the observation of increased levels of lin28a and lin28b in many cancers arising from different cell types, including lung, colon, prostate, breast, germ cell and cervical cancer (Cao et al., 2011; King et al., 2011; Nadiminty et al., 2012; Pan et al., 2011; Sakurai et al., 2012; Viswanathan et al., 2009; West et al., 2009), with these findings being shown in both primary tumours and transformed cell lines. Piskonuova and colleagues (2011) found that in transformed cell lines derived from tumours, lin28a or lin28b was expressed in most of the cell lines studied, but never together, leading them to conclude the expression of these was likely to be mutually exclusive in cancer cells. There have also been further reported cases of cancer cells expressing just one of the *lin28* genes (Pan et al., 2011); however, other work suggests that the two genes may be expressed together in some forms of cancer (Sakurai et al., 2012; Viswanathan et al., 2009), and it may simply be that there is a level of

redundancy in the genes, with both having the same effect upon cancer, and thus ectopic expression of either lin28a or lin28b is sufficient.

Correspondingly, a reduction in levels of let-7 has been found in cancer and, with the lin28 expression data, the oncogenic activities of lin28a and lin28b have largely been attributed to their suppression of let-7 levels. There is a correlation of high lin28 levels with low let-7 levels in many cancers studied, with downstream effects being rapid proliferation, increased migration and increased invasiveness of cells (Nadiminty et al., 2012; Pan et al., 2011; Viswanathan et al., 2009). As this function is attributable to both lin28a and lin28b, this may explain why expression of these are commonly mutually exclusive in cancer. This increase in lin28 and decrease in let-7 promotes the transformation of cancer cells, and their return to a more undifferentiated state (Viswanathan et al., 2009). Additionally, lin28 has been found in ovarian cancer cells in conjunction with Oct4, and this co-expression was found to mark the more stem cell-like cells in the cancerous population (Peng et al., 2010). There are instances, however, in which there is not a concomitant reduction in let-7 levels where lin28 is overexpressed (Sakurai et al., 2012), and instead lin28 may be influencing other targets. Work has shown that the lin28s also promote transformation by binding to mRNA targets as discussed above, with key targets including cell cycle genes (Feng et al., 2012; King et al., 2011; Li et al., 2012; Pan et al., 2011).

Expression of lin28 has also been investigated for a marker of cancer progression. The co-expression of Oct-4 and lin28 in ovarian cancer was seen as a marker of more advanced tumour classification with a poorer prognosis (Peng et al., 2010), with lin28a or lin28b expression alone an indicator of poor prognosis in other cancers (Feng et al., 2012; Hamano et al., 2012). A reduction in levels of let-7 was also linked to earlier deaths in lung cancer (Takamizawa et al., 2004). Further work into this may determine whether lin28 can be a valuable marker for classification of cancer progression and prediction of prognosis.

1.3 miRNAs

1.3.1 miRNAs: An important class of regulatory RNAs

The field of non-coding RNA study is becoming increasingly important, with the discoveries of multiple classes of such RNAs; this includes small nuclear and nucleolar RNA (snRNA, snoRNA), which are involved in the processing of RNA in the cell, and regulatory RNAs such as piwi-interacting RNA (piRNA), long intergenic non-coding RNA (lincRNA) and miRNAs.

Mature miRNA are ~22 nucleotide (nt) RNA sequences that provide post-transcriptional gene regulation in the cell, via RNA interference (RNAi). RNAi affects mRNA translation as a

form of post-transcriptional regulation, through the formation of double-stranded RNA (dsRNA). As mentioned above, the first miRNA to be discovered in animals was lin-4 in *C. elegans*, which was found to regulate mRNA, including *lin28*, by binding complementary elements in the 3' UTR (Lee et al., 1993; Wightman et al., 1993), preventing protein expression of targets (Olsen and Ambros, 1999).

miRNAs have been identified in both animals and plants, suggesting they represent an important regulatory mechanism. Over 1000 miRNA sequences are presently known in humans, with this list appearing to be ever expanding (Griffiths-Jones et al., 2008). Conservation of these sequences is high across many species (Suh et al., 2004; Watanabe et al., 2005), with some identical mature miRNA sequences being found in diverse species. For example, there are variants of the miRNA let-7 that are highly conserved from humans to *C. elegans,* and are conserved 100% at the seed region of the mature sequence (Figure 1.5). This demonstrates the biological importance of these short regulatory RNA sequences.

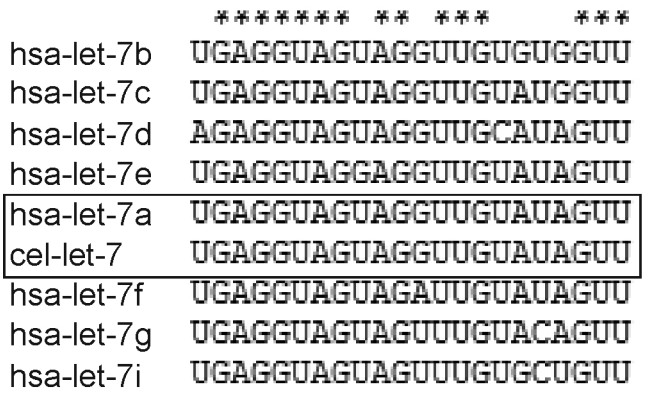


Figure 1.5 Conservation of let-7 sequences between humans and *C. elegans*
Clustal alignment of all human mature let-7 sequences and *C. elegans* let-7. Stars represent conserved residues in all sequences. Box is drawn around miRNA with 100% conservation between humans and *C. elegans*.

1.3.2 The biogenesis of mature miRNA

miRNAs undergo multiple processing events during their biogenesis to result in the mature active form, which is represented in Figure 1.6.

miRNAs are present in the genome in various different forms. Some are present in clusters, where numerous miRNAs are transcribed as polycistronic RNAs and cleaved to separate into their individual forms, with others present as individual miRNAs and transcribed alone (Lee et al., 2002); these routes provide the most common forms of

miRNA generation. Primary miRNA (pri-miRNA) from miRNA genes with independent promoters is synthesised by RNA polymerase II (Cai et al., 2004; Lee et al., 2004), the same RNA polymerase responsible for transcribing mRNA. The same modifications are further imparted in post-transcriptional modification onto miRNAs as for mRNA, including polyA tails, and 5' capping, although this was not found to be present in all cases of miRNA transcription (Lee et al., 2004). Pri-miRNA has been found of varying sizes, from around 100 nt to over 800 nt (Lee et al., 2002), which gives strength to the idea that some are transcribed as clusters. This pri-miRNA is processed to the mature, active form by a series of cleavage steps.

Within the nucleus, the primary miRNA is cleaved to form precursor miRNA (pre-miRNA) of ~70 nt in length, by the RNase III protein Drosha (Lee et al., 2003) as part of a Microprocessor complex with the involvement of other proteins including DiGeorge syndrome critical region 8 (DGCR8), also known as Pasha in invertebrates (Denli et al., 2004; Gregory et al., 2004). Neither DGCR8 nor Drosha can initiate pri-miRNA cleavage without the other, and the method by which they recognise and bind to RNA is still being uncovered. Work by Zeng and Cullen (2005) showed that to be recognised for Drosha cleavage, the pri-miRNA must have an extended stem beyond the pre-miRNA, and a loop of ≥10 nt; this makes the pri-miRNA stem around three RNA helical turns in length, and they further found that Drosha cleaved the RNA two helical turns from the terminal loop (Zeng et al., 2005). However, some of this has been contested in more recent work, which showed that although having an unstructured loop region was required it could be internal to the miRNA, and the cleavage site was not dependent upon distance from this (Han et al., 2006). Flanking single-stranded RNA was demonstrated to be required for Drosha recognition (Zeng and Cullen, 2005), and it was instead suggested to be one helical turn from the stem-single-stranded region that determined the cleavage site, which is bound by DGCR8 (Han et al., 2006). This cleavage results in a stem-loop structure with a 2 nt 3' overhang, a typical result of RNase III proteins, and this cleavage also determines one end of the mature miRNA sequence (Wu et al., 2009). In a number of cases, it was seen that Drosha was capable of alternative cleaving of pri-miRNA, which resulted in altered mature miRNA ends and perhaps modified functions (Wu et al., 2009), which may be a further method in which miRNAs can be regulated.

The pre-miRNA is exported from the nucleus, before any further modifications, by the Exportin-5 (Exp5) protein, a RanGTP-dependent exportin which acts primarily to export miRNA and short interfering (siRNA) (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003), although it can also export other non-coding RNA such as tRNA and adenovirus RNA. Exp5 directly binds to pre-miRNA without the requirement for adaptor proteins

(Lund et al., 2004). Exp5 also stabilises pre-miRNA within the nucleus and prevents degradation before export; a knockdown of Exp5 was discovered to prevent the presence of both mature and pre-miRNA in the cytoplasm, but also produced reduced levels of pre-miRNA in the nucleus without affecting pri-miRNA levels (Yi et al., 2003; Zeng and Cullen, 2004). Exp5 had previously been demonstrated to bind RNA for export via a mini-helix with a 3-8 nt 3' overhang (Gwizdek et al., 2004), but it was shown that this interaction with pre-miRNA is more flexible. A stem of ≥16 bp was required for the interaction, with a 3' overhang ranging from 1 nt to 5 nt or even a blunt end capable of binding Exp5, and with any 5' overhang inhibiting the miRNA binding (Zeng and Cullen, 2004).

Pre-miRNA is cleaved in the cytoplasm by the RNase III protein Dicer, which results in the duplex of two arms of the miRNA (Bernstein et al., 2001; Hutvagner et al., 2001). Dicer contains a dsRNA binding domain, and RNase domains which cleave double-stranded miRNA, around 22 nt from its terminus (Zhang et al., 2002). It also contains a PAZ domain, which binds to the 3' overhang generated by Drosha (Macrae et al., 2006). Dicer uses the length of the 3' overhang to count along nucleotides to determine the cleavage site (Ma et al., 2004; Vermeulen et al., 2005). As Drosha cleavage has already determined one end of the mature miRNA, processing by Dicer selects the other end and determines the final sequence length of the miRNA. The resulting duplex of both arms of the miRNA dissociate, leaving single-stranded RNA (ssRNA), with the incorporation of an active mature miRNA into an RNA-induced silencing complex (RISC).

There are alternative processes to generating miRNA; a small number are present within introns of genes, known more specifically as mirtrons, and are present in both invertebrates and vertebrates (Ruby et al., 2007; Sibley et al., 2011). There are further examples of miRNA generated from tRNA-like structures, or transcribed straight into short hair-pin sequences (reviewed in Curtis et al., 2012). Mirtrons are not processed by Drosha, but instead by splicing and debranching of the mRNA, which results in pre-miRNA-like structures (Ruby et al., 2007). These are then processed in the same manner as pre-miRNA. Some of the alternative pathways, such as tRNA-like structures, can be processed independently of Drosha and the spliceosome to pre-miRNA-like structures, or straight to mature miRNA sequences (reviewed in Curtis et al., 2012). A reduced dependence on Exp5/Dicer has been reported for some mirtrons, but an alternative pathway for these has not yet been determined (Sibley et al., 2011).

This increased understanding of endogenous miRNA processing is greatly benefitting RNAi research, which is now able to better design synthetic siRNA and short hairpin RNAs to be favourably processed in cells and target mRNAs in better understood ways. There is also

ongoing work to manufacture synthetic mirtrons for use in clinical therapy (Curtis et al., 2012).

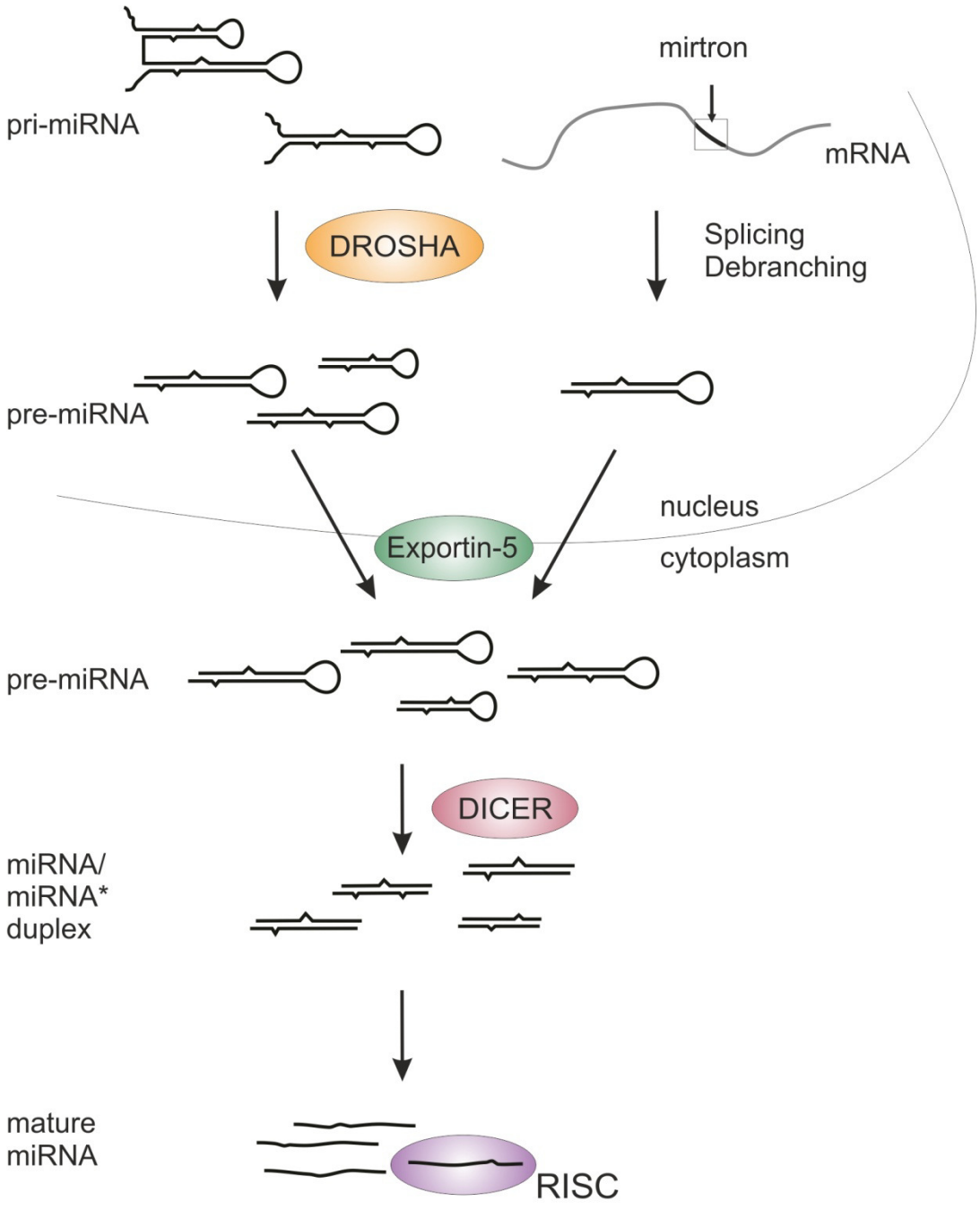


Figure 1.6 miRNA processing pathway

Pictorial representation of miRNA processing in the cell. Pri-miRNA and mRNA have been transcribed from DNA in the nucleus. These are processed by the Microprocessor complex containing Drosha, or the spliceosome to pre-miRNA. This is exported to the cytoplasm by Exportin-5, and processed further by Dicer. The resultant miRNA duplex dissociates and the active mature miRNA is incorporated into RISC.

1.3.3 RISC and translational inhibition by miRNAs

Critical components of RISC are ssRNA, to identify the sequence-specific target, and a nuclease protein from the Argonaute family; RISC typically cleaves target mRNA to prevent translation (Hammond et al., 2000; Hammond et al., 2001; Martinez et al., 2002). Mammals contain four Argonaute proteins (Ago1-4), but only Ago2 was shown to have catalytic activity in RISC (Liu et al., 2004). The RNA is loaded into RISC as a duplex, with Argonaute cleaving the non-active strand, which then allows activation of RISC (Rand et al., 2005). This interaction is likely to occur through the PAZ domain on the proteins, which is able to bind to the 3' overhang on the duplex (Song et al., 2003).

In contrast to siRNA, which shows full complementarity to target mRNA, miRNA recognises targets with only a short part of its sequence: a 7 nt region situated at nt 2-8 at the 5' end of the miRNA, known as the seed sequence. This acts as the complementary region to the mRNA and thus provides the miRNA specificity and function (Lewis et al., 2003). The miRNA targets sites typically found on the 3' UTR of mRNA (Saito and Saetrom, 2010; Wightman et al., 1993), but has also been shown to target open reading frames containing multiple binding site repeats and 5' UTRs, with this interaction occurring through the 3' end of the miRNA, as an addition to simultaneous interactions from the usual 5' seed with the 3' UTR (Lee et al., 2009; Schnall-Levin et al., 2011). In light of this, miRNAs have been shown to be most conserved between species at both the 3’ and 5’ ends of the miRNA (Okamura et al., 2008; Yang and Moss, 2003).

A single miRNA and RISC are capable of silencing multiple mRNA target strands (Hutvagner and Zamore, 2002); however, how the RISC functions with miRNA is still under investigation, and it appears multiple pathways of regulation are possible. Research has discovered that some miRNA targets, where protein expression has been inhibited, are not diminished at the RNA level (Olsen and Ambros, 1999), whereas there is also evidence of other mRNA targets showing modifications, such as deadenylation, and downregulation (Djuranovic et al., 2012; Lim et al., 2005). However, it would appear that translational inhibition is underway before any destabilisation of the mRNA begins to occur (Bazzini et al., 2012; Djuranovic et al., 2012; Mathonnet et al., 2007), and a large area of debate with miRNAs is how this translational repression occurs. Research has demonstrated examples of inhibition of protein translation after initiation has successfully occurred, blocking translational elongation, but also repression acting at the stage of translational initiation, with a requirement for a cap on the mRNA (Bazzini et al., 2012; Mathonnet et al., 2007; Olsen and Ambros, 1999; Petersen et al., 2006; Pillai et al., 2005). It is likely that different miRNAs can act through these different mechanisms to exert their effects on mRNA

translation, and the Bushell group found evidence to suggest that miRNA regulation may occur by either of these mechanisms, which they termed type I and type II, and have been attempting to identify what causes this selection. Their work has shown that the promoter responsible for the transcription of target mRNA appeared to direct whether translation initiation was slowed or there was translation repression postinitiation (Kong et al., 2008). Theories for how a mechanism of miRNA repression is selected have not yet been validated, but it is suggested that during transcription from particular promoters, mRNA may acquire factors which direct the miRNA repression to a time at translation, or that the RISC may interact with initiation factors.

1.3.4 Selection of active miRNA strand

The traditional view was that only one strand of the duplex was the active miRNA, termed the major miRNA, and was incorporated into RISC, with the minor miRNA (formerly universally represented by miRNA*, which will be used in this thesis) degraded. The so-called 'major' form can be found as both the 3p- arm and the 5p- arm of miRNAs. The 'major strand' classification was based on which miRNA was found to have the highest expression levels on preliminary identification, following the belief that the minor form was merely degraded. However, further research has shown that it is possible for miRNA* to have biologically repressive functions (Okamura et al., 2008; Yang et al., 2011). Both papers showed that some miRNA* strands accumulated in levels greater than some 'major' miRNA strands, dependent on cell type and developmental timing. An identical duplex can result in a different dominant miRNA strand, accumulating in higher levels than the alternative, when processed in different species (Griffiths-Jones et al., 2011), and can also switch in different cell types and upon developmental timing (Okamura et al., 2008; Yang et al., 2011). Griffiths-Jones et al. found that the number of target genes present in each of the cell types reflected the dominant miRNA arm choice, suggesting selective pressures, with the most biologically relevant miRNA being the chosen regulatory form (Griffiths-Jones et al., 2011).

Selection of dominant miRNA strand was initially theorised to be due to thermodynamic properties of the duplex, with the favoured miRNA strand showing less 5' thermodynamic stability (reviewed in Hutvagner, 2005). However this is now not believed to be the only decisive factor, as the dominant miRNA product can vary. The major miRNA was usually observed to be the more conserved strand of the duplex. Papers by Okamura and colleagues (2008) and Yang et al (2011) showed that active miRNA* was highly conserved between different species, producing a correlation between conservation and

miRNA* strand accumulation; this suggests that a high conservation of sequence between species reflects the biological importance.

1.3.5 miRNAs have important roles in development

The important role of miRNA through development can be shown by inhibition of the biogenesis pathway for all miRNAs, which studies have performed by interfering with Dicer expression. One group found that they could produce heterozygous($^{+/-}$) Dicer mutant mice, but not homozygous($^{-/-}$) Dicer mutants; Dicer $^{-/-}$ embryos, lacking mature miRNAs, showed abnormal early development and embryonic lethality (Bernstein et al., 2003). A separate study showed that Dicer could be reduced in ES cells using RNAi and cells continued to proliferate and express pluripotency markers. However, they did this at a slower rate than their Dicer-expressing counterparts, and did not contribute to chimeras, or form tumours or embryoid bodies (Kanellopoulou et al., 2005), showing that any differentiation requires expression of miRNA. To bypass this need for Dicer in early differentiation events, many studies have used transgenics to knock out Dicer in specific tissues in later development. A loss of Dicer induced in mesoderm tissue following gastrulation resulted in a failure of mouse axes to elongate, and a failure of hindlimb bud development (Zhang et al., 2011). Alternative transgenic models have shown that a loss of Dicer affected lung epithelial branching, reduced differentiation of skeletal muscle due to a failure in myofibre morphogenesis, impaired angiogenesis and reduced head size and craniofacial structures (Harris et al., 2006; O'Rourke et al., 2007; Yang et al., 2005; Zehir et al., 2010); all of these effects were found to involve an increased rate of cell death.

The vast number of miRNAs means, however, that miRNAs are involved in the numerous different roles that their target RNAs have. There is also a great variety in the expression of miRNAs; pluripotency-specific miRNAs have been characterised, along with differentiation-specific miRNAs, which can be widespread in their expression or restricted to very particular tissue types or times.

1.3.6 The let-7 miRNA family

Let-7 is one of these differentiation-specific miRNAs, and was discovered to be a heterochronic miRNA in *C. elegans*, with mutations causing retarded development (Reinhart et al., 2000; Slack and Ruvkun, 1997). The let-7 family of miRNAs is a highly conserved miRNA family – with one member in *C.elegans*, and 9 mature members in vertebrate species such as humans and *Xenopus*, as shown previously in Figure 1.5, although these arise from up to 12 different pre-let-7 forms (Pasquinelli et al., 2000; Peter,

2009). In *C. elegans*, mutations in let-7 were able to suppress the phenotype seen in lin-28 mutants, and binding sites for let-7 are present in the 3' UTRs of *lin28a* and *lin28b*; let-7 blocks the translation of both lin28a and lin28b, resulting in a negative feedback loop, with the lin28s also exerting a negative regulation on let-7 (Reinhart et al., 2000; Slack and Ruvkun, 1997). As discussed in section 1.2.6, let-7 expression is commonly repressed in cancer, which appears to allow cells to revert to an undifferentiated state, important in the progression of cancer.

1.3.7 The regulation of let-7 by lin28

One of the earliest examples of post-transcriptional regulation of miRNA was the discovery that lin28 inhibits let-7 processing. Both lin28a and lin28b proteins were found to bind to the let-7 terminal loop and prevent processing to the mature form. This interaction was first identified following a perceived block of the Drosha Microprocessor, cleaving pri-let-7 to pre-let-7, when incubated with cell extract, and lin28 was purified as bound to pre-let-7 (Viswanathan et al., 2008). There has been speculation since this discovery over whether this binding is to the pri-miRNA, pre-miRNA, or both, and therefore inhibiting processing either by Drosha or Dicer, as both methods appear possible *in vitro* (Newman et al., 2008; Piskounova et al., 2008; Viswanathan et al., 2008). This interaction between let-7 and lin28 and the inhibition of miRNA processing has since been repeated by a number of groups, using protein and RNA sequences from *C. elegans, X. tropicalis* and humans, and a wide range of binding affinities has been reported, from 0.15 nM to 2.1 μM (Desjardins et al., 2011; Lightfoot et al., 2011; Loughlin et al., 2012; Mayr et al., 2012; Nam et al., 2011). These studies have all used varying techniques to determine the binding efficiency between lin28 constructs and let-7 sequences, including electrophoretic mobility shift assay (EMSA), isothermal titration calorimetry, and fluorescence titrations, and the varied assay conditions, and protein and RNA constructs used are likely to have contributed to the ranging binding affinities reported.

Much research has since been dedicated to uncovering the mechanism by which this binding and the block on processing occurs. Early work in George Daley's group found that both the CSD and zinc fingers were required for binding let-7 (Piskounova et al., 2008), and subsequent work with truncated lin28 constructs, containing only one of the two RNA-binding domains on occasion, has determined separate binding sites for the two RNA-binding domains, and the crystallisation of the lin28a structure bound to let-7g demonstrated the involvement of both the CSD and zinc finger motifs in this binding interaction (Nam et al., 2011).

Piskounova and colleagues identified a cytosine residue which was conserved in the terminal loop of all human let-7 family members; this residue is exposed on the loop of the RNA, and when mutated reduced binding of the lin28 (Piskounova et al., 2008), but no importance has been attributed to this nucleotide by other work. It is possible that their mutation of the cytosine greatly altered the structure of the let-7 RNA, altering exposure of the true lin28 binding sites, and thus affected the binding interaction.

The binding of the zinc finger domains is suggested to provide the specificity to the interaction (Mayr et al., 2012), and CSDs have been shown previously to commonly bind co-operatively (Ermolenko and Makhatadze, 2002). The zinc fingers, similarly to NCp7, recognise G-rich single-stranded RNA (Desjardins et al., 2011); the simplest sequence identified in let-7 for the binding of lin28 via these domains is 'NGNNG', where N represents any nucleotide (Loughlin et al., 2012). The most common form by which this is seen to occur in let-7 RNA is in the sequence of 'GGAG', which was identified as a binding site by Heo et al (2009). Mutations in this site have shown reduced binding affinity, with mutations including the loss of the 'GG' and replacements of the 2nd and 3rd guanines; work has suggested that the final 'G' or the first two together produced the largest effect on binding, all supporting this proposed binding motif (Heo et al., 2009; Mayr et al., 2012; Newman et al., 2008). This motif is commonly found towards the end of the terminal loop, before the stem-loop junction (Figure 1.7A). There is more debate, however, surrounding a consensus sequence for a binding site for the CSD domain. A potential binding sequence was identified following crystallisation of the lin28-let-7 complex as 'NNGNGAYNN', where Y = pyrimidine (Nam et al., 2011). Interestingly, it was previously found that mutating only the adenine residue within this sequence in let-7g reduced binding affinity (Newman et al., 2008), lending strength to this motif. An alternative site has been proposed to be 'GNUNNUNNN' (Mayr et al., 2012); this was predominantly identified using a protein construct of the CSD only and may be less biologically representative. These sequences are typically found shortly after the miRNA stem region (Figure 1.7A). The work by the Sliz lab proposed that the lin28 protein would be capable of folding around the let-7 RNA and binding one molecule by the two different RNA-binding domains; however, they used a protein with a truncated linker region between the RNA-binding domains which required the binding of two protein molecules to each let-7, and may also not fully reflect the endogenous interaction between protein and miRNA. These papers have separately demonstrated that the prominent interaction between the RNA and lin28 proteins is occurring via base stacking, with hydrogen bonds providing further connections. Amino acids within the CSD and zinc fingers are capable of interacting with the nucleotides in the binding sites via pi-interactions; as these are largely non-specific this may account for the

degeneracy in proposed binding sequences (Loughlin et al., 2012; Mayr et al., 2012; Nam et al., 2011).

The mechanism by which lin28 inhibits the processing of let-7 is also under active discussion. It has been shown that the binding of lin28 both remodels the RNA to block the RNase cleavage site and drives uridylation of the miRNA. The binding of lin28 has been shown to remodel let-7 by partially melting the double-stranded RNA over the Dicer cleavage site, subsequently protecting the miRNA from such cleavage (Lightfoot et al., 2011; Mayr et al., 2012). The binding of the CSD to let-7 remodels the terminal loop, by wrapping the miRNA around a protruding section of the CSD, into a more open structure which also exposes the GGAG motif, for subsequent binding by the zinc finger domains (Figure 1.7B), and it is believed that this binding further impacts the processing (Mayr et al., 2012; Nam et al., 2011). The proximity of the 'GGAG' motif in let-7 miRNAs to the mature sequence and the Dicer cleavage site may also mean that the cleavage site is physically blocked when a lin28 molecule is bound to the miRNA (Nam et al., 2011), but another possible regulatory mechanism is by promoting uridylation. Work in ES cells demonstrated the binding of lin28 to let-7 recruits a noncanonical poly A polymerase, terminal uridylyltransferase 4 (TUT4, also known as ZCCHC11), which adds a uridine tail to the miRNA and thus targets it for degradation; TUT4 was not capable of binding to either lin28 or let-7 without the presence of the other (Heo et al., 2009). This firmly indicates that the binding of lin28 is crucial to this regulation, although it was shown that expression of both lin28 and TUT4 were necessary for the proper regulation of let-7 (Hagan et al., 2009). This work has also been confirmed in *C. elegans* (Lehrbach et al., 2009). Although reports have indicated that derivations from 'GGAG' were still capable of being bound by lin28 (Loughlin et al., 2012), they were not sufficient for uridylation (Heo et al., 2009), indicating some importance for this precise sequence. TUT4 is also a zinc finger protein (Minoda et al., 2006), and it has been suggested that of the 'GGAG' sequence, the second guanine residue is not bound by lin28, but may instead be exposed and available to be bound by TUT4 (Loughlin et al., 2012). TUT4 has also been implicated in the uridylation of another miRNA, miR-26a (Jones et al., 2009), but there has been no evidence that lin28 is involved in this interaction. A second factor that may additionally interact with lin28 to regulate let-7 is Musashi1, another RNA-binding protein; this is capable of binding to lin28 in the presence of RNA, and a knockdown of both proteins together was more effective at increasing levels of mature let-7 than a lin28 knockdown alone (Kawahara et al., 2011), although the process by how this modifies let-7 processing remains unknown.

The regulation of let-7 was shown to be a function of both lin28a and lin28b (Heo et al., 2009; Mayr et al., 2012; Piskounova et al., 2008) but a separate report has indicated that

lin28b does not act through TUT4 to inhibit let-7 biogenesis; it was shown that lin28b was present in the nucleus and nucleolus, and was acting independently of TUT4, blocking let-7 processing at the pri-miRNA stage (Piskounova et al., 2011). In a different cell type lin28b was still shown to bind let-7 with TUT4 (Heo et al., 2009), indicating a possible difference of regulation upon cell type. In *C. elegans* there is only one form of lin-28, which was found to predominantly bind to the pri-let-7 form (Van Wynsberghe et al., 2011), in contrast to earlier reports that indicated lin-28 primarily blocked Dicer processing (Lehrbach et al., 2009). The Pasquinelli group also found that within the nucleus, lin-28 was binding co-transcriptionally, allowing an early association with the let-7 transcripts that may allow the protein to successfully compete with the processing pathway (Van Wynsberghe et al., 2011).

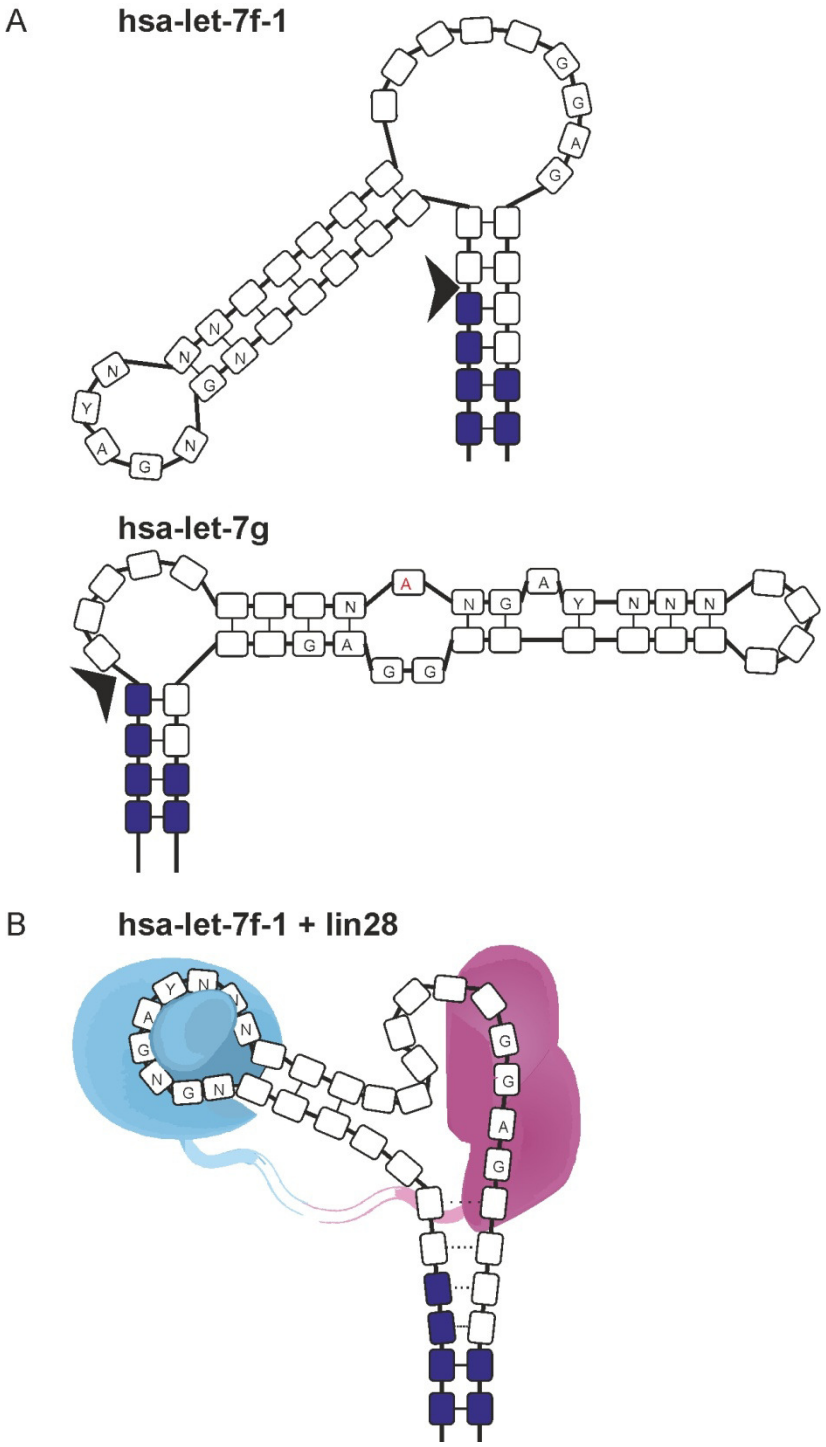


Figure 1.7 The structure of let-7 terminal loops and alteration upon lin28 binding

Schematic representation of let-7 terminal loop RNA. Boxes represent nucleotides, with blue shading indicating mature miRNA. Nucleotides given represent putative lin28 binding sites; those in red differ from proposed binding site. A) Loop regions of hsa-let-7f-1 and hsa-let-7g, to represent different possible structures. Arrow indicates Drosha cleavage site. B) Hsa-let-7f-1 miRNA loop bound to lin28, adapted from Nam et al (2011). Blue region represents CSD and purple regions represent the zinc finger domains, with flexible linker between. Dashed lines indicate proposed stem-melting.

1.4 Model systems

1.4.1 *Xenopus tropicalis*

X.tropicalis and *X. laevis* have been used as model systems for developmental studies for many years. A major advantage for the *Xenopus* species as a model system is the ease of manipulation and rapid development time for developmental studies. *X. tropicalis* is favoured as a genetic model due to its diploid genome, as opposed to the tetraploid *X. laevis*(Grainger, 2012). They have been used in recent years to characterise expression patterns of genes and identify previously unreported genes, amongst other uses (Lea et al., 2009; Sivak et al., 2005). To assess gene function, embryos can be induced to overexpress proteins by the injection of mRNA, and proteins can also be knocked down using anti-sense morpholinos (MOs), which can act in two different ways to prevent new protein production. Splice-blocker MOs target splice sites, and by binding over these prevent the processing of the mRNA to a translatable form. Translation-blockers bind somewhere between the 5' cap to ~25 bases 3' of the AUG translational start site, and block binding of ribosomes. MOs are known to persist, and effects can continue to be seen days after use in some systems, therefore they are a potent inhibitor of protein production in the *Xenopus* embryo. It is known that zygotic transcription begins in the embryo shortly after midblastula transition (MBT) (Newport and Kirschner, 1982), but RNA is present in the oocyte, having been maternally deposited; it is from this RNA only that proteins are synthesised in the earliest stages of amphibian development. This occurs just prior to gastrulation, when the genome is transcribed and inductive signalling events begin to occur, and patterning of the embryo takes place, making this a key event in the development of these organisms.

As much of the work on FGF function in gastrulation has already been carried out in *Xenopus* and is well characterised, as discussed earlier in this chapter, this is an ideal system to use to further study this stage of development with regards to new potential FGF targets. Additionally, attempts at generating *lin28a* knock outs in mice were lethal (Zhu et al., 2010), highlighting important embryonic processes for this gene. As *Xenopus* embryos develop externally, it is very easy to study changes occurring in the embryo at very early stages of development, such as gastrulation.

1.4.2 Human stem cells

Discoveries unearthed in vertebrate models such as *Xenopus* must also be investigated in humans to determine conservation and applicability of any findings. Carrying this out *in*

vivo in developing embryos is not possible, and *in vitro* cell culture models must instead be used.

The stem cells with the most potential are ES cells, isolated from the inner cell mass (ICM) of blastocysts (Evans and Kaufman, 1981; Martin, 1981). They are pluripotent and can develop into any type of cell, however are difficult to successfully culture long-term in this state. Additionally, as the understanding of pathways which control differentiation remains incomplete, driving development into required cell types can be difficult to achieve effectively and efficiently. Unlike early *Xenopus* embryos, the human zygotic genome is activated by the 4-8-cell stage (Braude et al., 1988; Tesařík et al., 1986), and thus transcription is active in mammalian ES cells. A simpler model of pluripotency can often be provided by using embryonal carcinoma (EC) cells, which are typically cell lines that have been isolated from germ line tumours. These are malignant transformed cells, and are much easier to maintain in culture, whilst expressing a similar set of genes to ES cells (Josephson et al., 2007). The major drawback, however, with EC cells is that very few lines are truly pluripotent - some, such as the 2102Ep line are nullipotent and incapable of differentiation despite being an EC line which shows a very highly similar transcriptome to ES cells, whereas others may be pre-disposed to particular lineages (Duran et al., 2001).

Adult stem cells offer another alternative for *in vitro* study. Multipotent in nature, they can be collected from different tissues of the body and show restrictions to germ layers and cell types. Mesenchymal stem cells (MSCs) are multipotent stem cells, isolated from primary sources such as bone marrow and adipose tissue. MSCs are already restricted to mesoderm, and they are capable of differentiating into diverse cell types such as bone, cartilage, fat, and muscle (Figure 1.8), although are not capable of haematopoiesis despite this also developing from mesoderm in the embryo. They can be used to model differentiation into these cell types, or to investigate functions within mesodermal cells, as they are intended for use in this thesis. As MSCs are derived from human patients, there can be a high degree in variability in the differentiation potential of cells. The cells obtained are a heterogeneous population, and due to the variability are best used to analyse for gross trends. It has also been shown that the site from which cells are derived can affect the potential for different tissue types (Strioga et al., 2012).

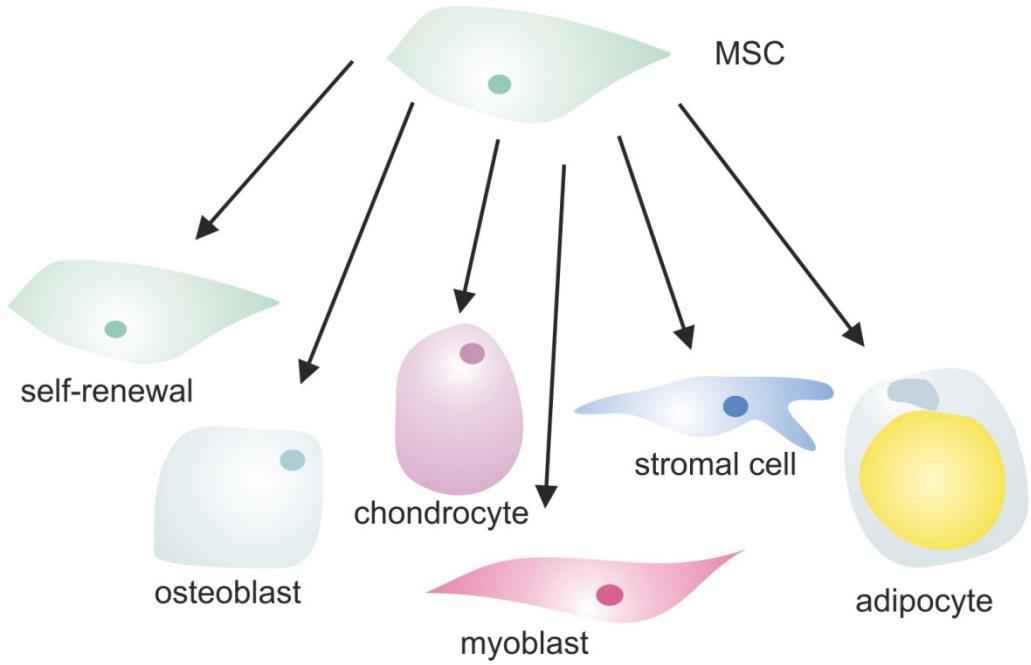


Figure 1.8 Mesenchymal stem cell potential
When dividing, MSCs are capable of self-renewal into more MSCs, or differentiation into cell types from mesodermal lineages such as osteocytes for bone development, chondrocytes for cartilage, myoblast cells to form muscle, connective stromal cells, or adipocytes for fat differentiation.

1.5 Aims for this thesis

This thesis aims to explore the potential of *lin28a* and its related gene *lin28b* as targets of FGF signalling. It is hoped that in exploring further the nature of this regulation, a closer move can be made to identifying the functions of FGF signalling in development, and whether this is positively regulating pluripotency through such factors. As little is known about the functions of lin28a and lin28b in development, the functions of these genes will be investigated; it is known that they are required for embryonic development but not what roles they are playing at this time. The inability to generate viable mammalian lin28-null specimens highlights this importance, but the difficulty of studying processes in mammalian embryonic development leads us to turn to the *Xenopus* model, where development and the influence of lin28 proteins can be easily studied across embryonic stages.

The interest in miRNAs is continually developing, and lin28 was one of the first discovered proteins to post-transcriptionally regulate miRNA. The only confirmed target so far has been the let-7 family, although as lin28 is a potent RNA-binding protein, it is possible that targeted binding sites are present in alternative miRNAs. let-7 is usually transcriptionally repressed at the earliest stages of development, and the investigation will aim to discover

if there are any other putative lin28-regulated miRNAs expressed during early development.

In order to begin the investigation into the conservation of these pathways in humans, this thesis aims to identify if MSCs can be used to model these early effects of FGF signalling upon undifferentiated mesoderm cells, and if lin28 is a developmentally relevant gene in these cells. For any identified targets of lin28, MSCs or, if these cells are not a suitable model, an alternative cell line model will be explored to determine their suitability for use in investigating conservation of identified targets.

Chapter 2. Materials and Methods

2.1 *Xenopus* techniques

2.1.1 *Xenopus tropicalis in vitro* fertilisation and embryo culture

Female *X.tropicalis* were primed by subcutaneous injection of 10 units of human chorionic gonadotrophin (hCG) (Chorulon; Intervet) 24-72 hours prior to laying. To induce ovum laying in 2.5-4 hours, the females were injected with 100 units of hCG, and kept at 27°C in the dark. Eggs were fertilised with a suspension of crushed testis in L15 with 10% heat treated foetal calf serum, collected from a male that had been treated with 100 units of hCG 3-4 hours before fertilisation. Embryos were cultured in MRS/9 ($1/9^{th}$ Modified Ringer's solution: 0.1 M NaCl, 1.8 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 5 mM HEPES-NaOH) on 1% agarose-coated (Sigma) 60 mm dishes (VWR) at 23-27°C. Embryos were treated prior to the first cleavage to remove jelly coats with 3% L-cysteine (Sigma) in MRS/9, pH 7.8-8. Before gastrulation, embryos were transferred to MRS/20 with 100 µg/ml gentamicin.

2.1.2 *Xenopus laevis in vitro* fertilisation and embryo culture

X. laevis females were primed by subcutaneous injection of 50 units of hCG 48 hours to 1 month before use. To induce laying in 16 hours, females were injected with 250-350 units of hCG and kept in the dark at 19°C. Eggs were fertilised with a suspension of freshly crushed testis in distilled water, collected from a male. Embryos were cultured in NAM/10 ($1/10^{th}$ Normal Amphibian Medium: 110 mM NaCl, 2 mM KCl, 1 mM $Ca(NO_3)_2$, 0.1 M EDTA, with 5 mM HEPES and 2.5 µg/ml gentamicin added to dilutions) throughout development at 12-24°C. Embryos were treated to remove jelly coats prior to first cleavage with 2.5% L-cysteine hydrochloride monohydrate (Sigma) in NAM, pH 7.8-8.

2.1.3 *Xenopus* injections

Injections of RNA were performed on both *X. laevis* and *X. tropicalis* embryos using pulled glass needles (Narishige) using a pneumatic microinjector (Harvard Apparatus). *X.tropicalis* embryos were transferred to MRS/9 + 3% Ficoll (Sigma) for injection, then to MRS/20 before gastrulation. *X. laevis* embryos were transferred to NAM/3 + 5% Ficoll (Sigma) for injection, and transferred to NAM/10 before gastrulation.

Injections were performed with *in vitro* transcribed mRNA (section 2.3.12), antisense morpholino oligonucleotides (MOs) (Gene Tools, LLC) or synthetic precursor miRNAs (Applied Biosystems), at required concentrations. Injections were carried out into all cells at either the two- or four-cell stage, with a maximum of 10 nl/embryo. *X. tropicalis* injections were targeted to the marginal zone. MOs were heated to 65°C for 3 minutes prior to needle-loading and injection. MOs were designed by Gene Tools, LLC, and sequences for those used in this project are given below (Table 2.1).

Table 2.1 Sequences for antisense morpholinos

Gene	MO sequence
Standard control	CCTCCTACCTCAGTTACAATTTATA
lin28a-1	GGTCTGCCTTGAAGTTGTCCCAGCT
lin28a-2	GAGTCTCTTCTATCTGAGGTCAGGC
lin28b	TCCTTCGGCCATGATGCCTCCTGCT

All MOs used in this study were obtained from GeneTools, LLC. lin28a MOs were targeted to the 5' UTR region of the gene, and lin28b MO to the region around the start site of translation (underlined). Sequences are given 5' to 3'.

2.1.4 Collection of embryos

Embryos were staged according to Nieuwkoop and Faber (NF) (Nieuwkoop and Faber, 1994). Embryos were collected for RNA and protein extraction by removing excess liquid and snap freezing on dry ice. For whole-mount *in situ* hybridisation, membranes were removed from embryos pre-NF stage 25 using forceps under a microscope. Embryos developed to the required NF stage were fixed in MEMFA (0.1 M MOPS, pH 7.4, 2 mM EGTA, 1 mM $MgSO_4$, 3.7% formaldehyde) for 1 hour at room temperature. Embryos were subsequently washed in methanol three times, before being stored in fresh methanol at -20°C. Embryos were fixed in the same manner for dpERK whole-mount immunostaining, but without prior membrane removal.

2.1.5 Photographing embryos

Embryos were photographed using a SPOT 14.2 Color Mosaic camera (Diagnostic Instruments Inc.) and SPOT Advanced software, on a Leica MZ FLIII microscope. Images were processed using Adobe Photoshop Elements 4.0.

2.2 Cell techniques

2.2.1 Isolation of MSCs from femoral heads

Femoral heads were collected from patients undergoing hip replacement surgery at Harrogate District Hospital, Harrogate, and Clifton Park Treatment Centre, York, following informed consent. Cells collected from donors were cultured separately from other donor cells, and cell lines were numbered sequentially using the prefix 'FH', for femoral head.

Trabecular bone was scraped from inside the femoral heads into Dulbecco's Modified Eagle's Medium (DMEM) with 10% foetal bovine serum (FBS). Trabecular bone was minced with scissors and supernatant removed to a new tube after fragments had settled. This was repeated two further times then supernatants were centrifuged at 500*g* for 5 minutes and the pellet resuspended in 16 ml of culture medium. To remove large debris, the suspension was passed through a 70 µm cell strainer, before layering over 12 ml of Ficoll-Paque Plus (Amersham Biosciences). The layered suspension was centrifuged at 350*g* for 30 minutes. The mononuclear cell fraction was harvested, and washed in 10 ml Buffer A (PBS, 0.2% bovine serum albumin (BSA), 5 mM EDTA) and seeded into a 75 cm^2 flask with normal growth medium.

2.2.2 MSC cell culture

MSCs were cultured in DMEM with 10% FBS (batch tested) and 1 U/ml penicillin/streptomycin (P/S) (Invitrogen) at 37°C with 5% CO_2. Medium was changed every 3-4 days. Cells were passaged using Trypsin-EDTA (0.05% Trypsin, 0.02% EDTA, Invitrogen) at ~80% confluency, and reseeded at a ratio of 1:3.

2.2.3 EC cell culture

Two lines of EC cells were grown, 2102Ep cell line and NTera2.D1 clonal line. Cells were cultured in DMEM with 10% FBS at 37°C with 5% CO_2. Media was changed every 2-3 days. Cells were passaged using Trypsin-EDTA (Invitrogen) at ~80% confluency, and reseeded at a ratio of 1:6.

2.2.4 FGF treatment of cells

MSCs were seeded in a confluent layer at a density of $5.26x10^4$ cells/cm^2 in 6-well plates (Corning) in DMEM with 10% FBS, 1% P/S. As FBS contains growth factors that could

mask effects from FGF signalling, cells were serum-starved before treatment to ensure a minimal presence of alternative signalling factors. 20 hours after seeding, the cells were serum-starved in DMEM, 0.5% FBS, 1% P/S. Cells were treated with a range of FGF-2 (Invitrogen) concentrations (1, 10, 25 ng/ml) in 3ml media (DMEM, 0.5% FBS, 1% P/S). At the relevant time points, medium was removed and cells were washed in ice cold 1x PBS, before being harvested in lysis buffer according to the subsequent protocol. Samples were collected at 6, 12, and 24 hours for RNA analysis; and 5, 15, 30 and 60 minutes for protein analysis.

2.2.5 ES cell samples

ES cell samples were kindly gifted by the Coles group (University of York); H7 and H9 cell lines, isolated by Thomson and colleagues (1998) were cultured in feeder-free conditions on Matrigel, collected in cell pellets and treated with lysis buffer for RNA extraction.

2.3 Molecular biology

2.3.1 Agarose gel electrophoresis

DNA and RNA samples were run on 0.8-2% agarose gels in TAE buffer (4 mM Tris acetate, 1 mM EDTA), stained with ethidium bromide or SYBR safe (Invitrogen) at 100-150 V. Samples were loaded with a 1:6 ratio of loading buffer (Promega) and run alongside a ladder: Hyperladder IV or V (bioline), 1Kb ladder (Promega), or Q-Step 4 ladder (YorkBio).

2.3.2 Quantification of nucleic acids

DNA and RNA samples were analysed on a Nanodrop ND-1000 spectrophotometer which measured the absorbance of samples at 260 nm, to determine nucleic acid concentration. 260/280 and 260/230 ratios were also measured to determine purity of nucleic acid samples.

2.3.3 mRNA extraction and cDNA synthesis

Total RNA extraction from human cells

To harvest cells, medium was removed and cells washed with 1x PBS. Cells were lysed directly in the culture dish with 1 ml of TRIzol (Invitrogen) and incubated for 5 minutes at room temperature. RNA was extracted following the manufacturer's instructions. Lysed cells were scraped into a tube and 200 µl of chloroform was added; samples were

vortexed vigorously for 30 seconds and incubated for 5 minutes at room temperature, followed by centrifugation at 13,000*g* for 20 minutes at 4°C. The upper aqueous phase was removed to a new tube, and 1 volume of isopropanol was added to precipitate the RNA with a 30 minute incubation at 4°C. Samples were centrifuged at 13,000*g* for 15 minutes at 4°C, and isopropanol removed. The pellet was washed in 70% ethanol, vortexed, and centrifuged for 5 minutes at 13,000*g*. Samples were then dried and resuspended in 50 µl of RNase-free water.

Total RNA extraction from *Xenopus*

Total RNA was extracted using TRI reagent (Sigma) according to the manufacturer's instructions with some modifications. 1 ml of TRI Reagent (Sigma) was added to samples on ice, which were homogenised by a pipette tip and incubated at room temperature for 5 minutes. The samples were centrifuged for 10 minutes at 4°C at 13,000*g*, and the supernatant removed to a fresh tube. 200 µl of chloroform was added and the remainder of the protocol was as described for human cells. Following resuspension in WATER, an additional precipitation step was undertaken. 7.5 M LiCl and 0.05 M EDTA (final concentrations) were added, and RNA was incubated at -80°C overnight. Samples were centrifuged as before, for 20 minutes, and the resulting pellet was ethanol washed and resuspended in 20 µl RNase-free water.

cDNA synthesis

cDNA was synthesised from total RNA using 1 µg RNA with either oligo (dT) (Invitrogen) or random hexamers (Invitrogen) using SuperScript II Reverse Transcriptase (Invitrogen) according to manufacturer's instructions. Random hexamers were used with *Xenopus* timecourse samples where any stages were pre-MBT, with oligo(dT) used in all other cases. RNA, primers and dNTPs were incubated at 65°C for 5 minutes for denaturation. First strand buffer and dithiothreitol (DTT) were added and reactions heated before the addition of 1 µl Superscript II; for oligo d(T) primers the reaction was heated to 42°C for 2 minutes, and for random hexamers the mixture was heated to 25°C for 2 minutes. Final reactions were incubated at 42°C for 50 minutes then 15 minutes at 70°C to inactivate reverse transcriptase. With random hexamers there was an initial incubation of 25°C for 10 minutes preceding this. cDNA was diluted accordingly in nuclease-free water and stored at -20°C.

2.3.4 Reverse transcriptase PCR (RT-PCR)

cDNA was diluted 1/5 for use in RT-PCR reactions. 25 µl reactions were set up with 12.5 µl 2x PCR Master Mix (Promega), 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 2.5 µl cDNA, and RNase-free water to 25 µl.

Samples were run in the following cycling conditions, with annealing temperature (AT) and elongation time (ET) given in the table below (Table 2.2) for each primer set.

94°C - 5 minutes
94°C - 20 seconds
AT°C - 30 seconds
72°C - ET
} 25 cycles for L8/GAPDH, 30 cycles for FGFR1/FGFR2
72°C - 5 minutes;

Primer sequences are given below (Table 2.2), which were designed across exon boundaries. Primers for FGFR1 and FGFR2 were designed to target multiple splice isoforms. PCR products were run out on an agarose gel as in 2.3.1.

Table 2.2 List of RT-PCR primers

Gene	Forward primer	Reverse primer	AT	ET
***GAPDH* (human)**	GGTGAAGGTCGGWGTCAACGG	GGTCATGAGYCCTTCCACGAT	54°C	1m
***FGFR1* (human)**	GAAGTGCCTCCTCTTCTGGG	ATCACTGCCGGCCTCTCTTC	62°C	1m30s
***FGFR2* (human)**	TGGTCACCATGGCAACCTTG	TTCATTCGGCACAGGATGAC	61°C	1m30s
***L8* (*Xenopus*)**	GGGCTRTCGACTTYGCTGAA	ATACGACCACCWCCAGCAAC	62.5°C	30s
***miR-17-92 cluster* (*Xenopus*)**	TGCAGTGAAGGCACTTGTAG	TAAACAGGCCGGGACAAG	55°C	55s
***miR-106-363 cluster* (*Xenopus*)**	TGCTGGACACCTGTACT	TTCTGCGGTTTACAGATGGA	55°C	55s

Sequences are given 5' to 3'. AT = Annealing temperature, ET = elongation time, where m = minutes and s = seconds.

2.3.5 mRNA real-time quantitative PCR (qRT-PCR)

All cDNA for qPCR was checked prior to use by RT-PCR for housekeeping genes (*Xenopus*: *L8*, human cells: *GAPDH*).

Primers were designed using Primer Express II (Applied Biosystems) software. Amplified sequences were checked by a BLAST search, to determine specificity. Primer sequences are found below for genes analysed in *Xenopus* (Table 2.3) and humans (Table 2.4).

Table 2.3 *Xenopus* qPCR primer sequences

Gene	Forward primer	Reverse primer
brachyury	CGGCACACACTGGGATGTT	CCACAAGGAAGGGTACTGACTTG
chordin	GGATTGCTCTCGCCCAAA	AATTTTCCACCAAACGTACAACCT
lin28a	CCAAAGTCCAAACCACATGGT	AGATTGGCTGTTCTTCTAAGTTAGCA
lin28b	AAAGGTTTTGAGTCATTACGGGTAA	TATCTCCCTTTGGTTTTCTTTTTTGA
ODC	AAAGCTTGTTCTGCGCATAGCAACT	AGGGTGGCACCAAATTTTAC
sox17b	ACGAAGAAAGAGGCTGGATCTG	GCCCACACCATAAACGCATT
sox3	CAGAGTAGCAGACTGCACAGTGTACA	AGTGAGCGGTACAGTGCCATT

Sequences for all qPCR primers for mRNA genes analysed in X.tropicalis. Sequences are given 5' to 3'.

Table 2.4 Human qPCR primer sequences

Gene	Forward primer	Reverse primer
DUSP6	AAGGTGGCTTCAGTAAGTTCCAA	CTGCTACACGAGCCGTCTAGATT
IL-6	CCGGGAACGAAAGAGAAGCT	GCGCTTGTGGAGAAGGAGTT
lin28a	CCCCCCAGTGGATGTCTTT	CCGGAACCCTTCCATGTG
lin28b	CCATGATAAACCGAGAGGGA	GGCTTCTAAATCCTTCCATGA
RPS27A	TGGATGAGAATGGCAAAATTAGTC	CACCCCAGCACCACATTCA
sprouty2	CACTCGCAGGTCCATTCTTCT	CCCTGAGCTGACCGTGCTT

Sequences for all qPCR primers for mRNA genes analysed in humans. Sequences are given 5' to 3'.

3 µl of cDNA was used per reaction, at 1:3 dilutions for *lin28a* and *lin28b* in *Xenopus*, and 1:5 dilutions for all other genes. Reactions were carried out in triplicate per sample in 96-well plates using SYBR green (Applied Biosystems). The reaction was set up with 12.5 µl 2x SYBR Green PCR Master Mix, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 3 µl cDNA, and made up to 25 µl with RNase-free water.

qRT-PCR was carried out on an ABI Prism 7000 Sequence Detection System (Applied Biosystems). Thermal cycling was undertaken at 50°C for 2 minutes, 95°C for 10 minutes, and followed by 40 cycles of 95°C for 15 seconds and 50°C for 1 minute. Data was analysed using ABI 7000 System software (Applied Biosystems), and relative expression levels of each gene calculated using the $2^{-\Delta\Delta Ct}$ method. The Ct value for each gene was normalised to the Ct value of the housekeeping gene (*Xenopus*: *ODC*, human cells: *RPS27a*), and then calibrated to the control sample for each experiment.

Primers were initially optimised using cDNA which was diluted 10-fold in four serial dilutions. Samples were run at 50°C for 2 minutes, then 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 50°C for 1 minute. To check for a single product, a dissociation step was run of 95°C for 15 seconds, 60°C for 20 seconds and 95°C for 15 seconds. Resultant Ct values from the four dilutions were used to create a standard curve. Primers were further used if efficiency was > 97%, and had an R_2 value of close to 1. cDNA dilution was then chosen based on the dilution with a Ct value close to 25.

2.3.6 miRNA extraction, reverse transcription and qRT-PCR

Total RNA extraction with miRNA enrichment from human cells

Samples to be used for miRNA analysis were isolated using the miRVana miRNA isolation kit (Applied Biosystems). The protocol was carried out according to manufacturer's instructions. Cells were lysed in lysis buffer and collected; 1/10th volume of homogenate additive was added and samples vortexed before being incubated on ice for 10 minutes. 1 volume of acid phenol-chloroform was added, then samples were vortexed and centrifuged for 5 minutes at 13,000*g* at room temperature. The upper aqueous phase was removed to a new tube and 1 volume of ethanol added. Samples were centrifuged through a filter cartridge, which was then washed with miRNA wash solutions. Samples were eluted in 100 µl RNase-free WATER which was pre-heated to 95°C.

Total RNA extraction with miRNA enrichment from *Xenopus*

The protocol was carried out as above, with slight modifications; after lysing the embryos in lysis buffer and addition of homogenate additive, samples were centrifuged for 10 minutes at 4°C and supernatant removed to a fresh tube. The rest of the procedure was as for human cells, with the addition of acid phenol-chloroform.

miRNA cDNA synthesis and qRT-PCR

This protocol is summarised schematically in Figure 2.1. cDNA was synthesised from 10ng RNA/RT reaction with miRNA-specific primers for TaqMan assays (Applied Biosystems) using the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems) as manufacturer's instructions. Sample RNA, reverse transcriptase, buffer, and RNasin were multiplexed and divided into separate RT reactions for each miRNA gene, with the specific RT-primer added (Figure 2.1), to ensure even quantities of RNA template for

each gene. Reactions were incubated at 16°C for 30 minutes, followed by 42°C for 30 minutes and 85°C for 5 minutes to deactivate reverse transcriptase.

qRT-PCR was carried out using TaqMan Universal Master Mix II (Applied Biosystems) with Taqman miRNA probes (Applied Biosystems) according to manufacturer's instructions (Figure 2.1). All reactions were performed in quadruplicate per sample in 96-well plates on an ABI Prism 7000 detection system (Applied Biosystems) with thermal cycling at 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Gene expression levels were normalised to snoRNA U6 using the $2^{-\Delta\Delta Ct}$ method as described in section 2.3.5.

Assays used were: hsa-miR-19b, hsa-miR-17-5p, hsa-miR-130b, hsa-miR-20b, hsa-miR-363#, hsa-miR-363, hsa-miR-18b, hsa-let-7a, hsa-let-7f, custom xtr-miR-106, custom xtr-let-7g, and U6 snRNA (all Applied Biosystems).

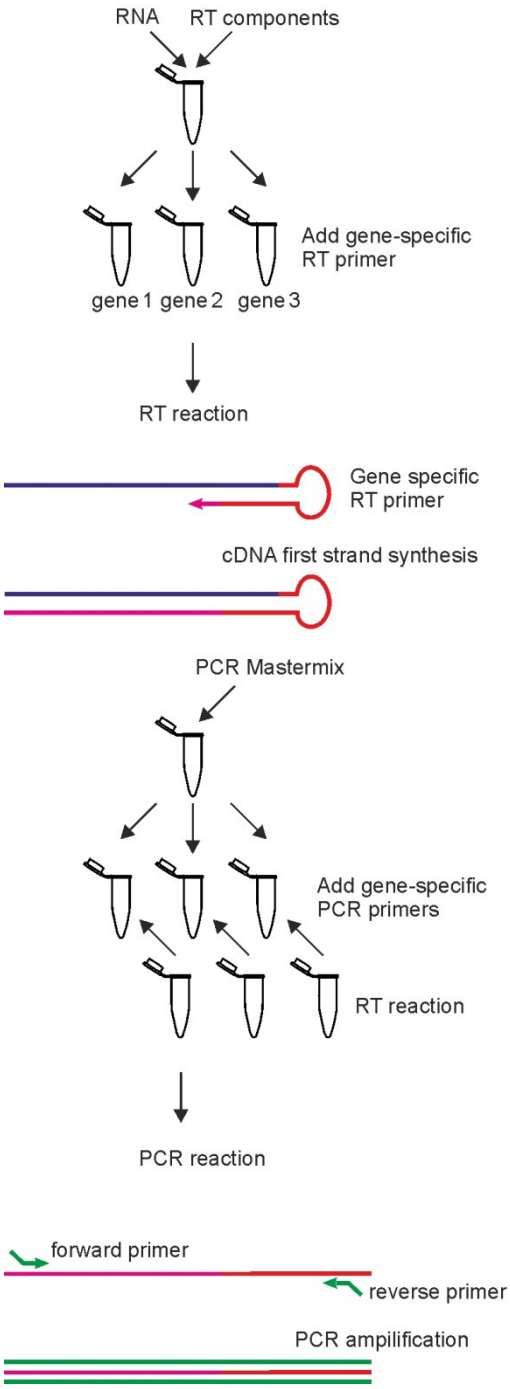


Figure 2.1 miRNA RT and qPCR set-up

Pictorial representation of set-up for miRNA RT and PCR, with multiplexing of reagents and RNA to ensure even divisions of all reaction components. Coloured lines represent different reaction components: blue = RNA, red = gene specific RT-primer, pink = cDNA extension, green = PCR primers, with non-complementary overhang to increase length of PCR product; these are then extended to result in the PCR products.

2.3.7 Transformations

Plasmids were transformed into DH5α competent cells. 5 µl of ligation reaction was added to 50 µl of competent cells and incubated on ice for 30 minutes; cells were heat shocked at 42°C for 90 seconds, and then chilled on ice before being incubated for 1 hour at 37°C with shaking in 1 ml of liquid broth (LB). After incubation, transformed cells were plated onto LB with 1 µg/ml ampicillin and incubated overnight at 37°C.

2.3.8 DNA minipreps

Plasmid DNA was prepared from colonies by inoculating a 5ml culture of LB with 1 µg/ml ampicillin and growing overnight at 37°C with shaking. Preparations were centrifuged at 7000*g* for 5 minutes to pellet the DNA, and extracted using a QIAprep Spin Miniprep kit (Qiagen).

DNA was cleaned up by adding an equal volume of phenol-choloform to samples and vortexing for one minute. Samples were centrifuged for 5 minutes at 13,000*g* at 4°C, and the top aqueous layer removed to a fresh tube; 0.1 volumes 3 M sodium acetate (Sigma) and 2.5 volumes 100% ethanol were added and the solution precipitated overnight at -20°C. Samples were centrifuged for 15 minutes as before, and the resulting pellet was washed in 70% ethanol, centrifuged for 5 minutes and air-dried. The DNA was resuspended in an appropriate volume of nuclease-free water and stored at -20°C.

2.3.9 Cloning of pri-miR-106 cluster into pCS2+ vector

The full sequence for the pri-miR-106 cluster was amplified from *X. tropicalis* genomic DNA using the following primers:
F: GCGUUGUGCAAAAGUGCUUA R: TTGTGATGCTCAGCCTTCAG

The reaction was set up in 50 µl with 100 ng *X. tropicalis* genomic DNA, 5 µl 10x Pfu polymerase buffer, 1 µl 10 mM dNTPs, 1 µl 10 µM forward primer, 1 µl 10 µM reverse primer, 0.5 µl Pfu polymerase (Promega), and made up to volume with RNase-free water.

The PCR programme was carried out as:

94°C - 5 minutes

94°C - 30 seconds
57°C - 30 seconds
72°C - 1 minute
} 30 cycles

72°C - 15 minutes; with the addition of 1 µl *Taq* polymerase (Invitrogen) for the final 10 minute extension to add an 'A' overhang, making the products compatible with the 'T' overhang in the pGEM vector.

PCR products were cleaned up as in 2.3.8 and ligated to the pGEM-T Easy Vector System (Promega). Ligation reactions were set up with 2 µl PCR product/control insert, 5 µl 2x buffer, 1 µl PGEM T-easy vector, 1 µl T4 Ligase (Promega) and 1 µl RNase-free WATER. Reactions were incubated at room temperature for 1 hour, and overnight at 4°C.

Ligations were transformed into DH5α competent cells as in 2.3.7. Colony PCR screens were performed to identify colonies containing an insert of the correct size. This was carried out as in 2.3.4 using primers to T7 and SP6 within the plasmid, as listed below:
SP6: AGGTGACACTATAGAATACTCA T7: GTA ATACGACTCACTATAGGGC

Colonies containing the correct insert were then selected for minipreps as carried out in 2.3.8, and subsequently sequenced in the Technology facility using the primers below, to check for sequence mutations:
SP6: CGATTTAGGTGACACTATAG T7: TAATACGACTCACTATAGGG

To subclone into pCS2+ plasmid, the pri-miR-106 cluster DNA insert was cut out of pGEM using EcoR1 and incubated at 37°C for 2 hours. The reaction was set up in 100 µl with 10 µl pGEM-pri-miR-106 cluster, 10 µl 10x H buffer (Roche), 2 µl (100 units) EcoR1 (Promega), 78 µl RNase-free WATER. The digest was checked on a 1% agarose gel, and then the full reaction run on a 0.8% pure agarose gel. The band for the insert was cut from the gel and DNA purified using Qiagen gel extraction kit (Qiagen) according to manufacturer's instructions, and eluted in 30µl of nuclease-free water.

1 µg of pCS2+ was linearised using EcoR1 in 100 µl volume as the reaction for pGEM plasmid above. The linearised vector was checked on a 1% agarose gel and was then treated with 1 µl of calf intestinal alkaline phosphatase (CIAP) for 5 minutes at 37°C, to prevent the vector ligating with itself. CIAP activity was stopped by incubation at 65°C for 15 minutes, and was cleaned up using Qiagen clean-up kit according to manufacturer's instructions.

Excised DNA insert was incubated with linearised pCS2+, with the insert in 3 times excess in the following reaction; 0.5 µl pCS2+ (CIAP-treated), 1.5 µl 10x ligase buffer, 12 µl DNA pri-miR-106 cluster insert and 1 µl T4 ligase (Promega). Ligation reactions were incubated at room temperature for 1 hour and at 4°C overnight, and transformed as 2.3.7. Colony PCRs and minipreps were carried out as in pGEM vector to check for correct insert size, and sequenced by the Technology Facility as above, but with the T7 primer: TGTCTGGATCTACGTAATACG.

2.3.10 Cloning of truncated pri-miR-17-92 clusters into pGEM vector

To generate whole-mount *in situ* hybridisation probes to use on the miR-17-92 and miR-106-363 clusters, truncated sections of the pri-RNA were cloned into a vector. Sequences to be cloned were produced using PCR from *X.tropicalis* genomic DNA and inserted into plasmid vector pGEM®-T Easy (Promega). PCR reaction was carried out using 100 ng genomic *X.tropicalis* DNA, and set up in 50 µl reactions as 2.3.4. Primer sequences, annealing temperatures and extension times are given in the table below (Table 2.5). PCR products were cleaned up as in 2.3.8, and ligation reactions, transformations, screening and sequencing as in 2.3.9.

Table 2.5 Primers used for cloning

Gene	Forward primer	Reverse primer	AT	ET
mir-17-92 cluster	TGCAGTGAAGGCACTTGTAG	TAAACAGGCCGGGACAAG	55°C	55s
mir-106-363 cluster	TGCTGGACACCTGTACT	TTCTGCGGTTTACAGATGGA	55°C	55s

Sequences are given 5' to 3'. AT = annealing temperature, ET = elongation time, where s = seconds.

2.3.11 Linearisation of plasmids

Plasmids must be linearised to allow susbsequent RNA transcription. Plasmids were linearised in 100 µl reactions; 5 µg DNA was incubated with 3 µl enzyme and 10 µl 10x buffer, and incubated for 2 hours at 37°C. Restriction enzymes used were dependent on insert and plasmid (Table 2.6). Digests were run on a 2% agarose gel to check linearisation was completed, and templates were cleaned up by phenol-chloroform extraction and sodium acetate precipitation.

Table 2.6 Details of plasmids used for functional RNA transcription

Plasmid	Linearisation enzyme	Origin
pCS2+dnFGFR4	Sal1	gift from Harumasa Okamoto (Hongo et al., 1999)
pCS2+lin28a1CR	Not1	Cloned by L.Faas from EST
pCS2+lin28a2CR	Not1	Cloned by L.Faas from EST
pCS2+lin28bCR	Not1	Cloned by L.Faas from EST

2.3.12 *In vitro* transcription of functional mRNA

Sense mRNA for injection was transcribed from linearised templates. Functional mRNA was transcribed using MEGAscript® SP6 transcription kit (Ambion) following manufacturer's instructions, with methylated GTP cap analog (Ambion) to provide capped RNA, in a 20 µl reaction consisting of 2 µl 50 mM ATP, 2 µl 50 mM CTP, 2 µl 50 mM UTP, 2 µl 5 mM GTP, 2.5 µl 40 mM G' cap, 2 µl 10x transcription buffer, 1 µl linearised template (1 µg/µl), 1 µl SP6 enzyme and 4.5 µl RNase-free water.

RNA was checked on a 2% agarose gel, and RQ DNase (Promega) was added for 2 minutes at 37°C to remove template DNA. RNA was then diluted with 115 µl water and 15 µl 3 M ammonium acetate. One volume of phenol-chloroform was added, and the sample was vortexed and centrifuged for 5 minutes at 13,000*g*. The aqueous phase was removed to a new tube, and one volume of chloroform was added. The sample was vortexed, centrifuged and the aqueous layer removed as before. One volume of isopropanol was added and RNA was left on dry ice for 30 minutes to precipitate. Samples were then centrifuged for 15 minutes at 13,000*g* and the pellet washed in 70% ethanol. After being dried, the RNA was resuspended in 20 µl RNase-free water and stored at -80°C.

2.3.13 *In vitro* transcription of pre-miRNA

Pre-let-7g and pre-mir-363 were synthesised *in vitro*. DNA templates from which to synthesise RNA were created from PCR products, with an SP6 site inserted at the beginning of the sequence. PCRs were undertaken from genomic DNA for pre-let-7g and from a pri-miR-106-363 cluster plasmid for pre-mir-363, generated in 2.3.9. Primers are given in Table 2.7 and PCR conditions are given below:

94°C – 3 minutes
94°C – 20 seconds
57°C – 20 seconds } 22 cycles
72°C – 10 seconds
72°C – 5 minutes

PCR products were cleaned up using a phenol-chloroform and precipitation procedure. RNA was synthesised using SP6 Megascript kit (Ambion) according to manufacturer's instructions, without using the G' cap as used for mRNA, and cleaned up as in 2.3.12.

Table 2.7 Primer sequences to synthesis templates for *in vitro* transcription of pre-miRNAs

Gene	Forward primer	Reverse primer
Pre-let-7g	ATTTAGGTGACACTATAGTTTTGTTGTTCGCGG	TTCTGCGGTTTACAGATGGATACCG
Pre-mir-363	ATTTAGGTGACACTATAGGGCTGAGGTAGTTGTTT	TAGGCAAGGCAGTGGCCTGTACAG

Sequences in grey corresponds to SP6 promoter recognition site. Remaining primer sequences cover the 3' and 5' ends of the pre-miRNA. Sequences are given 5' to 3'.

2.3.14 *In vitro* transcription of digoxigenin labelled antisense RNA

Antisense RNA was synthesised with digoxigenin (DIG) labelling for whole-mount *in situ* hybridisation. Plasmid templates were linearised as in 0. Synthesis was carried out using 10x DIG labelling mix (Roche Diagnostics). Details of linearisation enzyme and RNA polymerase are listed in Table 2.8.

Table 2.8 Clone information, restriction enzyme and polymerases used for *in situ* probes

Name	Restriction Enzyme	Polymerase for ISH probe	Reference/Source
pGEMmiR-17-92 cluster	Nsi1	T7	Cloned by F. Warrander
pGEMmiR-106-363 cluster	Nco1	SP6	Cloned by F. Warrander
pCS2+lin28a1CR	BamH1	T3	Subcloned by L. Faas from TNeu 101a16
pCS2+lin28bCR	BamH1	T3	Subcloned by L. Faas from TNeu104b13
pCMVFGF8 (pCMV-SPORT6.1)	Nco1	T7	Image:6983047

For *in vitro* transcription, reactions were set up with 10 µl 5x transcription buffer (Promega), 2.5 µl 10x DIG labelling mix (Roche), 5 µl 100 mM DTT, 2 µl (50 units) RNase inhibitor (Promega), 3 µl (150 units) SP6/T7/T3 polymerase as appropriate (Promega), 1-2 µg linear DNA template, with RNase-free water up to 50 µl.

The reaction was incubated initially at 37°C for 2 hours, with a further 2 hours after adding an additional 1 µl of the appropriate polymerase. 2 µl of the reaction was checked on an agarose gel, and template was then removed by incubating with 1 µl RQ1 DNase (Promega) for 5 minutes at 37°C. After re-checking on a gel, the probe was cleaned up and precipitated by adding 50 µl water, 25 µl 10 M NH_4OAc and 312.5 µl 100% ethanol and left on dry ice for 1 hour. Probes were centrifuged for 15 minutes at 13,000*g*, and the pellet was washed with 70% ethanol before being dried and resuspensed in 50 µl nuclease-free water.

2.3.15 Whole-mount *in situ* hybridisation for mRNA

Probes to be used were *in vitro* transcribed with DIG-labelling as in section 2.3.14. The method used was a modification of Harland (1991). Fixed embryos were allowed to warm to room temperature slowly at ambient temperature. Embryos were then washed for 10 minutes in 75% methanol/PBST (1x PBS, 0.01% Tween-20), followed by 10 minutes with 50% methanol/50% PBST, and then three times for 5 minutes in PBST. Embryos were treated with Proteinase K (Roche) at 10 µg/ml in PBST at room temperature for a

time dependent on stage. Embryos were then rinsed twice for 5 minutes in 0.1 M triethanolamine (pH7.8); to the second of these washes 12.5 µl acetic anhydride was added and after 5 minutes another 12.5 µl was added, for a further 5 minutes. Samples were washed twice in PBST and refixed in 3.7% formaldehyde for 20 minutes, which was followed by five further washes of 5 minutes in PBST. Embryos were pre-hybridised at 60°C for 10 minutes in pre-hybridisation buffer (50% formamide (Ambion), 5x SSC, 100 µg/ml heparin, 1x Denhart's, 0.1% Tween-20, 0.1% CHAPS (Sigma), 10 mM EDTA) and then for a further 2 hours at 60°C in hybridisation buffer containing 1 mg/ml yeast RNA. The embryos were incubated overnight at 60°C with hybridisation buffer with DIG-labelled probe. DIG-probes were heated to 80°C for 3 minutes before being added to hybridisation buffer.

The next day, probe was removed and embryos were washed twice for 10 minutes in hybridisation buffer, then three times in 2x SSC + 0.1% Tween-20 for 20 minutes, and three times in 0.2x SSC + 0.1% Tween-20 for 30 minutes. The above washes were carried out at 60°C. Embryos were washed at room temperature twice for 15 minutes in 1x maleic acid buffer (MABT) (100 mM maleic acid, 150 mM NaCl, pH 7.8, 0.1% Tween-20), and subsequently blocked in 1x MABT with 2% Boehringer Mannheim Blocking Reagent (BMB, Roche) and 20% heat treated lamb serum (HTLS, Fisher) for 2 hours at room temperature. This was replaced with fresh solution containing anti-DIG antibody coupled to alkaline phosphatase (AP) (Roche) at 1/2000, and incubated overnight at 4°C.

Embryos were then washed briefly three times in MABT, and then three times for 1 hour in MABT. Samples were washed for 10 minutes at room temperature in AP buffer (100 mM Tris, pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween-20), which was replaced with BM purple (Roche) and left at room temperature for colour to develop. Once developed, embryos were washed twice in PBST for 15 minutes, before being fixed overnight in 3.7% formaldehyde at room temperature.

Embryos were bleached for photography in 5% hydrogen peroxide in PBST until pigment was lost. Embryos were subsequently washed twice for 15 minutes in PBST, and stored in 3.7% formaldehyde.

2.3.16 Whole-mount *in situ* hybridisation for miRNA

Probes used were 5'-DIG labelled miRNA detection probes (Exiqon), named 'hsa-miR-363-3p', and 'xtr-miR-363*'.

Probes were pre-absorbed six times by repeating Day 1 with 40 embryos at stage 35 and hybridising with the probe overnight. Probe was used at 20 nM, in hybridisation buffer. Protocol was carried out according to Sweetman (2011), with adaptations made by G. Wheeler (UEA, personal communication).

Fixed embryos were brought up to room temperature in methanol. They were rehydrated into PBST by a series of washes for 5 minutes; 75% methanol/25% PBST, 50% methanol/50% PBST, and twice with PBST. Embryos were treated with Proteinase K at 10 µg/ml in PBST, for a time dependent on stage. Following two washes in PBST for 5 minutes, embryos were re-fixed in 3.7% formaldehyde in PBST for 45 minutes. This was followed by two washes in PBST. Embryos were equilibrated into hybridisation buffer (500 ml: 250 ml formamide, 32.5 ml 20x SSC (pH 5, with citric acid), 25 ml 10% CHAPS, 5 ml 20% Tween-20, 5 ml 0.5 M EDTA, 250 µl 200 mg/ml heparin, 1.25 ml 200 mg/ml yeast tRNA) with a 10 minute wash in 50% PBST/50% hybridisation buffer, and followed with a 10 minute wash in 100% hybridisation buffer. Hybridisation buffer was replaced and placed at hybridisation temperature for at least 2.5 hours (53°C for miR-363-3p, 55°C for miR-363-5p), which was then replaced with pre-warmed probe, pre-absorbed in hybridisation buffer, and incubated overnight.

The probe was removed and kept, and embryos were then washed once for 10 minutes in hybridisation buffer, at hybridisation temperature. This was followed with two washes of 15 minutes in Wash buffer (50 ml: 25 ml formamide, 2.5 ml 20x SSC (pH 5, with citric acid), 250 µl 20% Tween-20), and one 10 minute wash in 50% MABT/50% Wash buffer, all at hybridisation temperature. Samples were then transferred to room temperature and two 30 minute washes in MABT were undertaken. Embryos were blocked for an hour in 2% BMB in MABT, and then a further hour in 2% BMB, 20% HTLS, in MABT. This solution was then replaced with anti-DIG AP antibody (1/2000 dilution, in 2% BMB + 20% HTLS in MABT) and rocked overnight at 4°C.

Embryos were rinsed three times in MABT, and then washed six times for 1 hour in MABT, before replacing the solution and washing overnight at 4°C. Samples were washed for 10 minutes in NTMT (50 ml: 2.5 ml 1 M $MgCl_2$, 1 ml 5 M NaCl, 5 ml 1 M Tris, pH 9.5, 2.5 ml 20% Tween-20), before being replaced with colour mix (NBT/BCIP in NTMT; 9:1000 of NBT at 75 mg/ml in 70% DMF, 7:1000 of BCIP at 50 mg/ml in 100% DMF) and incubated until colour began to appear.

When colour began to develop, embryos were transferred to 5x TBST, and washed at room temperature or overnight at 4°C in fresh TBST. They were then put through repeat

cycles of colour mix and TBST washes until strong specific signal could be seen. Embryos were then fixed in 3.7% formaldehyde and bleached in 5% hydrogen peroxide to remove pigment, before being stored in fixative.

2.3.17 Radioactivity labelling

RNA oligonucleotides were labelled using the KinaseMax kit (Ambion) according to manufacturer's instructions. 2 µM RNA oligonucleotide was incubated with 1x kinase buffer, 1 µl ATP [γ -^{32}P]- 5 mCi/ml (Perkin Elmer), and 1 µl kinase at 37°C for one hour. The reaction was halted by placing on ice before continuing. For long RNA sequences (pre-miRNA), the reaction was placed at 95°C for 2 minutes before cooling slowly to allow proper annealing.

2.3.18 Electrophoretic mobility shift assay (EMSA) with recombinant protein

EMSAs were performed with recombinant protein, made in the the Anston lab (York Structural Biology Laboratory) in buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10% sucrose, 0.5 mM DTT). Recombinant protein sequences are shown in Table 2.9.

Table 2.9 Recombinant protein sequences

Recombinant protein	Sequence
Human full-length lin28a	MGSVSNQQFAGGCAKAAEEAPEEAPEDAARAADEPQLLHGAGICKWFNVR MGFGFLSMTARAGVALDPPVDVFVHQSKLHMEGFRSLKEGEAVEFTFKKSA KGLESIRVTGPGGVFCIGSERRPKGKSMQKRRSKGDRCYNCGGLDHHAKECK LPPQPKKCHFCQSISHMVASCPLKAQQGPSAQGKPTYFREEEEEIHSPTLLPEA QN
Human truncated lin28a	LLHGAGICKWFNVRMGFGFLSMTARAGVALDPPVDVFVHQSKLHMEGFRS LKEGEAVEFTFKKSAKGLESIRVTGPGGVFCIGSERRPKGKSMQKRRSKGDRC YNCGGLDHHAKECKLPPQPKKCHFCQSISHMVASCPLKAQQ
***Xenopus* truncated lin28a**	GSGVCKWFNVRMGFGFLTMTKKEGTDLETPVDVFVHQSKLHMEGFRSLKE GESVEFTFKKSSKGLESTRVTGPGGAPCIGSERRPKVKGQQKRRQKGDRCYN CGGLDHHAKECKLPPQPKKCHFCQSPNHMVAQCPAKASQAAN

RNA oligonucleotides (Dharmacon), listed in Table 2.10, and precursor miRNA (synthesised in 2.3.13) were labelled as in 2.3.17. Probe was diluted 1:2 and 1 µl used per 10 µl binding reaction. Reaction was carried out as in Piskounova *et al.* (2008). Radio-labelled RNA was incubated with protein in binding buffer (60 mM KCl, 10 mM HEPES,

pH 7.6, 3 mM $MgCl_2$, 5% glycerol, 1 mM DTT, 5 µg/µl heparin (Sigma) and 150 ng yeast total RNA competitor (Ambion)) for 30 minutes at room temperature.

Loading buffer (Promega) was added 1:6 to binding reactions, which were run on a 10% native gel (10% acrylamide, 0.5x TB, set with APS and TEMED) and run in 0.5x TB (1 L 5x TB buffer: 27.5 g boric acid, 53 g Tris, in water). Gels were run until unincorporated radioactivity was eluted from the gel with Orange G dye front. Gels were fixed in 10% methanol with 10% acetic acid in water, and dried down on Whatmann paper. Gels were exposed either to a Phosphor Screen (GE Healthcare) and were scanned, processed and analysed using a Bio-Rad Molecular FX Imager and Quantity One software (Bio-Rad); or exposed to Hyperfilm ECL film (Amersham) and films analysed using Image J. Analysis was performed on bands as a measure of volume or density, with background normalisation to the no-protein reaction 'bound' area. Proportion bound was calculated and plotted using SigmaPlot with the equation:

$$\text{Proportion bound} = \frac{B_{max}\,[\text{lin28}]}{K_d + [\text{lin28}]}$$

Table 2.10 RNA oligonucleotide sequences

RNA	Sequence
L-let-7g	UUUGAGGGUCUAUGAUACCACCCGGUACAGGAGAU
L-mir-138	UUGUGAAUCAGGCCGUGACCACUCAGAAAACGGCUACUUCACAAC
L-mir-363	UGCAAUUUUAUUUAGUUUGGUAGGAGAAAAAUUGC
mL-mir-363	UGCAAUUUUAUUUAGUUUGGUAUGAUAAAAAUUGC

Sequences are given 5' to 3'.

2.3.19 EMSAs using *in vivo* translated protein

EMSAs used embryo lysates from *X. laevis* embryos, either uninjected controls, or having been injected with 1 ng of mRNA to overexpress lin28a1, lin28a2 or lin28b.

Frozen *X. laevis* embryos were lysed in lysis buffer (50 mM Tris-HCl pH 7.9, 25% glycerol, 50 mM KCl, 2 mM DTT, 0.1 mM EDTA, 1/100 Protease inhibitor cocktail III (Calbiochem)) at 10 µl/embryo and homogenised by pipetting. Lysate was then centrifuged at 13,000*g* for 10 minutes at 4°C. The supernatant was centrifuged again for 5 minutes, and supernatant removed to a new tube, which was then used in binding reactions as for recombinant proteins. Extract was diluted as required in the lysis buffer, and 3 µl used per binding reaction. Supershifts were also performed when using embryo extract, to confirm that any shift seen was a result of the protein being overexpressed, and not due to changes in expression of alternative proteins. To perform these, anti-lin28 antibodies

(Enzo Life Sciences (UK) Ltd, custom produced, raised against peptides from the *X. tropicalis* protein sequences) were used at 1/20 dilution per binding reaction, with 1/20 dilution pre-immune bleed used as a serum control. 20 units of RNAsin (Promega) were added per binding reaction with embryo lysates, due to the presence of RNases in serum and extract. Antibody was pre-incubated with protein and binding buffer for 20 minutes on ice, before labelled probe was added for a further 20 minutes at room temperature. Binding buffer and remainder of protocol was as in 2.3.18.

2.3.20 Western blots

Protein extraction from human cells

Medium was removed from cells in culture and washed with ice cold 1x PBS. 300 µl PhosphoSafe (Novagen) was added per well of 6-well plates. Samples were incubated at room temperature for 5 minutes, and lysate was scraped into a tube. Samples were then centrifuged at 13,000*g* for 5 minutes and supernatant removed to a new tube.

Protein concentrations were measured using a BCA protein assay kit (Pierce) as per instructions. BSA standards were made up and spectrophotometric change at 562 nm read using a plate reader from 96-well plates after the addition of working solution, and a standard curve was drawn. Samples were read in triplicate, and average reading was used and compared to the standard curve to determine protein concentration in the cellular extracts. 8 µg per sample was used for a western blot.

Protein extraction from *X. tropicalis*

Frozen embryos were lysed in PhosphoSafe (Novagen) at 40 µl/10 embryos and homogenised using a pipette tip. Lysates were centrifuged at 13,000*g* for 5 minutes and 30 µl supernatant removed to a fresh tube. Protein extracted from cells or *Xenopus* was used in equal amounts.

SDS-PAGE and western blotting

SDS-PAGE was performed using the Biorad miniProtean II PAGE equipment. 5x SDS-PAGE loading buffer (final concentrations: 250 mM Tris, pH 6.8, 10% SDS, 30% glycerol, 250 mM DTT, bromophenol blue to colour) was added to protein samples, which were heated to 100°C for 2 minutes. Samples were loaded onto a 12% SDS PAGE acrylamide gel (10 ml resolving gel: 3.3 ml water, 4 ml 30% acrylamide mix (Sigma), 2. 5ml 1.5 M Tris, pH 8.8, 100 µl 10% SDS, 100 µl 10% APS, 10 µl TEMED; 5 ml stacking gel: 3.4 ml

water, 830 µl 30% acrylamide (Bio-Rad), 630 µl 1 M Tris, pH 6.8, 40 µl 10% SDS, 40 µl 10% APS, 5 µl TEMED) alongside 5 µl Pageruler prestained protein ladder (Fermentas), and run in 1x running buffer (1 L 10x buffer: 15 g Tris-base, 72 g glycine, 50 ml 10% SDS, pH 8.3). Gels were run at 100 V through the stacking gel, then 200 V for 50 minutes. The gel was subsequently transferred for 2 hours at 100 V onto a PVDF membrane (Millipore) using the Trans-Blot system (Bio-Rad). PVDF membrane was saturated in 100% methanol and equilibrated into 1x transfer buffer (1 L 1x buffer: 5.8 g Tris base, 4.35 g glycine, 100 ml methanol, pH 8.8) prior to transfer.

The membrane was washed twice in PBST for 5 minutes and subsequently blocked in PBST + 5% Marvel milk powder for one hour. Blocking solution was replaced with PBST + 5% milk containing primary antibody and rocked overnight at 4°C. Antibody concentrations are given below (Table 2.11). The membrane was washed three times for 15 minutes in PBSAT, and then rocked for 2 hours in PBST + 5% milk with a species-compatible horseradish peroxidase (HRP) secondary antibody (Table 2.12) at room temperature, with subsequent washing in PBST as before. The BM Chemiluminescent Substrate Kit (Roche) was used to detect protein expression according to manufacturer's instructions, and membrane was exposed to Hyperfilm ECL film (Amersham) to visualise expression.

When further antibody exposures were needed, the membrane was stripped in stripping buffer (100 ml: 0.3 g DTT, 2 ml 1 M Tris, pH 6.8, 500 µl 10% SDS, pH to 6.8 and made up to 100 ml with distilled water) at 55°C for 20 minutes, and washed five times in PBST. It was subsequently blocked and probed with antibody as previously.

Table 2.11 Primary antibody concentrations for western blots

Antibody	Concentration (*Xenopus*)	Concentration (human cells)	Species
dpERK (Sigma) (Cat # M9692	1/4000	1/4000	Mouse
Total ERK (Sigma) (Cat # M5670)	1/80,000	1/4000	Rabbit
GAPDH (Sigma) (Cat # G8795)	1/100,000	1/100,000	Mouse
lin28a (Abcam) (Cat # ab46020)	Not used	1/20,000	Rabbit
lin28b (Abcam) (Cat # ab71415)	Not used	1/2000	Rabbit
lin28a (Enzo Life Sciences (UK) Ltd, custom produced)	1/10,000	Not used	Rabbit
lin28b (Enzo Life Sciences (UK) Ltd, custom produced)	1/10,000	Not used	Rabbit

Table 2.12 Secondary antibody concentrations for western blots

Antibody	Concentration (*Xenopus*)	Concentration (human cells)
Anti-mouse HRP (GE Healthcare) (Cat # NA931)	1:4000	1:4000
Anti-rabbit HRP (Abcam) (Cat # ab6721)	1:4000	1:4000

2.3.21 *Xenopus* dpERK immunostaining

Fixed embryos were rehydrated in 1x PBS through 15 minute stage dilutions with methanol (75%, 50%, 0%) and rocked for 40 minutes in potassium dichromate in 5% acetic acid. Embryos were washed in PBS three times for 5 minutes, and three times for 30 minutes, until embryos were clear of orange stain; embryos were subsequently bleached in 5% H_2O_2 in PBS for 40 minutes, and washed again in PBS. The embryos were then blocked in BBT (1x PBS, 1% BSA, 0.1% Triton X-100) twice for one hour, and again for an hour in BBT + 5% horse serum (VectorLabs). Embryos were rocked overnight at 4°C with anti-dpERK antibody (Sigma) at 1/10,000 in BBT with 5% horse serum. Four washes of an hour with BBT, and one hour with BBT + 5% horse serum were undertaken, before incubating embryos with secondary anti-mouse IgG-AP-conjugated antibody

(VectorLabs) overnight at 4°C. Embryos were then washed for one hour in BBT, then for four hours in PBST. Protein expression was detected using BMPurple (Roche). Once a signal had developed, embryos were washed twice for 15 minutes in PBS and stored in 3.7% formaldehyde in PBS.

2.4 Microarray and bioinformatics analysis

2.4.1 miRNA affymetrix microarray

RNA to be used for Affymetrix miRNA microarray analysis was isolated according to 2.3.6, with 1 µg used per sample. All further laboratory work was carried out by the Departmental Technology Facility, which is provided as a full service. The quality of the RNA was verified using the Agilent 2011 Bioanalyzer (Agilent). RNA was labelled using HSR FlashTag Biotin RNA labelling kit (Genisphere) according to manufacturer's instructions, which included the addition of spike-in RNA controls to act as a method control. Samples were then hybridised to Genechips miRNA 2.0 (Affymetrix) overnight, and washed on a Fluidics Station 450 (Affymetrix), all carried out according to manufacturer's instructions. Scanning of the chips was carried out using an Affymetrix Genechip Scanner.

Data was analysed using QC tools (Affymetrix) to provide background detection and normalisation, and statistical comparisons were carried out using BRB-ArrayTools (developed by Dr. Richard Simon and BRB-ArrayTools Development Team, http://linus.nci.nih.gov/BRB-ArrayTools.html).

2.4.2 mRNA human genome Agilent microarray

Samples of human MSC RNA were isolated according to 2.3.6. All further laboratory work and analysis was carried out by the Departmental Technology Facility, which is provided as a full service. 100 ng of RNA was used per sample. This was labelled with Cy3 dye using the Low-Input RNA Amplification Kit (Agilent), according to manufacturer's instructions. The arrays used were SurePrint G3 Human GE 8x60k arrays (Agilent), and array scanning was carried out on the DNA Microarray Scanner (Agilent). Analysis was performed using GeneSpring (Agilent Technologies).

2.4.3 Sequence alignments

Protein homology percentages were calculated using Vector NTI (Invitrogen). Alignments of RNA and protein sequences were performed using Clustal X2.

Chapter 3. Characterising expression of *lin28* genes in *X. tropicalis*

3.1 Introduction

3.1.1 *lin28a* as a putative FGF target

The gene *lin28a* was identified as one of a number of genes that showed altered expression upon inhibition of FGF signalling in previous work, which was investigating novel FGF targets during gastrulation in *Xenopus* embryos using an Affymetrix array to allow a large scale investigation (Branney et al., 2009). In this study, dominant negative forms of both FGFR1 and FGFR4 were overexpressed, which individually had previously been found to successfully block FGF downstream signalling and resulted in embryos with varying developmental defects; overexpressing these constructs effectively reduced all FGF signalling (Amaya et al., 1991; Hongo et al., 1999). This is likely to occur through promiscuous heterodimerisation with available FGFRs within the cell despite any ligand-receptor preferences. The study by Branney et al. (2009) showed that overexpressing either of these receptors resulted in dysregulation of a similar set of genes. Of these, *lin28a* was identified as showing a 2.7 fold inhibition following the blocking of FGF signalling.

3.1.2 *lin28* expression

Within the literature, studies have identified *lin28a* as being expressed in pluripotent cells and downregulated upon differentiation, shown both in stem cell culture populations, with rapid downregulation of the expression of both RNA and protein upon forming embryoid bodies, and also *in vivo* (Darr and Benvenisty, 2009; Moss and Tang, 2003; Polesskaya et al., 2007; Richards et al., 2004); a more extensive study into lin28a protein expression, showing similar results, has been carried out during murine development (Yang and Moss, 2003). In comparison, *lin28b* has not been as widely studied, but was first identified in hepatocarcinoma cells and additionally found to be expressed in testis, fetal liver and placental tissues (Guo et al., 2006). Both genes have been shown to be expressed in numerous cancer cells, in both primary tumours and transformed cell lines (Viswanathan et al., 2009). However, there has been no definitive comparison of expression analysis between *lin28a* and *lin28b* during development. Therefore it is imperative to explore the developmental expression patterns of the two *lin28* genes, which may indicate both overlapping and distinct roles for the two proteins.

The possibility that FGF signalling regulates *lin28a* could potentially provide an exciting insight into the role of FGF in pluripotency, with the knowledge that *lin28a* is pluripotency-specific and the current understanding of the differentiation functions of FGF signalling being at odds with the requirement for its presence in stem cell culture. Alternatively, as *lin28a* was identified as a possible target of FGF signalling during gastrulation, perhaps the *lin28s* play a role at this time in mediating FGF signalling and subsequently influence the development of mesoderm and neural differentiation.

3.1.3 Aims

The aims of this chapter are to:

- To investigate the role of FGF signalling in expression of *lin28a* and *lin28b*
- To identify when and where *lin28* genes are expressed during early *Xenopus* development

3.2 Results

3.2.1 FGF signalling in the gastrula embryo

To follow up the findings by Branney et al. (2009), experiments were carried out at stage 10.5 in *X. tropicalis*. FGF signalling is known to occur at this stage in development, in tissue around the developing blastopore lip, known as the presumptive mesoderm. This is demonstrated by the presence of RNA coding for the FGF ligand FGF-8 (Figure 3.1A), and diphospho-ERK (dpERK) protein (Figure 3.1B), a rapid downstream response of FGF signalling, which is localised to this same area.

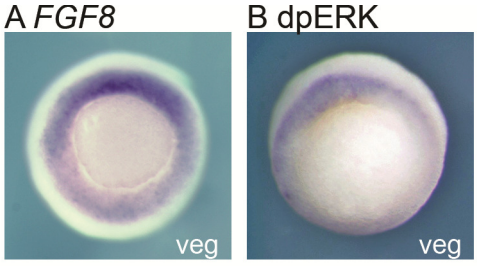


Figure 3.1 Analysis of FGF signalling in gastrula-stage embryos
Embryos showing expression of A) *FGF-8* ligand RNA by *in situ* hybridisation and B) dpERK protein, phosphorylated downstream of FGF signalling, by whole mount immunohistochemistry. Embryos are viewed vegetally (veg), with dorsal side to the top.

3.2.2 *lin28* genes as targets of FGF signalling

Despite only *lin28a* being identified as a novel putative target of FGF signalling in the Affymetrix screen (Branney et al., 2009), both related genes, *lin28a* and *lin28b*, were investigated to validate and confirm them as FGF targets. FGF signalling was inhibited in embryos through the use of a dominant negative FGF receptor. 1 ng of *dominant negative FGF receptor 4* (*dnFGFR*) mRNA was injected into *X. tropicalis* embryos which were then collected at gastrula stage 10.5, alongside uninjected controls.

The phosphorylation of ERK is a rapid downstream response to FGF signalling, and was used as an indicator of inhibition of FGF signalling. Levels were investigated in gastrula stage embryos by western blot and whole-mount immunohistochemistry, following injection with dnFGFR. In accordance with a decrease in FGF signalling, levels of dpERK were reduced in *dnFGFR*-injected embryos compared to controls (Figure 3.2).

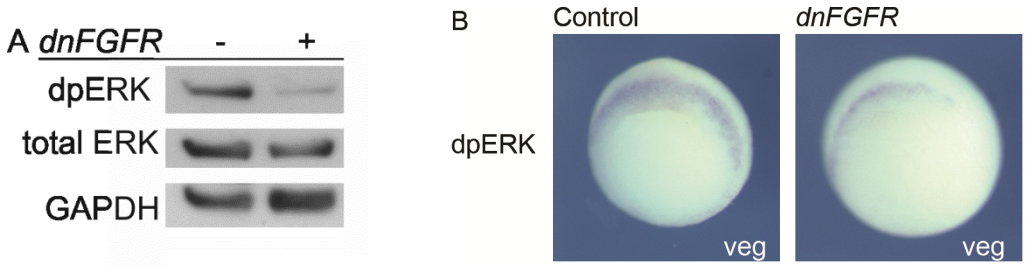


Figure 3.2 Effects of FGF signalling on dpERK levels
Embryos were injected with 1 ng *dnFGFR* mRNA and collected with uninjected controls at stage 10.5; protein was analysed by A) western blot and B) whole mount immunohistochemistry. Lower levels of dpERK indicate decreased FGF signalling. A) Total ERK (phosphorylated and unphosphorylated) and GAPDH were used as loading controls. Images represent three independent samples. B) Embryos are viewed vegetally (veg), with dorsal to the top.

Embryos were collected at gastrula stage and analysed for expression of the *lin28* genes by qRT-PCR and whole-mount *in situ* hybridisation. Inhibition of the FGF signalling pathway was confirmed at the transcriptional level by qRT-PCR for the known FGF target *brachyury*, which showed a significant decrease in expression in *dnFGFR*-injected embryos to 25% of control expression (Figure 3.3). Similarly, the *lin28* genes also showed a significant decrease in expression when FGF signalling was inhibited, however not to the same extent as seen with *brachyury*. The loss of *lin28a* was greater than that of *lin28b*, with a reduction of 45% of expression of *lin28a* following inhibition of FGF signalling when compared to controls, and a 33% reduction in *lin28b* levels (Figure 3.3).

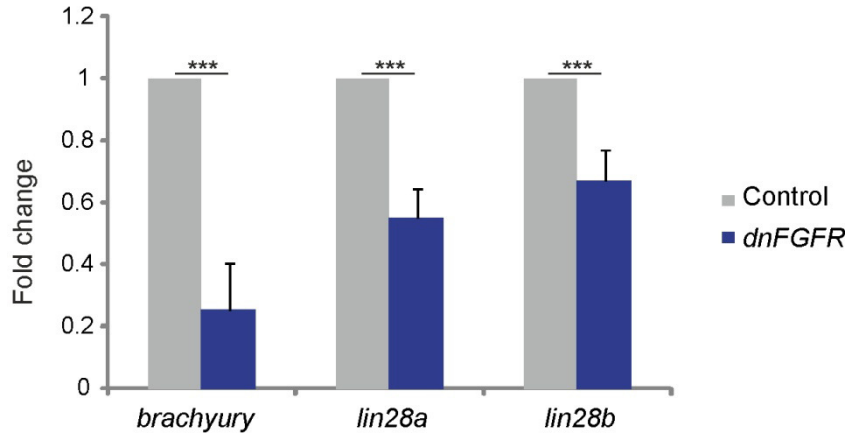


Figure 3.3 Effects of FGF signalling on *brachyury* and *lin28* genes, analysed by qRT-PCR
qRT-PCR was performed on RNA from embryos injected with 1 ng *dnFGFR* mRNA and uninjected controls at stage 10.5. Fold change in expression of RNAs is shown compared to controls and normalised to *ODC* by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of three biological replicates. Error bars represent standard error (SE). Significance calculated by *t*-test: ***p<0.001.

The location and any changes in pattern of RNA expression was observed using whole-mount *in situ* hybridisation. Interestingly, in wild-type embryos, the pattern of expression seen for both *lin28a* and *lin28b* was surrounding the blastopore lip in the presumptive mesoderm, an overlap with the known region of FGF ligand expression and signalling seen in Figure 3.1.

Following inhibition of FGF signalling, expression of both *lin28a* and *lin28b* was reduced around the blastopore lip, however some retention of expression was observed at the dorsal side (Figure 3.4). This dorsal expression remained at a similar level to expression in control embryos for *lin28a*, and appeared slightly enriched for *lin28b*.

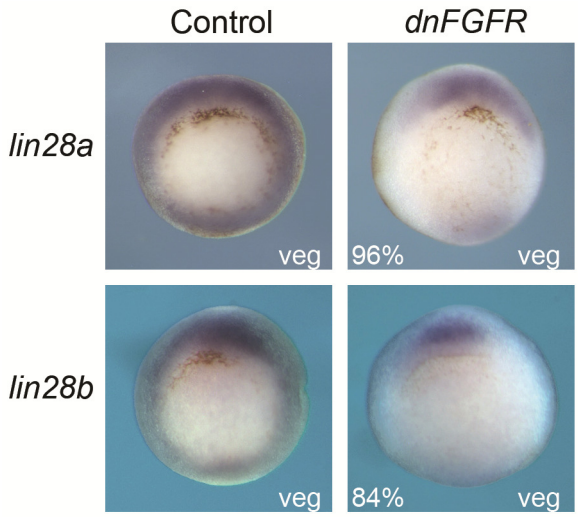


Figure 3.4 Analysis of *lin28* genes by whole-mount *in situ* hybridisation following FGF signalling inhibition

In situ hybridisation in uninjected control embryos or those injected with 1 ng *dnFGFR* mRNA, for expression of *lin28a* and *lin28b*. For *lin28a*, 96% from n=45 showed similar knockdown expression, for *lin28b* this was 84% from n=43. Representative images are shown; experiment was carried out in four biological replicates. Embryos are viewed vegetally (veg), with dorsal to the top.

3.2.3 Timing of expression of *lin28* genes during development

The timing and location of gene expression can give an indication to its function and role; these were investigated for the *lin28* genes using qRT-PCR and whole-mount *in situ* hybridisation during embryological development. As the *lin28a1* and *lin28a2* isoforms show very high sequence identity, these methods detected both forms of *lin28a*, due to design of primers and probes.

RNA levels for *lin28a* and *lin28b* in early stages of embryos following fertilisation, from 1-cell through to late gastrulation stages, were documented in sibling embryos using qRT-PCR. Expression was calibrated to the stage at which highest expression was detected, which varied for the two genes. *lin28b* was found to be expressed during the earliest stages, which would represent maternally deposited RNAs in the oocyte. However, *lin28a* levels remained low in the embryo until MBT and the activation of zygotic transcription, where levels were seen to rapidly increase and were maintained at this level of expression. *lin28b* also showed an increase in expression at this time, peaking in expression at stage 10, before levels were seen to reduce, but the gene remained expressed (Figure 3.5).

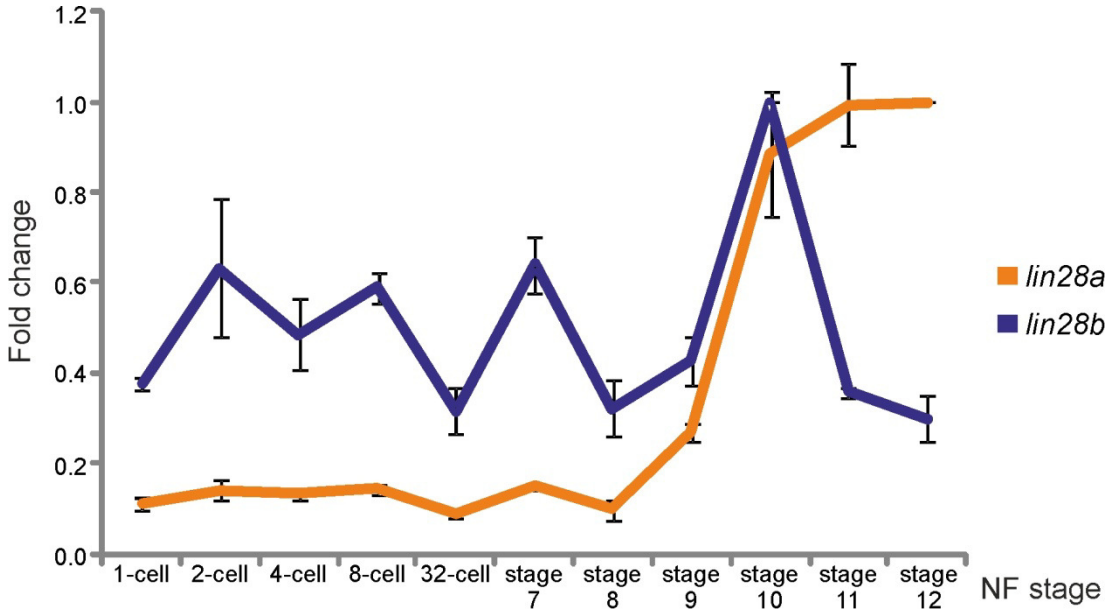


Figure 3.5 Analysis of *lin28a* and *lin28b* gene expression during early development by qRT-PCR

qRT-PCR was performed on sibling embryos from 1-cell through to stage 12. Expression was normalised to housekeeping gene *ODC* using the $2^{-\Delta\Delta Ct}$ method, and calibrated to the highest expression for either gene (*lin28a* – stage 12 , *lin28b* – stage 10). Error bars show SE of technical replicates. Expression trends are representative of n=2.

3.2.4 Localisation of *lin28a* expression during development

This trend in *lin28a* expression is further seen using whole-mount *in situ* hybridisation. Only a very low level of expression was detectable before zygotic transcription (Figure 3.6A-F). Expression began to show localisation to the animal hemisphere at stage 8 (Figure 3.6G), and at stage 9 expression was detected within the marginal zone (Figure 3.6H). From gastrulation, the RNA levels were much more readily detectable, in line with the qRT-PCR data. *lin28a* was found in the presumptive mesoderm at gastrula stage (Figure 3.6I), as shown in Figure 3.4; this is confirmed upon looking inside the embryo, this expression can be seen to be located in the deep sensorial cells of the embryo, below the surface ectoderm cells (Figure 3.6I'). During neurulation, expression of *lin28a* at stage 13 was through the dorsal midline, which may be localised to the notochord; additionally, expression was detected at stages 13 and 17 through the developing dorsal plate, together with anterior neural crest expression and posterior expression that may be in the neural plate or mesodermally derived tissue (Figure 3.6J, K). In the tailbud and early tadpole stages, expression of *lin28a* had become further restricted. Expression was still predominantly dorsal in tailbud embryos (Figure 3.6L, M), but by stage 30 was present in the branchial arches and heart, the pronephros and developing nephric duct, and dorsal of the cement gland, which may correspond to cranial mesoderm. Dorsal expression was found within the midbrain-hindbrain junction and at the very dorsal tip of the tailbud;

lower level expression which extended along the dorsal side of the embryo may correspond to neural or somite expression (Figure 3.6N). By stage 35 (Figure 3.6O), however, expression only persisted in the branchial arches.

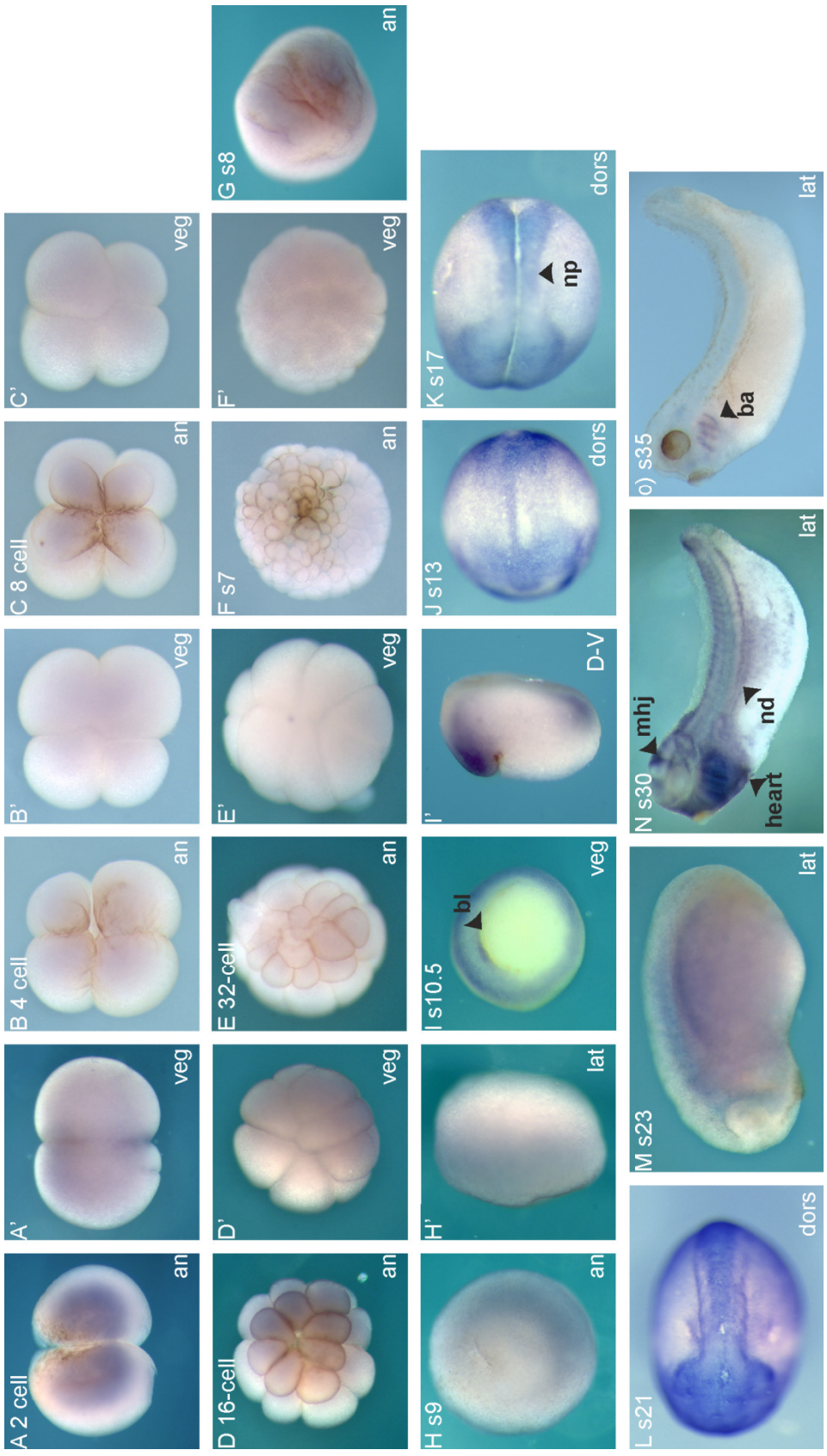


Figure 3.6 The expression of *lin28a* during embryonic development

Whole-mount *in situ* hybridisation showing expression of *lin28a* in wild-type embryos over development. (') refers to an alternative view of the same stage embryo. Code bottom right refers to view: an = animal, veg = vegetal, dors = dorsal, lat = lateral, D-V = dorsal-ventral axis. Key to structures: bl = blastopore lip, np = neural plate, mhj = midbrain-hindbrain junction, nd = nephric duct.

3.2.5 Localisation of *lin28b* expression during development

When studying expression of *lin28b*, there are both similarities and striking differences in the patterning of this gene compared to *lin28a*. The difference seen in maternal expression by qRT-PCR persisted here, with the detection of *lin28b* at stages prior to zygotic transcription, which was restricted to the animal hemisphere (Figure 3.7A-G). Expression in late blastula embryos was, as for *lin28a*, in the marginal zone (Figure 3.7H). As discussed previously, *lin28b* expression overlaps with *lin28a* in the developing gastrula, in the presumptive mesoderm surrounding the blastopore lip within the deep cells (Figure 3.7I). Upon neurulation, *lin28b* was located around the developing neural plate, but not through the dorsal midline, and with strong anterior and posterior patterning as for *lin28a* (Figure 3.7J, K). By stage 20, expression in the posterior embryo appeared restricted to mesoderm, with prominent anterior neural patterning (Figure 3.7L). In the later tailbud, expression of *lin28b* was more widespread than *lin28a*, and was surrounding the eye, in the forebrain extending through the neural system, and in regions that would go on to form the branchial arches (Figure 3.7M), where it remained present in the tadpole embryos (Figure 3.7N, O). By stage 30, some overlap in expression was again seen between *lin28a* and *lin28b*, in the branchial arches and along the embryo dorsally, which may represent neural tissue or the dorsolateral and ventromedial somite lips. The expression of *lin28b* in the latter area was stronger than of *lin28a*. The nephric duct, branchial arches and potential cranial mesoderm also showed a similar expression pattern for the two genes. Differences existed, however, as *lin28b* was absent in the cardiac tissue and instead featured in the otic vesicle and persisted in the eye (Figure 3.7N). By stage 35, as for *lin28a*, expression of *lin28b* was only detected in the branchial arches and at low levels along the nephric duct (Figure 3.7O).

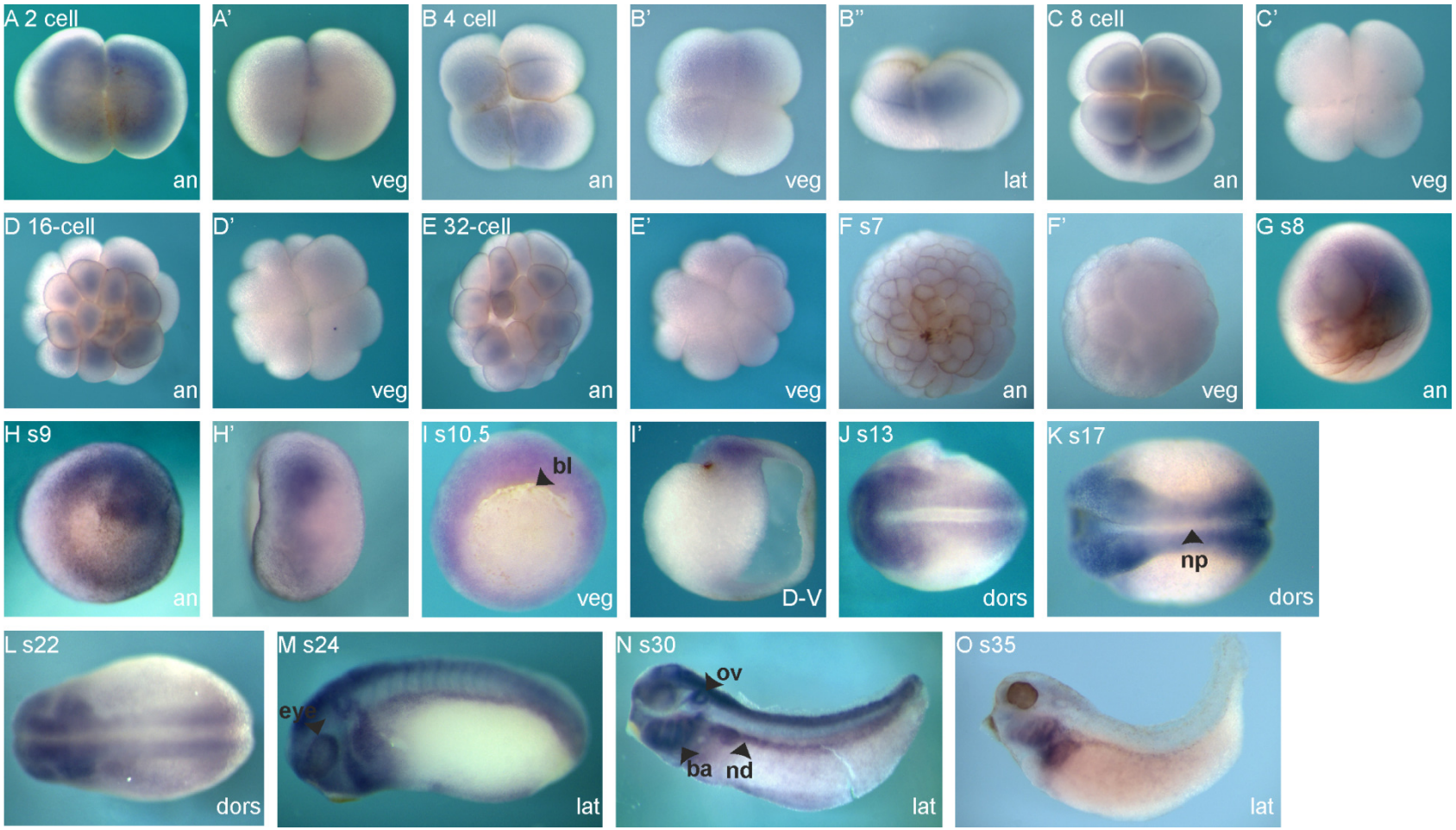


Figure 3.7 The expression of *lin28b* during embryonic development

Whole-mount *in situ* hybridisation showing expression of *lin28b* in wild-type embryos over development. (') refers to an alternative view of the same stage embryo. Code bottom right refers to view: an = animal, veg = vegetal, dors = dorsal, lat = lateral, D-V = dorsal-ventral axis. Key to structures: bl = blastopore lip, np = neural plate, ov = otic vesicle, ba = branchial arches, nd = nephric duct.

3.3 Discussion

3.3.1 FGF signalling regulates *lin28a* and *lin28b* during gastrulation

It is known that the overexpression of dnFGFR4 is a potent inhibitor of FGF signalling (Branney et al., 2009; Hongo et al., 1999), which has been confirmed here as being effective through an inhibition of FGF targets, both investigating dpERK and *brachyury* levels. dpERK is known to be downstream of multiple signalling pathways, but this early activity in gastrula stage embryos was discovered to be dependent upon only FGF expression (Christen and Slack, 1999), therefore the changes in expression seen at the gastrula stage were caused by dysregulation of FGF signalling. *Brachyury* similarly has been shown to be an FGF target at this stage, with expression reduced following the inhibition of FGF signalling through dominant negative receptor expression (Amaya et al., 1993; Isaacs et al., 1994), which was also seen here.

Experiments here found that the inhibition of FGF signalling resulted in a decreased expression of both of the *lin28* genes, which indicates that the *lin28* genes are positive regulatory targets of FGF, with FGF necessary to maintain their expression at a normal endogenous level. This result confirms that seen for *lin28a* in previous work, although the decrease in expression seen by Branney and coworkers (2009) was more dramatic, with a 63% drop in expression levels compared to 55% seen here. The overlap in expression of *lin28* RNAs with a domain of active FGF signalling in gastrula embryos increases the likelihood of an interaction. These experiments demonstrated localisation around the blastopore of *FGF-8*; previous research has shown the strong expression of this and more FGF ligands to the same restricted area, including *FGF-3*, *FGF-4* and *FGF-20* (Christen and Slack, 1997; Isaacs et al., 1995; Lea et al., 2009; Lombardo et al., 1998), and thus with this varied expression and the use of pathway inhibition it cannot be determined here which FGF ligand is responsible for the regulation of *lin28* expression.

There was a larger reduction in the level of *lin28a* expression compared to *lin28b*, which also showed stronger dorsal retention of expression following dnFGFR expression. These results may indicate that *lin28b* is less influenced by FGF transcriptional regulation than *lin28a*, and this is at least true of the earlier *lin28b* expression during the cleavage stages. Maternally deposited RNAs are outside of the transcriptional control of signalling pathways such as FGF signalling, and so it is possible that the reduced loss of *lin28b* mRNA seen when compared to *lin28a* or *brachyury* may be affected by the persistence of such maternal RNA. The failure to reduce expression of both *lin28a* and *lin28b* to the same extent as *brachyury* suggests that although FGF is important for correct expression of the

lin28 genes, it is not the only factor that is governing their expression; these additional regulatory pathways remain currently unidentified.

3.3.2 Potential for further FGF regulation of *lin28s* in development

Co-expression of *lin28a* and *lin28b* with FGF ligands continues to occur in later development. Areas of both *lin28a* and *lin28b* expression include the branchial arches, which have been found to express the ligands *FGF-2, FGF-3, FGF-7*, *FGF-8*, *FGF-10*, *FGF-20* (Lea et al., 2009). Further examples of co-expression of the *lin28s* with FGFs include the midbrain-hindbrain junction, where *FGF-3*, *FGF-4*, and *FGF-8* are localised; the cranial mesoderm also with *FGF-8* and *FGF-20* expression; and the tailbud, which contains expression of ligands *FGF-2*, *FGF-3*, *FGF-4*, *FGF-8*, and *FGF-20* (Christen and Slack, 1997; Lea et al., 2009). Weak expression in the pronephros was found for *FGF-2* and *FGF-8* which do not extend along the nephric duct, but instead expression of *FGFR1* and *FGFR2* mirrors the expression in these areas of the *lin28s* (Christen and Slack, 1997; Lea et al., 2009), indicating the potential of these tissues to respond to FGF signals. Expression of many FGF ligands overlaps with *lin28b* expression in the otic vesicle and the eye, but none of these ligands show the same dorsal expression along the length of the embryo (Lea et al., 2009). These multiple areas of co-expression strengthen the hypothesis that some regulation via the FGF pathway may occur. Work in chicken and mouse embryos has found dynamic expression of *lin28a* in the developing limb buds (Yokoyama et al., 2008), and with the process of limb development requiring FGF-8 and FGF-10 signalling (Mariani et al., 2008), the possibility that FGF signalling influences *lin28* expression is further strengthened.

Work here has not, however, determined if the *lin28* genes are subject to regulation by FGF during these later developmental stages. Using the current method of FGF signalling inhibition, any effects at later stages would be masked by the misregulation caused in *lin28a* and *lin28b* by gastrulation stages, and an alternative method would need to be employed to investigate any regulation in later development whilst allowing normal gastrula development. With no available drugs to target all FGF signalling, as the commonly used SU5402 and PD173074 both target only FGFR1 (Skaper et al., 2000), the best available option to study this is through the use of transgenics (Kroll and Amaya, 1996; Pownall et al., 1998). Kroll and Amaya (1996) were able to successfully overexpress a dominant negative FGF receptor during gastrulation or neurulation, depending upon the promoter used, without affecting the role of FGF in mesoderm specification, and could then investigate its role specifically during later stages of embryogenesis.

3.3.3 Expression of *lin28* genes during pluripotency

During the earliest stages of *Xenopus* development, it was found that *lin28b* was expressed maternally, but *lin28a* was not. There was, however, expression of *lin28a* from the onset of zygotic transcription. This has also been confirmed at the protein level (Faas et al., 2012), suggesting an early developmental role for these proteins.

In mammalian development, expression of the *lin28* genes has been detected in ES cells, and as such these genes are linked with pluripotency. In human development, the first zygotic transcription occurs much earlier than in *Xenopus*, with transcription occurring by the 4-cell stage (Braude et al., 1988; Tesařík et al., 1986), and ES cells are subsequently collected from the ICM of blastocyst embryos (Evans and Kaufman, 1981; Martin, 1981) and can differentiate into all tissue types. In *Xenopus*, zygotic transcription does not occur until MBT is completed, just before gastrulation (Newport and Kirschner, 1982), and work here has shown expression of *lin28b* and *lin28a* in late blastula embryos is first restricted to the animal hemisphere; this consists of cells that are subjected to inductive signalling events and would be capable of differentiation into a multitude of tissue types. Expression of both genes is then restricted to the marginal zone and presumptive mesoderm upon the beginning of gastrulation. This expression is more restricted compared to the mouse embryo, where expression of lin28a protein at the beginning of gastrulation during E6.5 was found across all three germ layers of the embryo (Yang and Moss, 2003).

This work suggests that although both *lin28a* and *lin28b* in *Xenopus* were expressed in multipotent cells, capable of a diverse range of differentiation dependent on inductive signals, the expression of at least *lin28a* was more restricted than in mammalian tissues, in which it is expressed in a larger number of cells and for a more prolonged period of time. This may indicate wider functions for lin28a in mammalian tissue, but further investigations are needed to determine what these could be.

3.3.4 The importance of *lin28* genes during gastrulation

An important role may be being played by these genes during gastrulation, indicated in part by an increase in RNA levels of both *lin28a* and *lin28b* at this time. This work supports and expands on findings by Moss and Tang (2003), who first detected *Xenopus* lin28a protein during gastrulation, as we also detected *lin28a* following MBT in gastrula stage embryos, but found this to occur earlier than the developmental stage that they had investigated (stage 11). The expression of both *lin28a* and *lin28b* in the gastrulating embryo is within the deep cells which are those responding to the inductive signals from

gastrulation. Additionally, FGF signalling is a key occurrence during germ layer development, and the regulation of *lin28s* by this pathway strengthens the hypothesis that a vital role is being undertaken by these genes at this stage of development, which will be further investigated.

3.3.5 Conservation of *lin28a* expression

The expression patterns of *lin28a* uncovered here show correlation to previously published work.

In *C. elegans* and *D. melanogaster*, expression of lin-28 RNA and protein was found in late embryogenesis and early larval stages, being absent from adults (Moss et al., 1997; Moss and Tang, 2003). Differences were seen, however, between the RNA and protein expression. Levels of protein were reduced after the first larval stage in each species, although *Drosophila* later showed a resurgence of protein by pupal development, whereas the level of RNA remained more constant, suggesting a possibility of post-transcriptional control over the protein production during their development.

Previous work on *Xenopus* found protein expression of lin28a was maintained from gastrulation though to stage 40 (Moss and Tang, 2003), and although stage 35 was the latest stage studied here, RNA expression was seen to persist in the *X. tropicalis* embryo until at least this stage for both *lin28a* and *lin28b*.

A more extensive study looking at the areas to which lin28a protein was located during development was carried out in mice. Research found that during organogenesis at E9.5, expression was beginning to show restriction to particular tissues, which included the branchial arches and the myocardium of the heart (Yang and Moss, 2003). During later development, lin28 was further restricted to tissue types including the myotome of somites, some tubules in the developing kidney and the foregut epithelium. Embryonic analysis ended at E17.5, at which stage lin28a expression remained in the myocardium and epithelia covering skeletal muscles and the body surface. This was then compared to adult expression, where lin28 remained in the myocardium, the transit cells in the gut epithelia, and restricted structures within the kidney (Yang and Moss, 2003), retaining expression in cells which have the capability to rapidly turnover and regenerate. The detection of *Xenopus lin28a* RNA within the heart, branchial arches and pronephros and nephric duct supports previous findings of lin28 protein in equivalent developing murine structures (Yang and Moss, 2003). Expression of *lin28a* does not appear to persist in the heart at the latest tadpole stage studied, however there may simply be stable and persistent

expression of the protein, which would explain that which was detected in the mouse study. Were the dorsal expression in late tailbud embryos to be located in the dorsolateral and ventromedial somite lips, as opposed to the neural tissue, this would suggest that proliferative muscle-forming cells express *lin28a*, consistent with previous work (Yang and Moss, 2003; Yokoyama et al., 2008). A study using embryoid bodies found *lin28a* expression in developing primordial germ cells, and the surrounding posterior mesoderm (West et al., 2009). This is in agreement with the mesodermal localisation seen here in the *Xenopus*, although there does not appear to be germ cell expression.

With differences between reported protein expression of lin28a and our observations at the mRNA level, characterisation of the protein locations for both lin28a and lin28b in *Xenopus* would further aid the understanding of their functions.

3.3.6 Differences between *lin28a* and *lin28b* expression

The large number of similar expression domains for *lin28a* and *lin28b*, and known overlapping roles such as in the regulation of let-7 (Viswanathan et al., 2008), suggest that a level of redundancy may exist between the proteins. The finding of the genes in separate cancer tissues may indicate the presence of some degree of redundancy in their functions, at least with regards to their role in cancer transformation; the same research also showed that lin28a and lin28b were found in different subcellular locations and acted in distinct ways to regulate let-7 (Piskounova et al., 2011), indicating important differences between the proteins. Differences in the expression patterns of the two genes were seen through *Xenopus* development, with *lin28b* showing a more widespread expression throughout much of development. As previously demonstrated in Chapter 3, *lin28b* is present maternally, suggesting it may play an additional role during the earliest stages of development without involvement of *lin28a*. There are also differences during tailbud development and early neurulation, with *lin28b* found to have stronger anterior patterning at stage 13, and a presence in structures such as the eye and otic vesicle along with the greater expression dorsally. However, the role of lin28b during development still remains largely unstudied and unknown, and we cannot yet determine the importance of this wider expression.

Chapter 4. The function of lin28 proteins in *X. tropicalis*

4.1 Introduction

4.1.1 lin28 function in early vertebrate development

It is known that lin28a is an important factor in pluripotent cells, although its precise role still remains under investigation. Early work in *C. elegans* was able to generate lin28 overexpressing and null mutants. *lin28* recessive mutants, with a loss of lin28 protein, showed precocious development (Ambros and Horvitz, 1984), whereas a gain-of-function mutation caused the reverse and resulted in retarded development (Moss et al., 1997). However, in vertebrates this work has not yet been fully replicated. *In vitro* research has yielded contrasting results; work in both human and mouse ES cells found that the use of lin28 siRNA reduced cell viability, with an increase in apoptosis in human but not murine cells (Peng et al., 2011; Xu et al., 2009), and lin28 overexpression increased cell number, effects they found were due to changes in the cell cycle (Xu et al., 2009). On the other hand, alternative research found a downregulation of lin28 did not affect cell proliferation or the cell cycle in human ES cells, and overexpression actually slowed self-renewal (Darr and Benvenisty, 2009). Within *in vivo* work, attempts to generate lin28a knockout mice resulted in mice that were underdeveloped, and died shortly after birth (Zhu et al., 2010), indicating a requirement of lin28a expression for successful development. However, a lack of phenotype or cause for premature death was not provided by Zhu and colleagues, and it is unknown what effects the lin28a knockout was having.

The effects on development following increased lin28a and lin28b expression have, however, been more effectively studied through both transgenic mice and genetic studies within the human population. Transgenic lin28a overexpressing mice were found to have a higher body mass and bone density shortly after birth, and displayed delayed sexual development (Zhu et al., 2010), indicating a delay of the onset of adult development. In human studies, the LIN28B locus has been linked with determining the age of menarche (He et al., 2009; Ong et al., 2009), further linking an increase in expression with a delay in the onset of traits of adulthood.

4.1.2 Potential redundancy between the lin28 proteins

Expression of *lin28a* and *lin28b* was shown in sections 3.2.4 and 3.2.5 to overlap during many stages of development. The homology between lin28a and lin28b in the active RNA-binding domains means that the proteins may potentially target the same RNAs and have

a high level of redundancy. Both proteins have been shown to be capable of regulating let-7 biogenesis (Viswanathan et al., 2008), and it is possible this is also the case for further targets. Therefore, care must be taken when manipulating expression of the proteins to ensure effects can be properly attributable to them.

4.1.3 Function of lin28s during gastrulation

The importance of FGF signalling during gastrulation, and the increase particularly in *lin28a* expression at this time, may indicate a function for the lin28s during this stage of development. The localisation of both *lin28* genes to the presumptive mesoderm in gastrula stage embryos highlights this as an area where the lin28 proteins could be functioning, but a role has not yet been identified. As development of, and markers for, lineage specification has been well documented in *Xenopus* embryos these are ideal models in which to investigate the effects that the lin28s are having upon differentiation.

4.1.4 Aims

The aims of this chapter are to:

- To determine the effect of a knockdown and overexpression of lin28 proteins on *Xenopus* development.
- To uncover the roles of the lin28 proteins with respect to FGF signalling during gastrulation.
- To investigate whether the lin28s are regulating the let-7s or other miRNAs at gastrula stage.

4.2 Results

4.2.1 Effects of a knockdown of lin28 on normal development

To analyse the effect of the lin28 proteins in *Xenopus* development, mRNA expression can be knocked down using MOs; the MOs used against all three lin28s, both lin28a isoforms and lin28b, are translation blockers which bind to the 5' end of mRNA and prevent the binding of ribosomes. Injections were carried out of all three MOs together to reduce protein levels of lin28a1, lin28a2 and lin28b in a compound knockdown; this was carried out to avoid any protein redundancy preventing a complete loss of function that might be seen with a loss of just one protein. Injections of 10 ng/embryo of each MO allowed successful gastrulation, but there was a very high rate of death in subsequent development, and no embryos survived to tadpole stages for phenotypic analysis. The use

of a lower dose, using 5 ng/embryo of lin28a MOs and 2.5 ng/embryo of lin28b MO to a total of 12.5 ng/embryo, resulted in a 54% survival rate to tadpole stage, whereas survival rates in the CMO-injected and uninjected controls exceeded 90% and showed normal development (Figure 4.1A, B). Thus the reduced knockdown still impaired development of the embryos to this stage, but allowed some phenotypic analysis.

Injection of the three lin28 MOs at the mild dosage resulted in abnormal development, with less than 5% of embryos showing a normal phenotype (Figure 4.1C-F). Mild phenotypes were observed in 45% of cases, with severe defects occurring in over 50% of surviving embryos. Dorso-anterior structures were affected, with a reduction in head size and eyes, often resulting in a complete loss. A shortened axis was observed, demonstrating a failure of the tail to elongate, which was also curved in a number of cases. There was a failure of the somites to develop normally, appearing reduced or disorganised in embryos. An enlarged proctodeum was also apparent, with a high level of failure in blastopore closure.

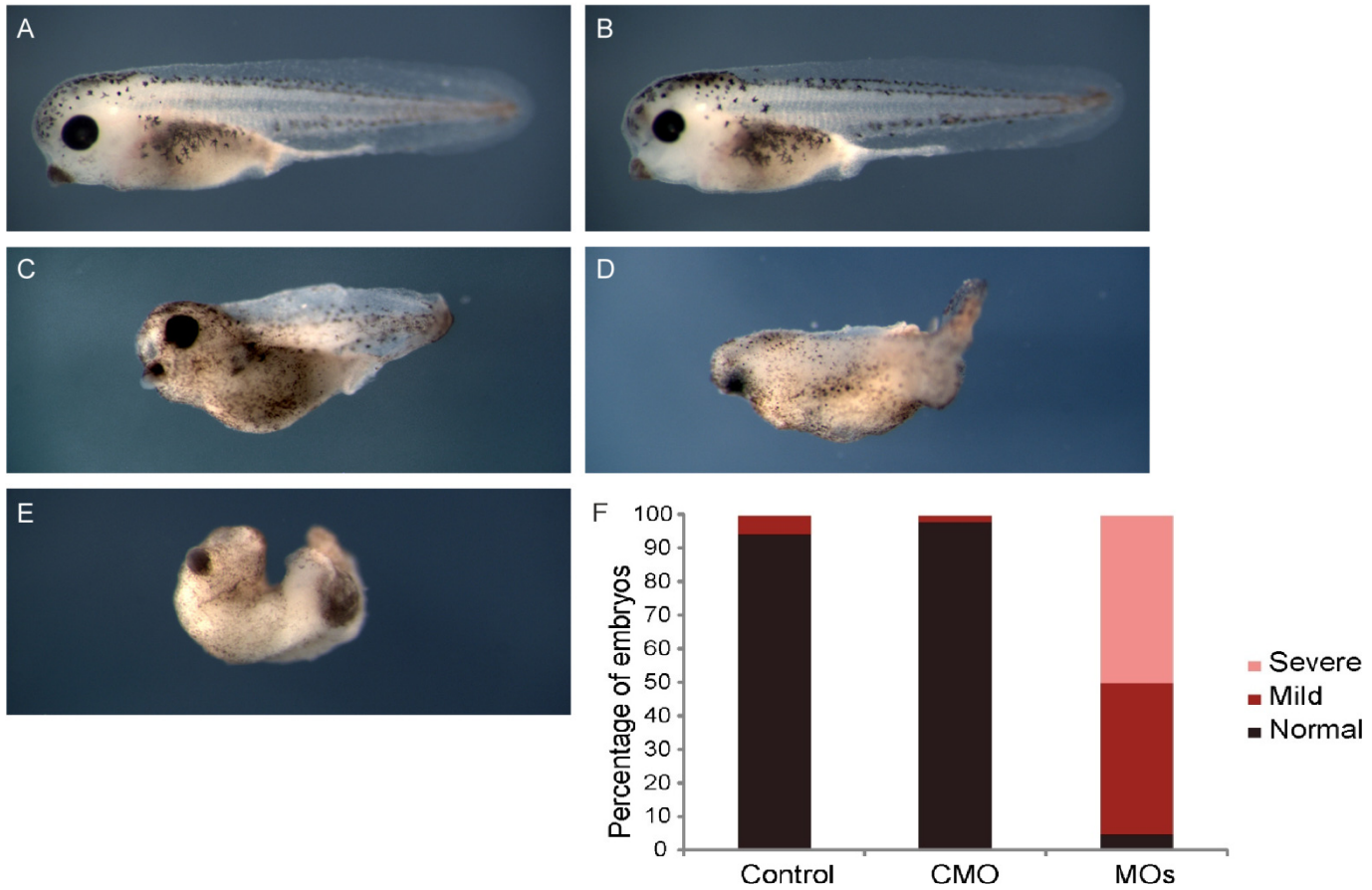


Figure 4.1 Effects of lin28 knockdown upon development

Phenotypes of *X. tropicalis* embryos at the same stage: A) uninjected control, B) CMO injected, C-E) injected with 12.5 ng total MO/embryo; all are representative of n = 3 experiments. F) Percentage of phenotypes classed as normal (A,B), mild (C) and severe (D, E), with numbers collated from 3 separate experiments (CMO n – 60, MOs n = 63).

The injections of MO were confirmed to be effectively targeting the lin28 proteins, with a loss of protein seen for the MO-injected embryos only, at stage 16, using western blot analysis (Figure 4.2). This stage was used to study knockdown as it was sufficiently late enough to detect high levels of endogenous lin28 proteins, and therefore see any reductions.

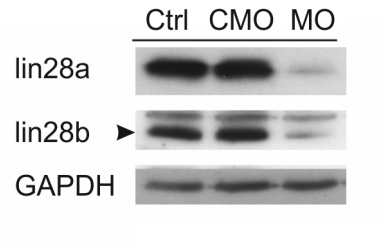


Figure 4.2 Analysis of lin28 expression upon MO treatment
Western blot analysis of embryos injected with a total of 12.5 ng/embryo of lin28 MOs in the compound knockdown compared to CMO injected and uninjected control embryos at stage 16. GAPDH was used as a loading control. For lin28b, arrowhead indicates relevant band.

4.2.2 Effect of lin28 proteins on germ-layer specification

FGF signalling is known to be important in patterning the embryo, particularly in specifying the early mesoderm. Due to the resultant phenotype following the knockdown of lin28 proteins together with the finding of the *lin28* genes being regulated by FGF, expression of early germ layer markers was investigated in gastrula stage embryos. This analysis was carried out using the dosage of 10 ng/each MO in a compound knockdown, with CMO at the same total amount/embryo.

Expression of the early mesodermal markers *brachyury* and *chordin* was examined in lin28 knockdown embryos, as both are also FGF targets (Amaya et al., 1993; Isaacs et al., 1994; Mitchell and Sheets, 2001). Analysis by qRT-PCR showed that expression of both genes was significantly downregulated in lin28 knockdown embryos compared to wildtype controls; *brachyury* showed a 58% reduction in expression, and 50% of the control expression of *chordin* was detected following lin28 knockdown (Figure 4.3). FGF signalling at this stage is also involved in neural specification; s*ox3* is a neuro-ectodermal marker and additional downstream target of FGF signalling (Rogers et al., 2008), which showed a significant decrease in expression with lin28 knockdown, being reduced to 36% of control RNA levels (Figure 4.3), indicating that this effect was not merely restricted to mesodermal targets. No change in expression was seen for *sox17b* (Figure 4.3), an endodermal marker (Hudson et al., 1997) which is not reported to be regulated by FGF signalling.

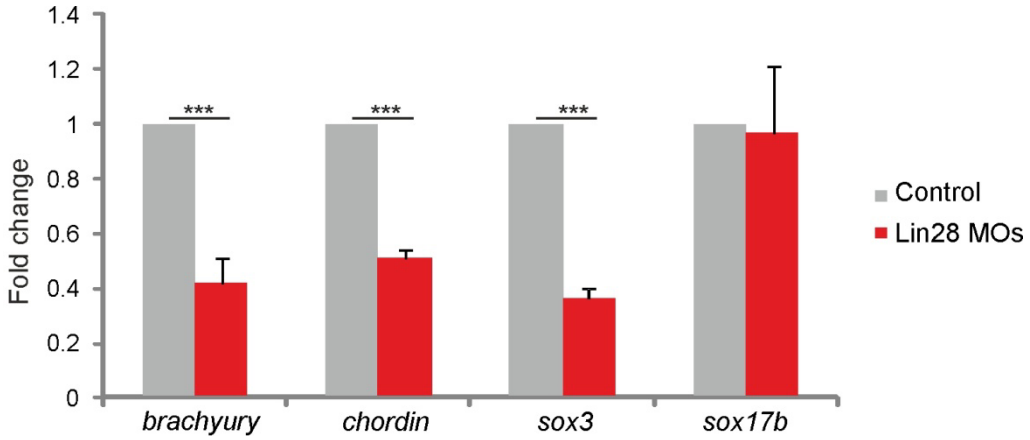


Figure 4.3 Effects of lin28 on RNA expression of germ layer markers

qRT-PCR was performed on embryos injected with 10 ng each/embryo of lin28a1, a2 and b MOs and control embryos, at stage 10.5. Fold change in expression of RNAs is shown compared to controls and normalised to *ODC* by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of three biological replicates and error bars represent SE. Significance calculated by *t*-test: ***p<0.001.

4.2.3 Identifying miRNA targets of the lin28s

To investigate whether the lin28 proteins were regulating the expression of any miRNAs at this time, an Affymetrix miRNA microarray was undertaken. Embryos subjected to the compound lin28 knockdown were collected and compared to control embryos at gastrula stage. To allow statistical analyses on the data this was carried out with three biological replicates, resulting in six microarray datasets. The microarray chips contained probes to all known miRNAs at the time, in all species, and as such a large number of probe sets were available for analysis, with 7626 probes in total on the chips.

Data was processed using QC tools (Affymetrix), with a background detection and adjustment and normalisation applied. This output provided a 'detection call', indicating which miRNAs were detected as being expressed above background; the detection call was then used to remove miRNAs that were not detected as expressed in any sample. Those expressed in any of the six samples were retained for analysis, as they may have shown interesting changes in regulation. The number of probes detected above background in any samples came to 4294. This, however, included multiple control probes and a small number of hits in some irrelevant species, such as *Zea mays*, the maize plant. Filtering the results for *X. tropicalis* genes, 87 miRNAs were detected across the microarrays, from 167 *X. tropicalis* probes present on the chips (Appendix 1). Of these, 33 were detected above background in all six microarrays, and a further three expressed in all control samples at least, suggesting that these miRNAs would be ones most likely to be expressed during this stage of development.

Normalised data were imported to BRB-tools to allow statistical analysis. Interestingly, when attempting to apply clustering analysis, the closest expression patterns were found between sibling embryos despite treatments, rather than those with the same manipulations, indicating that miRNA expression may differ vastly between unrelated embryos at this time. As such, limited clustering analysis could be carried out. Another common method of analysis with array datasets is to apply a 2-fold change threshold to the data, but this was not sufficient to detect many targets within this dataset as the changes seen between conditions were generally less than 2-fold altered. This highlights the fact that the majority of miRNAs detected in these embryos did not change with the knockdown of lin28.

After exhausting analysis options as above, an alternative method was employed, using paired statistical *t*-tests to compare the three data sets, with a filter applied for only the *X. tropicalis* genes. This was to ensure that any miRNAs detected were biologically relevant, for example multiple plant species were also present on the array, which may have been falsely detected. When applying a *t*-test using a significance value of 0.01, only 2 genes were detected as showing a significant difference, so this was expanded to a 0.05 significance level. At this level, 10 miRNAs were found to change significantly (Table 4.1). These were all miRNAs that were present above background levels in each of the microarray datasets, and therefore deemed to be truly expressed (Appendix 1).

Of these 10 miRNAs, 8 are members of the group of miR-17-92 clusters (Tanzer and Stadler, 2004) (miR-18b, miR-17-5p, miR-20b, miR-106, miR-20a, miR-18a*, miR-363-5p, miR-18a), which all showed downregulation upon a loss of lin28 proteins.

Due to the highly conserved nature of miRNA, where a sequence is completely conserved between multiple species the same probe sequence is therefore present multiple times on the array, under gene names that differ by their species prefix. Re-running the analysis without a species filter, at 0.01 significance, now detected 31 miRNAs (Table 4.2). Of these, 23 were from the miR-17-92 clusters, which were the same members as detected by *X. tropicalis* probes with the addition of miR-92. There was no further detection in any species of miR-16c or miR-24a, the additional miRNAs identified in *Xenopus*, which are conserved in other species on the chips. This lent strength to the hypothesis that the members of the clusters were targets of lin28 at this time.

Table 4.1 Fold change data for putative *X. tropicalis* miRNA targets of lin28

miRNA	Fold change	*t*-test p-value
xtr-miR-18b	-1.63	0.005196
xtr-miR-17-5p	-1.80	0.007282
xtr-miR-20b	-1.63	0.010716
xtr-miR-106	-1.57	0.017647
xtr-miR-24a	+1.20	0.021995
xtr-miR-20a	-1.39	0.025513
xtr-miR-18a-star	-2.08	0.02951
xtr-miR-363-5p	-2.88	0.033074
xtr-miR-18a	-1.54	0.036532
xtr-miR-16c	-1.18	0.039499

Fold change between control and lin28 compound knockdown embryos. Data calculated using BRB-tools, using paired *t*-test at p<0.05 significance, and probe list filtered for *X. tropicalis* genes. +/- show increase/decrease in expression following lin28 knockdown.

Table 4.2 Fold change data for putative miRNA targets in all species of lin28

miRNA	Fold change	*t*-test p-value	miRNA (cntd)	Fold change	*t*-test p-value
fru-miR-20	-1.54	4.90E-06	**ptr-miR-106a**	-1.84	0.004085
mne-miR-92	-1.72	0.000498	**hsa-miR-125a-3p**	-1.12	0.004162
mmu-miR-106a	-1.56	0.000546	**gga-miR-106**	-1.67	0.004618
bta-miR-18a	-1.75	0.000918	**sla-miR-17-5p**	-1.88	0.004663
rno-miR-20a	-1.52	0.001067	**mmu-miR-17**	-1.88	0.004726
mmu-miR-327	+1.23	0.001333	**xtr-miR-18b**	-1.63	0.005196
sla-miR-106a	-1.71	0.001347	**mml-miR-92a**	-1.77	0.005979
hsa-miR-1300	+2.04	0.00164	**lca-miR-18**	-1.78	0.006322
mmu-miR-26a	+1.12	0.001793	**ppa-miR-106a**	-1.74	0.00641
ssc-miR-106a	-1.93	0.001806	**cfa-miR-92a**	-1.63	0.007115
gga-miR-20a	-1.39	0.002636	**lla-miR-92**	-1.75	0.007158
mmu-miR-365	-1.47	0.003219	**xtr-miR-17-5p**	-1.8	0.007282
ptr-miR-92	-1.77	0.003239	**cfa-miR-193a**	+1.56	0.007537
dre-miR-184	+3.84	0.003511	**mml-miR-106a**	-2.02	0.008696
lla-miR-17-5p	-1.99	0.004034	**ppy-miR-17-3p**	-1.45	0.009442
bta-miR-148a	+3.23	0.004038			

Fold change between control and lin28 compound knockdown embryos. Data calculated using BRB-tools, using paired *t*-test at p<0.01 significance. +/- show increase/decrease in expression following lin28 knockdown.

4.2.4 Selection of miRNA housekeeping for qRT-PCR

qRT-PCR was carried out to validate these miRNAs as lin28 targets. A housekeeping gene is required for normalising expression between treatments, with mRNA targets being unsuitable for this process due to their length and the method of miRNA qRT-PCR.

However, when attempting to find a suitable gene for use with *X. tropicalis*, many of the published possible housekeeping genes, such as 5.8S rRNA (Galiveti et al., 2009), were found not to be conserved to *Xenopus*, and primers for the homologues were not commercially available and guaranteed for specificity. The snRNA U6 has also previously been validated as a housekeeping gene for miRNA analysis (Galiveti et al., 2009); U6 is a small RNA which is involved with controlling mRNA splicing along with U1, U2, U3, U4 and U5 (Thomas et al., 1990), and is conserved between humans and *Xenopus*, making these primers suitable for use. However, the suitability of U6 for a housekeeping gene in this experimental condition must be validated, as there was no probe to it on the microarray.

From the microarray, miR-130b was identified as a highly expressed RNA with expression not altering upon embryo manipulation. It is also independent of the identified miR-17-92 clusters and the let-7 family miRNAs, known lin28 targets. Expression of this was assayed using qRT-PCR in lin28 knockdown embryos compared to controls, and normalised using U6. Here, we saw no difference in expression between the conditions (Figure 4.4), implying that both U6 and miR-130b are suitable controls for these experiments.

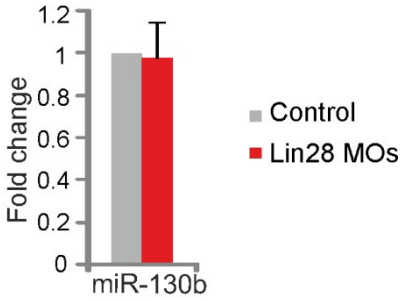


Figure 4.4 Testing the suitability of miRNA housekeeping genes
qRT-PCR was performed on embryos injected with 10 ng each/embryo of lin28a1, a2 and b MOs and control embryos, at stage 10.5. Expression of miR-130b was normalised using U6 by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of 3 biological replicates, with error bars representing SE.

4.2.5 Effects of lin28 knockdown on levels of mature let-7 during gastrulation

As the let-7 family of miRNAs are known negatively regulated targets of lin28, these were the first miRNAs of interest to be investigated, despite not showing significant changes on the microarray. Levels of mature let-7 were not detected above background levels on the microarrays, due to low level expression. As qRT-PCR is a more sensitive technique to investigate expression of single miRNAs, this was used to investigate levels of let-7 family members let-7a, let-7f and let-7g following knockdown of lin28.

No change in expression from control levels was detected for any of the let-7 family members at stage 10.5 following a loss of the lin28s (Figure 4.5). This method also found that expression of the mature let-7s was at a very low level, with expression only detected following a high number of amplification cycles.

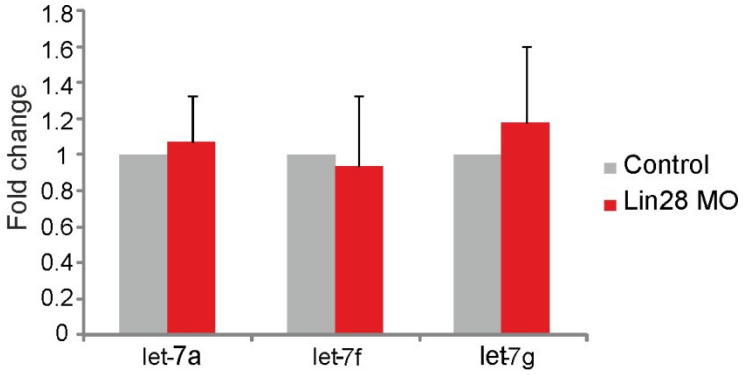


Figure 4.5 Effect of lin28s on let-7 family members by qRT-PCR analysis
qRT-PCR was performed on embryos injected with 10 ng each/embryo of lin28a1, a2 and b MOs and control embryos, at stage 10.5. Fold change in expression of miRNAs is shown compared to controls and normalised using U6 by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of 3 biological replicates, with error bars representing SE.

4.2.6 Effect of lin28 knockdown on members of the miR-17-92 clusters

As changes were seen in multiple members of the miR-17-92 clusters on the arrays, expression of miR-17-5p, miR-19b-3p, miR-106-5p, miR-18b-5p, miR-20b-5p, miR-363-3p and miR-363-5p was investigated using qRT-PCR. With the exception of miR-363-5p (which will be referred to as miR-363*), all miRNAs studied were the previously termed 'major' form of the pre-miRNA, and will be referred to without the 3p/5p reference from here.

Comparing expression after lin28 compound knockdown to that in control embryos, expression levels of all cluster miRNAs investigated were significantly decreased with the

loss of the lin28s (Figure 4.6). Smaller changes were seen in miR-106, with expression reduced to 80% of control with the use of lin28 MOs, with the largest effects being an expression level of less than 50% of control for miR-20b and miR-363, and a 75% reduction in expression of miR-363* (Figure 4.6).

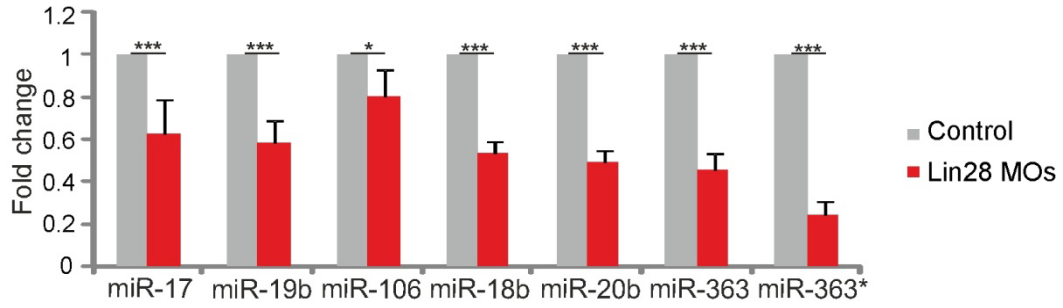


Figure 4.6 Effect of lin28s on miR-17-92 and miR-106-363 cluster members by qRT-PCR analysis

qRT-PCR was performed on embryos injected with 10 ng each/embryo of lin28a1, a2 and b MOs and control embryos, at stage 10.5. Fold change in expression of miRNAs is shown compared to controls and normalised using U6 by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of 3 biological replicates, with error bars representing SE.

4.2.7 Effect of FGF signalling on miR-17-92 cluster miRNAs

As FGF signalling was able to influence the expression of *lin28a* and *lin28b* at this point in development, this led to the investigation of whether these putative miRNA targets of lin28 are also downstream of FGF signalling at this time. The levels of these miRNAs were subsequently investigated in embryos with inhibited FGF signalling, through the overexpression of dnFGFR.

Following injection of 1 ng *dnFGFR*, the miRNAs are not uniformly reduced in expression as seen with lin28 MOs. The only miRNAs to show a reduction in expression were miR-18b and miR-363, which were reduced by 55% and 35% respectively, a smaller reduction in expression for miR-363 but a greater decrease for miR-18b compared to changes after lin28 manipulation. No changes were seen in miR-17, miR-19b, miR-18b and miR-363*, while a significant increase was seen in miR-106 expression (Figure 4.7).

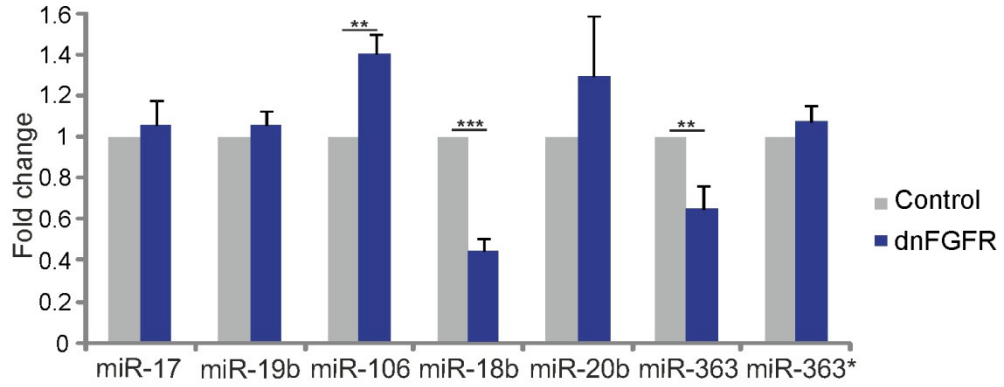


Figure 4.7 Effect of FGF signalling on miR-17-92 and miR-106-363 cluster members by qRT-PCR analysis

qRT-PCR was performed on embryos injected with 1 ng/embryo of *dnFGFR* and control embryos, at stage 10.5. Fold change in expression of miRNAs is shown compared to controls and normalised using U6 by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of 3 biological replicates, with error bars representing SE.

4.2.8 Effects of lin28 overexpression on development

lin28 proteins were overexpressed individually in embryos. 1 ng RNA for the coding region of each of lin28a1, lin28a2 and lin28b was used, to prevent UTRs allowing regulation within the embryo and preventing overexpression. The overexpression of proteins in gastrula embryos was confirmed using western blot (Figure 4.8).

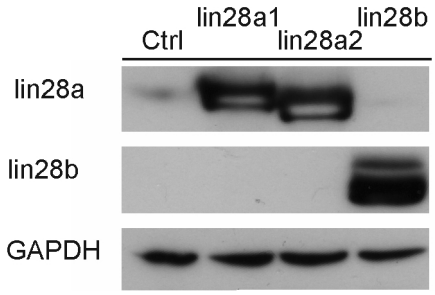


Figure 4.8 Analysis of lin28 protein overexpression by western blot

Western blot analysis of embryos injected with 1 ng/embryo of *lin28a1*, *a2*, or *b* individually compared to control embryos, at stage 10.5. GAPDH was used as a loading control.

However, in contrast to the knockdown of lin28, the overexpression of these proteins did not result in a disrupted phenotype. Development of the embryos to tadpole stage appeared to be unaffected when compared to the phenotypes of controls (Figure 4.9).

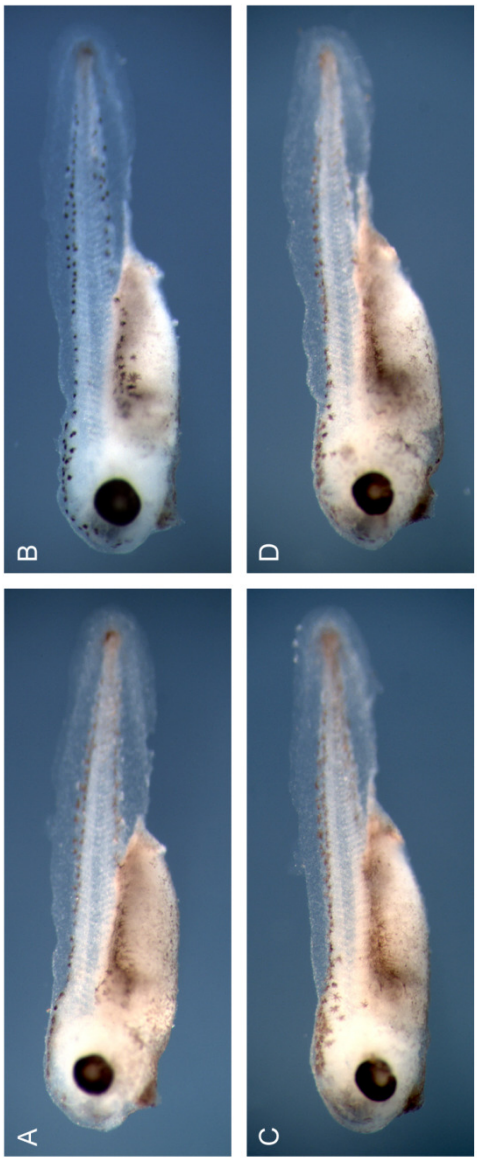


Figure 4.9 Effects of lin28 overexpression upon development
Phenotypes of *X. tropicalis* embryos at the same stage A) uninjected control, or embryos injected with 1 ng either B) *lin28a1* RNA C) *lin28a2* RNA, D) *lin28b* RNA. All are representative of n = 2 experiments.

4.2.9 Effect of lin28 overexpression on miR-17-92 clusters

Despite the lack of a developmental phenotype, it remained possible that an overexpression of lin28 could alter levels of the miR-17-92 and miR-106-363 cluster miRNAs. Levels of mature miRNA from the miR-17-92 and miR-106-363 clusters were analysed by qRT-PCR for each overexpressing condition. Expression of miR-130b was also used as an independent control, not being related to the clusters. Mature levels of all miRNAs were not altered significantly upon lin28 overexpression (Figure 4.10). miR-130b, previously used as a negative control, was actually the miRNA which showed the highest degree of change in level, being increased following lin28a overexpression.

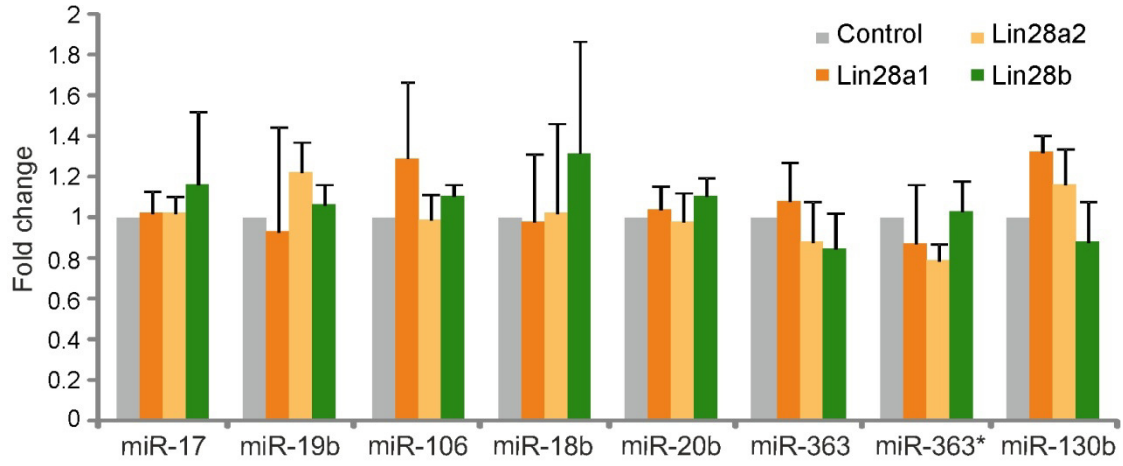


Figure 4.10 Effect of lin28 overexpression on miR-17-92 and miR-106-363 cluster members by qRT-PCR analysis

qRT-PCR was performed on embryos injected with 1 ng/embryo of *lin28a1, a2,* or *b* RNA and control embryos, at stage 10.5. Fold change in expression of miRNAs is shown compared to controls and normalised using U6 by the $2^{-\Delta\Delta Ct}$ method. Fold change is given as average of 3 biological replicates, with error bars representing SE.

4.3 Discussion

4.3.1 The lin28 proteins are important in germ layer specification

Due to the early lethality of lin28 mutations in vertebrates, and the ease of following *Xenopus* development, work here and further research from the group, presented in Faas et al. (2012), is a first early vertebrate *in vivo* study into the downstream effects of lin28 proteins following a loss of expression.

The knockdown in expression of lin28 proteins resulted in abnormal development, which is likely to result from dysregulation as early as gastrulation. Results here have shown that a loss of these proteins by stage 10.5 reduced expression of early germ layer markers for both mesoderm and neuro-ectoderm, indicating that these proteins are required for an

important inductive role by this early stage. Further work in our lab has demonstrated that significant changes in expression of mesodermal markers are also seen at stage 13, using a wider range of mesodermal markers(Faas et al., 2012). Whether expression of these markers is rescued later in development in these knockdown embryos has not yet been determined, however the resultant phenotypes suggest that this manipulation severely alters gene expression and subsequent development, and does not just result in retarded development, although a heterochronic effect can be seen upon other morphogens, such as Smad2 and some FGF ligands (Faas et al., 2012). This effect on development through the loss of key germ layer markers can be seen in the resultant phenotypes; a loss of mesodermal markers is reflected in problems seen in the development of the somites, which appeared reduced and disorganised. Cartilaginous structures were also affected, including the notochord, which would contribute to the shortened body axis, and the branchial arches, where *lin28* expression was detected. The reduction in *sox3* expression was reflected in a reduced development of neural and head structures, and the abnormal development of structures such as the eye reflects expression of *lin28b* here, as seen in Chapter 3.

As a heterochronic gene in *C. elegans*, lin28 is known to control the timing of the progression of larval fates, and carries out this function via multiple targets and mechanisms (Vadla et al., 2012). In vertebrates, its role has been less clear. As lin28 is downregulated upon differentiation (Moss and Tang, 2003; Yang and Moss, 2003), and with the ability to reprogramme cells (Yu et al., 2007), it has largely been classed as a pluripotency factor, with any role in specifying lineage differentiation largely unknown. Early work demonstrated a rapid increase in let-7 at the onset of differentiation (Thomson et al., 2006), which suppresses self-renewal (Melton et al., 2010); it was as such thought that the role of lin28 in specifying cell fate was merely preventative via regulation of let-7, but new research is beginning to indicate a wider role with an involvement in cell fate, as work has seen here.

The results from such investigations both support and contrast results shown in this chapter. Studies have found that lin28, known to be expressed in undifferentiated cells, actually increases during the first few days of embryoid body differentiation, and this is critical for subsequent cell fates (Kawahara et al., 2011; Wang et al., 2012). This is also consistent with our increased detection of lin28 expression during gastrulation, seen in Chapter 3. An involvement in neural development is likely, as lin28a has been found to persist in early neural tissues of the mouse embryo (Balzer et al., 2010; Yang and Moss, 2003), and overexpression of lin28a or lin28b were both found to increase neuronal differentiation and block glial specification, by separate mechanisms (Balzer et al., 2010).

Mutations in lin28 proteins actually decreased the number of neural stem cells differentiating into neurons, indicating that a positive role in the regulation of this specification is likely to be played by these proteins (Balzer et al., 2010), and this idea has been supported by work in embryoid bodies (Kawahara et al., 2011). The known role of FGF signalling in neural differentiation, with *sox3* an important gene for this, supports the possibility for a neural function of lin28. However, alternative results suggest that lin28 is not involved in the specification of neuronal differentiation (Wang et al., 2012), and is instead involved in the other germ layers of endoderm and mesoderm. Nevertheless, research in these alternative areas has also come to two conflicting conclusions; lin28 is regulated by miR-125a in embryoid bodies, and this regulation has been linked to subsequent endodermal, mesodermal and cardiac differentiation (Wang et al., 2012; Wong et al., 2012). Work by Wang et al (2012) showed an initial decrease in expression of miR-125 during days 3-5 of embryoid body differentiation, in tandem with the increase in lin28 at this time; they found that overexpressing miR-125, thus decreasing lin28, caused decreases in endodermal markers, such as *sox17*, mesodermal markers including *brachyury* and *mesoderm posterior 1 (mesp1)*, and cardiomyocyte markers, and expression of all could be rescued by re-introducing lin28. This indicated the necessity for some lin28 expression in early differentiation specification. Work here and in Faas et al. (2012) supports the role of lin28 in mesodermal development, however this work is in contrast to the published effect by Wang and colleagues (2012) upon endodermal development, even though investigating the same marker in *sox17*. On the other hand, Wong and colleagues (2012) were led by the discovery that miR-125 is substantially increased in expression from day 8 of differentiation compared to undifferentiated cells, although their results also showed the initial decrease in expression in the earlier days. Here they found that overexpressing miR-125 in undifferentiated cells, thus reducing lin28, had an opposite effect, prematurely increasing expression of cardiac markers, with increased expression in undifferentiated cells of *brachyury* and decreased expression of endoderm and pluripotency markers. Further investigation is therefore needed to clear the waters on the role of lin28 in promoting any particular cell fates.

Brachyury, chordin and *sox3* are known to be FGF regulated targets (Amaya et al., 1993; Mitchell and Sheets, 2001; Rogers et al., 2008), and it is possible that the ability to affect these genes is connected to the regulation by FGF, with lin28a and lin28b having a role in mediating FGF signalling at this time. Animal cap experiments have supported this notion; animal cap explants from the embryo normally form unlayered epidermis when left to develop alone, but the addition of FGF or activin to the caps results in the differentiation of mesoderm. Following knockdown of lin28 proteins, animal caps were inhibited in

mesoderm development with FGF or activin treatment, indicating a reduced response to inductive signals (Faas et al., 2012). The current hypothesis is that lin28 works as a competence factor to FGF signalling, and may be modulating FGF signalling. However, the mechanism by which this is occurring has not yet been fully confirmed; a reduction in both dpERK and FGF ligand expression is seen following a loss of lin28 proteins, but how precisely lin28 causes this effect is unknown (Faas et al., 2012).

Interestingly, an overexpression of the three proteins independently failed to result in an abnormal phenotype. It is possible that the endogenous levels of lin28 in the embryo at this time of gastrulation are sufficient for its function, such that an overexpression cannot further alter downstream effects. Any negative gene targets may already be suppressed to the maximum effect, and where involved with potentiating translation or any positive regulation, this may also be occurring already to maximum capacity. Further understanding of this mechanism by which lin28 functions at this time needs to be gained, to properly interpret this lack of phenotype. This does not match previous *in vivo* work, with mice showing a higher body mass and delayed sexual development with an overexpression of lin28a (Zhu et al., 2010). It is further possible that an overexpression of the lin28 proteins is unable to function in a usual manner within the embryos, and therefore we are not able to mimic accurately an actual gain-of-function.

4.3.2 Functional redundancy between the lin28 proteins

The overlap of expression patterns for *lin28a* and *lin28b* seen during particular stages of development in Chapter 3 may allow for functional redundancy, and due to this all experiments were carried out using a triple knockdown. Evidence of redundancy can be seen in knockdown embryo phenotypes, where a more severe phenotype resulted from a knockdown in the *lin28a*s and *lin28b* together, compared to knocking down either protein individually (Faas et al., 2012). With the multiple areas of co-expression of *lin28a* and *lin28b*, there is a high chance that the proteins may be expressed in the same cells, and therefore redundancy of function could be anticipated. lin28a and lin28b have previously been classed as functioning redundantly, due to their regulation of let-7 miRNAs, however recent research has indicated that they may be regulating this by different mechanisms: lin28b was found to act primarily in the nucleus, binding and sequestering the pri-miRNA, whereas lin28a acted in the cytoplasm on the pre-miRNA (Piskounova et al., 2011). These differences indicate that there may be further important distinctions in the roles and functions of these proteins, not least due to their possible dissimilarity in subcellular locations. As the effects seen here were due to a loss of all lin28 proteins, more careful

analysis would need to be carried out to determine if targets identified are downstream of both lin28a and lin28b, or if any are effects specific to either protein.

4.3.3 The function of lin28s in gastrula stage embryos is independent of let-7 regulation

lin28a and b proteins were not found to be inhibiting levels of let-7 miRNA during gastrulation. However, the levels of mature let-7 miRNAs were very low at this developmental stage, being difficult to detect by both microarray and qRT-PCR.

The current knowledge of let-7 expression in model systems varies. Prior analysis in *Xenopus* reported let-7f present in the blastula in precursor form, but by northern blot this appeared to be the precursor form, and levels of the mature RNA only began to increase from tailbud stages (Watanabe et al., 2005). In *C. elegans*, expression of let-7 was detected weakly from the L3 stage, and was increased in L4 and adult stages (Reinhart et al., 2000), much later stages than the embryonic gastrulation. Pasquinelli et al (2000) found that let-7 expression in *Drosophila* began just before metamorphosis, in the late third instar phase, and that zebrafish expression occurred between 24 and 48 hrs post fertilisation and remained into adulthood. However, work in mammalian stem cells has found early embryonic expression; pre-let-7a was detected in ES cells, but not in the mature form (Newman et al., 2008; Suh et al., 2004), indicating transcription was at least occurring. Further to this, a knockdown of lin28 in mouse ES cells proved sufficient to increase let-7 levels (Hagan et al., 2009).

It remains possible the let-7 precursors and primary forms are being expressed during gastrulation; however the situation may be that transcription of these is at a very low level in *Xenopus* at this time and the lin28s are providing minimal regulation of the miRNAs. In extension to this, other regulators of let-7 processing may be present which still inhibit its maturation despite the loss of the lin28s. Further studies to investigate levels of pri- and pre-let-7s at this stage, such as by northern blots, would provide a clearer understanding of the levels of processing occurring at this stage of development. During later development, however, it appears that a knockdown of lin28 can influence the levels of let-7; at stage 22, increases in the mature levels of let-7g were seen following knockdown of lin28 proteins (Faas et al., 2012).

Recent work has found, however, that in *C. elegans* lin28 carries out its earliest functions, the specifying of L2 fates, via a mechanism independent of let-7 (Vadla et al., 2012). This paper found that lin28 was supporting expression of an mRNA, *hbl-1*, involved in

regulating L2 fates, through its 3' UTR (Vadla et al., 2012). Indeed, in mammalian stem cell work, lin28 has been found to further act independently of let-7, by binding to mRNA and enhancing their translation, targets of which include IGF-2, histone H2A, stem cell genes such as Oct-4, and cell cycle genes including cdk4 and cyclin B (Balzer et al., 2010; Polesskaya et al., 2007; Qiu et al., 2010; Xu and Huang, 2009; Xu et al., 2009). The work by Vadla and colleagues (2012) investigated if lin28 was affecting the expression of other miRNAs to let-7 during this early stage of development, but did not find evidence to suggest that it was. In support of this, we did not find changes in many miRNAs expressed at this time; changes were only notable in members of the miR-17-92 clusters, which have not been discovered in *C. elegans.*

4.3.4 The miR-17-92 cluster miRNAs as novel targets of the lin28s

Work in this study identifies putative miRNA targets of the lin28 proteins are members of the miR-17-92 clusters. In the *Xenopus* genome there are four paralogous clusters - the miR-17-92 cluster, the miR-106a-363 cluster, and the two miR-93-25 clusters (Figure 4.11). These miRNAs are highly conserved and can be grouped into 4 families within the cluster, based on conserved seed sequences (reviewed in Mendell, 2008). These clusters are discussed further in Chapter 6.

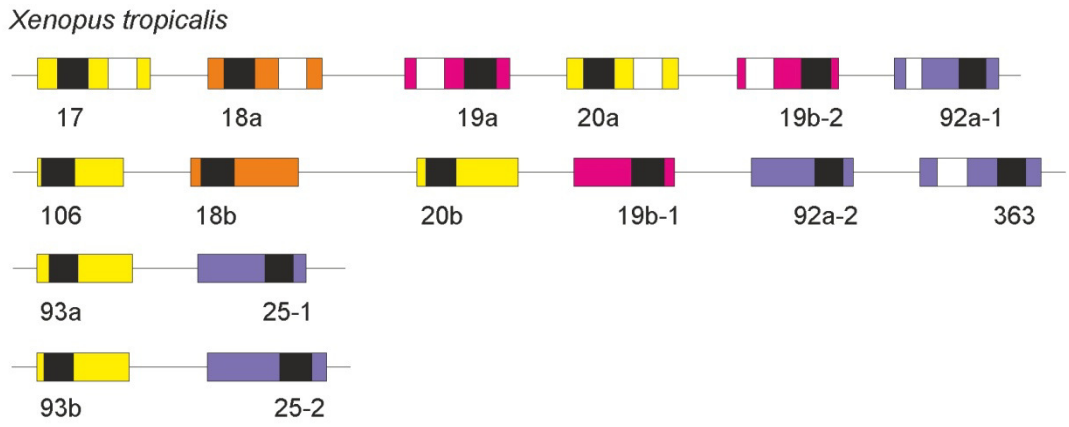


Figure 4.11 Arrangement of miR-17-92 clusters in *X. tropicalis*
Scale diagram of arrangement of pri-miRNA for miR-17-92 and paralogous clusters. Coloured boxes show precursor sequence, with colour corresponding to family groupings based on seed sequence. Black box represents the 'major' mature miRNA, and white box the 'minor' form, where known.

Multiple members of the miR-17-92 and miR-106-363 clusters were identified as lin28 targets following microarray analysis, and some of these were followed up using qRT-PCR. Targets were selected to ensure that members from each of the miRNA families were

pursued. The largest difference was seen in miR-363 and miR-363*, which are both from the same precursor.

Interestingly, a loss in expression of all of these miRNAs with a knockdown in lin28 protein suggests that the miRNAs may be positively regulated targets of lin28. The only known method of miRNA regulation currently undertaken by lin28 is negative, preventing processing of the let-7s to their mature form (Piskounova et al., 2008; Viswanathan et al., 2008), with effects on mRNA involving potentiating translation (Polesskaya et al., 2007), which is unnecessary with miRNA. This would indicate that lin28 must be acting in an alternative mechanism to any that have previously been determined.

4.3.5 The lin28 proteins are necessary for miR-17-92 and miR-106-363 miRNA expression

The loss of expression of miRNA at the mature level following a knockdown of lin28 protein suggests that the proteins are necessary to maintain correct expression of these miRNAs; however, mature miRNA levels were not altered upon overexpression of lin28 proteins, and the mechanism by which this regulation is occurring needs to be understood.

There is a possibility that lin28 is involved in promoting or assisting the processing of these miRNAs, and there may be insufficient levels of pri-miRNA or pre-miRNA for an overexpression of protein to increase expression levels at the mature stage; further investigation into the levels of pri-miRNA and pre-miRNA at this stage would be important to understand this. Future experiments could also involve overexpressing pri-miRNA for the clusters, both with and without lin28 proteins, and monitoring if this is able to affect levels of mature miRNA.

4.3.6 Effects of FGF signalling on miR-17-92 cluster miRNAs is unclear

Interestingly, inhibition of FGF signalling did not result in the same effects upon miRNA from the miR-17-92 clusters as for the knockdown of lin28 protein. Two members of the clusters were reduced in expression, which were also highly reduced with a loss of lin28s, but there were either no changes or an increase in the other members studied.

Inhibition of FGF signalling was seen in section 3.2.2 to decrease expression of *lin28a* and *lin28b* at the message level, but this could not be confirmed at the protein level with difficulties existing in detecting the endogenous lin28 proteins in gastrula stage embryos, with endogenous protein present but undetectable consistently by western blot, and as such no maintenance or loss of expression could be determined. It is possible, therefore,

that although some transcriptional regulation of the *lin28* genes is occurring as a result of FGF signalling, following pathway inhibition the lin28 proteins may still be present in sufficient abundance to carry out their normal roles to some degree, and are thus able to still carry out their role of regulating these miRNAs.

There is a contrast between the responses seen in miRNA genes and mRNA genes with upstream manipulations, with mRNA targets such as *brachyury* reduced similarly following both FGF signalling inhibition and a loss of lin28 proteins, with a potential that if lin28 is a mechanism by which these genes are regulated, the loss of FGF signalling was enough to drive this loss of regulation by the lin28s. However with miRNAs from the miR-17-92 and miR-106-363 clusters not showing the same reduction with FGF inhibition as for lin28 knockdown, there is a possibility that this is not linked to any role lin28 has in mediating the effects of FGF signalling, or that it is an easily mantained function of the lin28s, despite loss of other functions.

Chapter 5. The response of human MSCs to FGF treatment

5.1 Introduction

5.1.1 Are FGF signalling targets conserved to humans?

Much of the work on identifying FGF targeted gene expression has been undertaken in animal model systems, such as *Xenopus*, and it may be that genes identified as targets in these systems are conserved to humans and are contributing to the effects mentioned in previous chapters. The microarray based transcriptomic analysis by Branney *et al*. (2009) identified a large list of both putative and previously confirmed FGF targets in gastrula stage embryos. At this stage in development, specification of the germ layers is beginning, and FGF is predominantly involved with the differentiation of mesoderm. Particular genes of interest are *lin28a* and *lin28b*, with their known roles in pluripotency and their identification here as targets of FGF as described in Chapter 3. With a requirement of the lin28 proteins in *Xenopus* for correct patterning of the embryo during gastrulation, particularly in the control of expression of mesodermal markers as shown in Chapter 4, these genes may be important for competence to differentiate. The wider roles of FGF signalling in human cell culture are still under analysis, and the way in which FGF signalling is functioning in human cells remains unclear. More analysis is required to determine whether characterised FGF targets, with functions which are well understood, are also FGF targets in stem cell culture.

5.1.2 FGF-2 in cell culture

As discussed in section 1.1.6, the growth of human ES cells has been shown to require supplementation with FGF-2 to maintain pluripotency (Amit et al., 2000; Dvorak et al., 2005); this effect may be elicited through the antagonism of BMP signalling (Xu et al., 2005). However, the lack of requirement for FGF-2 in murine ES cell culture, and FGF treatment leading commonly to differentiation (Villegas et al., 2010; Ying et al., 2003), highlights both that FGF signalling is responsible for different outcomes dependent on the cell state, and that there are functions of FGF-2 which are yet to be fully understood in regards to proliferation and differentiation.

The culture of MSCs has not shown a requirement for exogenous FGF-2 to maintain undifferentiated growth. Work has found, however, that supplementing MSC growth media with FGF-2 increases the rate of proliferation of the cells and enhances

chondrogenic and osteogenic potential (Martin et al., 1997; Pri-Chen et al., 1998; Solchaga et al., 2005; Tsutsumi et al., 2001). The mechanisms behind these effects have not been fully determined, but are likely to be occurring through activation of a number of pathways, including phosphorylation of ERK downstream of the MAPK pathway (Choi et al., 2008).

MSCs were selected for the investigation of conserved FGF signalling, due to the flexible nature of their culturing both with and without FGF-2 supplementation. Also, these cells being already restricted to a mesenchymal fate means they may be of use in determining FGF targets that are important in further mesodermal development. Studies into the impact of FGF-2 on the culture of MSCs have previously investigated treatment over multiple days or weeks, and so rapid downstream targets of the pathway were likely to not have been detected. These rapidly activated targets could be vital for long-term roles and as such their identification may provide further insight into the functions of FGF-2 in cell culture.

5.1.3 Aims

The aims of this chapter are to:

- Determine if MSCs are capable of responding to exogenous FGF-2 treatment.
- Identify a suitable dosage and timing of FGF treatment of MSCs for activation of the signalling cascade and transcriptional changes to occur as a result.
- Investigate putative FGF targets in MSCs and conservation of these to *Xenopus*.
- Study expression of *lin28a* and *lin28b* in MSCs.

5.2 Results

5.2.1 MSCs are capable of responding to FGF

To respond to exogenous signalling ligands, cells must express the necessary receptors which in this case are members of the FGFR family. RT-PCR was carried out on three donor samples to determine expression of *FGFR1* and *FGFR2*. These were selected due to their common use as receptors for a number of FGF ligands, and are preferred by FGF2 (Zhang et al., 2006). Due to the existence of numerous alternative splice variants, primers were used that were capable of detecting multiple forms of each FGFR, resulting in different sized PCR products from these isoforms. Primers to *FGFR1* are capable of detecting transcript variants: 11(1162 bp), 1 and 10 (1094 bp), 2, 12 and 14 (1088 bp), and 3, 4 and 7

(827 bp). *FGFR2* primers can work to detect transcript variants: 2 and 3 (1170 bp), 1 (1167 bp), 9 (903 bp), 5 (900 bp), 4 (832 bp), and 6 and 8 (822 bp).

Due to the high degree of variability between donors, experimental replicates were carried out using MSCs from different patients, which are numbered sequentially using the prefix 'FH'. Expression of varying *FGFR* isoforms was detected across three individual MSC donors (Figure 5.1). Using *GAPDH* as a loading control, donor variation in gene expression was seen; this difference was strongest in *FGFR2* where higher expression was detected in MSC line FH403 compared to the other two. Only two discrete bands were detected for *FGFR2*, whereas *FGFR1* was detected as four distinct products, with one major form being present.

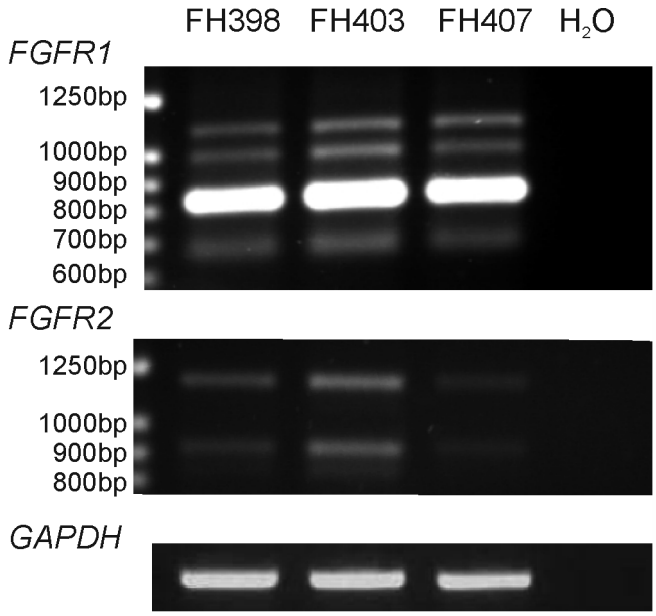


Figure 5.1 Expression of *FGFRs* in MSCs

RNA was collected from MSCs during normal growth, from three independent donors (FH398, FH403, FH407). RT-PCR was carried out for FGFR1 and FGFR2, to detect multiple isoforms. GAPDH (519 bp) was used as a loading control.

5.2.2 MSCs respond to exogenous FGF-2 treatment

To determine how rapidly MSCs respond to exogenous FGF-2 treatment, and at what levels, recombinant protein was added to culture medium and downstream effectors of the FGF pathway were examined. Western blot analysis for levels of dpERK protein was used to detect any rapid response to FGF in MSCs. Optimisation of levels of FGF-2 was established using three different concentrations of recombinant FGF-2 protein: 1, 10 and 25 ng/ml. Following the seeding of cells in normal growth medium, cells were serum-starved for 24 hours to reduce levels of growth factors in the medium, normally present in FBS, to prevent activation of alternative signalling pathways.

These data showed that all concentrations of FGF-2 were sufficient to increase phosphorylation of ERK compared to basal levels (Figure 5.2), suggestive of activation of the downstream signalling cascade of FGF. Increases in dpERK levels were seen after as little as 5 minutes with FGF-2 treatment, however this was subject to FGF-2 dosage, as seen by a smaller increase in ERK phosphorylation with 1 ng/ml FGF-2 compared to both 10 and 25 ng/ml treatments. By 15 minutes, levels of dpERK were greatly increased compared to untreated controls in all three treatment conditions. This may be the peak of activity, with dpERK levels at 30 minutes not at any greater abundance, and further decreased in level by 60 minutes compared to the earlier time points. Nevertheless, at this time the dpERK still remained above the basal levels prior to any FGF treatment.

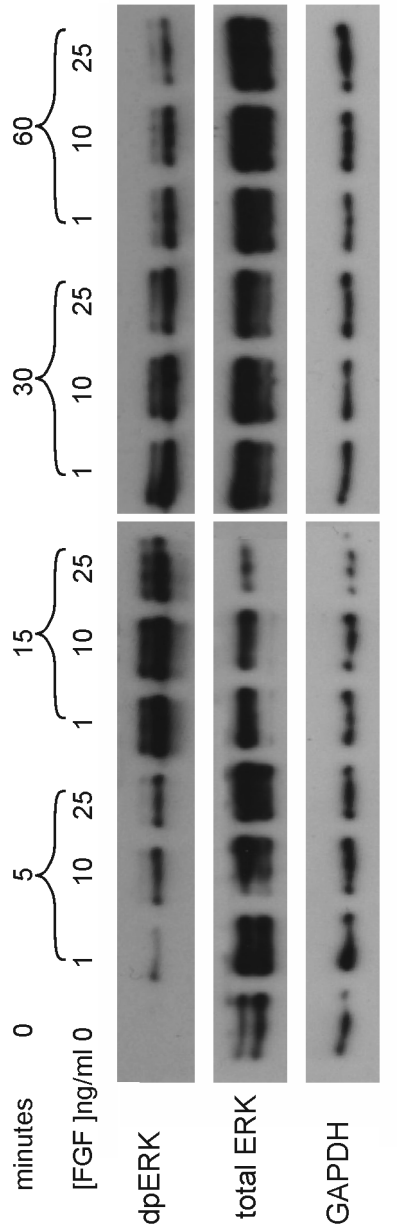


Figure 5.2 Levels of dpERK in MSCs following FGF-2 treatment analysed by western blot
Western blot showing levels of dpERK following treatment with differing concentrations of FGF-2 over a time period of 60 minutes. Total ERK and GAPDH were used as loading controls. Data from 1 donor, representative of n=2.

5.2.3 Analysis of putative FGF target genes

Effects on putative FGF target genes were investigated by qRT-PCR, again using differing FGF concentration levels over varying times of treatments. As this was investigating transcriptional changes the time frame was increased compared to that used for protein analysis. Samples were collected following 6, 12 and 24 hours of treatment. Due to the high degree of variability between human donors, generated fold-change data were not averaged across biological replicates, and are presented individually to observe any common trends in expression.

A number of putative FGF targets in MSCs, identified after longer-term exposure to FGF-2 in culture (Solchaga et al., 2005), were selected for analysis. Efficient primers and detectable expression was found for the genes sprouty 2 (*SPRY2)*, dual specificity phosphatase 6 (*DUSP6*) and interleukin-6 (*IL-6*), which were further pursued. Results showed that all three donors exhibited similar responses to FGF-2, with a general trend in upregulation of expression of these three genes following treatment.

Expression levels of *SPRY2* were increased at all times with all FGF treatments (Figure 5.3). There was no uniform trend across the three donors of any one treatment or time point that produced the largest upregulation, however donor differences were highlighted in the degree of changes seen. For Donor 1 (FH398), the changes seen at each time point were more variable, and the largest change of expression was an 8-fold upregulation at 6 hours. On the other hand, Donor 2 (FH403) showed the highest degree of change at 12 hours, with fold changes consistent with Donor 1 at this time of nearly 6-fold upregulation. Donor 3 (FH407), however, generally showed less of an induction of expression with FGF treatment; nevertheless all changes were significant when compared to untreated cells at the same time points. Dose-dependent effects were seen for each donor at varying time points, but this was not the case at all times.

An analysis of expression of *DUSP6* showed, similarly to *SPRY2*, that all treatments caused a significant increase in gene expression at all time points (Figure 5.4). Using the 10 ng/ml and 25 ng/ml dosage of FGF-2, levels of gene expression were maintained over the 24 hours; however the 1 ng/ml treatment resulted in a reduced induction of expression at 24 hours compared to earlier in the experiment, although this remained significantly above basal levels. Again, Donor 1 and Donor 2 showed a greater induction of *DUSP6* expression when compared to Donor 3. Dose-dependent effects were seen for Donor 1, whereas for Donors 2 and 3 the 10 ng/ml treatment resulted in the largest effect at all time points.

For *IL-6*, previously identified as a negatively regulated target of FGF following prolonged culture in FGF-2 (Solchaga et al., 2005), upregulation of RNA expression was seen across all three donors with all FGF-2 concentrations following 6 and 12 hours in culture (Figure 5.5). At 6 hours, all increases in expression were found to be significant, and occurred in a dose-dependent manner. By 12 hours, the effects were generally the greatest, with a maximum upregulation seen with 10 ng/ml FGF-2 in each donor. Donor variation is clear at this time, with expression increased by ~3-fold at 6 hours for all donors, but at 12 hours varying from 9-fold to 5-fold for 10 ng/ml treatment depending upon donor. Strong differences were seen at 24 hours; for Donor 1 FGF treatment still resulted in increased *IL-6* levels, however the alternative two donors did not show any increase when compared to *IL-6* levels in untreated cells, except for the 25 ng/ml dose for Donor 2. Interestingly, *IL-6* expression was found to increase in untreated cells over the 24 hour period, this effect being seen to occur more strongly in Donors 2 and 3.

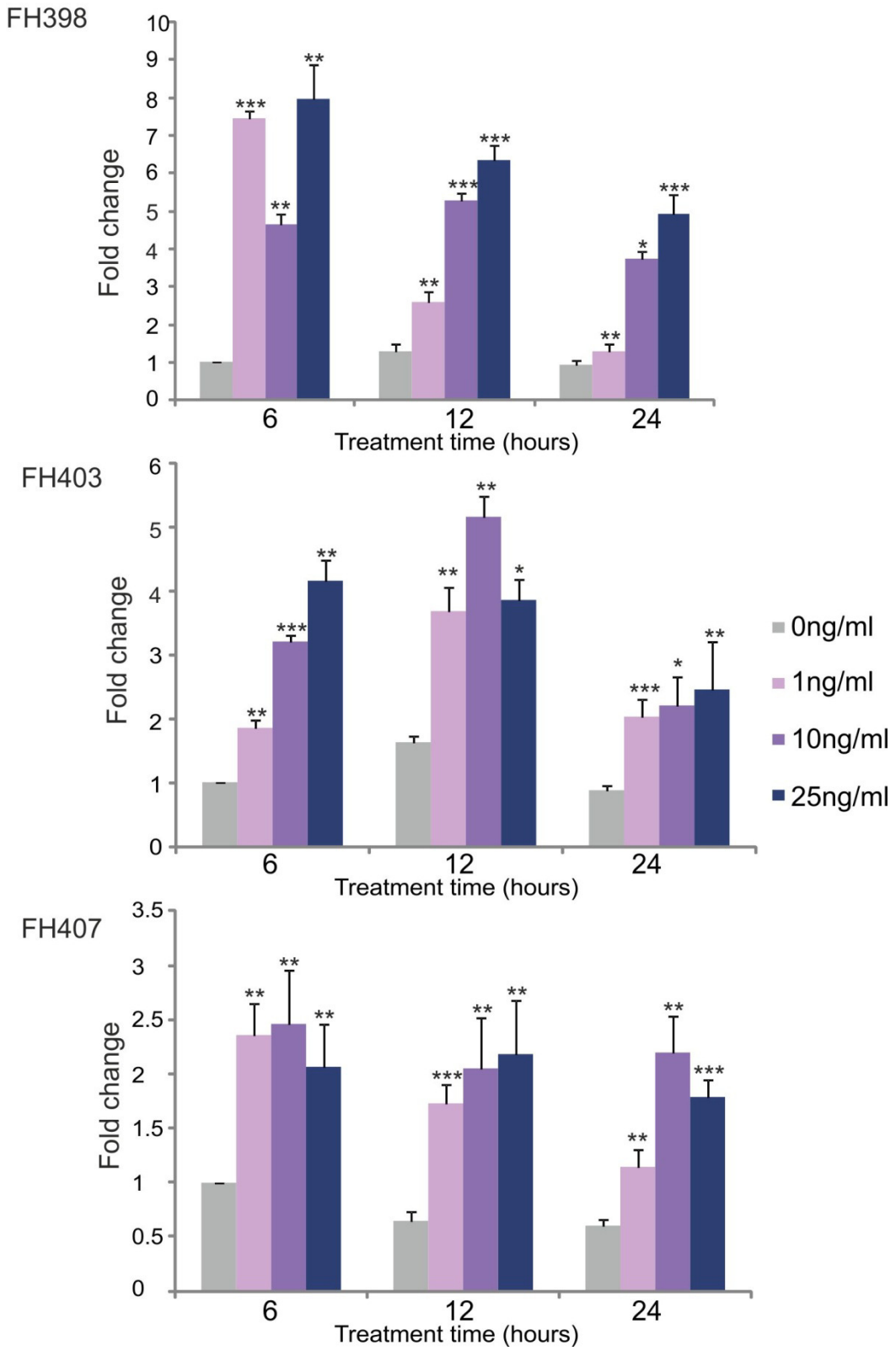


Figure 5.3 Analysis of *SPRY2* expression in MSCs following FGF treatment over 24 hours

Analysis by qRT-PCR of *SPRY2* expression following FGF-2 treatment, on separate donors. Error bar shows SE of three technical replicates. Fold change in expression of RNAs was calibrated to 0 ng/ml at 6 hours and normalised to RPS27a by the $2^{-\Delta\Delta Ct}$ method. Statistical analysis carried out by *t*-test, comparing treatments to 0 ng/ml condition at each time point; * p< 0.05, ** p< 0.005, *** p< 0.001.

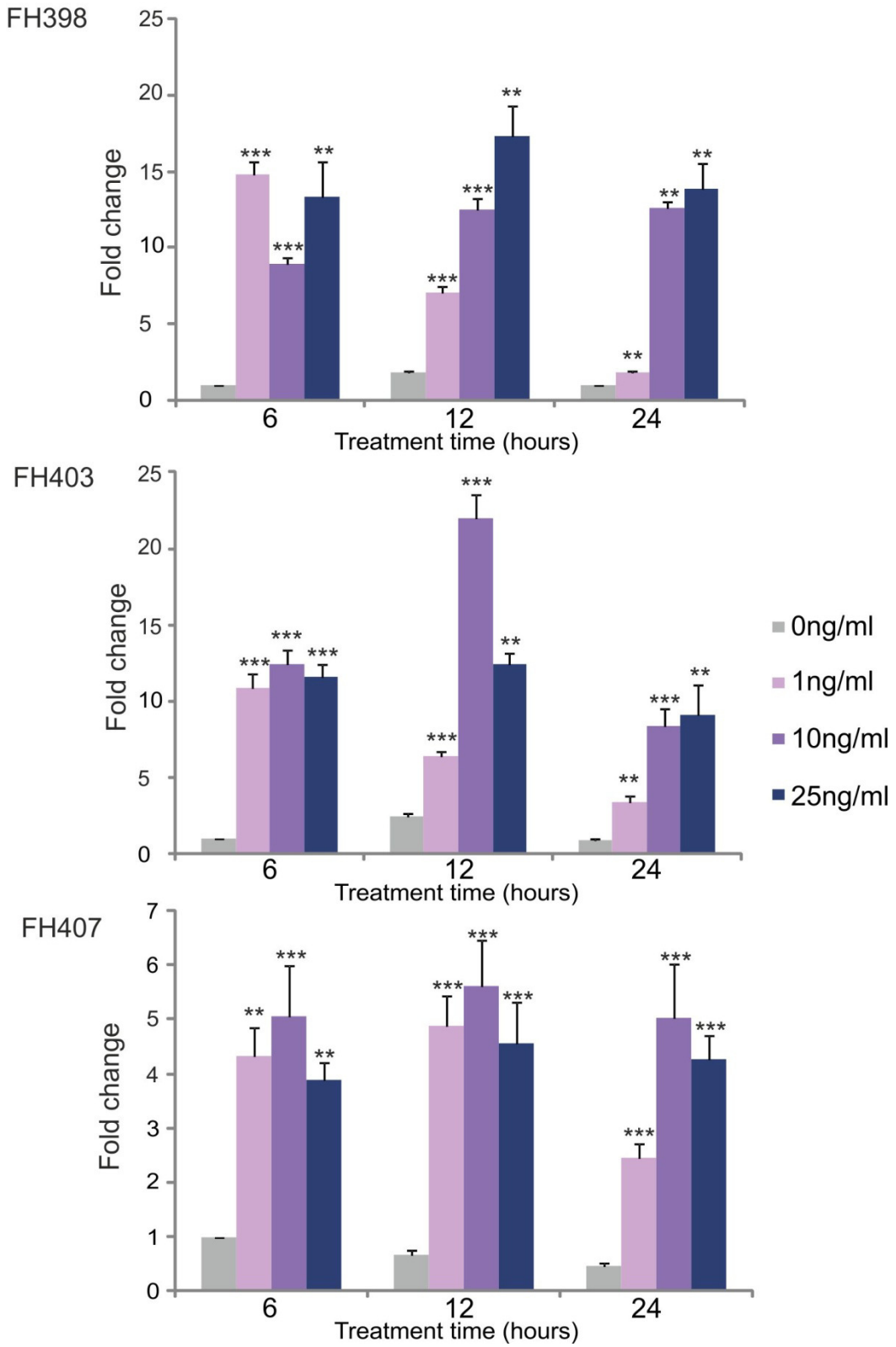


Figure 5.4 Analysis of *DUSP6* expression in MSCs following FGF treatment over 24 hours

Analysis by qRT-PCR of *DUSP*6 expression following FGF-2 treatment, on separate donors. Error bar shows SE of three technical replicates. Fold change in expression of RNAs was calibrated to 0 ng/ml at 6 hours and normalised to RPS27a by the $2^{-\Delta\Delta Ct}$ method. Statistical analysis carried out by *t*-test, comparing treatments to 0 ng/ml condition at each time point; ** p< 0.005, *** p< 0.001.

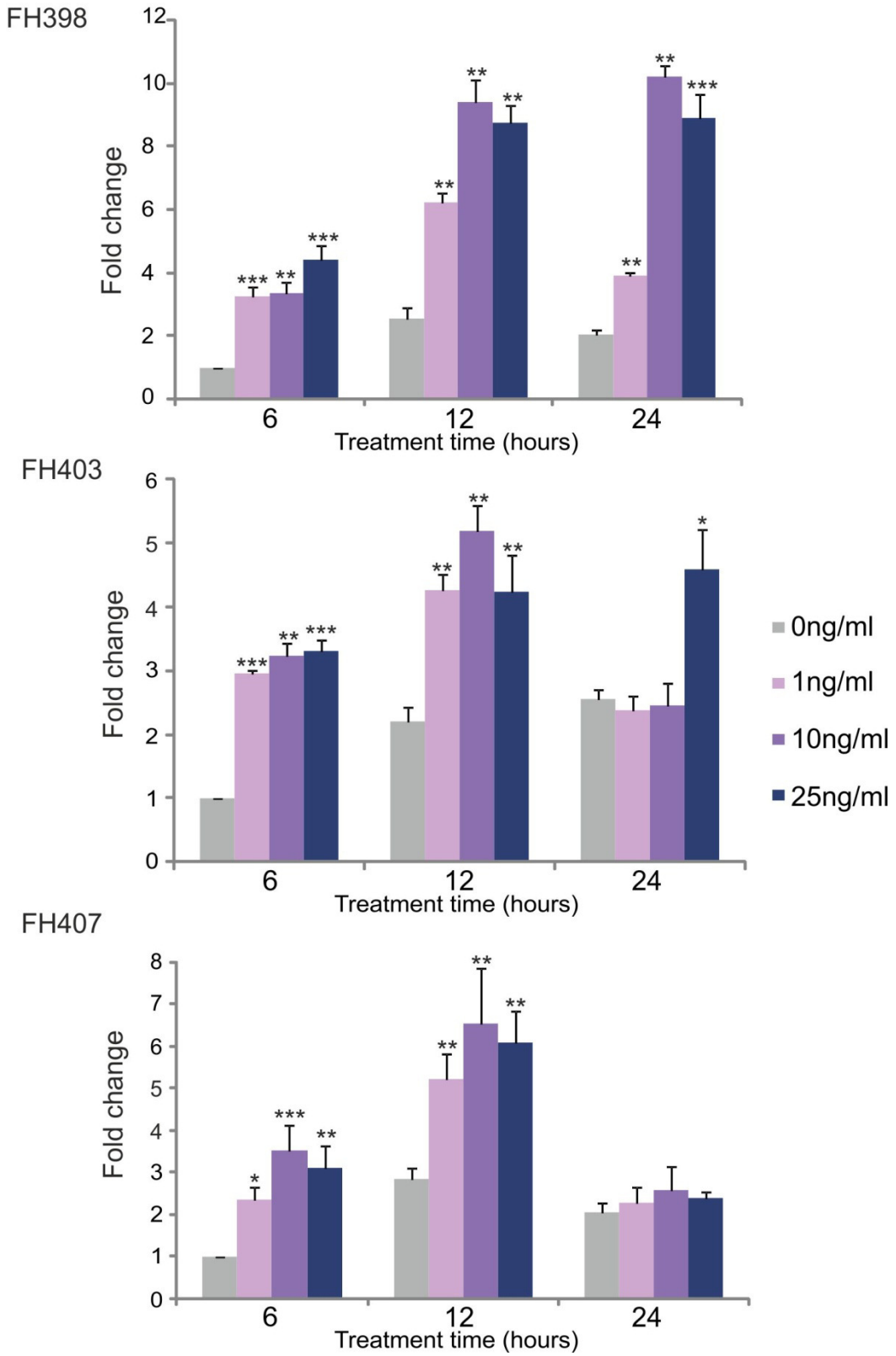


Figure 5.5 Analysis of *IL-6* expression in MSCs following FGF treatment over 24 hours

Analysis by qRT-PCR of *IL*-6 expression following FGF-2 treatment, on separate donors. Error bar shows SE of three technical replicates Fold change in expression of RNAs was calibrated to 0 ng/ml at 6 hours and normalised to RPS27a by the $2^{-\Delta\Delta Ct}$ method. Statistical analysis carried out by *t*-test, comparing treatments to 0 ng/ml condition at each time point; * p< 0.05, ** p< 0.005, *** p< 0.001.

5.2.4 The MSC transcriptome following FGF signalling

To study changes across a large range of genes, a human genome Agilent microarray was undertaken. This dataset could then be drawn in comparison to previously identified targets, including those that resulted from a whole genome microarray undertaken with *Xenopus*, where FGF signalling had been inhibited (Branney et al., 2009). Of the 83 genes identified, 40 of these were found to have confirmed homologues in humans.

The microarray was carried out using RNA from one donor, treated with 10 ng/ml FGF-2 for 6 hours. This concentration was selected as it had been shown to be more effective at activating MAPK signalling and inducing transcriptional effects compared to 1 ng/ml FGF-2, but no less effective than 25 ng/ml. Six hours was chosen as the time point as it appeared sufficiently early to see transcriptional changes, and using an early time point should allow the detection of rapid direct effects of FGF signalling.

Analysis of the data found that 3907 genes showed a >2-fold change following FGF-treatment. Of the transcripts identified as putative targets by the array, 1016 of these were lincRNAs and 428 were predicted/uncharacterised genes. From the remaining 2463 genes, 1063 were downregulated and 1398 were upregulated following FGF treatment. Genes of interest were identified and fold changes are given in Table 5.1 and Table 5.2. Selections were made on the basis of known FGF targets, by comparison to the *Xenopus* array, and from pathways that are known to crossover with FGF signalling.

The array data supported the prior qRT-PCR results for *SPRY2*, *DUSP6*, and *IL-6*, with fold increases of 3.95, 11.2 and 2.67 respectively (Table 5.1). This was in line with the fold changes seen previously for this treatment, although *IL-6* showed a slightly increased upregulation over the previous results.

In comparison to the *Xenopus* data, 11 putative gene targets were found to act in the same manner: 9 showed an increase in expression with FGF-2 treatment whereas they were reduced following FGF inhibition previously (*GJB2, DUSP5, EPHA4, SPRY2, TNFRSF1B, DUSP1, IRX3, PNP, EGR1*) (Table 5.1), and 2 were reduced upon FGF stimulation, but had shown increased expression upon FGF inhibition (*IRG1, TSC22D3*) (Table 5.2). Three genes showed significantly altered expression in the opposing direction (*PDGFA, GIGYF2, CHRDL1*) (Table 5.1, Table 5.2). Fold change data from the *Xenopus* array is shown in Appendix 2.

Some members of the FGF family showed altered expression on this array, with ligands *FGF-16* and *FGF-5* being increased (Table 5.1), but a reduction was seen in the ligand *FGF-*

22 and also *FGFR2* (Table 5.2). There are known interactions between the FGF and wnt pathways, and it was seen that wnt ligands *wnt5a* and *wnt9a* showed positive regulation (Table 5.1), with *wnt7b* and *wnt11* being negatively regulated (Table 5.2). The negative regulator *DKK1* showed an increase in expression following FGF stimulation (Table 5.1). Another signalling pathway known to interact with FGF is the BMP pathway, and, interestingly, treatment with FGF-2 also resulted in an increase in BMP-2 RNA levels (Table 5.1). *XPO5* codes for exportin-5, the protein responsible for exporting pre-miRNA from the nucleus to the cytoplasm, and levels were increased 2.06-fold following FGF pathway stimulation (Table 5.1).

Gene	Fold change	Gene description
DUSP6	11.2	Dual specificity phosphatase 6 , transcript variant 1, mRNA [NM_001946]
GJB2	7.78	Gap junction protein, beta 2, 26kDa, mRNA [NM_004004]
FGF-16	7.75	Fibroblast growth factor 16, mRNA [NM_003868]
BMP2	6.43	Bone morphogenetic protein 2, mRNA [NM_001200]
DUSP5	5.34	Dual specificity phosphatase 5, mRNA [NM_004419]
EPHA4	4.33	EPH receptor A4, mRNA [NM_004438]
WNT5A	4.04	Wingless-type MMTV integration site family, member 5A , mRNA [NM_003392]
SPRY2	3.95	Sprouty homolog 2 (*Drosophila*), mRNA [NM_005842]
SPRY4	3.81	Sprouty homolog 4 (*Drosophila*), transcript variant 1, mRNA [NM_030964]
SPRY4	3.61	Sprouty homolog 4 (*Drosophila*), transcript variant 1, mRNA [NM_030964]
FGF-5	3.26	Fibroblast growth factor 5, transcript variant 1, mRNA [NM_004464]
TNFRSF1B	3.12	Tumour necrosis factor receptor superfamily, member 1B , mRNA [NM_001066]
FGF-5	2.97	Fibroblast growth factor 5, transcript variant 2, mRNA [NM_033143]
PDGFA	2.74	Platelet-derived growth factor alpha polypeptide, transcript variant 1, mRNA [NM_002607]
IL6	2.67	Interleukin 6 (interferon, beta 2), mRNA [NM_000600]
DUSP1	2.62	Dual specificity phosphatase 1, mRNA [NM_004417]
IRX3	2.47	Iroquois homeobox 3, mRNA [NM_024336]
PNP	2.33	Purine nucleoside phosphorylase, mRNA [NM_000270]
DKK1	2.24	Dickkopf homolog 1 (*Xenopus laevis*), mRNA [NM_012242]
EGR1	2.14	Early growth response 1, mRNA [NM_001964]
XPO5	2.06	Exportin 5, mRNA [NM_020750]
WNT9A	2.02	Wingless-type MMTV integration site family, member 9A, mRNA [NM_003395]

Table 5.1 Candidate positively regulated targets of FGF-2 in MSCs

Fold change data for genes identified on a human genome Agilent array in MSCs treated with FGF-2 compared to no treatment. Genes in green and red were identified on the *Xenopus* array: green = same direction of regulation, red = different direction of regulation.

Gene	Fold Change	Gene description
WNT11	-4.52	Wingless-type MMTV integration site family, member 11, mRNA [NM_004626]
FGFBP3	-4.27	Fibroblast growth factor binding protein 3, mRNA [NM_152429]
FGF-22	-3.01	Fibroblast growth factor 22, mRNA [NM_020637]
IRG1	-2.92	Immunoresponsive 1 homolog (mouse) [Source:HGNC Symbol;Acc:33904] [ENST00000449753]
FGFR2	-2.72	Fibroblast growth factor receptor 2, transcript variant 2, mRNA [NM_022970]
GIGYF2	-2.36	GRB10 interacting GYF protein 2 [Source:HGNC Symbol;Acc:11960] [ENST00000410033]
CHRDL1	-2.15	Chordin-like 1, transcript variant 3, mRNA [NM_145234]
TSC22D3	-2.14	TSC22 domain family, member 3, transcript variant 2, mRNA [NM_004089]
WNT7B	-2.02	Wingless-type MMTV integration site family, member 7B, mRNA [NM_058238]

Table 5.2 Candidate negatively regulated targets of FGF-2 in MSCs
Fold change data for genes identified on a human genome Agilent array in MSCs treated with FGF-2 compared to no treatment. Genes in green and red were identified on the *Xenopus* array: green = same direction of regulation, red = different direction of regulation.

5.2.5 lin28 is undetectable in MSCs

The interest in *lin28* genes as targets of FGF signalling and as developmentally important genes, particularly in mesoderm specification, led to investigation of their expression in MSCs. However, both *lin28a* and *lin28b* did not appear to be expressed in MSCs; using mRNA levels to infer gene expression, data from the Agilent microarray did not detect expression of *lin28a* or *lin28b* above background levels, and attempts at measuring *lin28a* and *lin28b* using qRT-PCR also failed to reliably show expression of either gene, both endogenously and following FGF-2 treatment (data not shown). The 2102Ep EC cell line expresses both of the lin28 proteins and was used as a positive control. Expression of lin28a and lin28b was not seen in MSCs at the protein level, analysed by western blot, in comparison to 2102Ep cells (Figure 5.6).

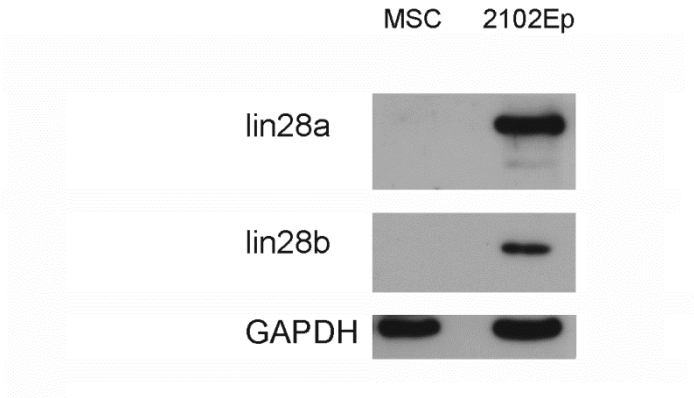


Figure 5.6 Analysis of lin28a and lin28b by western blot in MSCs compared to EC cells
Protein was collected and analysed from MSCs and 2102Ep cells. GAPDH was used as a loading control. MSC blot is representative of three separate donors.

5.3 Discussion

5.3.1 Investigating FGF signalling in MSCs

Exposing MSCs to exogenous FGF protein showed promising data but did limit the manipulations that could be carried out on the cells. With 22 FGF ligands expressed in humans (Ornitz and Itoh, 2001), it would be impractical to attempt to treat with all of these. FGF-2 was selected here as it is the most common FGF supplement used in cell culture, and identifying targets of this ligand in particular may help identify stem cell mechanisms for both proliferation and differentiation.

This does contrast with the methodology used in the *Xenopus* array with which data comparisons were drawn. In this, use of both dnFGFR1 and dnFGFR4 overexpression found that either was able to effectively inhibit all FGF signalling (Branney et al., 2009), and therefore would have been able to identify targets of all FGF ligands functional at the time. Differences between datasets may arise where genes detected in the *Xenopus* work are not downstream of the FGF-2 ligand.

A drawback with the over-activation of a signalling pathway is that some transcriptional targets are likely to be missed. The control of gene expression is often under the influence of multiple factors, and therefore if FGF is necessary to drive expression, a decrease in levels may be seen with a loss of signalling, but levels would not alter upon over-stimulation due to endogenous FGF signalling already activating expression. There is evidence for some autocrine FGF signalling in MSCs, which decreases as the cells increase in passage number (Zaragosi et al., 2006), and this may have masked the effects of adding FGF-2. Nevertheless, much stem cell work involves treatment with FGF-2, and by also following this method relevant targets are likely to be identified.

5.3.2 MSCs respond rapidly to FGF treatment

Work here has found that MSCs are capable of responding to FGF-2 treatment, through the expression of FGFRs, which have recently been shown to be developmentally regulated and may affect cellular responses to FGF (Coutu et al., 2011). The response is rapid, with activation of the signalling cascade apparent by as little as 5 minutes, seen by a phosphorylation of ERK protein, and the induction of transcriptional changes by as rapidly as 6 hours. The effects of FGF signalling were looked at in this study in a time frame much shorter than those in which previous studies have been carried out, which have normally used 7 days as their first time of observation (Martin et al., 1997; Solchaga et al., 2005). However, as previous reports were looking for alterations in rates of proliferation, such changes would take longer to become apparent compared to transcriptional changes. The earliest of the time points used here, 6 hours, was pursued for further investigation to enhance the likelihood that any alterations in expression detected were due directly to an increase in FGF signalling and to attempt to limit the indirect effects, which would usually take longer to be initiated.

The transcriptome analysis was carried out using the 10 ng/ml dosage of FGF-2, as results implicated that this generally resulted in a more increased activation of FGF signalling compared to using 1 ng/ml, but no reduction compared to 25 ng/ml treatment. Previous work found that different donors responded to different doses of FGF-2, with threshold effects where all treatments resulted in equivalent results, but an alternative donor showed increasing stimulation following increasing FGF-2 (Solchaga et al., 2005). Results like these highlight the effects donor difference can have on results, which were also demonstrated by the different fold changes seen in genes by qRT-PCR. In light of this, it is important that transcriptional targets identified by microarray are confirmed using cells from other donors.

Reduction in dpERK by 60 minutes compared to 30 minutes with all levels of FGF-2 suggests that perhaps there is a cyclic regulation occurring, and an initial increase in ERK phosphorylation cannot be sustained, and must decrease again. A similar response to this here was previously seen in MSCs, when using a much higher dosage of 50 ng/ml FGF-2, where a peak in ERK phosphorylation was seen at 5 minutes which decreased through to 60 minutes (Choi et al., 2008). It would be interesting to see if dpERK levels are maintained at this lower level, perhaps with the activation of MAPK repressors which is discussed later, or whether there are ongoing cyclic fluxes in ERK activation.

5.3.3 Novel targets of FGF signalling in MSCs

With over 1000 lincRNAs and over 2000 mRNA candidate targets showing changes in expression in this study, it is likely that further new targets of FGF signalling have been identified by the microarray carried out here. However, as mentioned above, true statistical analysis cannot be applied to this dataset alone, being carried out in only one donor, and further replicates and validation would be needed.

lincRNAs are a class of RNA which are still being progressively understood. They have been found to show some evolutionary conservation between mammals, and thus are concluded likely to have important biological functions (Guttman et al., 2009). By correlating lincRNA and protein expression, the same study was able to infer potential roles for a number of lincRNAs, including identifying a subset that were activated by p53 signalling, as well as other pathways (Guttman et al., 2009). Further work has connected particular lincRNAs with the reprogramming of stem cells, the maintenance of pluripotency being regulated by Nanog and other pluripotent proteins, and lincRNAs that were involved in germ layer specification (Guttman et al., 2011; Loewer et al., 2010). Unfortunately, due to differing terminology between both the papers and the Agilent microarray for the lincRNAs, no comparison has been drawn here between the lincRNAs altered upon FGF treatment in these cells and those involved in pluripotency from the studies; the Agilent array termed lincRNAs by chromosomal location, Loewer et al (2010) named the lincRNAs according to "their 3' protein-coding gene neighbour", and Guttman et al (2011) numbered the lincRNAs, as is the consensus for miRNAs. A more extensive comparison between these datasets, with a determination of corresponding names, may help elucidate further the effect that FGF treatment is having on these cells.

Our interest in the lin28s in the *Xenopus* system led to investigating for any conservation of this regulation in human development, using MSCs for a mesodermal model. However, expression of neither lin28a nor lin28b could be detected by methods at the protein level and at the RNA level in MSCs. Treatment of the cells with FGF-2 did not result in detectable levels of *lin28a* or *lin28b* either, and thus the conservation of these as targets of FGF signalling in humans could not be determined. 2102Ep cells were used as a positive control for expression of the lin28 genes, which confirmed that the techniques were able to detect expression of the RNA and protein in cells.

5.3.4 FGF targets may respond differently to short and long-term FGF-2 exposure

The initial finding that FGF-2 increased *IL-6* expression contrasted with those in the literature that reported a negative regulation by FGF signalling (Solchaga et al., 2005). The positive regulation that was seen here, however, was at the earliest time points, and was reducing in effect by 24 hours, being non-significant in two donors by this time. This could indicate that perhaps a later activity of the FGF signalling is a repression of *IL-6* expression, although this gene is initially a positive target. This is supported by work in other models, for example treatment of Schwann cells with FGF-2 found an increase in IL-6 at 5 and 10 hours, which had reduced to untreated levels by 24 hours (Grothe et al., 2000). Another report found that after 24 hours of treatment, exogenously applied FGF-2 resulted in no change in IL-6 levels in HeLa cells (Delrieu et al., 1999), which may support the result found here, with the upregulation maybe already having been lost by this time.

The study by Solchaga et al (2005) involved a microarray after 14 days in culture with FGF-2 supplementation, and a further comparison between these data and the microarray data generated here may identify if there are further targets responding in a different manner to long and short term culture. This coupled with analysis of differentiation capabilities following different FGF treatments may give an insight into what targets of the FGF pathway are important for different roles.

5.3.5 Conserved FGF signalling targets in MSCs

A number of mRNA targets for FGF that were identified on the microarray have been previously identified in other species, including the *Xenopus* array that comparisons have been drawn to. This highlights these genes as good candidates for being FGF-regulated in MSCs, but again must be verified with alternative donors and more specific techniques such as qRT-PCR, to determine that they are true FGF targets in these cells.

FGF signalling is known to cross-talk with other pathways. The FGF and wnt pathways are known to affect one another in the cell, as both are capable of signalling through the calcium pathway (reviewed in Bottcher and Niehrs, 2005; Kühl, 2002), and increased osteogenesis in MSCs as a result of activated FGFRs was found to occur by activation of both the MAPK and PKC pathways (Miraoui et al., 2009). The PKC pathway activation by wnt is a non-canonical pathway and wnt ligand activators of the non-canonical pathway, wnt7, wnt5a and wnt11 (Brade et al., 2006; Montcouquiol et al., 2006) were both upregulated and downregulated upon FGF treatment. wnt5a and wnt11 can activate PKC

via calcium pathway (Kühl, 2002). wnt9a, however, can function canonically and has been shown to be involved with mesenchymal differentiation (Brade et al., 2006), and expression of this was upregulated following FGF treatment. DKK1 is a wnt antagonist of the canonical pathway (Glinka et al., 1998), and was also increased with FGF signalling. These results show both activation and inhibition possibilities of the wnt pathway following FGF signalling, indicating there is no clear directional regulation of the wnt pathway by FGF here. Further investigation would need to look at protein levels of the wnt ligands, and activation of pathway factors such as PKC and β-catenin, for non-canonical and canonical effects.

FGF often plays an antagonistic role to BMP signalling (Buckland et al., 1998; Streit and Stern, 1999), but the only change in mRNA levels of key players in the BMP pathway was an over 6-fold increase in expression of BMP-2. Surprisingly, no downregulation of other pathway components were seen. It may be the case that there is a low level of BMP activity in these cells, and an increase in FGF signalling does not suppress activity any further, or that any regulatory effects are not occurring at the transcriptional level.

FGF stimulation in MSCs produced changes in members of the DUSP and SPRY families, which have been found to act in a negative feedback loop with FGF signalling. DUSP1, DUSP5 and DUSP6 are known to be positively regulated by FGF (Dickinson and Keyse, 2006; Li et al., 2007), but function to inactivate ERK phosphatase (Dickinson and Keyse, 2006), thus inhibiting FGF signalling via the MAPK pathway. Additionally, both sprouty2 and sprouty4 prevent activation of the MAPK pathway by targeting Ras (Gross et al., 2001; Lee et al., 2001), whilst also being activated by the FGF pathway (Lai et al., 2011; Sasaki et al., 2001). Work here showed a decrease in the activation of ERK following 60 minutes of treatment. It is possible that transcriptional changes as a result of FGF signalling were rapidly induced within this timeframe, and targets such as the DUSP and sprouty families activated and functioning to inhibit the MAPK pathway, resulting in this reduction of dpERK. This puts forward the case that although 6 hours is treated here as a rapid time point for changes in these cells, there may be alterations in the transcriptome from much earlier stages still. Work on sprouty proteins in *Xenopus* has suggested that these proteins are not involved in the specification of mesoderm, but instead function to inhibit FGF signalling through PKCδ and Ca^{2+} release (Nutt et al., 2001; Sivak et al., 2005), whereas MKP1, homologous to DUSP1, does affect mesoderm specification (Gotoh et al., 1995). Whether these roles are conserved in MSCs, however, remains unknown. Interestingly, *DUSP5* and *DUSP6* showed larger changes in expression than the *Sprouty* genes, which might indicate that the function of these at this time is more important than the *Sprouty* genes.

CHRDL-1 expression in MSCs was decreased following FGF treatment; this gene shows homology to *chordin,* which is known to be a positively regulated FGF target in species including *Xenopus* and zebrafish. How closely the function of CHRDL1 matches chordin has not been determined, nor has it been shown to be regulated by FGF, but it is active as a BMP-4 antagonist (Kane et al., 2008), which would be consistent with a function downstream of FGF. The reason for this difference in regulation is unknown, and for results like this it is important for further validation to be carried out to ensure that this gene is working differently in the MSC system.

A number of genes identified from the *Xenopus* array that were also identified as putative targets here have also been linked with FGF in a variety of roles: for example, EGR-1 has been identified as a target of FGF-1 and FGF-2, which can further work to activate *PDGFA* gene expression (Delbridge and Khachigian, 1997; Santiago et al., 1999). These genes look like further strong candidates as FGF-2 regulated genes in MSCs. However, there are additional targets identified here such as *GJB2* and *GIGYF* that were linked with FGF regulation in *Xenopus* (Branney et al., 2009), but have not otherwise been linked with FGF signalling. These genes would need further confirmation as true FGF targets in MSCs and investigation into their function to determine the importance of any interaction. This microarray was only carried out with one sample, and as such no statistical analysis can be carried out. The genes of interest have not here been further verified either, and this would need to be carried out, using qRT-PCR, to be certain that these are targets of FGF signalling in this system.

The lack, or very low level, of lin28 expression in MSCs, and the absence of any increase in expression with treatment of FGF resulted in no further investigation of lin28 as an FGF target in human cells, and no further experiments were undertaken with MSCs to look at a conserved signalling pathway involving lin28. The decision was taken not to pursue with 2102Ep cells as an alternative model, due to their lack of an ability to differentiate (Duran et al., 2001). It was identified that this nullipotency results as a loss of differentiation function (Duran et al., 2001) but the factors responsible for this remain currently unknown, therefore it is possible that the signalling pathway we are interested in is misregulated in these cells, and thus involved in the prevention of differentiation of 2102Ep cells. From this, experiments into mesodermal development and the FGF pathway may not be representative of true endogenous signalling.

Chapter 6. Physical interactions between lin28 proteins and miRNAs

6.1 Introduction

6.1.1 The miR-17-92 clusters

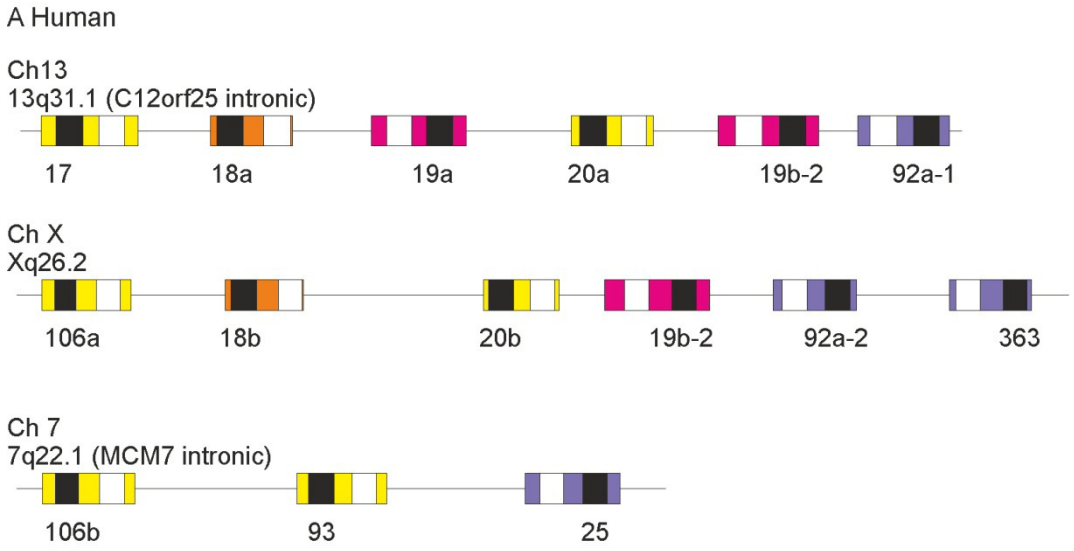
The miR-17-92 clusters are a highly conserved set of miRNA clusters that have been found in vertebrates from zebrafish to humans (Tanzer and Stadler, 2004). The arrangement on the human genome is highly similar to that in *Xenopus* as shown in section 4.3.4, but contains an additional miRNA, miR-106b, and is present as three clusters (Figure 6.1A), which is a more typical arrangement for other species (Tanzer and Stadler, 2004). Chromosomal locations for the clusters have been mapped in humans: the mir-17-92 cluster is expressed on chromosome 13 and has been found to be intronic to *C13orf25*, which does not seem to have another function in addition to producing this transcript (Mendell, 2008; Ota et al., 2004). miR-106a-363 is X-linked, but has not been shown to be linked to any other genes, and miR-106b-25 is intronic to *MCM7* on chromosome 7 (Mendell, 2008).

The clusters are believed to have evolved through both tandem duplications and cluster duplications (Tanzer and Stadler, 2004), and as such many of the miRNAs present within the cluster are highly related. This is reflected in their seed sequences, which allow the 'major' miRNAs from the clusters to be grouped into four subfamilies due to their conservation (Figure 6.1B).

6.1.2 Possible physical interaction between lin28 and miRNA

The lin28 proteins are known to bind to mRNA and let-7 miRNA targets to exert their effects, through their dual RNA-binding domains. It was hypothesised that in positively regulating multiple members of the miRNA cluster, the lin28s may be binding to the miRNAs to influence their processing. In order for a physical interaction to occur, there must be co-expression of the miRNAs with the lin28s. Using ES cells, with a known requirement for lin28a, research has uncovered expression of a number of miRNAs across the clusters (Foshay and Gallicano, 2009; Houbaviy et al., 2003; Suh et al., 2004; Tang et al., 2006; Ventura et al., 2008). Similarly to the lin28s, there are reports of upregulation of these cluster miRNAs in cancer, perhaps reflecting a role in pluripotency or an undifferentiated state. miR-17-92 is the most highly studied of the clusters; it was termed 'oncomiR-1', due to its oncogenic capabilities (He et al., 2005), and has been linked to

cancers including lung, colon, prostate, breast and pancreatic cancer, leukaemia, medulloblastoma and retinoblastoma (reviewed in Concepcion et al., 2012; Hayashita et al., 2005; Volinia et al., 2006), with the miR-106a-363 cluster having a possible role in T-cell leukaemia (Landais et al., 2007).



B

```
hsa-miR-17-5p            CAAAGUGCUUACAGUGCAGGUAG
xtr-miR-17-5p            CAAAGUGCUUACAGUGCAGGUAGU
hsa/xtr-miR-20a-5p       UAAAGUGCUUAUAGUGCAGGUAG
hsa-miR-106a-5p          AAAAGUGCUUACAGUGCAGGUAG
hsa-miR-106b-5p          UAAAGUGCUGACAGUGCAGAU
xtr-106-5p               AAAAGUGCUUAUAGUGCAGGUAGA
hsa/xtr-miR-20b-5p       CAAAGUGCUCAUAGUGCAGGUAG
hsa-miR-93-5p            CAAAGUGCUGUUCGUGCAGGUAG
xtr-miR-93a-5p           AAAGUGCUGUUCGUGCAGGUAG
xtr-miR-93b-5p             AAGUGCUGUUCGUGCAGGUAG

hsa/xtr-miR-19a-3p       UGUGCAAAUCUAUGCAAAACUGA
hsa/xtr-miR-19b-1/2-3p   UGUGCAAAUCCAUGCAAAACUGA


hsa/xtr-miR-18a-5p       UAAGGUGCAUCUAGUGCAGAUAG
hsa/xtr-miR-18b-5p       UAAGGUGCAUCUAGUGCAGUUAG

hsa-miR-92a-1/2-3p       UAUUGCACUUGUCCCGGCCUGU
xtr-miR-92a-3p           UAUUGCACUUGUCCCGGCCUG
hsa/xtr-miR-363-3p       AAUUGCACGGUAUCCAUCUGUA
hsa/xtr-miR-25-3p        CAUUGCACUUGUCUCGGUCUGA

hsa/xtr-miR-363-5p       CGGGUGGAUCACGAUGCAAUUU
```

Figure 6.1 Organisation of miR-17-92 clusters and conservation of seed regions

A) Arrangement of miRNAs across miR-17-92 cluster and paralogous clusters. Whole strand refers to pri-miRNA, coloured boxes represent pre-miRNA, the colour of which indicates miRNA families based upon seed sequence. Black and white boxes represent mature miRNA: black = 'major' miRNA, white = 'minor' miRNA. B) Human and *X. tropicalis* miRNA sequences grouped into families by seed sequence.

6.1.3 The binding of lin28 to let-7

lin28 is known to interact with let-7 via binding to the terminal loop, which inhibits its processing (Piskounova et al., 2008; Viswanathan et al., 2008). Research has since attempted to uncover what the binding motifs for lin28 recognition are; one such identified motif is a sequence consisting of 'GGAG' (Heo et al., 2009). Further work confirmed this, although more recent work has expanded this to a simpler sequence of 'NGNNG', where N = any base (Loughlin et al., 2012), and this site is believed to be targeted by the zinc finger domains (Loughlin et al., 2012; Nam et al., 2011). A binding site for the CSD has been identified as single-stranded RNA with the sequence 'NNGNGAYNN' in pre-let-7e, where Y = pyrimidine (Nam et al., 2011), and alternatively 'GNUNNUNNN' (Mayr et al., 2012). Data indicated that binding of the CSD to such sites remodelled the RNA, exposing the binding site for the zinc fingers, which were suggested as likely to provide the specificity for the interaction.

6.1.4 Aims

The aims of this chapter are to:

- Identify any conserved putative lin28 binding sites in miR-17-92 and miR-106-363 clusters.
- Investigate if a direct interaction occurs between lin28 and putative binding sites using an *in vitro* binding assay.
- Establish any specificity in binding preferences.
- Discover if any differences occur with full-length lin28a proteins compared to lin28b protein.

6.2 Results

6.2.1 Presence of putative lin28 binding motif in pre-mir-363

A published specific binding site for lin28 in let-7 is GGAG (Heo et al., 2009; Nam et al., 2011), and any presence of this in our RNAs of interest was investigated. As the *Xenopus* miRNAs that were showing the largest changes on the microarray belonged to the miR-17-92 and miR-106-363 clusters, the search for binding sites was only carried out in these two. This GGAG sequence is present in both the miR-17-92 cluster and miR-106-363 cluster; in one instance in each cluster, the motif is present between pre-miRNA: between miR-19b-2 and miR-92a-1 in the miR-17-92 cluster, and between miR-19b-1 and miR-92a-2

in the miR-106-363 cluster (Figure 6.2). Notably, there is one other GGAG motif in the miR-106-363 cluster, which is within the terminal loop region of pre-mir-363, the last miRNA in this cluster (Figure 6.2). Additionally, it has been suggested that GGUG may also provide a binding site for lin28 (Heo et al., 2009; Mayr et al., 2012), and there are further GGUG motifs present within these clusters (Figure 6.2). The identification of putative CSD binding sites proved more difficult to identify due to the structural importance in addition to sequence, so only potential zinc finger sites were followed up, as it has been hypothesised that these sites confer the specificity to the interaction.

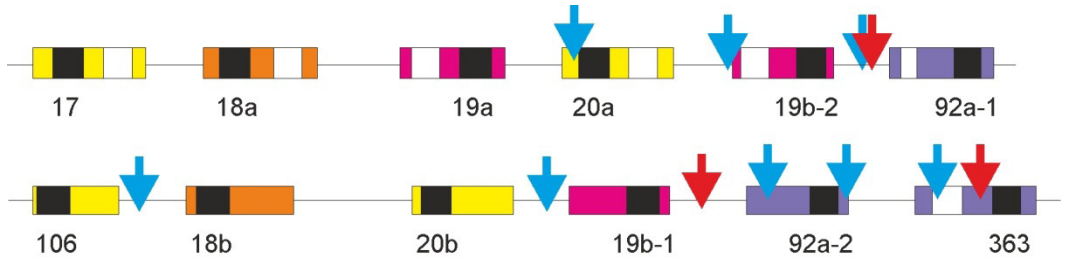


Figure 6.2 Location of GGAG motifs in *Xenopus* miRNA clusters
Arrows indicate location of possible lin28 zinc finger binding sites in *X. tropicalis* miR-17-92 and miR-106-363 clusters. Red arrows = GGAG sequences, blue arrows = GGUG sequences.

As the GGAG sequence within pre-mir-363 was in the terminal loop, similarly placed to the sequence in let-7, this was pursued. Seventeen vertebrate species have been identified so far as having a copy of the miR-363 miRNA (Figure 6.3). The GGAG motif is highly conserved between these species, with the only exceptions being the green anole lizard, *Anolis carolinensis*, which instead contains the sequence GGAC, and zebrafish. The zebrafish sequence shows low conservation in the loop region, and much more sequence identity is seen between the other species in this region (Figure 6.3). Conservation between all species is highest within the mature miRNAs, with the -3p 'major' form showing the highest conservation. The pre-miRNA sequences identified for zebrafish and the Tasmanian devil are also much longer sequences than in the other species, longer at both the -3p and -5p ends, although they continue to not show high sequence conservation to each other in these regions (data not shown).

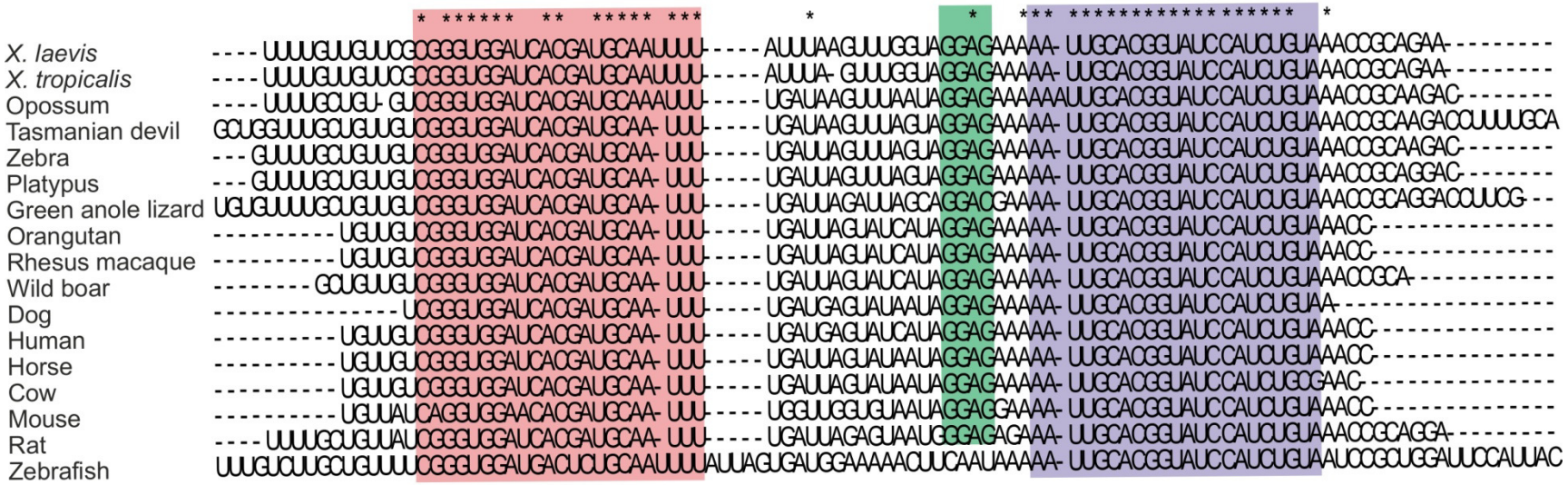


Figure 6.3 Alignment of pre-mir-363 miRNAs in species

Clustal alignmenet of all known pre-mir-363 sequences. Full-length sequences are shown for all but the Tasmanian devil, which is missing bases 1-34 and 141-157, and zebrafish missing bases 1-10 and 116-128. * denotes fully conserved nucleotide. Red = miR-363*, green = GGAG motif, purple = miR-363.

6.2.2 Truncated lin28 protein can bind RNA

The capability of lin28 protein to bind mir-363 was determined using electrophoretic mobility shift assay (EMSA). Recombinant lin28 protein was synthesised by the Anston group (York Structural Biology Laboratory). Due to difficulties in generating and purifying the protein, a truncated version of the *X. tropicalis* lin28a recombinant protein was provided (Xrt-lin28a). The protein lacked both the N- and C- 'disordered' termini, but still contained the RNA-binding domains (Figure 6.4).

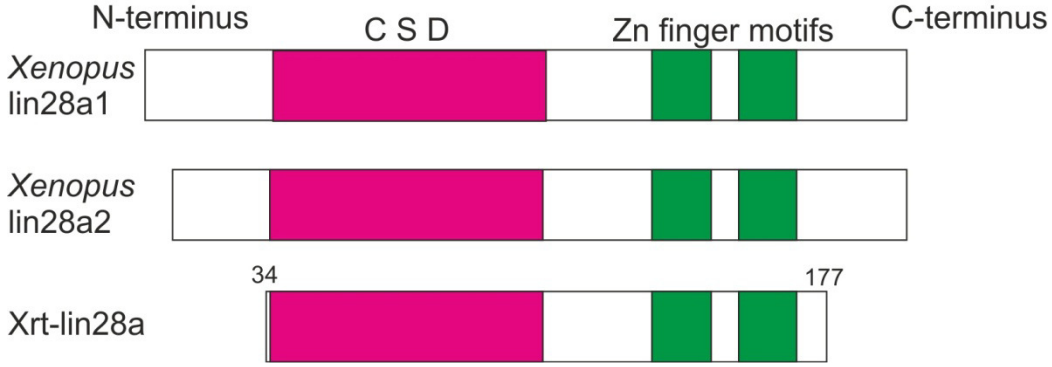


Figure 6.4 Truncation of recombinant lin28 protein

Scale diagrams of lin28a proteins. Pink region represents CSD, and green indicates the zinc finger motifs. Numbers on the recombinant protein denote the first and last amino acid residues of the truncated form, as counted in lin28a1.

To ensure that this truncation was not impeding the RNA-binding properties of the protein, binding reactions were carried out using recombinant human full-length (FL) and truncated termini (TT) lin28a protein, which again lacked the terminal regions. The human truncated protein was composed of residues 37-180, which covers the same areas of the protein as the *Xenopus* truncation of 34-177.

As the terminal loop of pre-let-7 is a known binding target of lin28 proteins, the proteins were tested by EMSA with RNA corresponding to the terminal loop of human let-7g (L-let-7g). Both human proteins were found to cause a shift in the let-7g RNA, indicating that binding was occurring; concentrations between 0.2 and 0.8 µM were sufficient to cause the shift in RNA, showing that RNA-binding function was maintained in the TT protein. A similar binding efficiency was maintained between the two proteins; however this was slightly improved with the FL protein (Figure 6.5).

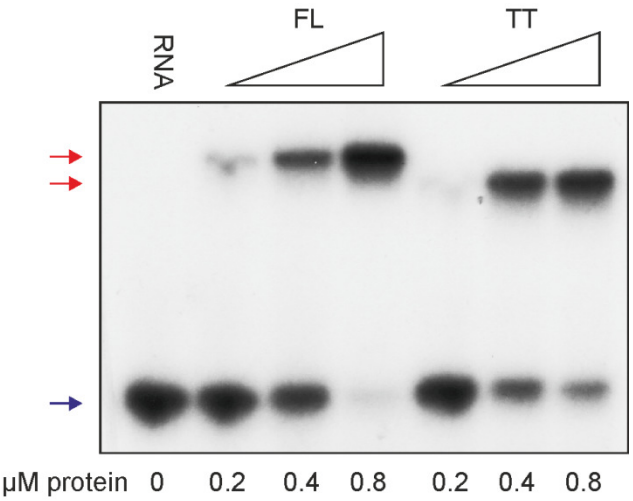


Figure 6.5 Binding of recombinant full-length and truncated lin28a protein to let-7
EMSA performed with ^{32}P-L-let-7g and indicated concentrations of human recombinant lin28a protein, either full-length (FL) or truncated (TT). Arrows indicate labelled RNA (blue) and lin28-RNA complex (red).

6.2.3 Binding to terminal loop regions

To analyse binding with the *Xenopus* lin28a protein, a range of concentrations were tested with the L-let-7g RNA. The human RNA shows slight differences to the *Xenopus* sequence, but retains lin28 binding sites, and so was used for EMSA. Binding of the RNA was observed, with an increase in the proportion of RNA shifting with increasing amounts of protein (Figure 6.6A). The calculation of the dissociation constant (K_d) signifies binding efficiency, which at 0.314 µM indicated a high affinity interaction. A binding of all RNA was predicted to occur (B_{max} = 1.017) (Figure 6.6B).

As the GGAG motif was found within the pre-mir-363 terminal loop, binding was tested with an RNA sequence for the loop of the *Xenopus* precursor (L-mir-363). Using the same range of protein concentrations as for L-let-7g, a shift in the radio-labelled RNA was seen to occur with a calculated K_d of 0.448 µM and a maximum proportion of RNA bound at 96.2% (Figure 6.7), indicating further effective binding.

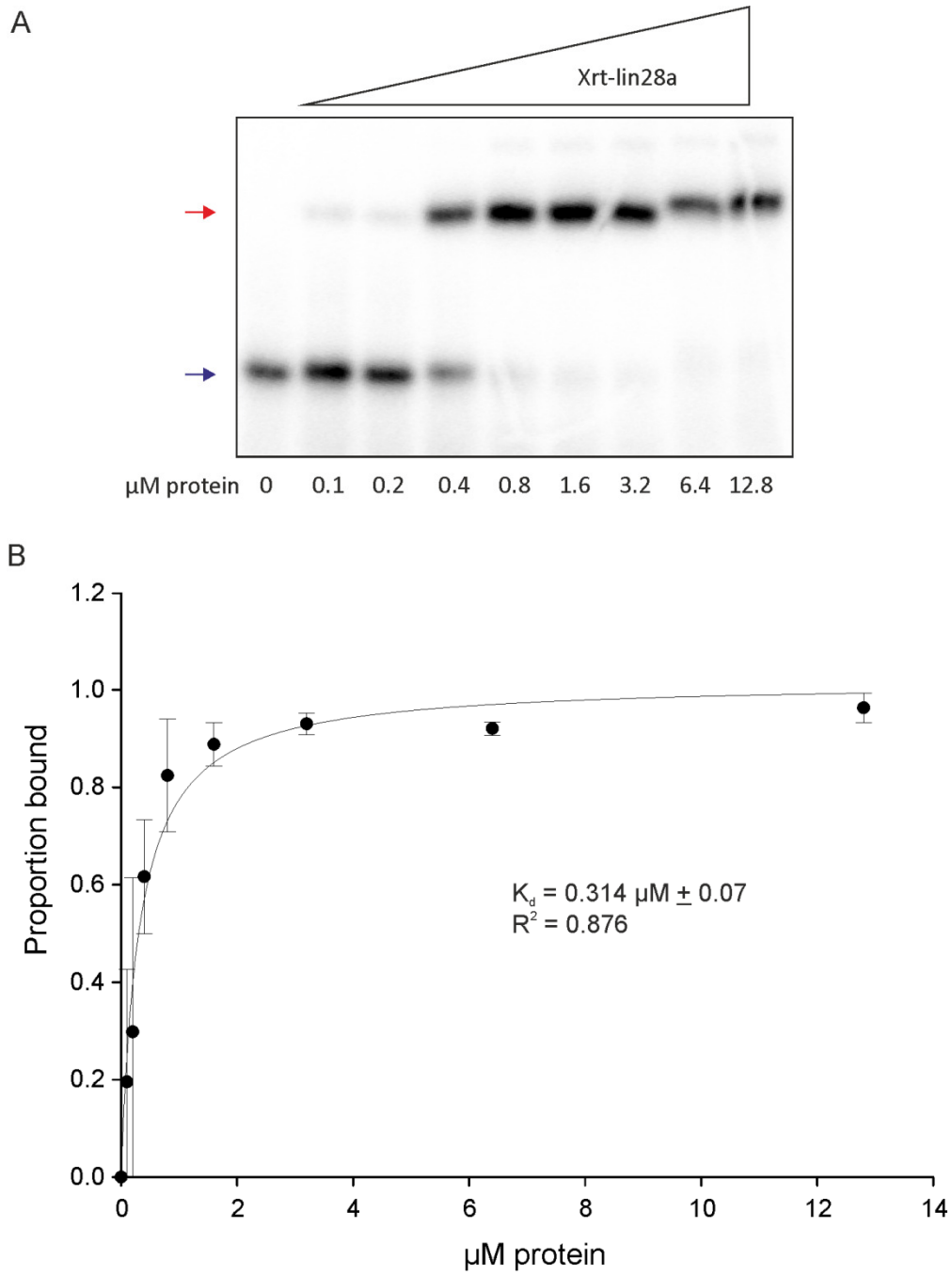


Figure 6.6 Binding of *Xenopus* lin28a with let-7g terminal loop

A) EMSA performed with ^{32}P-L-let-7g and indicated concentrations of Xrt-lin28a. Gel shown is representative of n=3. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red). B) Band intensities were quantified from three independent experiments and the proportion bound was calculated. Data were used to plot a binding curve. B_{max} = 1.017.

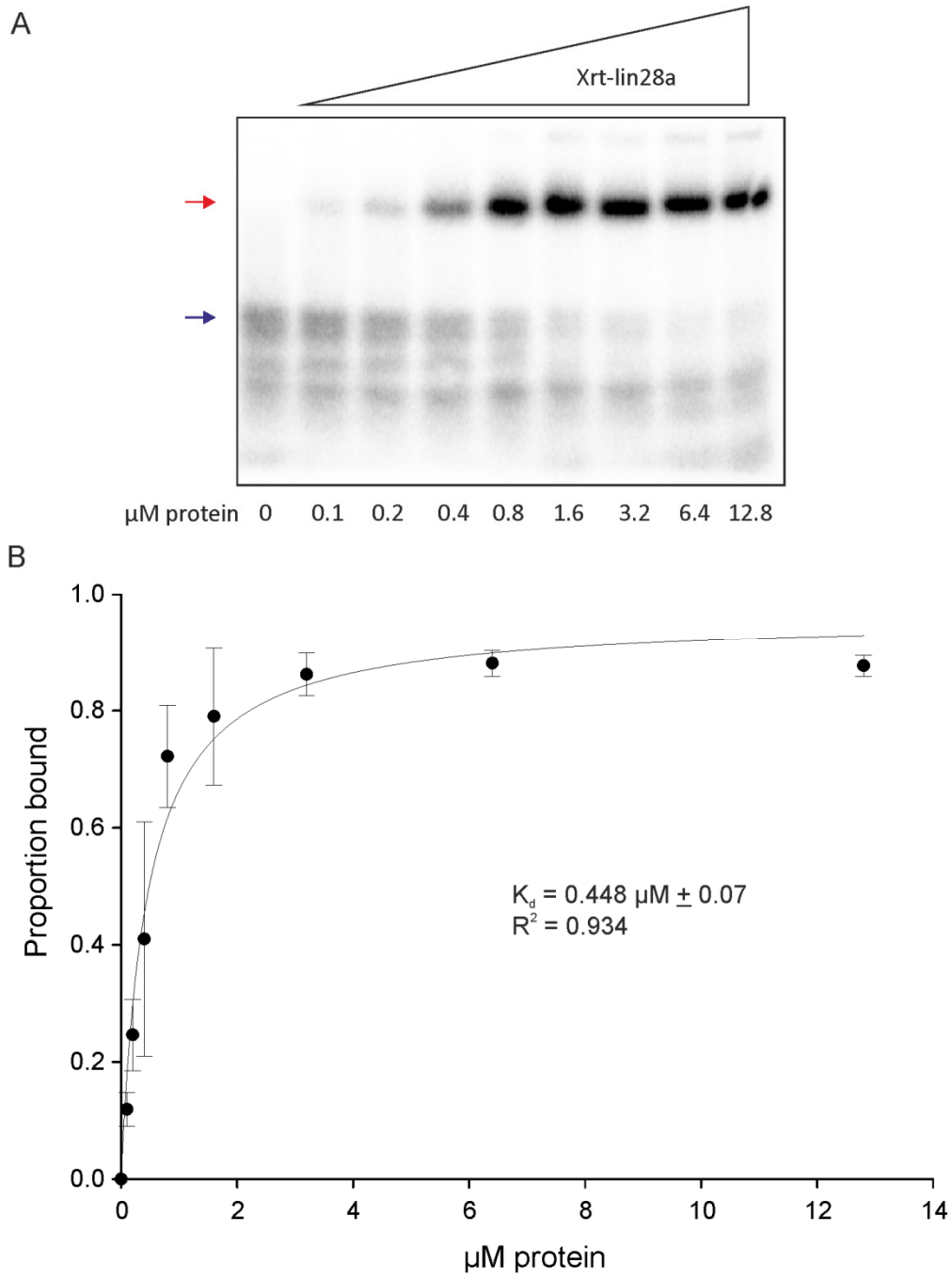


Figure 6.7 Interaction of *Xenopus* lin28a with mir-363 terminal loop

A) EMSA performed with ^{32}P-L-mir-363 and indicated concentrations of Xrt-lin28a. Gel shown is representative of n=3. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red). B) Band intensities were quantified from three independent experiments and the proportion bound was calculated. Data were used to plot a binding curve. B_{max} = 0.962.

6.2.4 The specificity of lin28 targets

To determine the specificity of this interaction, it must be shown that lin28 is not capable of binding to all RNA. pre-miR-138 has previously been shown to not bind with lin28 (Piskounova et al., 2008), and RNA for the terminal loop region of this *Xenopus* miRNA was used (L-mir-138). Interestingly, some binding was detected between lin28 and the mir-138 RNA (Figure 6.8); however, this required much greater concentrations of protein than either of the previous experiments. Over the concentrations used, the B_{max}, calculated as 0.22 indicated that the majority of mir-138 RNA would not be bound by the protein, and therefore an appropriate K_d could not be determined (plot not shown).

Cold competition assays were also used to demonstrate any preference of lin28 for pre-mir-363. 1 μM of Xrt-lin28 was used, as this was over the calculated K_d for the interaction with L-mir-363, and should bind a high proportion of the RNA, but at levels without an excess of protein and thus true competition for binding would occur. Indeed, when used without any unlabelled competitor RNA, 61% of the labelled RNA was shifted (Figure 6.9, Lane 2). Non-radiolabelled 'cold' RNA was used in excess amounts of 50x, 100x and 200x to compete with ^{32}P-L-mir-363 for binding of the lin28 protein. Cold competition using the same sequence in excess successfully competed for binding to lin28, resulting in up to 98% of the labelled RNA remaining unbound with 200x cold excess (Figure 6.9, Lanes 3-5). This indicates that the protein was preferentially binding to the unlabelled RNA. Cold L-mir-138 RNA was unable to compete to the same degree, and larger amounts of L-mir-363 were bound to the lin28: 13% remained bound to lin28 with 200x excess of L-mir-138 (Figure 6.9, Lanes 6-8). Nevertheless, a reduction in bound L-mir-363 was seen with excess L-mir-138, suggestive of some competition.

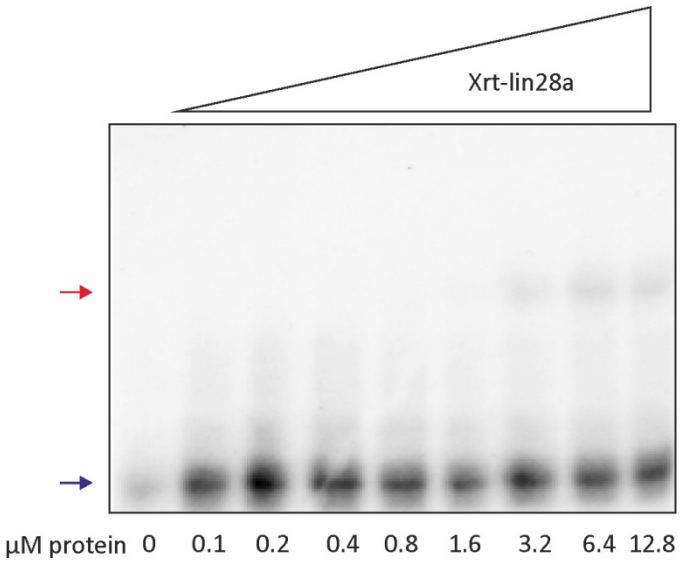


Figure 6.8 Interaction of mir-138 terminal loop with *Xenopus* lin28a
EMSA performed with ^{32}P-L-mir-138 and indicated concentrations of Xrt-lin28a. Arrows indicate RNA and lin28a-RNA complex. Gel shown is representative of n=3. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red).

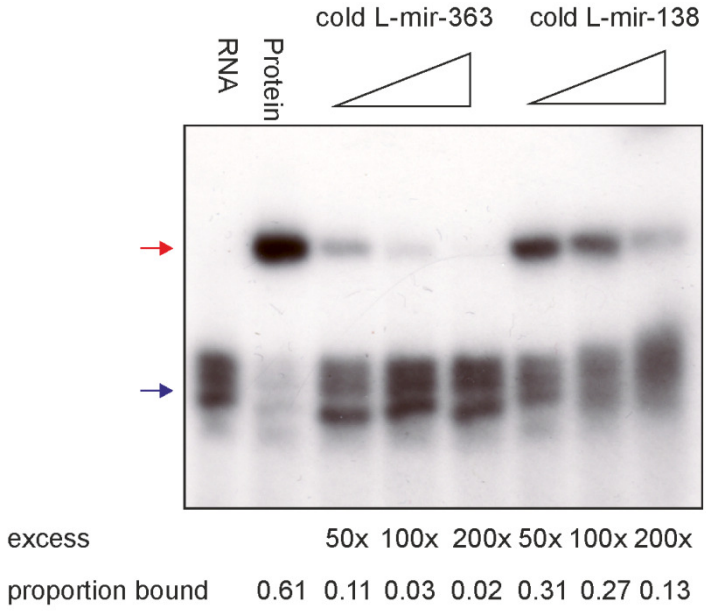


Figure 6.9 Cold competition of mir-363 loop binding by mir-363 compared to mir-138
EMSA performed with ^{32}P-L-mir-363 and 1 µM of Xrt-lin28a (except for RNA only lane). Arrows indicate RNA and lin28a-RNA complex. Reactions were competed with unlabelled RNA of L-mir-363 or L-mir-138 in excess levels as indicated. Band intensities were quantified and proportion of RNA bound was calculated. Gel shown is representative of n=2. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red).

6.2.5 The GGAG sequence in pre-mir-363 as a lin28 binding site

To determine if the GGAG motif is important for the binding of lin28 with mir-363, as it is with let-7, an RNA sequence was used with this sequence mutated to GUAU (mL-mir-363), because Heo et al (2009) found that this sequence reduced the binding interaction. Binding of Xrt-lin28a to this mutated sequence was seen, nevertheless this was reduced compared to the normal sequence, with higher amounts of protein required to see a shift (Figure 6.10A). This result was reflected in the reduction of the maximum proportion of RNA bound to 58% (Figure 6.10B).

Performing a cold competition with the mutated mir-363 against the wildtype sequence also demonstrated a preference for the sequence containing the GGAG motif. As before, competition of ^{32}P-L-mir-363 with unlabelled L-mir-363 resulted in up to a complete loss of bound RNA (Figure 6.11, Lanes 3-5), whereas competition with unlabelled mL-mir-363 was slightly less effective at competing at the highest excess used, although showed greater competition at lower excesses (Figure 6.11, Lanes 6-8), which would have to be investigated further to determine if there is any significance in this. The mL-mir-363 competed much more successfully than the mir-138 sequence previously.

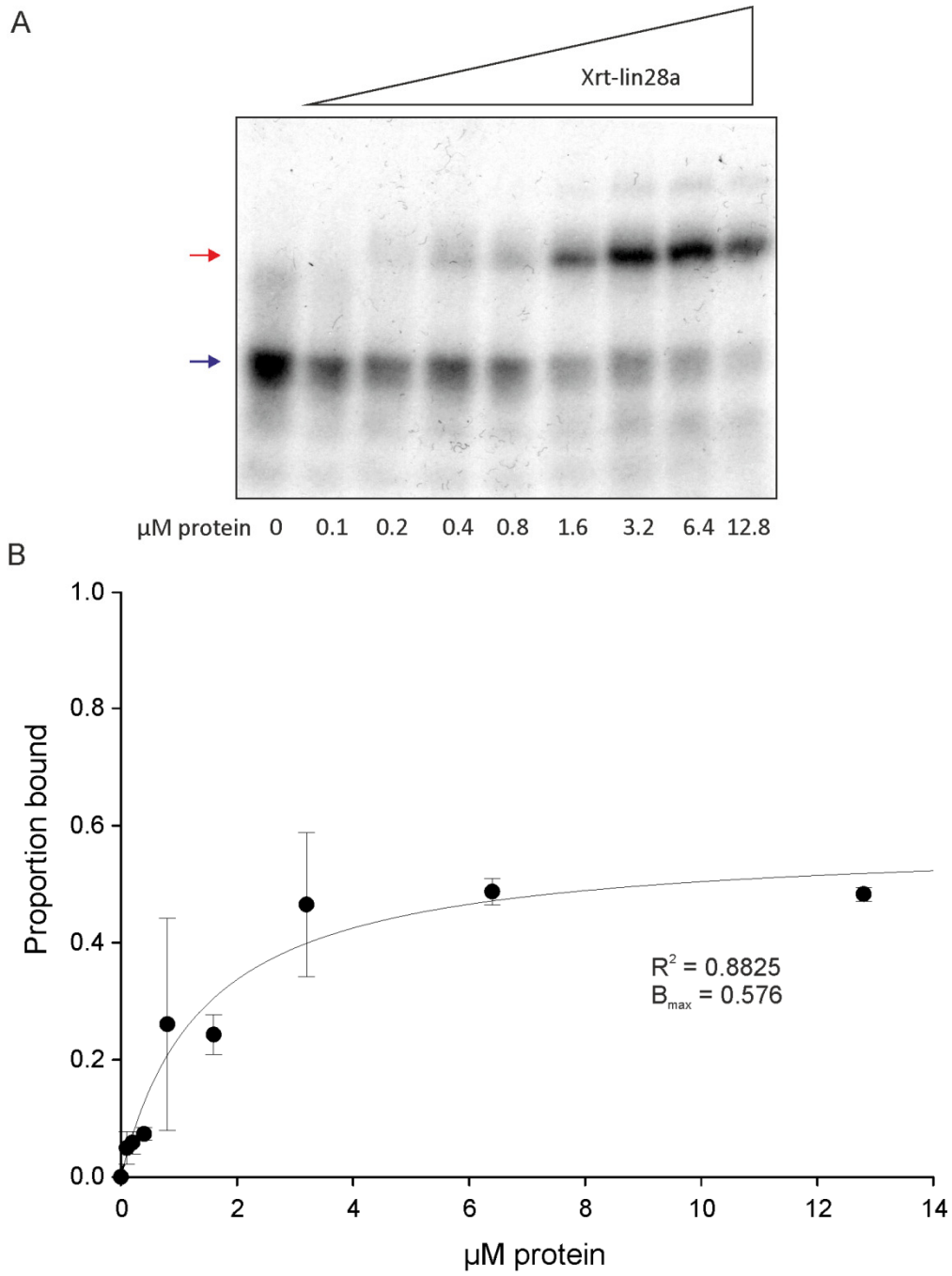


Figure 6.10 Interaction between *Xenopus* lin28a and a mutated mir-363 terminal loop

A) EMSA performed with ^{32}P-mL-mir-363 and indicated concentrations of Xrt-lin28a. Gel shown is representative of n=3. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red). B) Band intensities were quantified from three independent experiments and the proportion bound was calculated. Data were used to plot a binding curve.

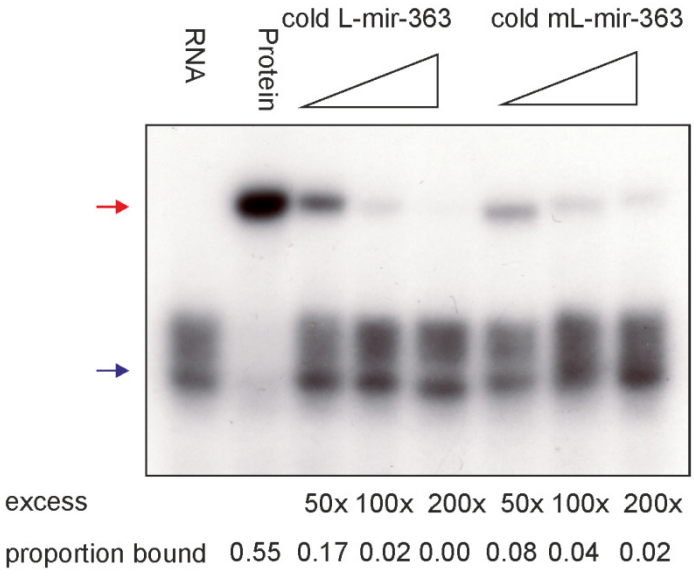


Figure 6.11 Cold competition of mir-363 loop binding by mutated mir-363
EMSA performed with ^{32}P-L-mir-363 and 1 µM of Xrt-lin28a (except for RNA lane). Reactions were competed with unlabelled RNA for L-mir-363 or mL-mir-363 in excess levels as indicated. Band intensities were quantified and proportion of RNA bound was calculated. Gel shown is representative of n=3. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red).

6.2.6 Expressing full-length protein endogenously

It is possible that the C- and N- termini of the lin28a protein can play a part in the recognition of binding targets, with the function of these domains currently unknown. Additionally, it was necessary to identify if there were any obvious differences in the binding specificity of lin28a and lin28b. To investigate these possibilities endogenously translated protein was used in whole embryo extract. *X. laevis* embryos were utilised for this, due to the production of larger amounts of proteins from these embryos compared to *X. tropicalis*. Embryos were injected to overexpress lin28a1, lin28a2, or lin28b. This overexpression was confirmed against control uninjected embryos, which would be expressing normal levels of the lin28s and other proteins. Using western blot analysis, no endogenous protein was detected when embryos were not induced to overexpress either protein, with greatly increased levels following RNA injections (Figure 6.12). The lin28 proteins were not purified from the embryo extract, and as numerous other proteins would be present, quantification of the lin28 protein levels was not possible. Protein was collected from embryos at stage 10.5, as this was the stage at which miRNA regulation had been identified and therefore other proteins normally expressed at this stage would continue to be present, which may be able to affect binding affinity.

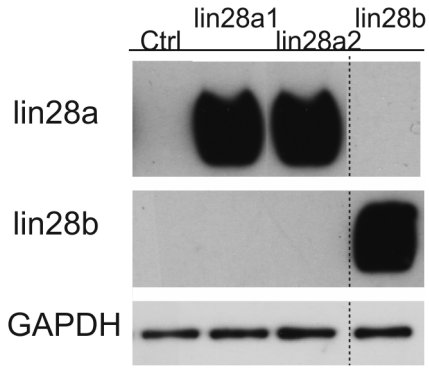


Figure 6.12 Analysis of lin28 protein overexpression by western blot
Western blot analysis of embryos injected with 1 ng/embryo of *lin28a1*, *a2*, or *b* individually, compared to control embryos, at stage 10.5. GAPDH was used as a loading control.

6.2.7 Assessing binding ability of full-length lin28 proteins

The capability of the embryo extracts to bind RNA was tested using L-let-7g. Interestingly, it was seen that embryo extract overexpressing lin28a1 produced a shift of the RNA, whereas normal control extract did not (Figure 6.13A), as endogenous levels of the lin28 proteins would be in the control extract. This was also seen for lin28a2 (Figure 6.13B) and lin28b (Figure 6.13C). To confirm that this shift was due to the lin28 proteins themselves and not another factor which had been upregulated following the lin28 overexpression, a supershift was performed using α-lin28a/b antibodies. For all three proteins a supershift was observed following the addition of the relevant antibody, with a depletion in the intensity of the shift which indicated the protein-RNA complex (Figure 6.13A-C, Lane 5). This supershift did not occur to completion, and more diluted extract was used for future assays due to a probable overload of protein. A pre-immune bleed serum control was also used to control for antibody specificity, and this did not produce a supershift (Figure 6.13A-C, Lane 6).

EMSA with the same embryo extract was also carried out using labelled L-mir-363. lin28a1 and lin28a2 overexpressing extract showed binding to the RNA, which decreased with a dilution in the embryo extract, and these proteins were also supershifted with lin28a antibody (Figure 6.14A,B). Binding to the RNA was observed with lin28b, but this did not appear to be as strong an interaction (Figure 6.14C). Interestingly, with lin28b another shifted band can be seen. This does not alter in any of the conditions, and appears to be a lin28-independent interaction with the RNA. This band was also present in the EMSAs performed with lin28a proteins, but due to stronger binding to L-mir-363 with the lin28as, this additional band was weaker and only observed over long exposures. We cannot quantify the amount of lin28 proteins in the embryo extract, and so it cannot be

quantitated here if the lin28as are better at binding the mir-363 RNA compared to lin28b, or whether less lin28b protein is merely available for the interaction. These EMSAs were carried out using a more diluted concentration of embryo extract than with the L-let-7g, to allow any supershift to be seen to completion. Additionally, it was found that when using more concentrated embryo extract, in all embryo extract conditions, we were unable to detect the unbound RNA, perhaps indicating that this was possibly unstable and being degraded or processed by proteins in the extract, and the true proportions of bound RNA could not be calculated (data not shown).

As before, the importance of the GGAG motif in this binding interaction was tested. The same embryo extract and conditions were used as with the L-mir-363 for labelled mL-mir-363. It was found that both lin28a proteins still bound the RNA, confirmed by the supershift (Figure 6.15A,B). Less RNA appeared shifted in this interaction than with the wildtype sequence, and as the same extract at the same dilutions was used here as previously used with the non-mutated sequence, it can be seen that the loss of the GGAG sequence does affect binding. For lin28b, the interaction was greatly weakened and a very small proportion of RNA was shifted, and also supershifted (Figure 6.15C).

As binding was observed using the mir-138 RNA with recombinant protein, this was also used with embryo extract which should indicate a more biologically relevant protein concentration. Only in one lane with embryo extract overexpressing lin28a1 was any observable shift detected (Figure 6.16), and as such this demonstrates a clear preference of the lin28 proteins for the other RNAs studied here.

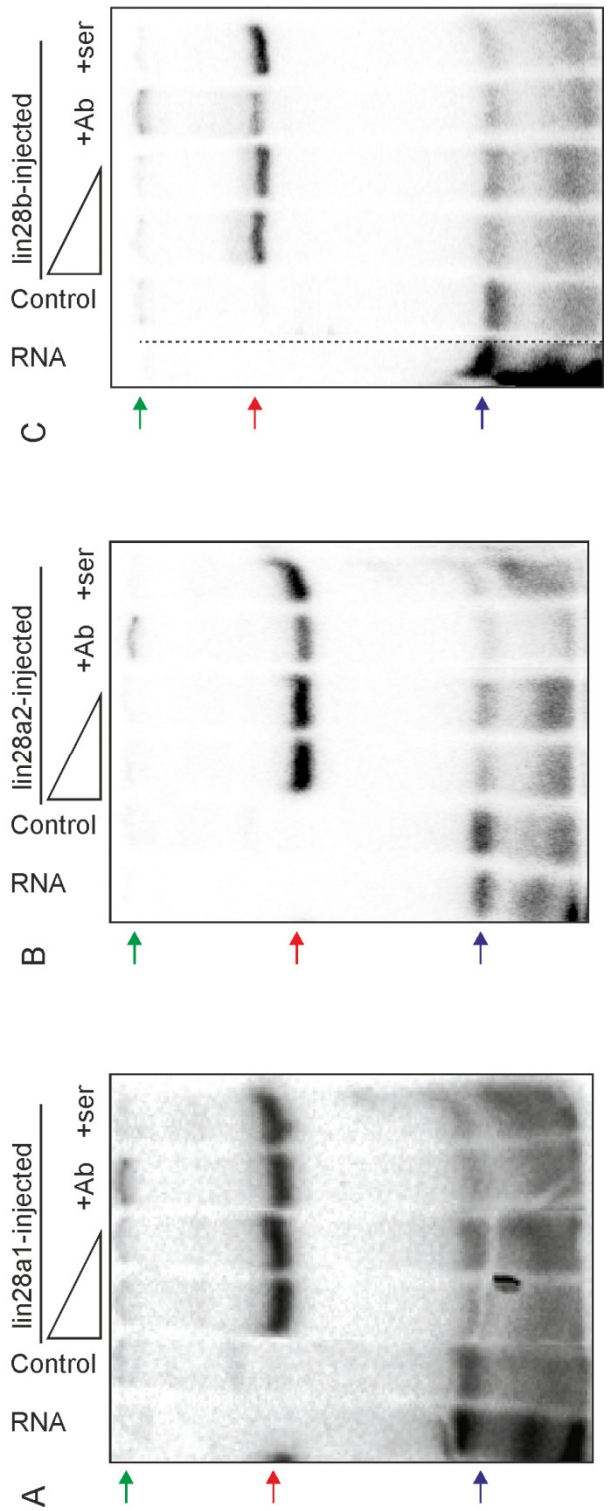


Figure 6.13 Interaction between the let7 terminal loop and endogenously translated lin28 proteins
EMSAs performed with ^{32}P-L-let-7g and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*. Embryo extract was used at 1/4 dilution for lanes 2-3, 5-6, and at 1/8 dilution for lower concentration of overexpressing extract in lane 4. Arrows indicate unbound RNA (blue), lin28-RNA complex (red), and supershift complex of antibody-lin28-RNA (green). +Ab = 1/20 dilution α-lin28a/b, +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

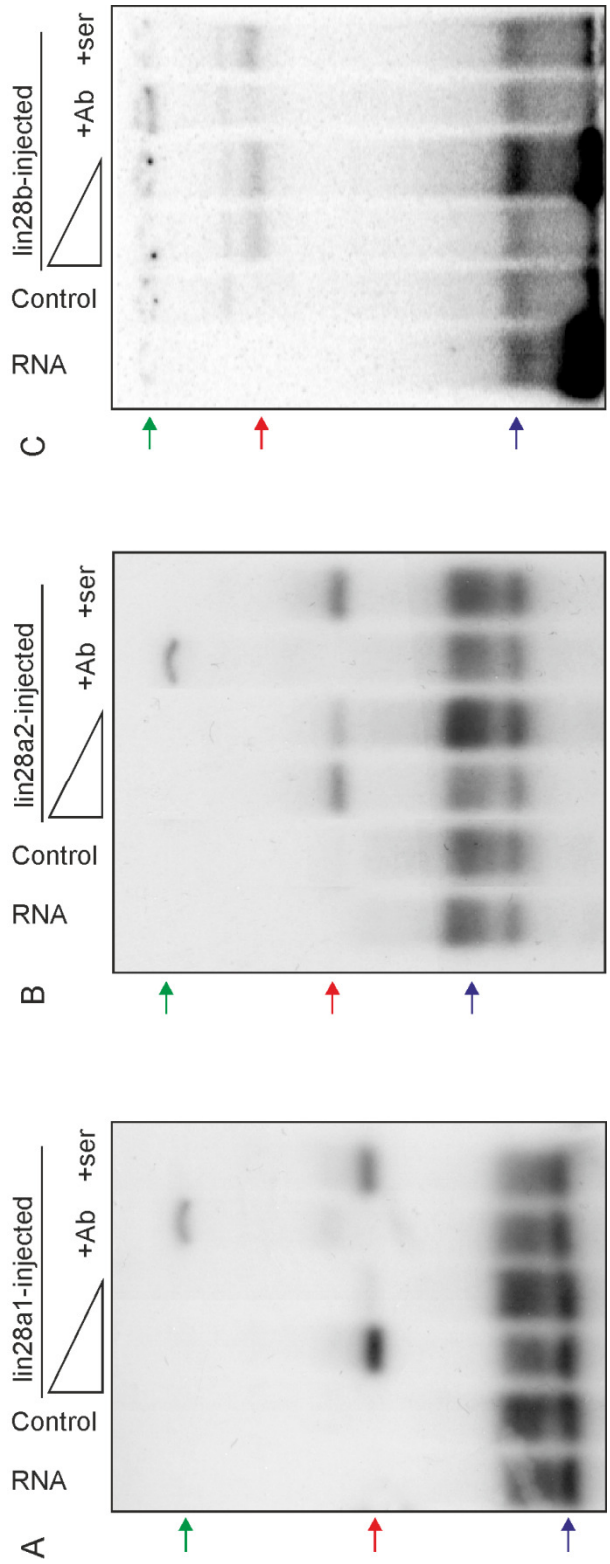


Figure 6.14 Interaction between the mir-363 terminal loop and endogenously translated lin28 proteins
EMSAs performed with ^{32}P-L-mir-363 and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*. Embryo extract was used at 1/16 dilution for lanes 2-3, 5-6, and at 1/32 dilution for lower concentration of overexpressing extract in lane 4. Arrows indicate unbound RNA (blue), lin28-RNA complex (red), and supershift complex of antibody-lin28-RNA (green). +Ab = 1/20 dilution α-lin28a/b, +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

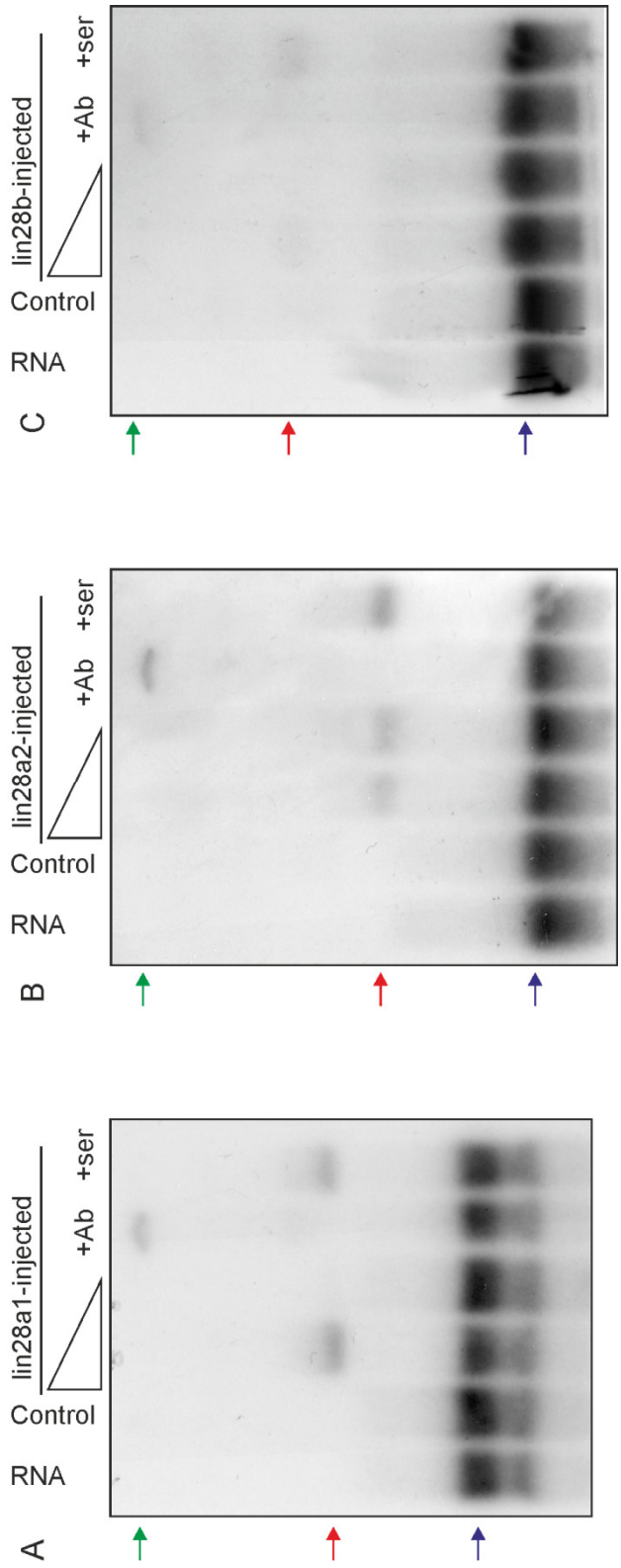


Figure 6.15 Interaction between a mutated mir-363 sequence and endogenously translated lin28 proteins
EMSAs performed with ^{32}P-mL-mir-363 and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*. Embryo extract was used at 1/16 dilution for lanes 2-3, 5-6, and at 1/32 dilution for lower concentration of overexpressing extract in lane 4. Arrows indicate unbound RNA (blue), lin28-RNA complex (red), and supershift complex of antibody-lin28-RNA (green). +Ab = 1/20 dilution α-lin28a/b, +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

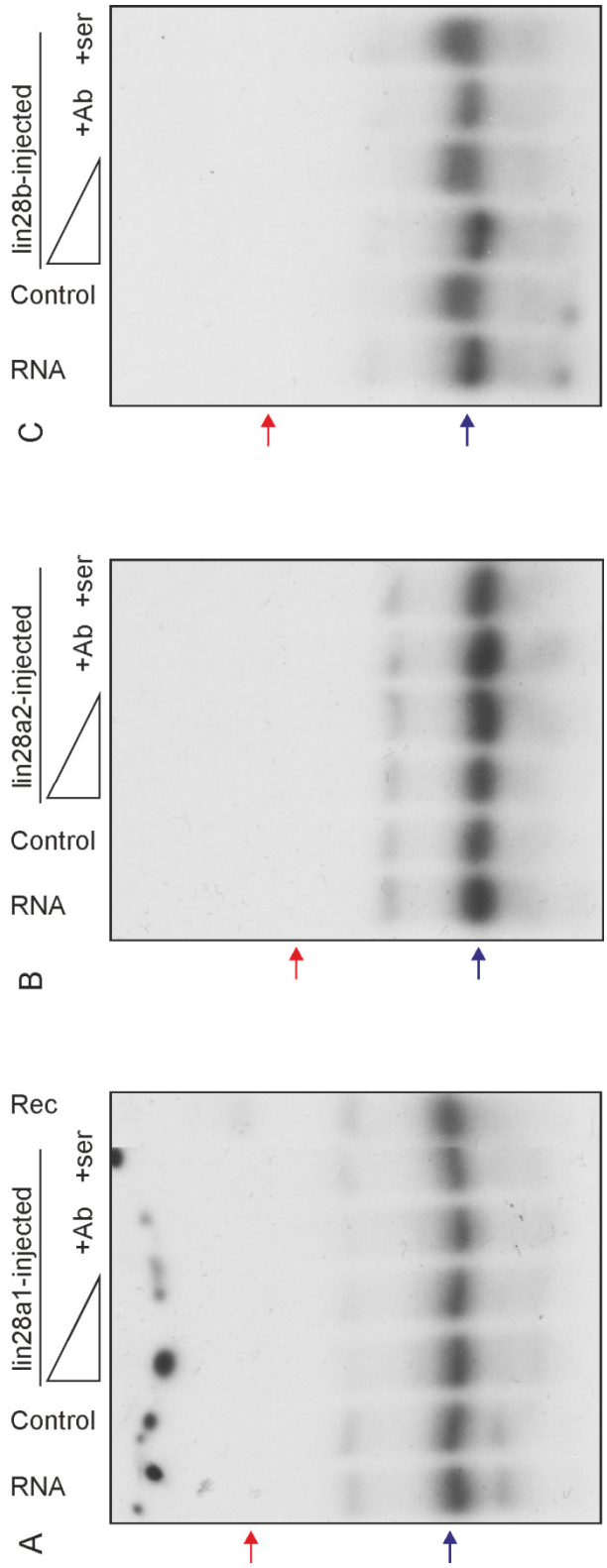


Figure 6.16 Interaction between the mir-138 terminal loop and endogenously translated lin28 proteins
EMSAs performed with ^{32}P-mL-mir138 and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*. Embryo extract was used at 1/16 dilution for lanes 2-3, 5-6, and at 1/32 dilution for lower concentration of overexpressing extract. in lane 4 Arrows indicate unbound RNA (blue) and expected height of lin28-RNA complex (red). +Ab = 1/20 dilution α-lin28a/b, +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

6.2.8 Binding of full-length pre-miRNA

Binding reactions were then carried out with full-length pre-miRNA, using EMSA with both the recombinant and *in vivo* translated full-length protein. Use of the full-length endogenous sequence would also address the potential miRNA instability seen earlier with L-mir-363. Precursor miRNA sequences for both xtr-let-7g and xtr-mir-363 were cloned by RT-PCR to allow *in vitro* transcription of pre-miRNA. These transcripts were then labelled and used as for RNA sequences above.

With let-7, previous work has suggested that the binding efficiency for the loop region is comparable to the whole RNA, due to the presence of the binding sites within this area (Desjardins et al., 2011; Piskounova et al., 2008). However, when using the recombinant Xt-lin28a protein against the pre-let-7g, a much increased binding affinity was seen (Figure 6.17A), with a K_d of 0.014 µM; however a B_{max} value of only 0.883 was calculated (Figure 6.17B), although it was seen with multiple experiments that 100% of RNA was bound at a number of alternative concentrations (data not shown). This increase in binding efficiency may be partially due to using the human sequence when looking at the interaction with the loop, and the *X. tropicalis* RNA for the precursor, which do differ slightly in sequence and involve differences in the sequence over the supposed binding site for the lin28 CSD (Figure 6.18). The low apparent B_{max} value may result from fitting the one-site binding equation over a longer range of protein concentrations, where a multiple-site binding equation might be more appropriate.

Using the *in vivo* translated protein, pre-let-7 was bound by embryo extract overexpressing each of the three lin28 proteins, and not by control extract (Figure 6.19). The supershift again produced a reduction in the shifted band relating to the lin28-let-7 interaction, but proved difficult to resolve as an extra band, with more radioactivity trapped high up the gel than in previous cases. The serum control did not differ to the extract-only, indicating that the lin28s were still responsible for the majority shift. There were a larger number of alternatively shifted bands with all embryo extracts and this RNA compared to the loop alone, which appeared to be lin28-independent. This hints at a possibility of additional binding sites in the full precursor that are targeted by other endogenous proteins.

Pre-mir-363 was, however, bound by Xtr-lin28a with a similar affinity as the L-mir-363. The RNA was seen to shift over the same range of protein concentrations as previously (Figure 6.20A), and a K_d was calculated as 0.46 µM, which compares to the dissociation constant of the terminal loop, calculated earlier as 0.448 µM (Figure 6.20B).

Pre-mir-363 showed different results with the full-length proteins. The control extract did not result in a shift attributable to lin28, but binding was still evident when overexpressing both isoforms of lin28a (Figure 6.21A,B), and again a supershift demonstrated the role of the lin28 proteins themselves in this interaction. However very little binding was apparent with lin28b (Figure 6.21C). In all embryo extract used, binding independent of lin28 was seen for this RNA, and appeared to be a more dominant binding partner than the lin28s for this sequence. No loss of detection of the unbound RNA was seen with the precursor at higher concentrations of embryo extract, as with the loop earlier, indicating that such a result earlier would have been an artefact of the shortened sequence and not a precursor miRNA stability issue.

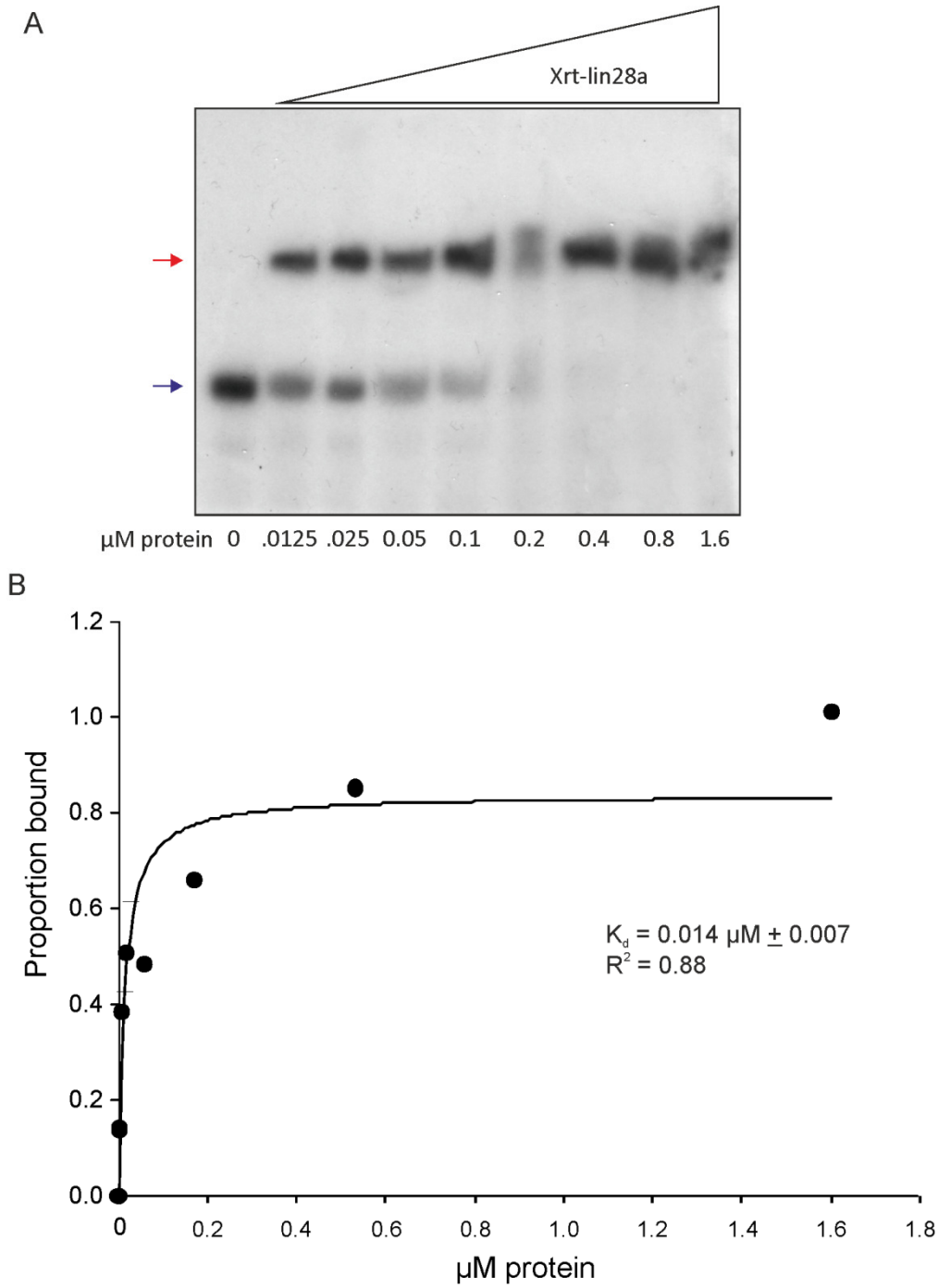


Figure 6.17 Binding of *Xenopus* lin28a with pre-let-7g

A) EMSA performed with ^{32}P-pre-let-7g and indicated concentrations of Xrt-lin28a. . Arrows indicate labelled RNA (blue) and lin28-RNA complex (red). B) Band intensities were quantified and the proportion bound was calculated. Data were used to plot a binding curve. B_{max} = 0.8331.

Figure 6.18 Alignment of pre-let-7g sequences

Alignment of *X. tropicalis* (Xtr) and human (Hsa) pre-let-7g sequences. * denotes conserved nucleotide. Red = let-7g miRNA. Yellow = L-let-7g RNA sequence. Green boxes = suggested lin28 binding sites.

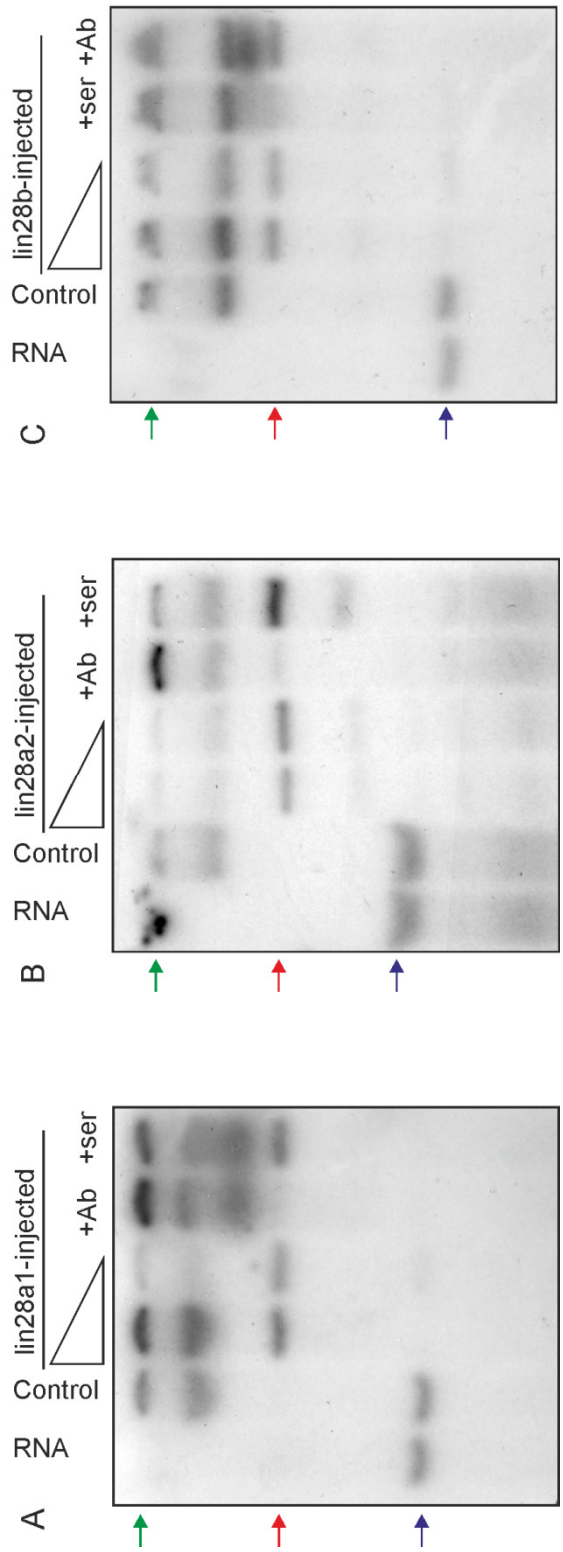


Figure 6.19 Interaction between pre-let-7g and endogenously translated lin28 proteins
EMSAs performed with ^{32}P-pre-let-7g and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*. Embryo extract was used at 1/8 dilution for lanes 2-3, 5-6, and at 1/16 dilution for lower concentration of overexpressing extract in lane 4. Arrows indicate unbound RNA (blue), lin28-RNA complex (red), and supershift complex of antibody-lin28-RNA (green). +Ab = 1/20 dilution α-lin28a/b , +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

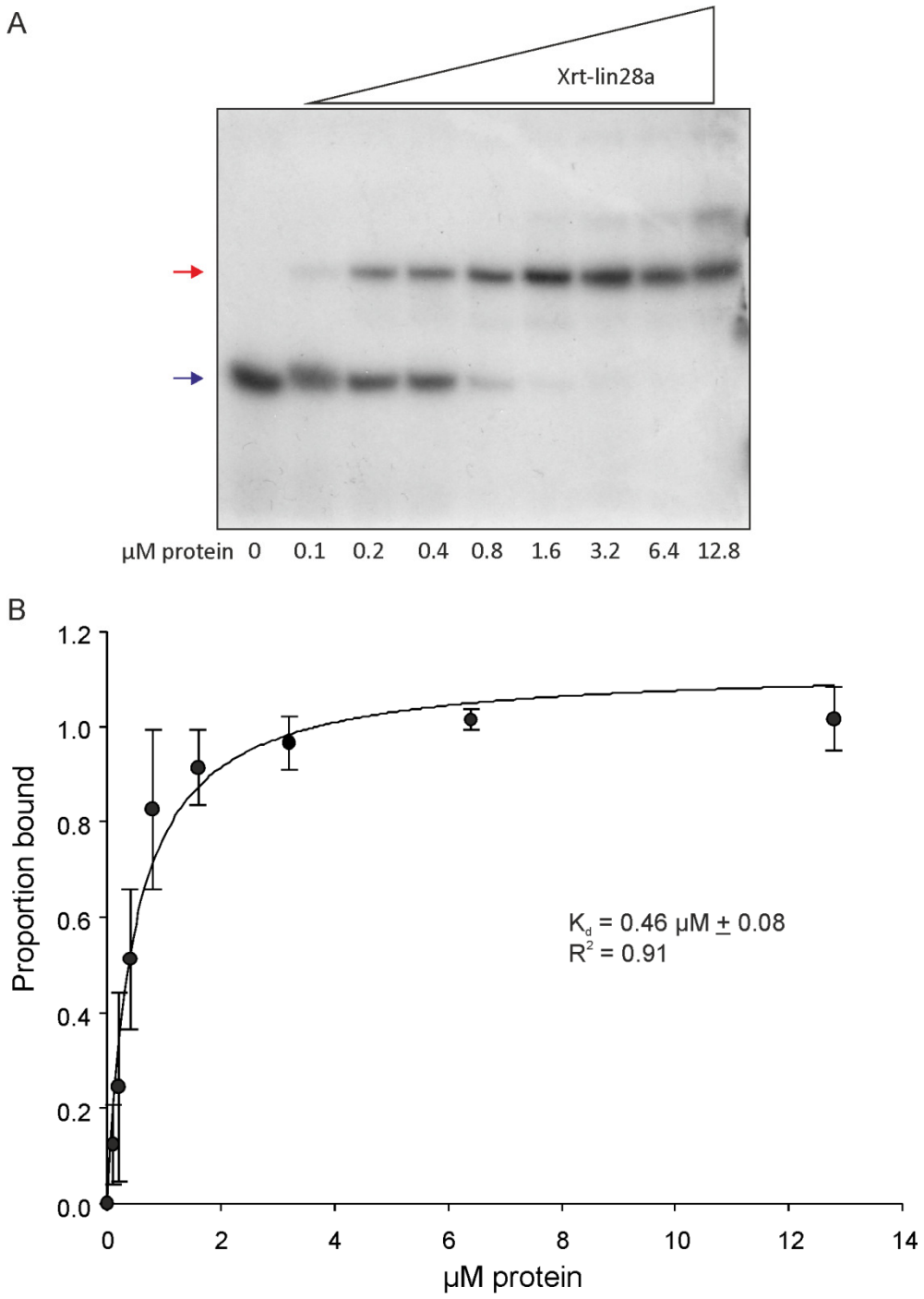


Figure 6.20 Binding of *Xenopus* lin28a with pre-mir-363

A) EMSA performed with ^{32}P-pre-mir-363 and indicated concentrations of Xrt-lin28a. Gel shown is representative of n=2. Arrows indicate labelled RNA (blue) and lin28-RNA complex (red). B) Band intensities were quantified from two independent experiments and the proportion bound was calculated. Data were used to plot a binding curve. B_{max} = 1.12.

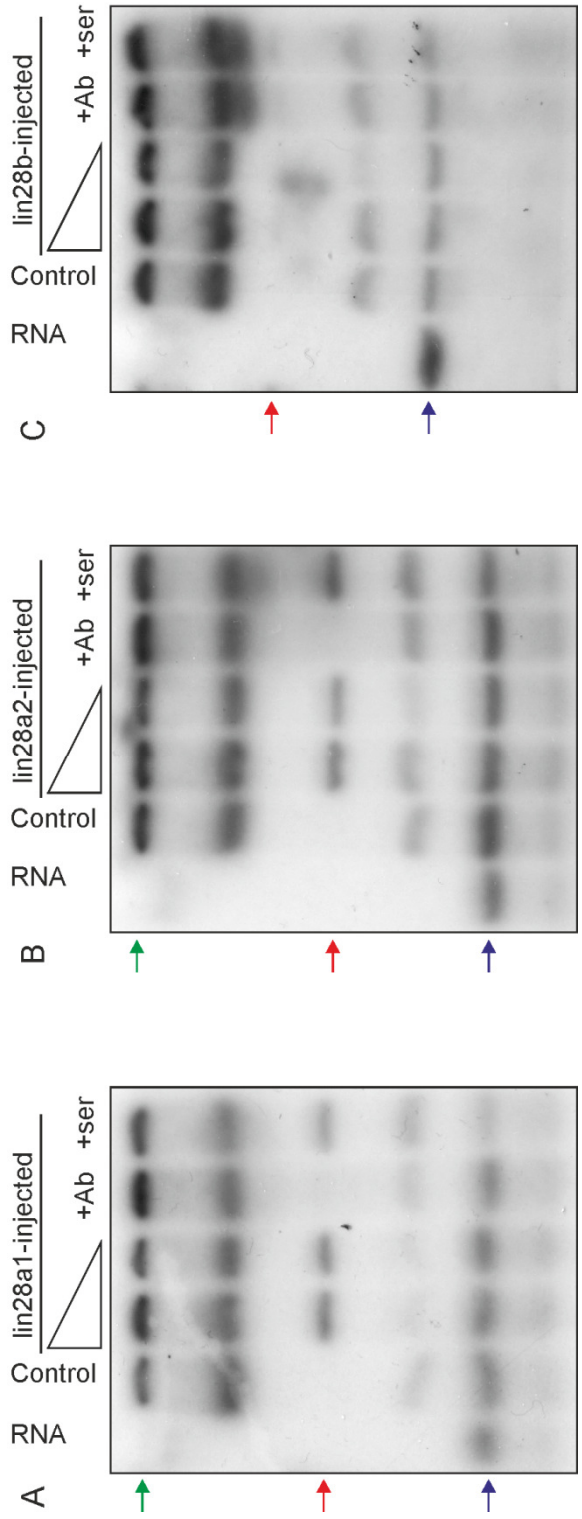


Figure 6.21 Interaction between pre-mir-363 and endogenously translated lin28 proteins

EMSAs performed with ^{32}P-pre-mir-363and embryo extract from uninjected controls or embryos injected with 1 ng of either A) *lin28a1*, B) *lin28a2* or C) *lin28b*.Embryo extract was used at 1/8 dilution for lanes 2-3, 5-6, and at 1/16 dilution for lower concentration of overexpressing extract in lane 4. Arrows indicate unbound RNA (blue), lin28-RNA complex (red), and supershift complex of antibody-lin28-RNA (green). +Ab = 1/20 dilution α-lin28a/b , +ser = 1/20 dilution pre-immune bleed serum, both incubated with protein on ice for 20 minutes before addition of probe.

6.3 Discussion

6.3.1 *Xenopus* lin28 proteins bind to let-7

This work has further demonstrated binding between let-7 and lin28 proteins. The K_d for this interaction with the terminal loop was determined to be 0.314 µM, and for the full length precursor was 13.7 nM. These are both between the published K_d values for this interaction, which range from 0.02 nM to 2.1 µM (Desjardins et al., 2011; Heo et al., 2009; Lei et al., 2012; Lightfoot et al., 2011; Loughlin et al., 2012; Nam et al., 2011; Newman et al., 2008; Piskounova et al., 2011; Piskounova et al., 2008). The use in these experiments, however, of a range of protein constructs, RNA sequences and experimental conditions may have affected binding affinity. Additionally, the binding coefficients were calculated here using a one-site binding equation, as carried out by others including Piskounova et al (2008) and Lightfoot and colleagues (2011). However, since the discovery of multiple binding sites, applying the Hill equation to determine dissociation constant may be more appropriate, such as was performed by Desjardins et al (2011), as this allows for and calculates any multiple binding sites. Nevertheless, they also applied the one-site binding site equation and found K_d values to be in a similar range.

Using the *in vitro* transcribed let-7 precursor, binding efficiency was increased greatly. This does differ from previous results, which found similar binding efficiencies between the loop and pre-let-7, which allowed the identification of the terminal loop as the lin28 binding site (Piskounova et al., 2008). However, one explanation for this may be that the loop RNA sequence used here was human, and the pre-let-7 and protein were from *Xenopus.* A potentially key difference between the human and *Xenopus* sequences was present within the putative CSD binding site. The binding site published by the Sliz lab as 'NGNGAYNNN' (Nam et al., 2011) appears, in human let-7g, to be the sequence 'UAUGAUACC', missing the first 'G' residue, which they stated made the interaction less energetically stable. The *Xenopus* sequence is altered to 'UGUGACACC', featuring a different pyrimidine at position 6, and additionally has the first G residue. The increased binding to the *Xenopus* RNA may indicate that this is an improved binding site for the lin28. Then again, both of these sequences differ by only one important nucleotide from the alternatively proposed 'GUNNUNN' binding site (Mayr et al., 2012), with the human pre-let-7g sequence missing the first 'G' and the *Xenopus* sequence lacking the second 'U' residue which, if the more correct sequence, may not explain such different binding specificities between the proteins, with each showing a simple sequence mutation. Further analyses using truncations and mutations of both the human and *Xenopus* sequences may

help determine which of these binding motifs is more accurate, and whether any evolutionary differences exist between the species.

There may also have been folding artefacts from the RNA, allowing exposures of different residues for binding. The sequence for the loop was suggested to fold into the endogenous conformation using mFOLD, and we would expect the same favoured folding from the *in-vitro* synthesised precursor; however, in either case there may have been alternative folding structures that altered binding specificity and therefore efficiency.

Both lin28a and lin28b have been shown to bind and inhibit let-7 processing (Viswanathan et al., 2008), and binding was seen here with all of the full-length proteins overexpressed in embryo extract. As we cannot quantitate the concentrations of protein in these extracts, we cannot conclude whether one protein shows improved binding over the others, but can conclude that all show capable interactions.

At the exposures used, no binding to any RNA sequences was seen by lin28 in the control extract, and using longer exposures did not vastly improve this. A failure to detect much binding of the RNA from endogenous protein was a little surprising, particularly with the mir-363 RNA, following our proposal that lin28 is binding this miRNA during gastrulation, from which stage the embryos were collected. The levels of lin28 in extract cannot be quantified accurately, and so cannot be determined whether the endogenous concentration is similar to the K_d for the interaction; additionally, the use of assay conditions may not completely accurately reflect physiological conditions, which could affect the ability to observe endogenous interactions without gross overexpression of one element.

6.3.2 *Xenopus* lin28 proteins bind to pre-mir-363 terminal loop

Work here has shown that the lin28 proteins are capable of binding to elements within the mir-363 terminal loop. This may happen at a lower affinity than the interaction with let-7, evidenced by a weaker dissociation constant when using the full-length precursors, but nevertheless shows preferential binding compared to alternative RNA such as mir-138. Work here has shown that mir-138 is capable of binding to lin28, but does so at a much reduced affinity, nevertheless this reduced affinity was effective for the needs here. The high levels of protein required may also not reflect levels that are found endogenously, and thus may not be a likely interaction *in vivo*. Likewise, this may also be the case regarding the amounts of RNA used at high levels of excess for successful cold competition. As the protein concentration was selected to ensure that this was the limiting

factor in the interactions, an excess of any RNA would have an advantage, particularly when capable of binding to the protein at some level.

All of the overexpressed full-length lin28a and lin28b proteins were able to effectively bind the terminal loop of the mir-363 RNA. The same embryo extract was used to investigate this interaction as was used with the let-7, and a slightly increased ability to shift the mir-363 RNA was seen with the lin28a proteins over the lin28b, perhaps indicating a stronger interaction from these proteins. This difference was further seen when using the full length precursor and mutant RNA. The lin28a proteins were both able to result in a prominent shift of the RNA, whereas overexpression of lin28b did not. It remains unknown what may have caused this apparent difference in binding preference. It may be that there was less available or overexpressed lin28b protein compared to the lin28as, or it may indeed have a lower binding affinity for mir-363. The truncated RNA used to investigate binding in the loop was anticipated to fold in the same way as the endogenous loop on the precursor, causing exposure of the same residues, but this is possibly not the case. The stem of the precursor may be altering binding preference to this RNA, but this is not seen in the recombinant lin28a. It is therefore possible that the presence of the N- and C-termini of the lin28 proteins affects their binding affinity, perhaps in combination with the RNA stem. One way to test this would be to use a recombinant truncated lin28b protein, and determine if binding to the precursor is affected to the same degree.

6.3.3 Possible lin28 binding sites in pre-mir-363

A similar K_d for mir-363 between the terminal loop and full length precursor binding to recombinant protein suggests that the binding sites are within this domain.

The identification of 'GGAG' as a minimum binding site for uridylation activity of lin28 on let-7 (Heo et al., 2009) led to the conclusion that it was a minimum binding site for all interactions, as mutation of this sequence to 'GUAU' weakened binding, although it did not fully inhibit it. This sequence was consistent with the discovery that the 'GG' motif was important for binding (Newman et al., 2008), along with other work that demonstrated a loss of either of the first 'G' residues reduced binding, but not as greatly as a loss of the fourth 'G' or the first two in combination (Mayr et al., 2012). The high conservation of the GGAG sequence between species containing pre-mir-363 suggests that it may have some biological importance. Indeed, results here have shown that mutating this sequence does have a negative effect on the binding between lin28 and the pre-mir-363 terminal loop; however it did not abolish all interaction. This also supports the more degenerate sequence proposed by Loughlin and colleagues (2012) of 'NGNNG', as the sequence used

retained the first 'G' residue. The alteration of 'GGAG' to 'GUAU' does, however, alter the expected folding of the RNA. This effect on the structure may have also affected the binding of the protein.

Further strength can be given to an interaction between mir-363 and lin28 with an identification of a binding site for the CSD. Interestingly, zebrafish pre-mir-363 contains the proposed motif 'NGNGAYNNN' near to the beginning of the terminal loop, but no other species contains this in a complete form. A conserved 'GAU' sequence is seen in 13 other species, including humans although not *Xenopus*, but this is not preceded by another 'G' residue and features at the very beginning of the loop, and as such its availability for binding to the lin28 protein is unknown. Alternatively, to fit the motif of 'GUNNUNN', there are multiple occurrences of 'GU', and 'UNNU' present within the loop region, but none of these match the full sequence. Nevertheless, the CSD has been described as showing flexibility with a broad spectrum of potential binding sequences and binds to pyrimidine-rich single-stranded sequences. With a large number of 'U' residues, the mir-363 loop could be defined as this, leaving the possibility that a binding site is present intact. To identify any specific binding site, further mutations would need to be carried out on the RNA sequence to investigate how they affect binding. Determining a crystal structure of the RNA bound to lin28 would also identify in the nature of this interaction.

6.3.4 The same binding interactions result in different regulatory mechanisms?

The current understanding of the effects that binding via the above sites has on the let-7s is explained by two main mechanisms: the uridylation of the RNA and the blocking of Dicer cleavage. Both of these negatively regulate processing to the mature form, whereas lin28 may be having opposite effects on the miR-17-92 and miR-106-363 clusters, and thus must be functioning in a different manner.

It was discovered that both lin28a and lin28b bound to a modified form of pre-let-7 in cells, which was elongated with a uridylation tail, and unable to be processed by Dicer (Heo et al., 2008). The authors further demonstrated that the uridylation of this RNA was dependent upon the binding of both lin28 and TUT4 (Heo et al., 2009). They went on to identify multiple miRNAs other than the let-7s that have a GGAG sequence in their terminal loop; they found that others, such as mir-107, could be both uridylated and bound by lin28, however mir-363 did not show uridylation by lin28, and a binding assay was not pursued (Heo et al., 2009). Work has shown here that pre-mir-363 can be bound by lin28, and work in Chapter 4 indicated that there is a positive regulation of the miRNA,

supporting the lack of uridylation seen previously. It was postulated that as TUT4 also contains zinc knuckles, the second G in the 'GGAG' motif may be exposed by lin28 binding and provide a binding site for the uridylase (Loughlin et al., 2012). The mir-363, however, contains this same sequence, and this does not appear to be the mechanism of action here.

The binding of the RNA by lin28 has been shown to remodel the RNA, which is suggested to block the Dicer cleavage site (Nam et al., 2011). The remodelling of the RNA may involve a partial melting of the RNA stem over the Dicer cleavage site, which renders it unrecognisable to Dicer (Lightfoot et al., 2011), as Dicer cleaves dsRNA (Zhang et al., 2004). This is believed to come about following the binding of the zinc fingers to the GGAG motif in the let-7 RNAs, which were present four nucleotides from the 3' miRNA (Loughlin et al., 2012; Nam et al., 2011). Again, the position of the GGAG motif in mir-363 is only three nucleotides from the 3' miRNA, which would perhaps be expected to show the same melting effect, and thus inhibition of processing, although this is not seen and the opposite appears to be possible. Remodelling of the miRNA loop to assist processing has previously been seen, with the action of hnRNPA1 on another member of the cluster, miR-18a; hnRNPA1 binds the RNA, whilst as pri-miRNA, aiding Drosha processing (Guil and Caceres, 2007). Mayr and coworkers (Mayr et al., 2012) found that certain proteins with mutations within the CSD were unable to remodel the RNA, but were still capable of binding the RNA. It may be that with a different CSD binding site to the let-7s, the mir-363 is bound by the lin28s but is not remodelled in the same manner. This could allow exposure of different areas of the RNA, that may in fact support further processing or cleavage. The nature of any endogenous binding of lin28 to mir-363 needs to be determined to understand the effect and mechanism of this interaction.

6.3.5 Putative binding sites within the miR-17-92 and miR-106-363 clusters

Numerous putative binding sites for the zinc finger binding domains have been identified here within the clusters, and one of these has been followed up here. This was the GGAG motif within the terminal loop of pre-mir-363.The GGUG motif within this precursor, and both motifs within other precursors, were present within the stem regions. With the remodelling of RNA upon lin28 binding (Lightfoot et al., 2011; Mayr et al., 2012), it is possible that these would be exposed for further binding. Additionally, the structure is not known of the inter-premiRNA sequences to determine whether the putative binding sequences here would be accessible by lin28s. The alternative binding sites for the CSD are highly dependent on the structure of RNA, as research has suggested that this is the

primary binding site, which remodels the RNA and allows the zinc fingers to bind (Mayr et al., 2012). Numerous occurrences of the two proposed CSD binding motifs were present within the clusters, but the structure of the pri-miRNA is not determined sufficiently to implicate them as binding sites, along with the current lack of agreement over a definitive CSD binding site.

Performing binding assays with different lengths and sections of the pri-miRNA would enable identification of any further interactions between these clusters and lin28, and may help explain why changes in expression are seen in multiple members of the cluster following a loss of lin28 proteins. lin28 has also been shown to bind co-transcriptionally to let-7 in *C. elegans* (Van Wynsberghe et al., 2011), and it remains possible that a similar mechanism may be in play here, thus affecting all members of the cluster together. lin28 is reported to bind to mRNA by a different motif, with a stem structure displaying an adenine bulge, flanked by GC pairs, and this attracted alternative co-factors to the interaction on miRNA, for a different effect on the RNA (Lei et al., 2012). It is a further possibility that lin28 may bind the pri-miRNA in a similar way, and whilst it would not be promoting translation, it may recruit further proteins that promote processing.

6.3.6 Current understanding of the regulation of miR-17-92 clusters

Current understanding of the regulation of the miR-17-92 cluster is mostly at the transcriptional level. The cluster is directly regulated by myc, which may additionally regulate the miR-106a-363 cluster, with both c-myc and n-myc increasing miRNA levels (O'Donnell et al., 2005; Schulte et al., 2008). The miR-17-92 and miR-106b-25 clusters are also regulated by E2F transcription factors (Petrocca et al., 2008b; Sylvestre et al., 2007; Woods et al., 2007), which are also negatively regulated targets of members of one of the miRNA families, with miR-17, miR-20a and miR-106b proven to interact with the mRNA (O'Donnell et al., 2005; Petrocca et al., 2008a; Sylvestre et al., 2007).

With the level of pri-miRNA not always reflecting the levels of mature miRNA present, and with different patterns of expression of the miRNAs within clusters, it would appear that the post-transcriptional regulation of the cluster is highly important. Known examples of individual processing of the cluster miRNAs include the targeting of miR-18b alone by hnRNPA1, which positively regulated expression (Guil and Caceres, 2007), along with VEGF, which increased expression of only three members of the cluster: miR-17, miR-20a and miR-18a (Suarez et al., 2008).

To determine if lin28 is directly interacting with, and regulating, the miRNA clusters, immunoprecipitation of the lin28 proteins could be carried out. A subsequent use of northern blots or RT-PCR analysis would be able to determine whether RNA from these miRNAs is bound *in vivo*. Due to the ability to resolve the different forms of miRNA on northern blots, of mature, precursor and primary miRNA, this technique would be able to demonstrate at what level any binding was occurring. This would then provide details to the stage in the miRNA biogenesis at which regulation is occurring.

6.3.7 The co-expression of miR-17-92 clusters with lin28s in development

Work in this chapter has shown that lin28 is capable of binding to miR-363 RNA *in vitro*, but has not demonstrated this in an endogenous setting. Any occurrence of co-expression between the RNA and these proteins strengthens the hypothesis that a direct interaction takes place, and this may occur both in development and cancer. The co-expression during development between the lin28s and these miRNAs is further explored in Chapter 7.

Chapter 7. Characterisation of miR-cluster expression

7.1 Introduction

7.1.1 The miR-17-92 cluster family in development

Investigations of the expression of miR-17-92 clusters have been carried out both in the developing embryo and using *in vitro* systems of ES cells. Much of the work has focused on the miR-17-92 cluster, which through the use of knockout mice was concluded to be the most important of the miR-17-92 related clusters for development, with miR-106a-363 and miR-106b-25 cluster manipulations not resulting in adverse developmental phenotypes (Ventura et al., 2008). Members of the miR-106a-363 cluster have generally been reported at either low levels of mature expression or have not been found to be expressed. On the other hand, there is a high level of conservation between species of the miR-106a-363 cluster, an evolutionary retention that may be indicative of an important function (Suh et al., 2004; Watanabe et al., 2005). A closer focus on this cluster may identify key roles that can be ascribed to its individual members.

In the analysis in Chapter 4, miR-363 and miR-363* were the miRNAs that showed the largest changes upon a loss of lin28 expression and the potential has been demonstrated for a physical interaction with pre-mir-363 and the lin28s, as seen in Chapter 6. Thus these are the miRNAs that are of most interest at this time for investigation of their expression and function.

Both miRNAs derived from pre-mir-363 are relatively unstudied, as this was the last member of the cluster to be identified, along with it being present on the less characterised cluster of the three. Much of the work determining expression of the clusters during development and in cancer studies was prior to pre-mir-363 being identified on the miR-106a-363 cluster. The few studies that have investigated miR-363 expression as part of the cluster have also found some contrasting results. In T-cell leukaemia, when retroviral inserts were present upstream of or within the miR-106a-363 cluster, research has found miR-363 both absent (Landais et al., 2007), and highly upregulated compared to non-cancer cells (Lum et al., 2007). Likewise, expression was not detected during murine development, from E. 8.5 to adulthood (Ventura et al., 2008), but subsequently has been reported in the chick embryo from E3.5 (Huang et al., 2010).

7.1.2 Interest in miR-363 vs miR-363*

Following the processing of pre-mir-363, the 'major' mature miRNA was classed as miR-363-3p. However, with no validated functions for miR-363, it is possible that miR-363* is also functional. Studies that have shown functionality in both precursor miRNA strands found that 'minor' strands that were highly conserved were more likely to be functional (Okamura et al., 2008; Yang et al., 2011), and the sequence analysis shown on the precursor in section 6.2.1 demonstrated a high level of conservation in both the -3p 'major' strand, and the alternative -5p arm. In support of pre-mir-363 producing two functional miRNAs, expression of both miR-363 and miR-363* have been reported in the literature (Huang et al., 2010; Jian et al., 2011), although functions have not been attributed to them.

7.1.3 Aims

The aims of this chapter are to:

- Identify timing and patterns of expression of miR-17-92 and miR-106-363 clusters in *Xenopus* development.
- Characterise the expression of miR-363 and miR-363* in *Xenopus* development.
- Investigate the function of miR-363 in *Xenopus* development.
- Determine expression of miR-17-92 and miR-106a-363 miRNAs in human cell culture systems, and any co-expression with lin28.

7.2 Results

7.2.1 Expression of pri-miRNA during *Xenopus* development

RT-PCR was carried out to investigate when in development the miR-17-92 and miR-106-363 clusters were being transcribed. Primers were designed close to the ends of the pri-miRNA, to ensure that the product amplified would be the full length pri-miRNA, rather than detecting a form of processed miRNA such as precursors. Embryos were collected from early development, post-fertilisation, to late gastrula stages. It was seen that there was no maternal RNA for these clusters. Expression of both pri-miR-17-92 and pri-miR-106-363 began shortly after MBT, with RNA detected from stage 8 for both pri-miRNAs. This expression continued through gastrulation, and was still present at stage 12, the latest stage studied here (Figure 7.1).

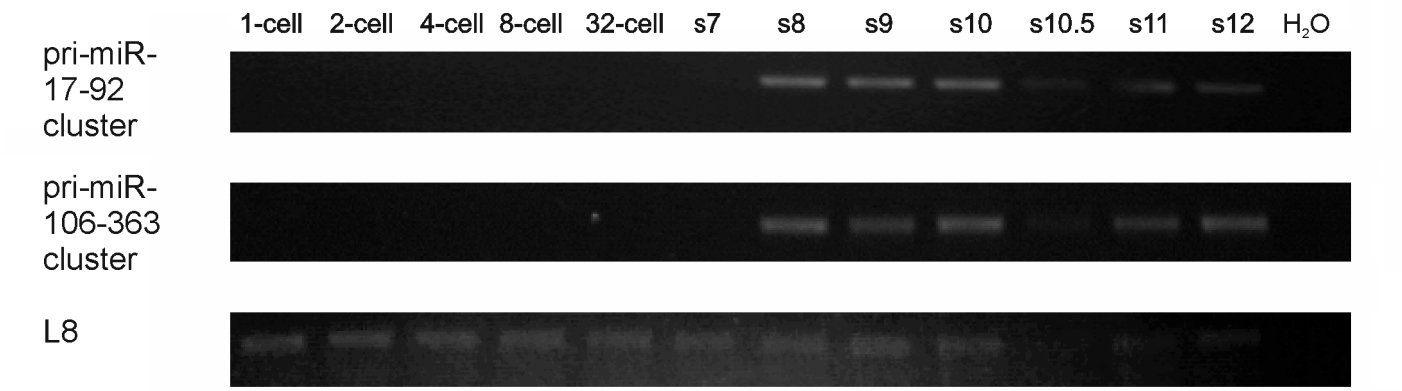


Figure 7.1 Analysis of cluster pri-miRNA expression during *X. tropicalis* development

RNA was collected from *X. tropicalis* embryos throughout early development. RT-PCR was carried out for pri-miR-17-92 and pri-miR-106-363. L8 was used as a loading control. Image is representative of n=2. s= stage. miR-17-92 = 669 bp, miR-106-363 = 639 bp, L8 = 435 bp.

Due to the interest in these miRNAs as potential lin28 targets during gastrulation, and detection of expression at this stage by RT-PCR, their expression patterns were investigated using whole-mount *in situ* hybridisation. RNA probes were used for the full length of the pri-miR-17-92 and pri-miR-106-363 clusters, covering the same RNA sequence as the RT-PCR primers. In early gastrula stage embryos, RNA for the two clusters was detected on the dorsal side of the embryo, with some RNA around the developing blastopore lip (Figure 7.2). Using this method, only a faint detection of expression was seen, however this was reproducible over multiple and non-sibling embryos. Both of the clusters showed an overlapping area of expression to one another.

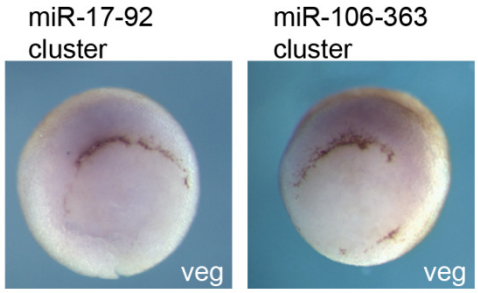


Figure 7.2 Analysis of cluster pri-miRNA expression in gastrula stage embryos
Gastrula stage embryos showing expression of pri-miR-17-92 and pri-miR-106-363 clusters, by whole-mount *in situ* hybridisation. Representative of 2 separate experiments.

7.2.2 Timing of miR-363 and miR-363* expression during *Xenopus* development

Expression of miR-363 and miR-363* were investigated using qRT-PCR. As expression of the cluster was found post-MBT, sibling embryos were collected after the onset of zygotic transcription for a time course; embryos were sampled through early gastrulation, and then once an hour after stage 11.5, prior to the onset of neurulation, until tailbud stages were beginning at stage 22. Both mir-363 and miR-363* miRNAs increased in expression at gastrulation, around stage 10.5. This was a large increase in miR-363* expression, which steadied at the onset of neurulation, before showing a small increase through neurulation and peaking in tailbud embryos. In contrast, miR-363 showed a small but sustained increase in expression throughout neurulation, also peaking in tailbud embryos at stage 22 (Figure 7.3).

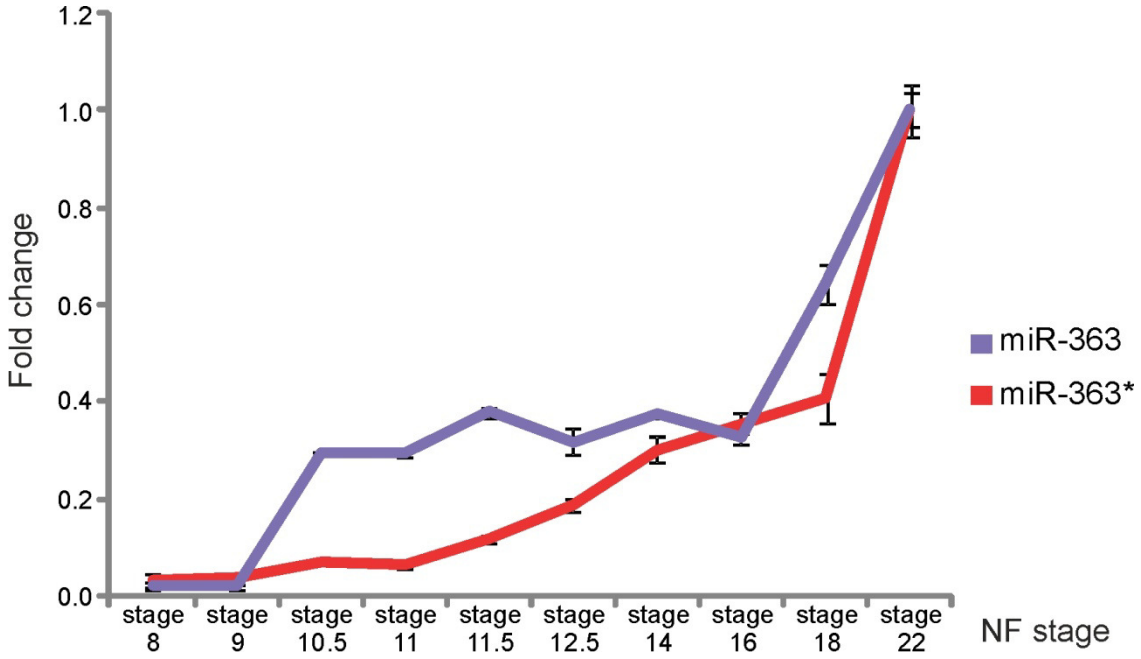


Figure 7.3 Analysis of mature miR-363 and miR-363* expression during *Xenopus* development by qRT-PCR

qRT-PCR was performed on sibling embryos from stage 8 through to stage 22. Expression was normalised to housekeeping gene U6 using the $2^{-\Delta\Delta Ct}$ method, and calibrated to the highest expression for either gene (both stage 22). Error bars show SE of technical replicates. Expression trends are representative of n=2.

Due to the high degree of primer efficiency, the comparative levels of miRNAs at these stages can be deduced, using the ΔCt values. At this point in analysis, data have been normalised to U6, the housekeeping gene, and should represent the relative level of RNA in the cells. Carrying this analysis further in directly comparing gene samples would allow accurate determination of the extent to which expression differs between genes; however, to do this the primer efficiencies must be equal. As we cannot guarantee a completely equal efficiency between the different miRNA primers, the data from different miRNAs should not be directly calibrated to one another. These data can still be used, though, to broadly infer which miRNAs are likely to be expressed at a higher abundance.

In carrying this analysis out for the data on miR-363 and miR-363* expression during development, it was seen that at all stages except during gastrulation, miR-363 was expressed at higher levels, appearing to be the more dominant miRNA; miR-363* was higher at stages 10.5 and 11 only, with both miRNAs expressed at the same levels at stage 11.5 (Figure 7.4).

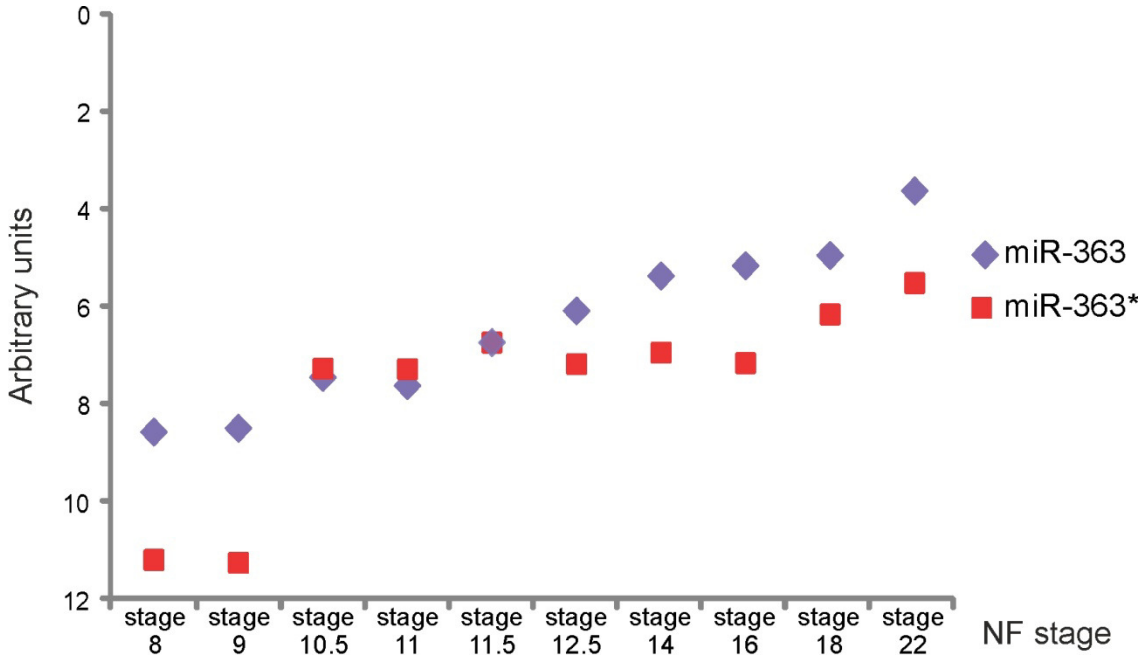


Figure 7.4 Comparative expression of miR-363 to miR-363* during development
qRT-PCR data was collected from sibling embryos. Comparative expression levels based upon ΔCt values after normalisation to U6. Low number of units indicates higher expression. Expression trends are representative of n=2.

7.2.3 Localisation of miR-363 expression

An analysis of the location of miR-363 and miR-363* in development was carried out to determine any areas of overlap with the lin28s, and to further identify possible functions of the miRNAs. Using whole-mount *in situ* hybridisation to visualise this, expression was investigated at a range of stages over development.

Embryos at late blastula stage, the stage at which data from Figure 7.1 indicated should be the earliest stage of expression, exhibited a presence of miR-363 within the animal hemisphere, which was further confirmed by a cross-section of the embryo, where miR-363 was detected around the blastocoel in cells below the surface layer (Figure 7.5A, A'). With the increase in expression, and potential link with lin28 during gastrulation, expression was studied through gastrula stages. miR-363 expression was initially found dorsal to the blastopore lip, with some retention in the animal hemisphere. As the blastopore developed, the domain of miR-363 expression expanded with this, surrounding the blastopore lip in stages 10.5 and 11. The cells expressing miR-363 were found in the deep sensorial layer below the surface of the embryo, in presumptive mesoderm cells (Figure 7.5B-D).

The qRT-PCR data also suggested that there was a second major increase in miR-363 expression late in neurulation. Expression of miR-363 at stage 19 was found to be predominantly on the dorsal side of the embryo, in the developing dorsal plate. It was also found to the anterior of the embryo, in developing neural tissue (Figure 7.5E, E'). This dorsal-neural expression was further seen at stage 21, although a small stripe of ventral expression also appeared to be present. Upon taking a cross-section of the embryo, this ventral expression can be more clearly seen, along with the dorsal expression which appeared to be localised within the notochord and surrounding somites (Figure 7.5F-F'').

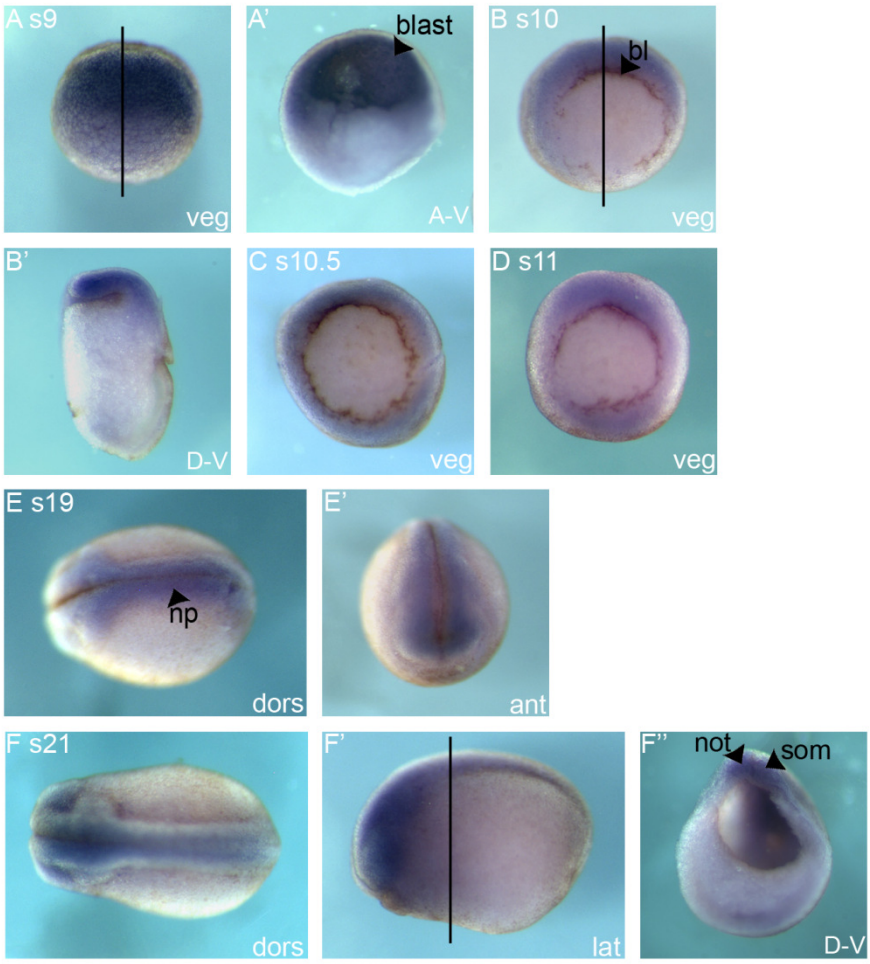


Figure 7.5 The expression of miR-363 during embryonic development

Whole-mount *in situ* hybridisation showing expression of miR-363 in uninjected embryos during development. (') refers to an alternative view of the same stage embryo. blast = blastocoels, bl = blastopore lip, np = neural plate, not = notochord, som = somites, s = stage. Bottom right refers to view: an = animal, veg = vegetal, dors = dorsal, lat = lateral, A-V = animal-vegetal axis section, D-V = dorsal-vegetal axis section.

7.2.4 Localisation of miR-363* expression

The expression patterns detected for miR-363* matched those seen for miR-363. miR-363* was detected in blastula embryos in the animal hemisphere, surrounding the

blastocoel and expressed below the surface cells (Figure 7.6A, A'). Expression in gastrula embryos began on the dorsal side, with the developing blastopore lip, and extended vegetally as the blastopore formed, and again was found in the deeper cells (Figure 7.6B-D).

Expression in the late neurula embryos was once more found on the dorsal side of the embryo, with some anterior neural patterning; internally, a stripe of ventral expression could again be seen, with dorsal expression found in and around the notochord (Figure 7.6E-F'').

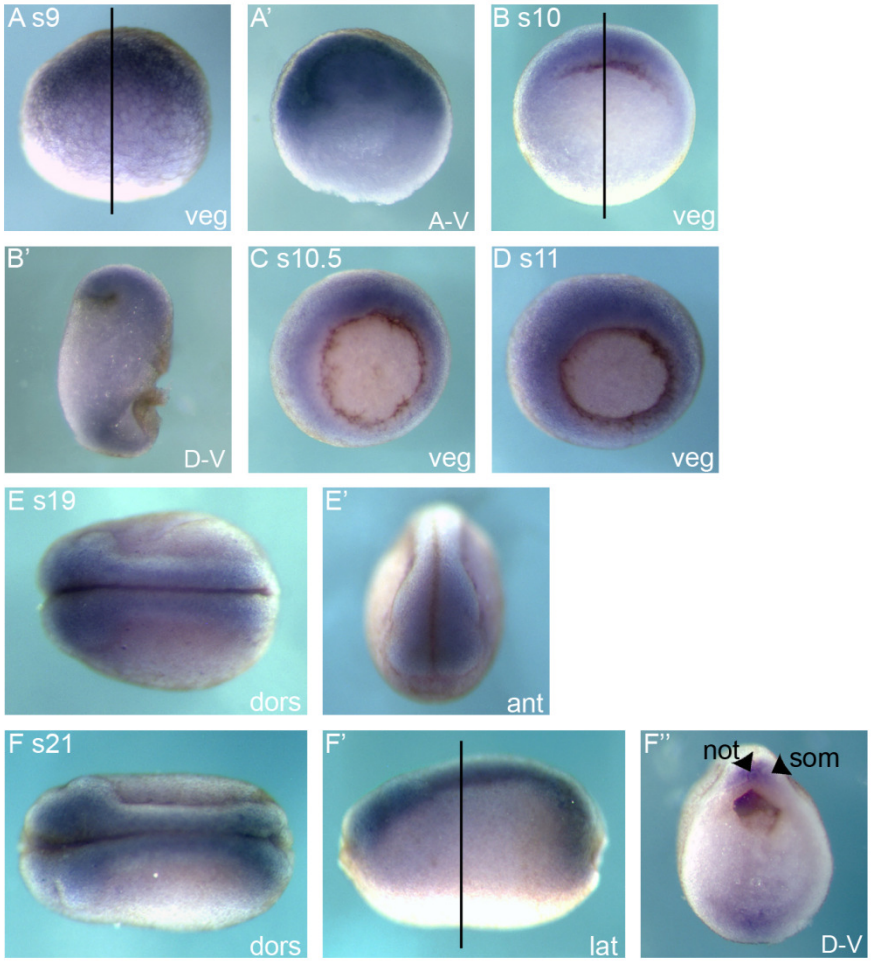


Figure 7.6 The expression of miR-363* during embryonic development
Whole-mount *in situ* hybridisation showing expression of miR-363* in wild-type embryos during development. (') refers to an alternative view of the same stage embryo. Not = notochord, som = somites, s = stage. Bottom right refers to view: an = animal, veg = vegetal, dors = dorsal, lat = lateral, A-V = animal-vegetal axis section, D-V = dorsal-vegetal axis section.

7.2.5 Function of miR-363

The function of mature miR-363 was investigated by overexpressing the mature miRNA, through the use of a synthetic precursor with a mutated miR-363* which allow the RNA to be favourably processed and only express miR-363. Injections of synthetic pre-mir-363

were carried out at 1 pmol/embryo and 0.5 pmol/embryo concentrations. At 1 pmol, there was an increased severity of the phenotype, but survival rates were much lower: a survival rate of 52% was obtained with 0.5 pmol injections, whereas only 38% of embryos survived to phenotype stage with injections of 1 pmol pre-mir-363. A control miRNA precursor was also injected; it was not predicted to target any mRNA but should be processed in the same manner as the mir-363 precursor, to account for any cellular processes altered by this. The negative control also showed a reduced survival rate, matching that of the 0.5 pmol-injected embryos at 52%, compared to survival in uninjected controls of 87%. However, the negative control precursor showed no abnormal development, and displayed a normal phenotype upon development to late tailbud stages (Figure 7.7B).

Overexpression of miR-363, however, resulted in an abnormal developmental phenotype, the severity of which increased with an increased dosage of the RNA (Figure 7.7C-G). The axis was commonly truncated, and was also bent in the more severe cases, which can indicate problems in development of the somites and notochord. A reduction in head size was seen, which can occur following neural defects, with a concurrent reduction in eye size, and two embryos subjected to 1 pmol of miRNA precursor developed no eyes. In a small number of embryos, cyclopia was observed, with reduced intra-ocular distance seen in others. An open blastopore, common following problems during gastrulation, was more frequently observed following injection with 1 pmol pre-miR-363, but did not occur in all embryos. Pigmentation was also seen to spread abdominally in some of the more affected embryos (Figure 7.7F).

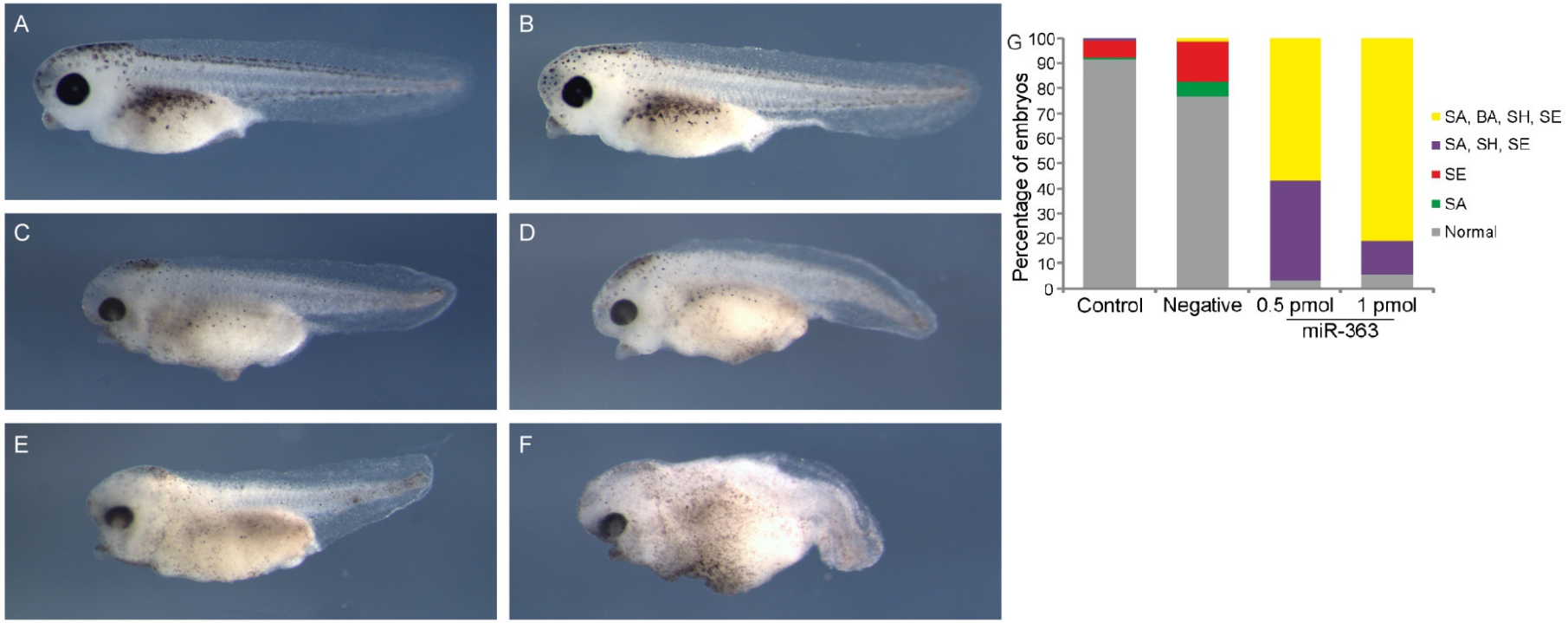


Figure 7.7 Effects of miR-363 upon development

Phenotypes of *X. tropicalis* embryos at the same developmental stage. A) Uninjected control, B) Injected with 0.5 pmol negative miRNA control, C-D) Injected with 0.5 pmol premiR-363, E-F) Injected with 1 pmol pre-miR-363. All are representative of n = 3 experiments. G) Percentage of phenotypes observed; classed as normal (A, B), or presence of short axis (SA), bent axis (BA), small eye (SE), and small head (SH). SA, SH, SE shown in B, E; SA, BA, SH, SE shown in D, F. Numbers collated from 3 separate experiments (Control n = 156, Negative n = 69, 0.5 pmol n = 93, 1 pmol n = 52).

7.2.6 Comparative levels of miR-17-92 clusters

The technique exploited in section 7.2.2 to infer comparative levels of miR-363 to miR-363* can also be used to further exploit the qRT-PCR data generated in section 4.2.6 looking at the cluster miRNAs in gastrula embryos, and provide comparative levels of the different miRNAs studied during gastrulation. miR-19b, which is contained on both the miR-17-92 and miR-106-363 clusters, appeared to be the miRNA with the highest expression. This was followed by miR-20b then miR-17, which are both members of the same family. As seen in Figure 7.4, more miR-363* was detected at this stage than miR-363, with miR-106 expressed at levels between these two. The least detected miRNA was miR-18b (Figure 7.8). All of these appeared to be expressed at higher levels than let-7g, which was included for an example of very low-level expression.

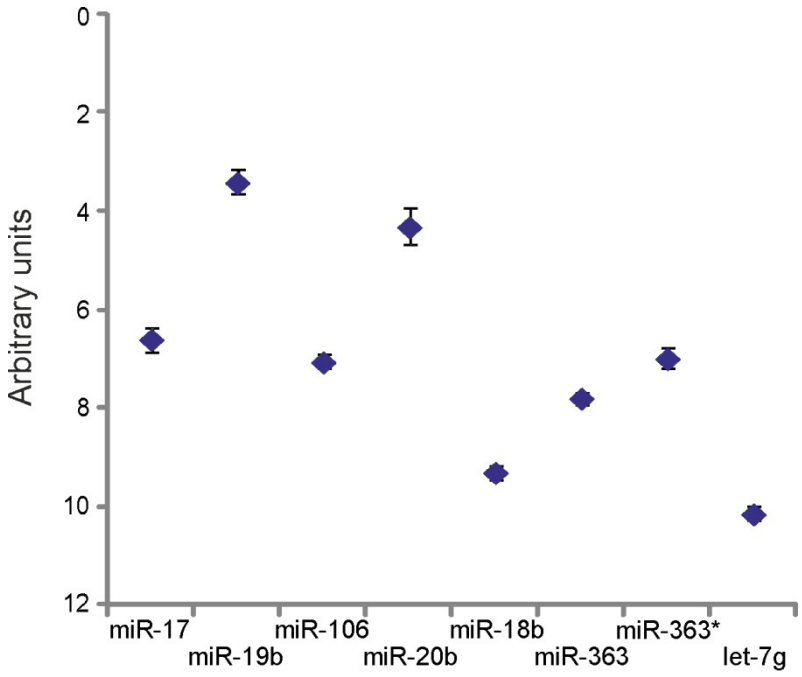


Figure 7.8 Comparative expression pattern of miRNAs at gastrulation

qRT-PCR data was collected from uninjected embryos at stage 10.5. Comparative expression levels based upon ΔCt values after normalisation to U6, averaged from 3 biological replicates. Low number of units indicates high level of expression. Error bars show SE of biological replicates.

7.2.7 miRNA expression in human cells

To identify if any human cellular systems are suitable models to look for conservation of regulation of miRNAs from the miR-17-92 clusters by lin28, any expression of these miR-17-92 and miR-106a-363 cluster miRNAs must be determined.

Despite their lack of *lin28a* and *lin28b* expression, as shown in section 5.2.5, MSCs were analysed for expression of the miR-17-92 and miR-106a-363 cluster miRNAs. Additionally, ES and EC cells were also examined, due to their pluripotent state and known expression of lin28s, and some acknowledged expression of miRNAs from these clusters, including miR-106a and miR-17 amongst others (Houbaviy et al., 2003; Suh et al., 2004; Thomson et al., 2004). There is a known role of some of these miRNAs in cancer, which also links the miRNAs to important expression within undifferentiated cell types. Two ES cell lines, H7 and H9, were characterised; these are amongst the earliest derived human ES cell lines generated by Thomson and colleagues (1998). The samples were a gift from M. Coles (University of York). Care was taken with the human ES cells to grow them in feeder-free conditions; this was to ensure that no material from MEFs was present due to the high conservation of miRNA sequences. It is possible that any miRNA expression in MEFs would be accidentally detected if the cells were to contaminate the ES cell sample. The EC cell line NTera2.D1 (NT2), a clonal cell line derived from the teratocarcinoma Tera2, was chosen over the 2102Ep cells used in 5.2.5, as the NT2 cells have a better differentiation potential (Ackerman et al., 1994; Andrews, 1988), and may therefore act as a better model for looking at pluripotency and differentiation mechanisms.

Expression of the mature miRNAs from the miR-17-92 and miR-106a-363 clusters varied greatly between the cells studied, with ES cells showing high expression of all miRNAs, which were at very low levels in MSCs. To display this appropriately, data were compared to expression in NT2 cells in two separate figures for the different cell types (Figure 7.9, Figure 7.10). The data from ES and EC cells was collected from cells at different passage numbers on separate occasions, to ensure they were appropriate as biological replicates, with replicates in MSCs from three separate donors. Levels of the miRNAs within the H9 samples were comparable, with the third replicate consistently showing slightly reduced expression compared to the other two samples of this ES cell line. There was a large degree of variability, however, in the H7 cells; one sample showed greatly increased expression of miR-18b, miR-20b and miR-363 over the other cell types, and expression was consistently higher for all other miRNAs too. In comparison to the EC cells, expression of all miRNAs except for miR-363* was generally increased in the ES cell lines (Figure 7.9). In the MSCs, expression of miR-363 was not detected, and miR-20b was expressed at greatly reduced levels compared to EC cells, with modest reductions in the expression of all other cluster miRNAs (Figure 7.10). When comparing the ES cells to MSCs, the levels of all miRNAs are hugely reduced in MSCs. In spite of differences in the level of expression, the same pattern of expression was seen within the miRNAs between ES and EC cells, with miR-19b expressed at the highest levels, and miR-363* at the lowest. miR-17 and miR-20b,

from the same miRNA family, were also relatively abundant miRNAs, commonly followed in expression levels by another member of the family, miR-106a. miR-363 was only the fifth most abundant miRNA, and only miR-363* was expressed lower than miR-18b (Figure 7.11). A similar expression pattern was also seen in the MSCs, with miR-19b the most dominant miRNA, followed by miR-17. Major differences, however, were seen in the expression of miR-20b, which was found in these cells to be a weakly expressed miRNA from the cluster, along with the lack of any detectable miR-363 (Figure 7.11).

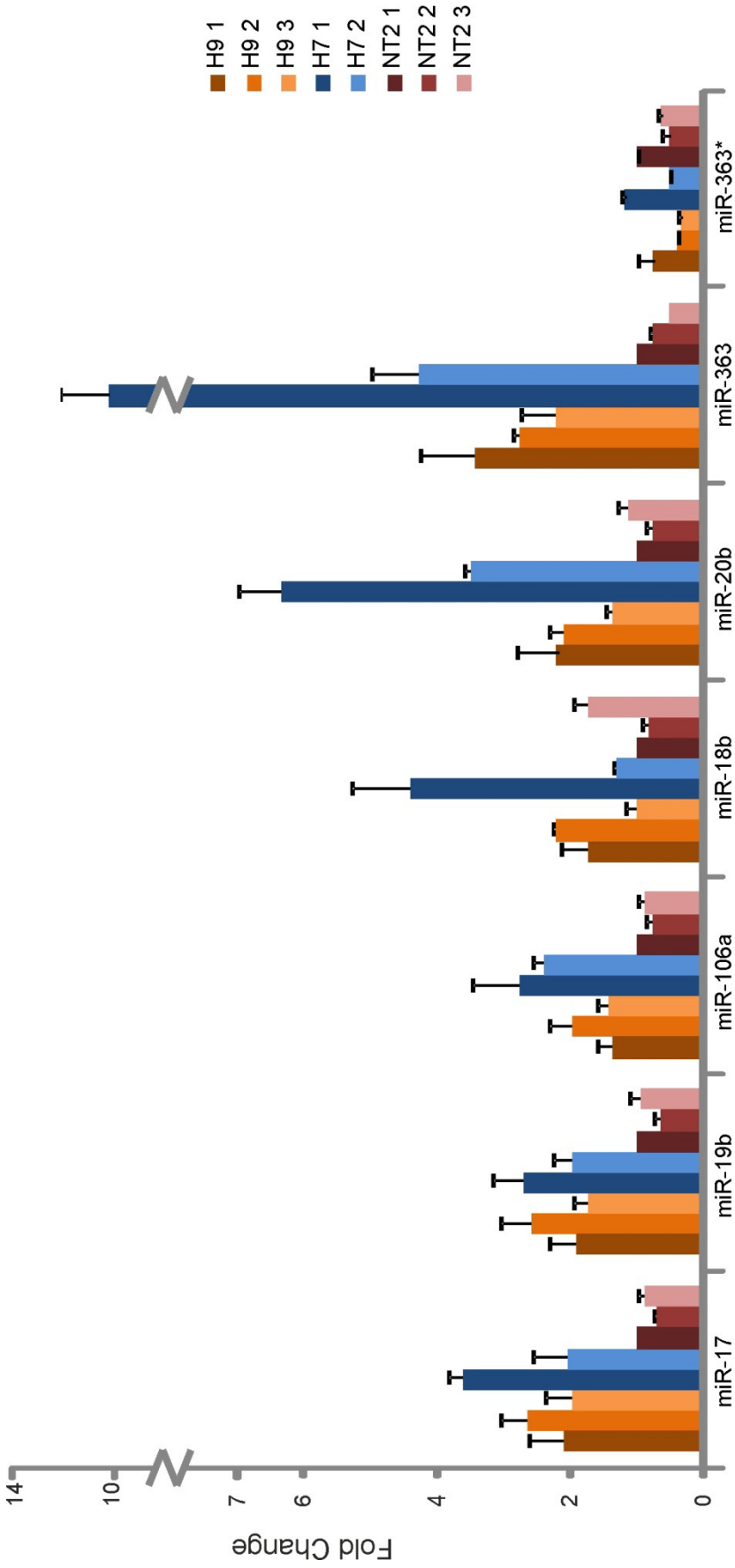


Figure 7.9 Analysis of miRNA expression in ES cells compared to EC cells
qRT-PCR was performed on RNA from ES cell lines H9 (3 biological replicates) and H7 (2 biological replicates), and the EC cell line NT2 (3 biological replicates). Expression was normalised to housekeeping gene U6 using the $2^{-\Delta\Delta Ct}$ method, and calibrated to one sample of EC cells (NT2 1). Error bars show SE of technical replicates.

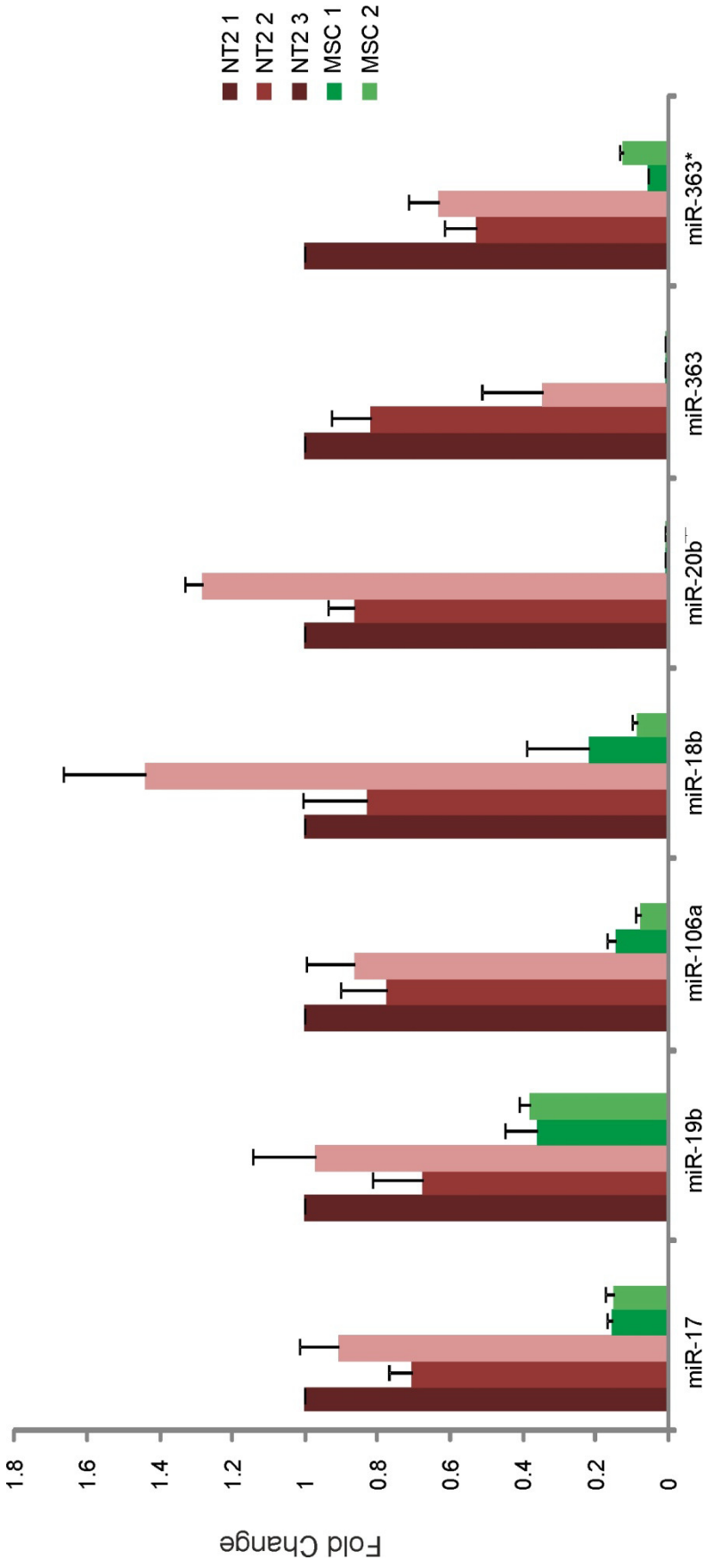


Figure 7.10 Analysis of miRNA expression in MSCs compared to EC cells
qRT-PCR was performed on RNA from MSCs from three separate donors and the EC cell line NT2 (3 biological replicates, data in Figure 7.9). Expression was normalised to housekeeping gene U6 using the $2^{-\Delta\Delta Ct}$ method, and calibrated to one sample of EC cells (NT2 1). Error bars show SE of technical replicates.

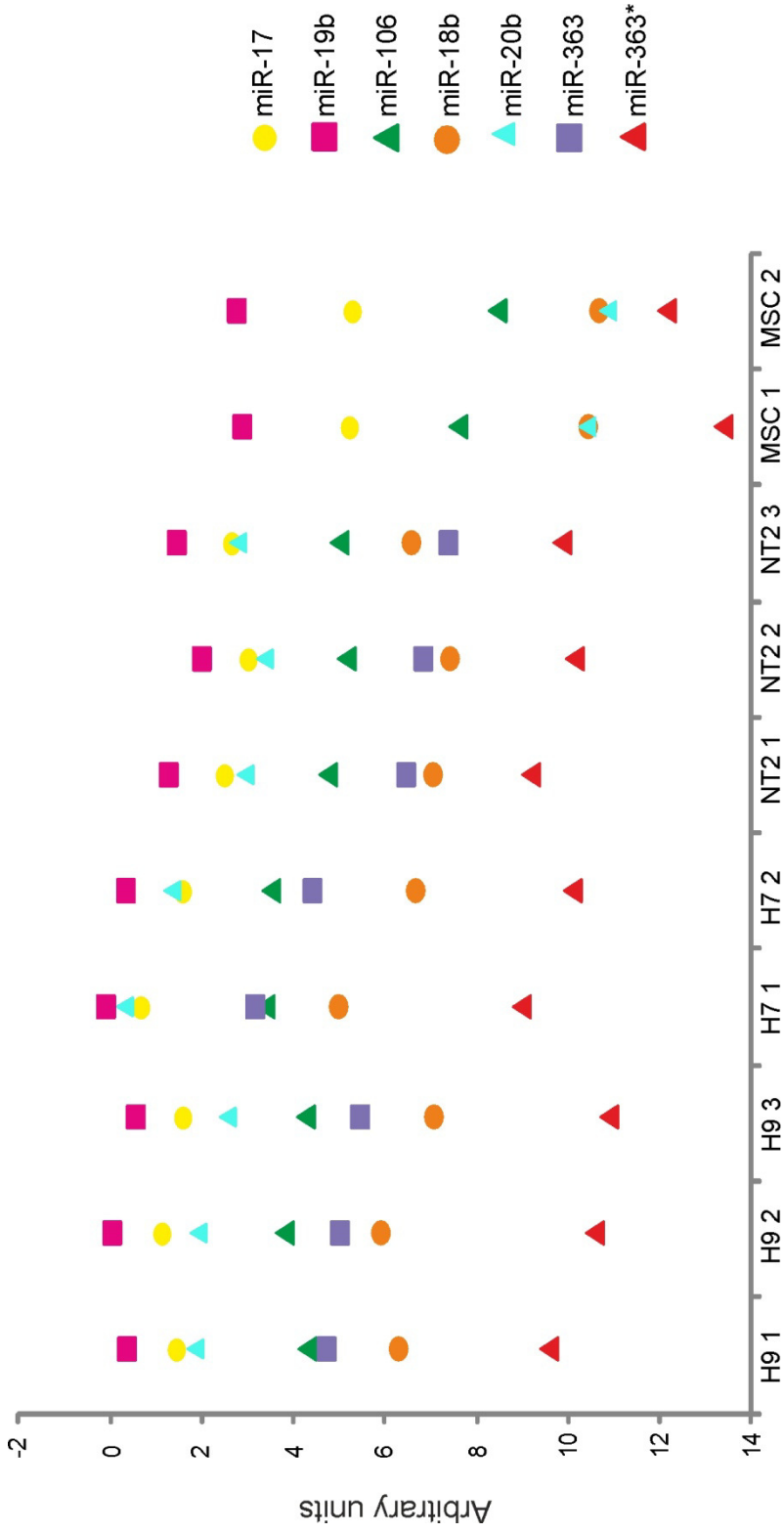


Figure 7.11 Comparative expression pattern of miRNAs in ES cells, EC cells and MSCs
qRT-PCR data was collected from ES cells, EC cells and MSCs (from Figure 7.9 and Figure 7.10). Comparative expression levels based upon ΔCt values after normalisation to U6. Low number of units indicates high level of expression.

7.2.8 Expression of lin28 in pluripotent cells

To determine whether there were similar trends in miRNA and lin28 expression, qRT-PCR was carried out to determine lin28 levels in the ES and EC cell lines. Interestingly, the levels of *lin28a* were more similar between ES and EC cell lines than the miRNA expression, with NT2 cells showing higher levels of *lin28a* compared to some samples of ES cells (Figure 7.12). The levels of *lin28b* were also consistently higher in NT2 cells compared to the ES cells. The same trend in expression was seen across the cell samples, where those with low *lin28a* expression also showed low levels of *lin28b* RNA. Interestingly, H7 1, the sample which had shown greatly elevated levels of miRNA, did not display higher levels of the *lin28* genes compared to other samples (Figure 7.12).

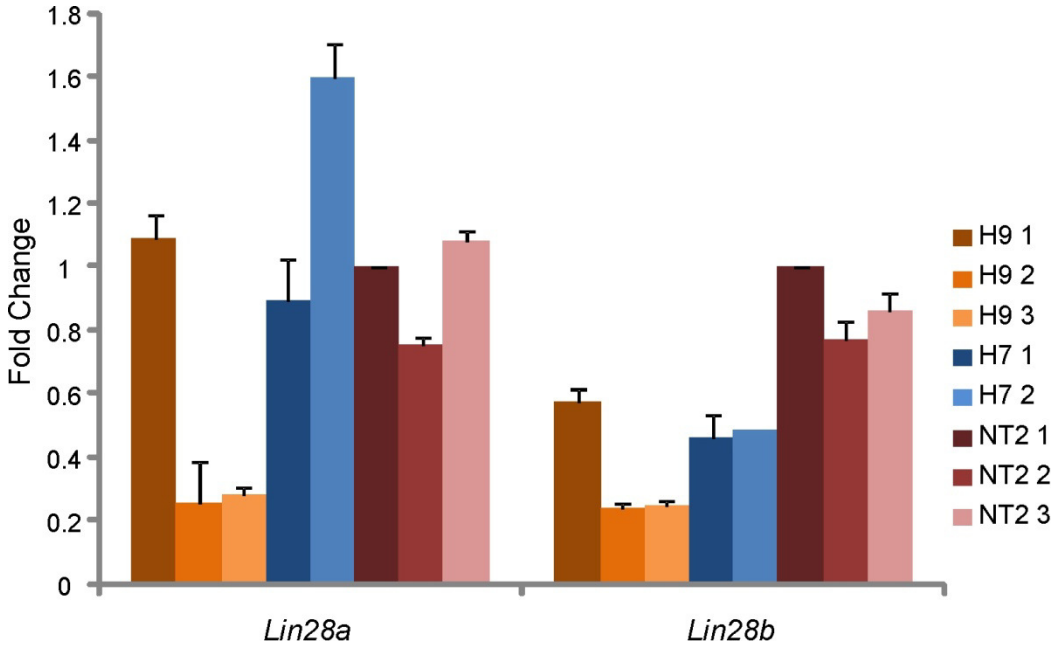


Figure 7.12 Analysis of *lin28a* and *lin28b* expression in pluripotent cells
qRT-PCR was performed on RNA from ES cell lines H9 (3 biological replicates) and H7 (2 biological replicates), and the EC cell line NT2 (3 biological replicates). Expression was normalised to housekeeping gene *ODC* using the $2^{-\Delta\Delta Ct}$ method, and calibrated to one sample of EC cells (NT2 1). Error bars show SE of technical replicates.

A summary of the data from figures 7.9 – 7.12 is included in Table 7.1, indicating which cells had the highest levels of expression of both the miRNAs and *lin28* genes, demonstrating that levels of miRNAs were slightly correlated with levels of the cluster miRNAs.

Cell type	Level of miR-17-92 clusters miRNAs	Level of lin28s
ES cells (H7 and H9 lines)	+++	++
EC cells (NTera2 line)	++	++
MSCs	+	-

Table 7.1 Summary of miRNA and *lin28* expression in different human cell types
Summary of expression based upon qRT-PCR data. - = no/very low expression, +-+++ = increasing level of expression.

7.3 Discussion

7.3.1 The co-expression of miR-17-92 clusters with lin28s in development

The data shown here from *X. tropicalis* suggests that there is endogenous *in vivo* co-expression of the lin28s with miR-17-92 and miR-106-363 cluster miRNAs. The switching on of pri-miR-17-92 and pri-miR-106-363 with MBT suggests they are early transcriptional targets, similarly to *lin28a* as was seen in section 3.2.3. All members of the clusters studied were shown to be expressed at varying levels in gastrula embryos, a stage at which work in this thesis has shown there is an important role for the lin28 proteins, with demonstrable increases in miR-363 and miR-363* at the onset of gastrulation. To strengthen further this notion of co-expression, *in situ* hybridisation data have shown a clear overlap in expression in both miR-363 and miR-363* with both *lin28a* and *lin28b*, expression of which was previously shown in sections 3.2.4 and 3.2.5, a key area of overlap is in the embryos undergoing gastrulation, with expression of the *lin28s* and miRNAs both found surrounding the blastopore, in the deep sensorial layer of cells. There are further similarities in the later stages studied, with overlapping dorsal and anterior expression, although miR-363 and miR-363* both showed additional expression in ventral areas, that were not found to express *lin28a* or *lin28b*.

In determining *in vivo* expression during mouse development, analyses of the miR-17 family (miR-17, miR-20a, miR-93 and miR-106a) found that expression in blastocysts varied for the miRNA: miR-17 and miR-20 were expressed throughout the embryo, with enrichment in the trophoectoderm, while miR-106a was elevated in the ICM and miR-92 in the primitive endoderm (Foshay and Gallicano, 2009). By E5.5, the miRNA were restricted to the differentiating extraembryonic cells, as opposed to the proliferative epiblast (Foshay and Gallicano, 2009). lin28a is also expressed in the blastocyst (Kumar et al., 2012; Xu et al., 2011), and although a characterisation of where lin28a is localised has not been carried

out in these mammalian embryos, expression can be assumed to be present in the ICM, due to expression in ES cells collected from this area. Work in later mouse embryos suggested that the miR-17-92 cluster miRNAs were at a low level of expression until E8.5, from when levels increased until E11.5, then steadily decreased through to the postnatal stages (Jevnaker et al., 2011); changes in *lin28* levels have also been seen during early differentiation events in the embryo, with an increase in expression during the first differentiation events, before a steady decline in RNA levels from around E9 (Kawahara et al., 2011; Wang et al., 2012). This precedes the miRNA expression, but it is likely that there is some persistence of the protein. There is a trend of lower levels of miR-106a-363 expression compared to the other two clusters in more adult mouse tissues (Ventura et al., 2008), and this was reflected by these authors' work using knockout mice: Ventura and colleagues were able to target deletions of the miR-17-92, miR-106-363 and miR-106b-25 clusters and observe how this affected development. They found that deletions of either the miR-106a-363 or miR-106b-25 clusters alone resulted in no apparent defects, with normal viable mice. In contrast to this, mice lacking the miR-17-92 cluster died in the minutes following birth, exhibiting problems including underdeveloped lungs and ventral septal defects, along with decreased numbers of B-cells, and these abnormalities were further compounded by the additional knockout of only the miR-106b-25 cluster (Ventura et al., 2008). Expression of both the miR-17-92 and miR-106b-25 clusters has also been found in the developing cerebellum, suggesting a neural role (Uziel et al., 2009). Studies into the chicken embryo found multiple members of these miRNA clusters were expressed at high levels generally throughout embryos at multiple stages from day 5, with reduced levels in a few tissues, and with high levels of dorsal neural expression; these members included miR-17, miR-106, miR-20b, miR-18a, miR-19a, miR-92 and miR-363 (Darnell et al., 2006), highlighting the likely importance of many of these miRNAs in development.

The expression of lin28a in ES cells, used for modelling pluripotency, has been well documented (Darr and Benvenisty, 2009; Moss and Tang, 2003; Richards et al., 2004), and has been replicated here in two ES cell lines. Despite varying reports on the different miRNAs, numerous members of the mir-17-92 clusters have been identified as expressed in ES cells, using both cloning and sequencing, and qRT-PCR. miRNAs with detected expression in both murine and human ES cells include miR-106a, miR-18, miR-19b, miR-20, miR-92, miR-93, miR-17 and miR-106b (Foshay and Gallicano, 2009; Houbaviy et al., 2003; Suh et al., 2004; Tang et al., 2006; Thomson et al., 2004; Ventura et al., 2008). Differences in techniques may be responsible for differences in observations; for example, miR-20a was found to be expressed at high levels, over miRNAs such as miR-106a and miR-92, in work by Foshay and Gallicano (2009) who used qRT-PCR, but was not detected by other

investigators using sequencing analysis, although miR-106a expression was seen (Houbaviy et al., 2003). There have not yet been reports of miR-363 expression in ES cells, but the discovery of this miRNA as a member of the cluster occurred subsequent to much of this previous work being carried out and it is possible any expression may merely have been missed; results here have shown that all miRNAs studied were expressed in ES cells and where miR-18 has been detected previously, this was found here to be less abundant than miR-363. Analysis of the EC cell line NT2 also demonstrated co-expression of *lin28a*, *lin28b* and the miRNAs under investigation. However, there did not appear to be a correlation between levels of *lin28* genes and expression of mature miRNAs.

On the other hand, a correlation of omission was observed in MSCs. A low level of expression of miR-17-92 and miR-106a-363 miRNAs was detected in MSCs, with no measurable miR-363 found in the cells. We were also unable to detect expression of lin28a and lin28b in these cells in section 5.2.5, and this data may support the hypothesis that lin28 assists the expression or processing of at least pre-mir-363.

Previous literature has concluded that the miR-17-92 cluster is the most vital in early development. It was supposed that due to the high similarity between the miR-17-92 and the miR-106a-363 clusters, potentially resulting in the same function of the miRNAs, the miR-106a-363 cluster is expressed at low levels so as to not interfere with the regulation and function of the miR-17-92 cluster (Ventura et al., 2008). On the other hand, the miR-106a-363 cluster remains highly evolutionarily conserved, and this maintenance in the genome often suggests some crucial role, which perhaps has not yet been identified. Work here has taken a step closer to showing there is a role undertaken by this cluster in development.

7.3.2 miR-363 vs miR-363* expression

The expression of miR-363 and miR-363* during development shows they are transcribed in their cluster as part of an early transcription event in the developing embryo.

The increased levels of miR-363 compared to miR-363* support the previous classification of miR-363-3p as the 'major' miRNA. The sequence alignment shown in section 6.2.1 illustrates a slightly higher level of conservation in this miRNA, which was previously believed to indicate the dominant miRNA, although conservation in both strands correlates with both arms having biological functions (Okamura et al., 2008). The change in dominant miRNA form at gastrula stages may suggest that if miR-363* is functional, it may be more important during this time than the alternative form, although the increased

expression in later stages, despite being at levels below miR-363-3p expression, may also be indicative of a functional role later on. In the human cells, miR-363 was found in ES and EC cells to be expressed at levels greater than miR-363*, again highlighting its role as the major miRNA, however it was absent in MSCs. miR-363* levels did not change greatly between the different stem cells, but this was at much lower levels of abundance than the other miRNAs present in all cases. The overlapping locations of miR-363 and miR-363* when compared to each other is unsurprising, as both are processed from the same pre-miRNA, although it is common that only one mature miRNA is expressed from a precursor.

7.3.3 Possible roles of miR-363 and miR-363*

Work here has shown that miR-363 and miR-363* are not the most highly expressed miRNAs from the cluster at gastrulation stage, but nevertheless may be playing a vital role.

The detection of their changing expression in development is indicative of a function of these miRNAs. Expression of both miRNAs increases at gastrulation, a stage of development where important inductive signals are taking place, with lots of new gene transcription, and miRNAs such as miR-363 and miR-363* may be required to regulate protein expression at this time.The increase, particularly in miR-363, during neurulation correlates with the areas of expression being throughout the developing neural system in the embryo, and developmental problems in the embryo following miR-363, such as a shortened axis. Expression of miR-363 was found to be located in the cartilaginous notochord (Darnell et al., 2006), which may also affect axis development, and in anterior neural structures that could have caused the small head and eyes phenotype. Expression has also been seen in the chick embryo of miR-363 in the developing brain (Darnell et al., 2006; Huang et al., 2010). These results are indicative of a potential role of miR-363, and perhaps also miR-363*, in neural development. However, the targets through which they may be acting have not been identified.

Possible roles for miR-363 that have previously been published involve the specification of the gonads, and the developing limb bud. Huang et al (2010) found expression of both miR-363 and miR-363* in chicken gonads, which showed different expression dependent upon the sex of the chicken, and postulated a potential role in their specification. Interestingly, expression of *lin28a* was found in developing PGCs (West et al., 2009), which precede the development of the gonads. Expression of miR-363 has also been shown at the developing limb buds (Darnell et al., 2006; Huang et al., 2010), with targets having been confirmed as *Hand2* and *Tbx3*, RNA coding for transcription factors that are critical

for limb development (Zhang et al., 2011). *lin28a* has also found in the developing limb bud (Yokoyama et al., 2008).

7.3.4 Determining the function of miR-363

The resultant phenotype from overexpression of miR-363 suggests that this miRNA is able to target mRNA during development. However, care must be taken when investigating the function of a miRNA by overexpression. It has been proposed that a fine balance exists in the cell between RISC and miRNA levels, and overexpression of a miRNA may outcompete other miRNAs that would normally be effective at that time. This could result in an alteration of cellular processes as a function of the de-repression of mRNA targets from an alternative miRNA, rather than further repression of the intended miRNA targets (Olejniczak et al., 2010). As such, alternative methods must also be carried out to determine proper function.

Target identification is key to understanding function. Available software can also be used to provide a list of putative mRNA targets for miRNA; however, outputs from multiple sources should be combined to provide as accurate a candidate gene list as possible, because many programs use different algorithms to identify targets, taking into account seed sequences and conservation, locations of binding sites and structural accessibility. For example, TargetScan (Lewis et al., 2005) results in 893 candidate targets for miR-363. miRanda (John et al., 2004) reports 6,386 possible targets, and DIANA-microT (Maragkakis et al., 2009) suggests only 381 genes. Cross-referencing these for common suggestions provides a stronger list of candidates, which must then be validated experimentally to confirm this function.

These alternatives could include a knockdown of the miRNA activity. This can be carried out using anti-sense morpholinos or commercially available anti-miRs, designed to block miRNA activity. A lack of miRNA activity would relieve the repression on mRNA targets, with any alteration of developmental phenotype then resulting as an effect of this de-repression. Analysis of the expression of putative targets at the protein level, and less reliably at the mRNA level, following a knockdown of miR-363 or blockade of the miRNA binding site would be able to confirm genes as miRNA targets; those genes with known functions would then provide insight to the function of miR-363. Additionally, if any genes regulated by miR-363 or miR-363* were also regulated by the lin28s with expression investigated following lin28 knockdown, this would strengthen the possibility of miRNA regulation by the lin28s.

Data here have identified human ES and EC cells as possible systems for further analysis of the interactions and functions between lin28 and miRNAs, with these cell types expressing both the *lin28* genes and miRNAs from the miR-17-92 clusters. Future work would involve a knockdown of the lin28 proteins, and analysis of any effect on expression of the miRNAs. These systems could also be used to analyse miR-363 function and gene targets; luciferase assays are a successful way of confirming binding sites for miRNA, and can be easily used in cell culture systems. Constructs can be engineered that have the 3' UTR containing possible binding sites, and output luciferase activity can then be measured and compared to alternative UTRs, or mutated sequences. The miRNA themselves can also be targeted as discussed above, and any effects upon proliferation and differentiation may be monitored.

Chapter 8. General discussion

8.1 lin28 is important in patterning the vertebrate embryo

FGF signalling is one of the inductive pathways that is highly involved with early germ layer specification in the vertebrate embryo. A number of targets of this pathway that mediate these effects have been identified, but the knowledge of what roles these targets all play is currently insufficient to provide the full picture. Work in Chapter 3 has shown both *lin28a* and *lin28b* to be genes that are positively regulated by FGF signalling, supporting the earlier finding for *lin28a* by Branney et al (2009). The identification of this regulatory pathway has significant potential; this could allow the expansion of both the understanding of the function of lin28 proteins, by providing new pathways known to be downstream of FGF for investigation of any involvement of these proteins, and also the function of FGF signalling, with the opportunity to study how manipulating this can affect the currently known roles of lin28s.

This regulatory pathway is supported by co-expression of *FGF*-ligands and active FGF signalling with *lin28a* and *lin28b* in the developing embryo, seen by comparing previous work documenting *FGF* expression (Lea et al., 2009) with *lin28* expression as seen in sections 3.2.4 and 3.2.5.

8.1.1 Function of lin28 in promoting differentiation

This thesis has identified a requirement for the lin28 proteins during development for correct germ layer patterning, indicating a wider role for these proteins than simply the promotion of self-renewal and pluripotency, particularly in the specification of mesoderm and neuroectoderm (Figure 8.1). This widening of the roles for lin28 is supported by previous work, which found that suppression of lin28a in mouse ES cells reduced subsequent mesoderm development, with effects on markers such as *brachyury* and *Mesp1*, which were also reduced in *Xenopus*, along with reductions in adipogenesis (Faas et al., 2012; Wang et al., 2012), although the mouse study also found the effect extended to endoderm development (Wang et al., 2012). Broader support is provided by observations that *lin28a* levels increase in developing embryoid bodies, and that lin28a and lin28b have roles in promoting development of skeletal muscle tissue, neurogenesis and lymphopoiesis (Balzer et al., 2010; Polesskaya et al., 2007; Wang et al., 2012; Yuan et al., 2012), all indicative of a requirement for lin28s to drive particular lineage differentiation.

Alternatively, lin28 may not actively promote specific lineages, but merely act as a competence factor for inductive FGF signalling. This appears to be the case in the *Xenopus* embryo, where the induction of mesoderm in animal cap experiments following treatment of inductive signals, FGF and Activin, is inhibited following knockdown of lin28 proteins (Faas et al., 2012). The way in which lin28 expression primes these cells for response to signalling factors is unconfirmed, although it was seen that there was a reduced induction of pERK.

lin28 was first identified as a heterochronic gene, and it was thought that it simply acted to promote the timing of developmental stages (Moss et al., 1997), but recent work has suggested that lin28 acts as in two stages in *C. elegans*, controlling the timing of progression between multiple larval stages, and that it first functions by targeting mRNA, and then later through the repression of let-7 (Vadla et al., 2012). Some heterochronic effects were also seen in the *Xenopus* embryo, with a loss of the lin28s resulting in altered gene expression at gastrulation; earlier peaks in expression was seen in *nodal-related 5, 6* and *1*, along with *FGF-4* and *FGF-8* (Faas et al., 2012). It is possible that the heterochronic effects of lin28 may explain some conflicting results discussed earlier, such as the striking differences seen by Wong et al. (2012) and Wang et al. (2012). When looking at marker expression, Wang and colleagues presented work from a longer and later sample of days, showing more accurately changes in expression over time (Wang et al., 2012). Experimental differences, such as the length of time for which cells were cultured following miR-125 manipulation before differentiation was induced would alter the levels of lin28 present during early specification events, which could result in varied outcomes due to the important role lin28 plays in controlling developmental timing.

8.1.2 Potential implication for FGF culture of stem cells

The confirmation of *lin28a* and *lin28b* as FGF targets could be an important discovery in aiding the ongoing understanding of human stem cell culture requirements. If this regulatory pathway can be clarified as being conserved in this human system, it may identify a mechanism by which FGF-2 supplementation aids the maintenance of ES cell pluripotency and self-renewal (Figure 8.1). As supplementing media with FGF-2 is the major difference in the culturing of mouse and human ES cells (Xu et al., 2005; Ying et al., 2003), a target could be provided in lin28 to examine for any differences in expression between these species, to help identify the differences in nature of these cells. Greater knowledge of the pathways required for stem cell culture will assist our ability to do so, leading to improvements and expansion in the clinical and therapeutic applications of stem cells.

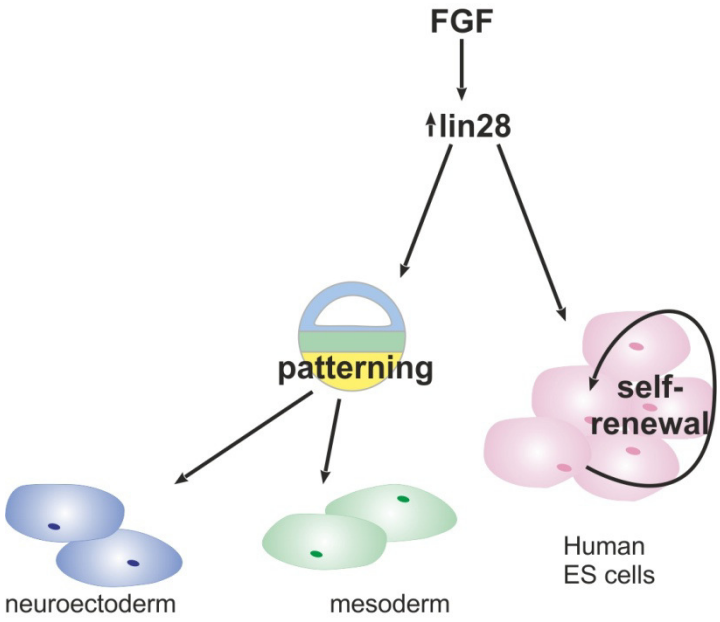


Figure 8.1 Putative pathway for functions of FGF regulation of lin28s
The regulation of lin28 expression by FGF may control its downstream effects, involving embryonic patterning, particularly of mesoderm and neuroectoderm, and its role in ES cells in maintaining self-renewal and pluripotency.

8.2 A novel set of miRNA targets for lin28

This work has highlighted that multiple members of the miR-17-92 and miR-106-363 clusters require lin28 expression for their correct endogenous mature expression. This is supported by the co-expression of *lin28a* and *lin28b* with pri-miR-17-92 and pri-miR-106-363 in both timing and location during gastrulation, and also the spatial co-expression specifically with miR-363 and miR-363* during development.

Although there is no strong evidence currently, it is possible that this is conserved in humans, with support from a correlation of expression in humans; work in Chapters 5 and 7 has shown there is a presence of both elements of this putative regulatory pathway, the lin28s and the miRNAs, in pluripotent ES and EC cells, and both are notably absent or reduced in the multipotent MSCs.

8.2.1 A positive miRNA regulatory mechanism of lin28?

The big question that has not been able to be answered here is how the lin28 proteins are exerting these effects on the levels of mature miR-17-92 and miR-106-363 cluster miRNAs. A loss of lin28 protein caused a loss of mature miRNA expression, and there are multiple levels at which this regulation may be occurring.

Work in Chapter 6 has shown that direct binding can occur between the lin28 proteins and pre-mir-363 from the clusters, indicating it is a distinct possibility that this requirement of lin28 expression for normal levels of the cluster miRNAs is a result of a direct interaction, either at the pri-miRNA or pre-miRNA stage. The prior knowledge of binding sites on let-7 RNA was applied to these RNA, and at least one conserved binding site, 'GGAG', was shown in Chapter 6 to be important for the interaction of the lin28 proteins with pre-mir-363. The proposed mechanism for let-7 regulation following binding to this particular site includes a melting of the dsRNA and blockage of Dicer cleavage site (Lightfoot et al., 2011; Mayr et al., 2012), which would not be presumed to promote any processing of the miRNA. Nevertheless, if regulation was occurring at the pri-miRNA level, this effect on the Dicer cleavage would be of little consequence. Further to this, the GGAG sequence was deemed sufficient for lin28 to drive uridylation of let-7, however mir-363 was found not to be uridylated (Heo et al., 2009), and thus another mechanism must be in force following binding to this site. Another possibility is that the binding of lin28 to the miRNA promotes stability of the miRNA until Drosha/Dicer processing, protecting it from targeting for degradation and maintaining levels of the RNA for processing to the mature form. Work has shown that, similarly to mRNA, miRNA levels can be regulated by decay, with different stabilities being demonstrated that can be regulated by external factors (Bail et al., 2010). A weaker affinity was seen for lin28a for binding of pre-mir-363 compared to let-7; if the interaction was to provide stability for the miRNA, this would perhaps need to be an interaction that could be overcome by the RNAse proteins, and thus may be of a weaker affinity than an interaction designed to block processing and prevent the binding of alternative proteins.

A key technique to understand the timing of regulation is northern blotting, as this allows monitoring of the levels of all forms of the miRNA, from pri-miRNA to mature, and shifts in relative abundances of the different forms can be detected. This was used in early papers to identify let-7 as a target of lin28 (Newman et al., 2008; Viswanathan et al., 2008). Processing assays are also an important next step in order to understand if and how lin28 is assisting biogenesis. *In vitro* assays can be performed with pri-miRNA, which can be transcribed and labelled *in vitro*, to monitor processing by Dicer and Drosha, and whether this is aided by lin28 incubation (Mayr et al., 2012; Viswanathan et al., 2008). The use of embryo extract, as used in EMSAs, can provide a more *in vivo* environment which may contain other factors that affect this miRNA biogenesis.

Positive regulation has been documented for cluster member miR-18a by interaction with hnRNP A1, a nuclear protein involved in mRNA metabolism and cytoplasmic shuttling. hnRNP A1 was found to bind to the loop region of miR-18a, which could bind at both the

pri- or pre-miRNA level, and affected processing of pri-miRNA (Guil and Caceres, 2007). Interestingly, their work found that no other members of the cluster were affected by hnRNP A1, and this interaction was dependent upon both the sequence of miR-18a and the positioning between miR-17 and miR-19a. Two binding sites were identified for hnRNP A1 within pre-mir-18: within the terminal loop and at the bottom of the stem; the binding to the stem melted the base pairing in this region, producing a bulge-like structure, which was preferentially processed by Drosha (Michlewski et al., 2008). hnRNP A1 protein has a further role in miRNA regulation, as it inhibits processing of let-7a, similarly to the lin28s (Michlewski and Caceres, 2010). This makes it possible that both lin28 and hnRNP A1 target overlapping miRNAs, both positively and negatively. A further case of positive miRNA regulation is exhibited by KH-type splicing regulatory protein (KSRP). KSRP targets let-7a, binding to the terminal loop (Michlewski and Caceres, 2010; Trabucchi et al., 2009). A knockdown of KSRP also reduced mature levels of miR-20 and miR-106a, but no physical interaction has been demonstrated (Trabucchi et al., 2009). The exact mechanism is currently unknown, but KSRP was seen to associate in complexes with Drosha and Dicer, and be exported from the nucleus with pre-miRNA (Trabucchi et al., 2009).

It remains a possibility that there are further indirect effects of lin28 in regulating these miRNAs. The interaction between mir-363 and the lin28 proteins has not yet, however, been proven *in vivo*, so it could also be possible that this regulation is indirect. This loss of mature expression was not seen to be a global effect on all miRNAs following lin28 knockdown and was specific to members of the clusters, so it is unlikely that the loss of lin28s is affecting a fundamental part of the miRNA biogenesis pathway. However, in altering levels of multiple members present on different clusters, it is a possibility that the effect is occurring at the transcriptional level; lin28 is capable of binding miRNA co-transcriptionally (Van Wynsberghe et al., 2011), but this could also perhaps be due to another target of lin28. The use of qRT-PCR and northern blots to investigate levels of the different forms of miRNA are therefore of great importance to developing a fuller understanding of the nature of this interaction.

8.2.2 Roles of lin28 regulation of miR-17-92 clusters in development

In order to understand the significance of the proteins regulating these miRNAs, the wider functions and operations of both the miRNAs and the lin28s still need to be understood.

Any early roles of the miRNAs from the miR-17-92 cluster families are largely unknown, with the only major investigation into this having been carried out by the Jacks group using knockout mice which concluded that only the miR-17-92 cluster is vital during early

embryogenesis, as loss of the other clusters resulted in normal development (Ventura et al., 2008); this work identified B-cell development as a process that required normal miR-17-92 expression, and lin28b has been shown to be involved in foetal lymphopoiesis, although it predominantly affected T-cells (Wang et al., 2012). It is possible that these miRNAs are further targets in mediating the effects of lin28 in early development, with other activities known to occur through direct mRNA interaction and let-7 repression (Figure 8.2).

The major target of lin28 proteins from these clusters as shown here are products of pre-mir-363, miR-363 and miR-363*. Relatively little is known about the function of these miRNAs, being the last to be discovered from the clusters and thus the least studied. The timing and location of expression could be indicative of a neural development role for miR-363, although this needs further identification of targets to be conclusive in this area. With lin28a having been previously linked with neurogenesis, with no primary target identified to mediate such effects (Balzer et al., 2010), these miRNAs may be candidates for this.

Identification of targets of these miRNAs, in addition to those already discovered, would help further the understanding of their role in development. Using a bioinformatics approach to suggest targets was discussed in section 7.3.4, and must be backed up with experimental validation. An expansion of this work would be to carry out specific knockdowns of the miRNAs, both individually and in full clusters, to identify redundancies and specific roles, with the use of anti-miRs or morpholinos in embryos. Expression of putative targets, with binding sites within their 3' UTR, could be assayed at the protein level following knockdown, and direct interactions tested through manipulations of the UTR; These experiments can include luciferase assays, where the 3' UTR may be compared to a mutated version, with the reporter only activated when a mutation prevents miRNA repression of the gene (Lytle et al., 2007; Tokumaru et al., 2008).

8.2.3 Possible roles for both lin28 and the miR-17-92 clusters in cancer

Both lin28 proteins and miRNAs from the miR-17-92 clusters have been implicated in the development of cancer in a range of tissues. Similarly to the results in development, cancer studies have so far attributed more importance as oncogenes to the miR-17-92 miRNAs than the other clusters, classing it as 'oncomiR-1' (He et al., 2005). The miR-17-92 cluster was first discovered to be overexpressed in B-cell lymphoma, and co-operated with myc to produce highly malignant tumours (He et al., 2005; Ota et al., 2004). This finding is consistent with a role of this cluster in the development of B-cells, with impaired numbers of differentiating B-cells in mice lacking expression of the miR-17-92 cluster (Ventura et al.,

2008). Much of the effect of this cluster in B-cell lymphomas has been attributed to miR-19a and miR-19b activity, by preventing apoptosis and promoting transformation (Mu et al., 2009; Olive et al., 2009). Further investigation into the origins of T-cell leukaemia has mapped integration points for retroviruses. These have identified both the miR-17-92 and miR-106-363 clusters as common areas for integrations, resulting in increased pri-miRNA levels (Landais et al., 2007; Lum et al., 2007; Wang et al., 2006).

There are numerous other types of cancer which have also reported increased expression of miRNAs from these clusters, and it is the case that the miRNAs believed to be responsible for the oncogenic effects of cluster overexpression differs between cancers. There have also been different findings reported for the same cancer types for some miRNAs; for instance, one study found that over 70% of T-cell tumours with an insertion at the miR-106a-363 cluster had an overexpression of both miR-106a and miR-363 (Lum et al., 2007), whereas another detected increased levels of miR-106a, but no miR-363 expression (Landais et al., 2007). A summary of cancers reported to overexpress both the whole pri-miRNA for clusters and individual miRNAs is shown in Table 8.1.

The importance of the clusters in pluripotency or differentiation specification is perhaps reflected in their implication in the development of multiple blastomas. The miR-17-92 and miR-106b-25 clusters were found to promote development of retinoblastomas (Conkrite et al., 2011), and the miR-17-92 cluster has been linked with medulloblastomas; the overexpression of the cluster in granule neural progenitors was sufficient to promote tumour formation (Uziel et al., 2009), indicating, as with the retinoblastomas, that a regulated level of miRNA expression is required for normal development. Members of the clusters are again increased in neuroblastoma and correlated with a poor prognosis in patients (Mestdagh et al., 2010; Schulte et al., 2008); the expression of miR-17-92 cluster miRNAs increased proliferation and tumourigenicity, with a knockdown abolishing tumourigenesis and increasing apoptosis (Fontana et al., 2008).

Table 8.1 Overexpression of miRNAs in cancers

miRNA	Cancer
miR-17-92 cluster	B-cell lymphoma, T-cell leukaemia, retinoblastoma, medulloblastoma, neuroblastoma
miR-106a-363 cluster	T-cell leukaemia, T-lymphoma
miR-106b-25	Retinoblastoma
miR-17	Pancreas, colon, prostate, breast, lung, neuroblastoma
miR-19	B-cell lymphoma, T-cell leukaemia
miR-106a, miR-20	Colon, pancreas, prostate, lung, T-cell leukaemia
miR-92	Pancreas, prostate, stomach, T-cell leukaemia

Table listing cancers which feature currently identified overexpression of miRNAs. Data was collated from multiple studies (Anzick et al., 1997; Bautista et al., 1998; Conkrite et al., 2011; Fontana et al., 2008; He et al., 2009; Landais et al., 2007; Lum et al., 2007; Mestdagh et al., 2010; Mu et al., 2009; Olive et al., 2009; Ota et al., 2004; Schulte et al., 2008; Uziel et al., 2009; Volinia et al., 2006; Wang et al., 2006).

A number of putative targets that mediate these oncogenic effects have been suggested. PTEN is a tumour suppressor that is involved in cell cycle regulation and inhibits the PI3K/Akt pathway; miR-19 was found to repress PTEN protein, and a gain of PTEN increased cell apoptosis (Mu et al., 2009; Olive et al., 2009). Bim is a member of the BCL-2 protein family, and is a pro-apoptotic regulator that works at the mitochondrial membrane; targeting of Bim translation by multiple members of the miR-17-92 clusters has been demonstrated, including miR-19, the miR-17 family, and miR-92 and miR-25 from the miR-363 family (Fontana et al., 2008; Petrocca et al., 2008a; Petrocca et al., 2008b; Ventura et al., 2008). Members of the miR-17 family have been shown to further regulate p21 and retinoblastoma (RB1) protein (Conkrite et al., 2011; Fontana et al., 2008; Petrocca et al., 2008a; Petrocca et al., 2008b; Volinia et al., 2006); p21 is a cyclin-dependent kinase inhibitor which prevents cell cycle progression and is a target of the tumour suppressor p53, and RB1 is another well known tumour suppressor that also inhibits cell cycle progression, and the function of this protein is commonly lost in cancer, showing in some cases heterogeneous protein expression despite ubiquitous RNA expression, indicative of post-transcriptional regulation (Ali et al., 1993; Murphree and Benedict, 1984). Further potential targets of the miRNAs, which are implicit in cancer development, were identified in the work by Volinia and colleagues (2006), although few of these have yet been confirmed. Their work did, however, validate TGF-β-2 receptor (TGFBR2), commonly mutated in gastric and colon cancer (Myeroff et al., 1995), as a target of miR-20, and possibly also other members of this miRNA family (Volinia et al., 2006).

Overexpression of lin28s in cancer was discussed in section 1.2.6, and although largely found to be implicated in cancers such as germ line, cervical and ovarian cancers, which

have not been investigation with reference to the cluster miRNAs above, lin28 overexpression has similarly been reported in principal solid tumours for breast, non-small cell lung, prostate and colon cancer (Inamura et al., 2007; King et al., 2011; Nadiminty et al., 2012; Pan et al., 2011; Sakurai et al., 2012; Takamizawa et al., 2004). There is overwhelming evidence for the involvement of multiple of these cluster miRNAs in leukaemia development, including miR-19 and miR-106a, and there are new fields of work linking the lin28s to lymphoid development. With *lin28b* expressed in foetal HSCs and promoting lymphopoiesis (Yuan et al., 2012), and miR-125, a negative regulator of lin28, driving the opposing lineage of myeloid cells, and also being expressed in myeloid leukaemia (Chaudhuri et al., 2012), it appears possible that a misregulation in lin28b could therefore be linked to malignancies. Recent work in the mouse has demonstrated that an overexpression of lin28b in HSCs resulted in T-cell malignancies, including peripheral T-cell lymphoma (Beachy et al., 2012). The role of lin28 in blastomas has also not been heavily explored, but there is research linking the lin28s to neuroblastomas. lin28b has been shown to be expressed in some neuroblastoma samples, as a target of myc (Cotterman and Knoepfler, 2009), and interestingly, cultured neuroblastoma cells show cell cycle arrest when treated with retinoic acid, with increases in miR-125 suggesting that this miRNA is misregulated in neuroblastoma, which would allow increased lin28 (Laneve et al., 2007).

With both the lin28 proteins and the miR-17-92 clusters being implicated in a number of cancers, it is possible that co-expression of these proteins and miRNAs can contribute to oncogenic properties of cancers (Figure 8.2). Determining the extent and mechanism of any regulation between these may be key to aiding the ongoing understanding of cancer development and maintenance. Additionally, important cancer targets of the miRNAs have already been identified, as discussed above, and would be good markers to investigate following lin28 knockdown in embryos or cells, to see if these are indirect downstream effects that may further play a role in cancer development.

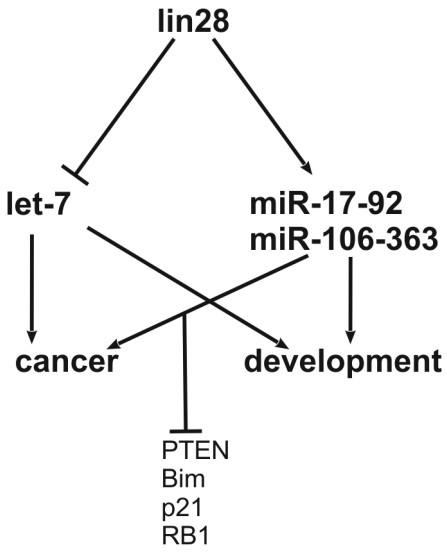


Figure 8.2 Potential downstream targets and effects of lin28
lin28 functions to repress let-7, and is required for expression of miR-17-92 and miR-106-363 cluster miRNAs. These regulations may then lead to effects in both normal development and cancer. Known cancer targets of the miRNAs include PTEN, Bim, p21 and RB1, which may in this pathway be downstream of lin28.

8.3 Future directions

8.3.1 Functions of lin28a versus lin28b

In order to avoid protein redundancy, knockdown of protein expression was carried out in combination of all three *Xenopus* lin28 proteins; however, it is possible that functions of these proteins clarified here may not be attributable to both lin28a and lin28b, or indeed the two lin28a isoforms. The difference in maternal expression between the genes is suggestive of different early developmental roles for the proteins. Knockdowns carried out with MOs to target either lin28a or lin28b should be analysed to see to what extent the functions identified here are lost; this may assist in determining lin28a and lin28b-specific roles. As lin28b is expressed maternally, the technique of maternal depletion should be investigated to fully understand the earliest functions of this protein. In this technique, the mRNA is targeted with antisense oligonucleotides pre-fertilisation and, when depleted, ensures that the protein is absent until transcription begins just prior to gastrulation (Wylie and Heasman, 1997). Any abnormalities arising as a result of maternal depletion of lin28b would provide an indication as to what role the protein is playing during the earliest stages of development.

Interestingly, although work in Chapter 3 showed earlier expression of *lin28b* compared to *lin28a*, it is *lin28a* that has been documented to be strongly expressed in ES cells. This may be partially due to the fact that when ES cells are derived, zygotic transcription has been

activated in the mammalian embryo (Braude et al., 1988; Tesařík et al., 1986), and work here has demonstrated *lin28a* expression following activation of the zygotic transcriptome. Additionally, work in this thesis showed a wider spread of tissue types were expressing *lin28b*, and *lin28a* showed restrictions to cells consistent with previous work, such as to proliferative muscle tissue and pluripotent areas (Yang and Moss, 2003).There was a degree of overlap in expression, which may suggest some redundancy in function, but also different areas of expression, which highlights differences between the proteins. Intracellularly, there are further differences, with lin28a being predominantly cytoplasmic and lin28b showing nuclear and nucleolar localisation (Piskounova et al., 2011). The majority of studies investigating direct targets have also focused more heavily on lin28a, and have not clarified these targets as also being downstream of lin28b. A known function of both lin28a and lin28b is the repression of let-7 biogenesis, although they operate to control this in different ways (Viswanathan et al., 2008).

A potential difference in binding affinity was shown in this work between lin28a and lin28b for pre-mir-363, which did not appear to exist for let-7, already known to be a target of both lin28a and lin28b. This might therefore be an indicator of a lin28a-specific, or lin28a-preferred role. Further binding and processing assays, with the additional resource of a recombinant lin28b, would contribute to the understanding of this novel miRNA interaction.

8.3.2 Conservation of regulatory pathway in humans

The implications discussed for these findings will hold greater merit if this regulation of miRNAs by lin28 is shown to be conserved to humans.

As this regulatory pathway was identified following a loss of lin28 protein, with less apparent results as a result of lin28 overexpression, it would therefore be most suitable to study this in human cell culture by the same methodology: investigating miRNA expression levels following a loss of the lin28s. This is supported by work in ES cells, where an overexpression of lin28a did not result in expression changes in miRNAs from the miR-17-92 clusters, both in ES cells and multipotent cells (Balzer et al., 2010; Yuan et al., 2012), as we saw following overexpression in *Xenopus* in Chapter 4.

There has been, though, one study which carried out a knockdown of lin28a in mouse ES cells and a subsequent global miRNA analysis. This work only identified major changes in let-7 miRNAs, with no recognition of miRNAs decreased in expression as hypothesised for the miR-17-92 clusters (Viswanathan et al., 2008). This is not suggestive that this

regulation is conserved in these cells, however it is possible that the levels of the cluster miRNAs were low in this study, or that changes did occur but not to a large degree of fold change, which is what the authors were examining.

Work here has clarified that both ES cells and EC cells express not only the *lin28a* and *lin28b* genes, but also multiple members of the miR-17-92 and miR-106a-363 clusters. This makes them suitable models for investigating conservation of this pathway. One point to be noted is that lin28a may be required for ES cell survival and undifferentiated state, and thus a reduction in expression of this may render the cells incapacitated and not allow accurate analysis of downstream lin28 effects. The advantage of EC cells, however, is that they are not as sensitive to such requirements, and are more likely to survive the loss of any essential ES cell markers, and could provide the ideal model for investigating this potential regulatory pathway. These cell systems can be treated with siRNA or shRNA to reduce expression of the lin28s, allowing for subsequent analysis of miRNAs. These knockdowns can be carried out using transient transfections, or through the use of stable manipulations such as lentivirus.

8.4 Conclusions

In summary, work in this thesis has highlighted new regulatory relationships during development, both the regulation of *lin28a* and *lin28b* by FGF signalling, and the identification of miRNAs that may be positively regulated by the lin28s. This study is a first instance of a successful knockdown of both lin28a and lin28b in the vertebrate embryo, which has provided the opportunity for identification of downstream targets and functions, characterising the genes as important not only for pluripotency but also in driving differentiation. The potential for lin28 proteins to be involved in a previously undiscovered alternative mechanism for regulatory control over miRNAs expands the capabilities of these proteins, and their roles can begin to be investigated more broadly. The potential regulation of miR-17-92 and miR-106-363 cluster miRNAs by lin28 could have important implications for cancer studies and developmental work in the future.

Chapter 9. Appendices

Appendix 1 Putative miRNA expressed in stage 10.5 *X. tropicalis* embryos

miRNA	# arrays detected	miRNA (cntd)	# arrays detected	miRNA (cntd)	# arrays detected
Xrt-miR-103	6/6	**Xrt-miR-140**	5*	**Xrt-miR-365**	2
Xrt-miR-106	6	**Xrt-miR-181a**	5*	**Xrt-miR-27b**	2
Xrt-miR-107	6	**Xrt-miR-19a**	5^	**Xrt-miR-146b**	2
Xrt-miR-130a	6	**Xrt-miR-148a**	5^	**Xtr-miR-128**	2
Xrt-miR-130b	6	**Xrt-miR-130c**	5^	**Xtr-miR-148b**	2
Xrt-miR-15b	6	**Xrt-miR-16a**	4^	**Xtr-miR-200b**	2
Xrt-miR-15c	6	**Xrt-miR-222**	4^	**Xtr-miR-182***	1
Xrt-miR-16b	6	**Xrt-miR-30e**	4^	**Xtr-miR-489**	1
Xrt-miR-16c	6	**Xrt-miR-367**	4*	**Xtr-miR-10b**	1
Xrt-miR-17-5p	6	**Xrt-miR-155**	4	**Xtr-miR-10c**	1
Xrt-miR-181b	6	**Xrt-miR-429**	4	**Xtr-miR-135**	1
Xrt-miR-18a	6	**Xrt-miR-17-3p**	4	**Xtr-miR-153**	1
Xrt-miR-18a*	6	**Xrt-miR-126**	4	**Xtr-miR-184**	1
Xrt-miR-18b	6	**Xrt-miR-218**	3	**Xtr-miR-189**	1
Xrt-miR-19b	6	**Xrt-miR-124**	3	**Xtr-miR-22***	1
Xrt-miR-202	6	**Xrt-miR-23a**	3	**Xtr-miR-31b**	1
Xrt-miR-20a	6	**Xrt-miR-15a**	3	**Xtr-miR-96**	1
Xrt-miR-20b	6	**Xrt-miR-9b***	3	**Xtr-miR-183**	1
Xrt-miR-214	6	**Xrt-miR-142-5p**	3	**Xtr-miR-208**	1
Xrt-miR-22	6	**Xrt-miR-30a-5p**	3	**Xtr-miR-20a***	1
Xrt-miR-223	6	**Xrt-miR-204**	3	**Xtr-miR-9a***	1
Xrt-miR-23b	6	**Xrt-miR-205a**	3		
Xrt-miR-24a	6	**Xrt-miR-129**	3		
Xrt-miR-26	6	**Xrt-miR-92b**	3		
Xrt-miR-30b	6	**Xrt-miR-31**	2		
Xrt-miR-30c	6	**Xrt-miR-181a-2***	2		
Xrt-miR-30d	6	**Xrt-miR-133d**	2		
Xrt-miR-363-3p	6	**Xrt-miR-191**	2		
Xrt-miR-363-5p	6	**Xrt-miR-150**	2		
Xrt-miR-427	6	**Xrt-miR-206**	2		
Xrt-miR-92a	6	**Xrt-miR-451**	2		
Xrt-miR-93a	6	**Xrt-miR-202***	2		
Xrt-miR-93b	6	**Xrt-miR-210**	2		

List of miRNA detected above background in arrays. Where less than 6, * indicates expressed in all control arrays, ^ expressed in all lin28 knockdown arrays.

Appendix 2 Changes in expression for FGF targets in *X. laevis*

Gene	Fold Change
Egr1	-12.4
Purine phosphorylase (PNP)	-7.9
Ephrin receptor A4 (EPHA4)	-4.8
Xiro3 (Irx3)	-4.4
P75-like fullback receptor (TNFRSF1B)	-4.2
Sprouty2	-3.1
DUSP5	-3.0
Chordin (CHRDL1)	-2.8
MKP1 (DUSP1)	-2.7
Gr10 interacting protein2 (GIGYF2)	-2.1
Connexin29 (GJB2)	-2.1
XIRG protein (IRG1)	+13.5
PDGF A Chain (PDGFA)	+5.6
Glucocorticoid inducible leucine zipper (TSC22D3)	+4.1

Data from Branney et al (2009), following inhibition of FGF signalling. Human homologue gene name is given in brackets where different to *Xenopus*. +/- indicates up/downregulation of gene
List of abbreviations

List of abbreviations

(dp-)ERK	(diphospho-)Extracellular regulated kinase
AIB1	Nuclear receptor coactivator 3
AP	Alkaline phosphatase
APS	Ammonium persulfate
BCIP	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
BLAST	Basic local alignment search tool
BMB	Boehringer Mannheim blocking reagent
BMP	Bone morphogenetic protein
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate
CIAP	Calf intenstinal alkaline phosphatase
DAG	Diacylglycerol
DGCR8	DiGeorge syndrome critical region 8
DIG	digoxigenin
DMEM	Dulbecco's Modified Eagle's Medium
DMF	Dimethylformamide
dnFGFR	Dominant negative fibroblast growth factor receptor
dNTPs	2'-deoxynucleoside 5''-triphosphates
dsRNA	Double stranded RNA
DTT	Dithiothreitol
EBs	Embryoid bodies
EC	Embryonal carcinoma (cell)
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ES	Embryonic stem (cell)
Exp5	Exportin 5 protein
FBS	Foetal bovine serum
FGF(R)	Fibroblast growth factor (receptor)
FRS2	FGFR substrate 2
hCG	Human chorionic gonadotrophin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
HSC	Haematopoietic stem cell
HSPGs	heparan sulphate proteoglycans
HTLS	Heat treated lamb serum
ICM	Inner cell mass

Ig	Immunoglobulin
IGF	Insulin-like growth factor
IP3	Inositol-1,4,5-triphosphate
iPS	Induced pluripotent stem (cell)
kDa	kiloDalton
KSRP	KH-type splicing regulatory protein KSRP
LB	Liquid broth
LIF	Leukaemia inhibitory factor
lincRNA	Long intergenic non-coding RNA
MAB(T)	Maleic acid buffer (with 0.01% Tween-20)
MAPK	Mitogen-activated protein kinase
MBT	Midblastula transition
MEFs	Mouse embryonic fibroblasts
Mesp1	Mesoderm posterior 1
miRNA	microRNA
MO(s)	Anti-sense morpholino(s)
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
MRS	Modified Ringer's saline
MSC	Mesenchymal stem cell
NAM	Normal amphibian medium
NBT	Nitro-blue tetrazolium chloride
NF	Nieuwkoop and Faber (stage)
nt	Nucleotide
NT2	NTera2.D1
P/S	Penicillin/streptomycin
PBS (T)	Phosphate buffered saline (with 0.01% Tween-20)
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
piRNA	Piwi interacting RNA
PI3K	Phosphoinositide-3 kinase
PKC	Protein kinase C
PLCγ	Phospholipase Cγ
Pre-miRNA	Precursor microRNA
Pri-miRNA	Primary microRNA
q(RT-)PCR	Quantitative PCR
RB1	Retinoblastoma (gene/protein)
RHA	RNA helicase A
RISC	RNA-induced silencing complex
RNAi	RNA interference
rRNA	Ribosomal RNA

RT	Reverse transcriptase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Standard error
SH2	Src homology 2
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SOS	Son of sevenless
ssRNA	Single-stranded RNA
TAE	Tris-acetate-EDTA (buffer)
TB	Tris-borate (buffer)
TBS(T)	Tris-buffered saline (with 0.01% Tween-20)
TEMED	Tetramethylethylenediamine
TGF	Transforming growth factor
TGFBR2	TGF-β-2 receptor
TUT4	Terminal uridylyltransferase 4

References

Ackerman, S. L., Knowles, B. B. and Andrews, P. W. (1994). Gene regulation during neuronal and non-neuronal differentiation of NTERA2 human teratocarcinoma-derived stem cells. *Brain Res Mol Brain Res* **25**, 157-162.

Alberio, R., Croxall, N. and Allegrucci, C. (2010). Pig epiblast stem cells depend on activin/nodal signaling for pluripotency and self-renewal. *Stem Cells Dev* **19**, 1627-1636.

Ali, A. A., Marcus, J. N., Harvey, J. P., Roll, R., Hodgson, C. P., Wildrick, D. M., Chakraborty, A. and Boman, B. M. (1993). RB1 protein in normal and malignant human colorectal tissue and colon cancer cell lines. *FASEB J* **7**, 931-937.

Amaya, E., Musci, T. J. and Kirschner, M. W. (1991). Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos. *Cell* **66**, 257-270.

Amaya, E., Stein, P. A., Musci, T. J. and Kirschner, M. W. (1993). FGF signalling in the early specification of mesoderm in Xenopus. *Development* **118**, 477-487.

Ambros, V. and Horvitz, H. R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. *Science* **226**, 409-416.

Amit, M., Carpenter, M. K., Inokuma, M. S., Chiu, C. P., Harris, C. P., Waknitz, M. A., Itskovitz-Eldor, J. and Thomson, J. A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol* **227**, 271-278.

Andrews, P. W. (1988). Human teratocarcinomas. *Biochim Biophys Acta* **948**, 17-36.

Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X. Y., Sauter, G., Kallioniemi, O. P., Trent, J. M. and Meltzer, P. S. (1997). AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**, 965-968.

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L. A., Hart, R. P. and Kiledjian, M. (2010). Differential regulation of microRNA stability. *RNA* **16**, 1032-1039.

Balzer, E., Heine, C., Jiang, Q., Lee, V. M. and Moss, E. G. (2010). LIN28 alters cell fate succession and acts independently of the let-7 microRNA during neurogliogenesis in vitro. *Development* **137**, 891-900.

Balzer, E. and Moss, E. G. (2007). Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol* **4**, 16-25.

Bautista, S., Vallès, H., Walker, R. L., Anzick, S., Zeilinger, R., Meltzer, P. and Theillet, C. (1998). In breast cancer, amplification of the steroid receptor coactivator gene AIB1 is correlated with estrogen and progesterone receptor positivity. *Clin Cancer Res* **4**, 2925-2929.

Bazzini, A. A., Lee, M. T. and Giraldez, A. J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233-237.

Beachy, S. H., Onozawa, M., Chung, Y. J., Slape, C., Bilke, S., Francis, P., Pineda, M., Walker, R. L., Meltzer, P. and Aplan, P. D. (2012). Enforced expression of Lin28b leads to impaired T-cell development, release of inflammatory cytokines, and peripheral T-cell lymphoma. *Blood* **120**, 1048-1059.

Berardi, A. C., Wang, A., Abraham, J. and Scadden, D. T. (1995). Basic fibroblast growth factor mediates its effects on committed myeloid progenitors by direct action and has no effect on hematopoietic stem cells. *Blood* **86**, 2123-2129.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366.

Bernstein, E., Kim, S. Y., Carmell, M. A., Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. A., Elledge, S. J., Anderson, K. V. and Hannon, G. J. (2003). Dicer is essential for mouse development. *Nat Genet* **35**, 215-217.

Bohnsack, M. T., Czaplinski, K. and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185-191.

Bottcher, R. T. and Niehrs, C. (2005). Fibroblast growth factor signaling during early vertebrate development. *Endocr Rev* **26**, 63-77.

Brade, T., Manner, J. and Kuhl, M. (2006). The role of Wnt signalling in cardiac development and tissue remodelling in the mature heart. *Cardiovasc Res* **72**, 198-209.

Branney, P. A., Faas, L., Steane, S. E., Pownall, M. E. and Isaacs, H. V. (2009). Characterisation of the fibroblast growth factor dependent transcriptome in early development. *PLoS One* **4**, e4951.

Braude, P., Bolton, V. and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459-461.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-195.

Buckland, R. A., Collinson, J. M., Graham, E., Davidson, D. R. and Hill, R. E. (1998). Antagonistic effects of FGF4 on BMP induction of apoptosis and chondrogenesis in the chick limb bud. *Mech Dev* **71**, 143-150.

Cai, X., Hagedorn, C. H. and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-1966.

Cao, D., Allan, R. W., Cheng, L., Peng, Y., Guo, C. C., Dahiya, N., Akhi, S. and Li, J. (2011). RNA-binding protein LIN28 is a marker for testicular germ cell tumors. *Hum Pathol* **42**, 710-718.

Carballada, R., Yasuo, H. and Lemaire, P. (2001). Phosphatidylinositol-3 kinase acts in parallel to the ERK MAP kinase in the FGF pathway during Xenopus mesoderm induction. *Development* **128**, 35-44.

Chaudhuri, A. A., So, A. Y., Mehta, A., Minisandram, A., Sinha, N., Jonsson, V. D., Rao, D. S., O'Connell, R. M. and Baltimore, D. (2012). Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proc Natl Acad Sci U S A* **109**, 4233-4238.

Choi, S. C., Kim, S. J., Choi, J. H., Park, C. Y., Shim, W. J. and Lim, D. S. (2008). Fibroblast growth factor-2 and -4 promote the proliferation of bone marrow mesenchymal stem cells by the activation of the PI3K-Akt and ERK1/2 signaling pathways. *Stem Cells Dev* **17**, 725-736.

Christen, B. and Slack, J. M. (1997). FGF-8 is associated with anteroposterior patterning and limb regeneration in Xenopus. *Dev Biol* **192**, 455-466.

---- (1999). Spatial response to fibroblast growth factor signalling in Xenopus embryos. *Development* **126**, 119-125.

Chuai, M., Zeng, W., Yang, X., Boychenko, V., Glazier, J. A. and Weijer, C. J. (2006). Cell movement during chick primitive streak formation. *Dev Biol* **296**, 137-149.

Ciruna, B. and Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev Cell* **1**, 37-49.

Ciruna, B. G., Schwartz, L., Harpal, K., Yamaguchi, T. P. and Rossant, J. (1997). Chimeric analysis of fibroblast growth factor receptor-1 (Fgfr1) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829-2841.

Concepcion, C. P., Bonetti, C. and Ventura, A. (2012). The MicroRNA-17-92 family of microRNA clusters in development and disease. *Cancer J* **18**, 262-267.

Conkrite, K., Sundby, M., Mukai, S., Thomson, J. M., Mu, D., Hammond, S. M. and MacPherson, D. (2011). miR-17~92 cooperates with RB pathway mutations to promote retinoblastoma. *Genes Dev* **25**, 1734-1745.

Cornell, R. A. and Kimelman, D. (1994). Activin-mediated mesoderm induction requires FGF. *Development* **120**, 453-462.

Cornell, R. A., Musci, T. J. and Kimelman, D. (1995). FGF is a prospective competence factor for early activin-type signals in Xenopus mesoderm induction. *Development* **121**, 2429-2437.

Cotterman, R. and Knoepfler, P. S. (2009). N-Myc regulates expression of pluripotency genes in neuroblastoma including lif, klf2, klf4, and lin28b. *PLoS One* **4**, e5799.

Coutu, D. L., Francois, M. and Galipeau, J. (2011). Inhibition of cellular senescence by developmentally regulated FGF receptors in mesenchymal stem cells. *Blood* **117**, 6801-6812.

Curtis, H. J., Sibley, C. R. and Wood, M. J. (2012). Mirtrons, an emerging class of atypical miRNA. *Wiley Interdiscip Rev RNA **3***, 617-632.

Darnell, D. K., Kaur, S., Stanislaw, S., Konieczka, J. H., Yatskievych, T. A. and Antin, P. B. (2006). MicroRNA expression during chick embryo development. *Dev Dyn* **235**, 3156-3165.

Darr, H. and Benvenisty, N. (2009). Genetic analysis of the role of the reprogramming gene LIN-28 in human embryonic stem cells. *Stem Cells* **27**, 352-362.

De Rocquigny, H., Ficheux, D., Gabus, C., Allain, B., Fournie-Zaluski, M. C., Darlix, J. L. and Roques, B. P. (1993). Two short basic sequences surrounding the zinc finger of nucleocapsid protein NCp10 of Moloney murine leukemia virus are critical for RNA annealing activity. *Nucleic Acids Res* **21**, 823-829.

Delbridge, G. J. and Khachigian, L. M. (1997). FGF-1-induced platelet-derived growth factor-A chain gene expression in endothelial cells involves transcriptional activation by early growth response factor-1. *Circ Res* **81**, 282-288.

Delrieu, I., Faye, J. C., Bayard, F. and Maret, A. (1999). Inhibition of interleukin-6 promoter activity by the 24 kDa isoform of fibroblast growth factor-2 in HeLa cells. *Biochem J* **340 (Pt 1)**, 201-206.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-235.

Dent, P., Haser, W., Haystead, T. A., Vincent, L. A., Roberts, T. M. and Sturgill, T. W. (1992). Activation of mitogen-activated protein kinase kinase by v-Raf in NIH 3T3 cells and in vitro. *Science* **257**, 1404-1407.

Desjardins, A., Yang, A., Bouvette, J., Omichinski, J. G. and Legault, P. (2011). Importance of the NCp7-like domain in the recognition of pre-let-7g by the pluripotency factor Lin28. *Nucleic Acids Res* **40**, 1767-1777.

Dickinson, R. J. and Keyse, S. M. (2006). Diverse physiological functions for dual-specificity MAP kinase phosphatases. *Journal of Cell Science* **119**, 4607-4615.

Djuranovic, S., Nahvi, A. and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**, 237-240.

Duran, C., Talley, P. J., Walsh, J., Pigott, C., Morton, I. E. and Andrews, P. W. (2001). Hybrids of pluripotent and nullipotent human embryonal carcinoma cells: partial retention of a pluripotent phenotype. *Int J Cancer* **93**, 324-332.

Dvorak, P., Dvorakova, D., Koskova, S., Vodinska, M., Najvirtova, M., Krekac, D. and Hampl, A. (2005). Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells* **23**, 1200-1211.

Eiselleova, L., Matulka, K., Kriz, V., Kunova, M., Schmidtova, Z., Neradil, J., Tichy, B., Dvorakova, D., Pospisilova, S., Hampl, A., et al. (2009). A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem Cells* **27**, 1847-1857.

Ermolenko, D. N. and Makhatadze, G. I. (2002). Bacterial cold-shock proteins. *Cell Mol Life Sci* **59**, 1902-1913.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.

Faas, L., Warrander, F., Maguire, R., Ramsbottom, S., Quinn, D., Genever, P. and Isaacs, H. (2012). lin28 proteins are required for germ layer specification in Xenopus. *Stem Cells* **(In submission)**.

---- (2013). Lin28 proteins are required for germ layer specification in Xenopus. *Development* **(Epub ahead of print)**.

Farre, J., Roura, S., Prat-Vidal, C., Soler-Botija, C., Llach, A., Molina, C. E., Hove-Madsen, L., Cairo, J. J., Godia, F., Bragos, R., et al. (2007). FGF-4 increases in vitro expansion rate of human adult bone marrow-derived mesenchymal stem cells. *Growth Factors* **25**, 71-76.

Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246-249.

Feng, C., Neumeister, V., Ma, W., Xu, J., Lu, L., Bordeaux, J., Maihle, N. J., Rimm, D. L. and Huang, Y. (2012). Lin28 regulates HER2 and promotes malignancy through multiple mechanisms. *Cell Cycle* **11**, 2486-2494.

Fontana, L., Fiori, M. E., Albini, S., Cifaldi, L., Giovinazzi, S., Forloni, M., Boldrini, R., Donfrancesco, A., Federici, V., Giacomini, P., et al. (2008). Antagomir-17-5p abolishes the growth of therapy-resistant neuroblastoma through p21 and BIM. *PLoS One* **3**, e2236.

Foshay, K. M. and Gallicano, G. I. (2009). miR-17 family miRNAs are expressed during early mammalian development and regulate stem cell differentiation. *Dev Biol* **326**, 431-443.

Funa, N. S., Saldeen, J., Akerblom, B. and Welsh, M. (2008). Interdependent fibroblast growth factor and activin A signaling promotes the expression of endodermal genes in differentiating mouse embryonic stem cells expressing Src Homology 2-domain inactive Shb. *Differentiation* **76**, 443-453.

Furthauer, M., Thisse, C. and Thisse, B. (1997). A role for FGF-8 in the dorsoventral patterning of the zebrafish gastrula. *Development* **124**, 4253-4264.

Gabbianelli, M., Sargiacomo, M., Pelosi, E., Testa, U., Isacchi, G. and Peschle, C. (1990). "Pure" human hematopoietic progenitors: permissive action of basic fibroblast growth factor. *Science* **249**, 1561-1564.

Galiveti, C. R., Rozhdestvensky, T. S., Brosius, J., Lehrach, H. and Konthur, Z. (2009). Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA* **16**, 450-461.

Glinka, A., Wu, W., Delius, H., Monaghan, A. P., Blumenstock, C. and Niehrs, C. (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**, 357-362.

Gomez, N. and Cohen, P. (1991). Dissection of the protein kinase cascade by which nerve growth factor activates MAP kinases. *Nature* **353**, 170-173.

Gotoh, Y., Masuyama, N., Suzuki, A., Ueno, N. and Nishida, E. (1995). Involvement of the MAP kinase cascade in Xenopus mesoderm induction. *EMBO J* **14**, 2491-2498.

Grainger, R. M. (2012). Xenopus tropicalis as a model organism for genetics and genomics: past, present, and future. *Methods Mol Biol* **917**, 3-15.

Greber, B., Wu, G., Bernemann, C., Joo, J. Y., Han, D. W., Ko, K., Tapia, N., Sabour, D., Sterneckert, J., Tesar, P., et al. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6**, 215-226.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-240.

Griffiths-Jones, S., Hui, J. H., Marco, A. and Ronshaugen, M. (2011). MicroRNA evolution by arm switching. *EMBO Rep* **12**, 172-177.

Griffiths-Jones, S., Saini, H. K., van Dongen, S. and Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-158.

Gritti, A., Parati, E. A., Cova, L., Frolichsthal, P., Galli, R., Wanke, E., Faravelli, L., Morassutti, D. J., Roisen, F., Nickel, D. D., et al. (1996). Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. *J Neurosci* **16**, 1091-1100.

Gross, I., Bassit, B., Benezra, M. and Licht, J. D. (2001). Mammalian sprouty proteins inhibit cell growth and differentiation by preventing ras activation. *J Biol Chem* **276**, 46460-46468.

Grothe, C., Heese, K., Meisinger, C., Wewetzer, K., Kunz, D., Cattini, P. and Otten, U. (2000). Expression of interleukin-6 and its receptor in the sciatic nerve and cultured Schwann cells: relation to 18-kD fibroblast growth factor-2. *Brain Res* **885**, 172-181.

Guil, S. and Caceres, J. F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* **14**, 591-596.

Guo, Y., Chen, Y., Ito, H., Watanabe, A., Ge, X., Kodama, T. and Aburatani, H. (2006). Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. *Gene* **384**, 51-61.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227.

Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., Young, G., Lucas, A. B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300.

Gwizdek, C., Ossareh-Nazari, B., Brownawell, A. M., Evers, S., Macara, I. G. and Dargemont, C. (2004). Minihelix-containing RNAs mediate exportin-5-dependent nuclear export of the double-stranded RNA-binding protein ILF3. *J Biol Chem* **279**, 884-891.

Hadari, Y. R., Gotoh, N., Kouhara, H., Lax, I. and Schlessinger, J. (2001). Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways. *Proc Natl Acad Sci U S A* **98**, 8578-8583.

Hagan, J. P., Piskounova, E. and Gregory, R. I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* **16**, 1021-1025.

Hamano, R., Miyata, H., Yamasaki, M., Sugimura, K., Tanaka, K., Kurokawa, Y., Nakajima, K., Takiguchi, S., Fujiwara, Y., Mori, M., et al. (2012). High expression of Lin28 is associated with tumour aggressiveness and poor prognosis of patients in oesophagus cancer. *Br J Cancer* **106**, 1415-1423.

Hammond, S. M., Bernstein, E., Beach, D. and Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293-296.

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146-1150.

Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T. and Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887-901.

Hanna, J., Cheng, A. W., Saha, K., Kim, J., Lengner, C. J., Soldner, F., Cassady, J. P., Muffat, J., Carey, B. W. and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci U S A* **107**, 9222-9227.

Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol* **36**, 685-695.

Harris, K. S., Zhang, Z., McManus, M. T., Harfe, B. D. and Sun, X. (2006). Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A* **103**, 2208-2213.

Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., Yatabe, Y., Kawahara, K., Sekido, Y. and Takahashi, T. (2005). A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res* **65**, 9628-9632.

He, C., Kraft, P., Chen, C., Buring, J. E., Pare, G., Hankinson, S. E., Chanock, S. J., Ridker, P. M., Hunter, D. J. and Chasman, D. I. (2009). Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat Genet* **41**, 724-728.

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., et al. (2005). A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828-833.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J. and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**, 276-284.

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., Yeom, K. H., Han, J. and Kim, V. N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708.

Herrmann, B. G. and Kispert, A. (1994). The T genes in embryogenesis. *Trends Genet* **10**, 280-286.

Hongo, I., Kengaku, M. and Okamoto, H. (1999). FGF signaling and the anterior neural induction in Xenopus. *Dev Biol* **216**, 561-581.

Houbaviy, H. B., Murray, M. F. and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. *Dev Cell* **5**, 351-358.

Huang, P., Gong, Y., Peng, X., Li, S., Yang, Y. and Feng, Y. (2010). Cloning, identification, and expression analysis at the stage of gonadal sex differentiation of chicken miR-363 and 363*. *Acta Biochim Biophys Sin (Shanghai)* **42**, 522-529.

Hudson, C., Clements, D., Friday, R. V., Stott, D. and Woodland, H. R. (1997). Xsox17alpha and -beta mediate endoderm formation in Xenopus. *Cell* **91**, 397-405.

Hutvagner, G. (2005). Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. *FEBS Lett* **579**, 5850-5857.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T. and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838.

Hutvagner, G. and Zamore, P. D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-2060.

Inamura, K., Togashi, Y., Nomura, K., Ninomiya, H., Hiramatsu, M., Satoh, Y., Okumura, S., Nakagawa, K. and Ishikawa, Y. (2007). let-7 microRNA expression is reduced in bronchioloalveolar carcinoma, a non-invasive carcinoma, and is not correlated with prognosis. *Lung Cancer* **58**, 392-396.

Isaacs, H. V. (1997). New perspectives on the role of the fibroblast growth factor family in amphibian development. *Cell Mol Life Sci* **53**, 350-361.

Isaacs, H. V., Pownall, M. E. and Slack, J. M. (1994). eFGF regulates Xbra expression during Xenopus gastrulation. *EMBO J* **13**, 4469-4481.

---- (1995). eFGF is expressed in the dorsal midline of Xenopus laevis. *Int J Dev Biol* **39**, 575-579.

Itkin, T., Ludin, A., Gradus, B., Gur-Cohen, S., Kalinkovich, A., Schajnovitz, A., Ovadya, Y., Kollet, O., Canaani, J., Shezen, E., et al. (2012). FGF-2 expands murine hematopoietic

stem and progenitor cells via proliferation of stromal cells, c-Kit activation and CXCL12 downregulation. *Blood* **120**, 1843-1855.

Itoh, N. and Ornitz, D. M. (2004). Evolution of the Fgf and Fgfr gene families. *Trends Genet* **20**, 563-569.

Jevnaker, A. M., Khuu, C., Kjole, E., Bryne, M. and Osmundsen, H. (2011). Expression of members of the miRNA17-92 cluster during development and in carcinogenesis. *J Cell Physiol* **226**, 2257-2266.

Jian, P., Li, Z. W., Fang, T. Y., Jian, W., Zhuan, Z., Mei, L. X., Yan, W. S. and Jian, N. (2011). Retinoic acid induces HL-60 cell differentiation via the upregulation of miR-663. *J Hematol Oncol* **4**, 20.

Jin, J., Jing, W., Lei, X. X., Feng, C., Peng, S., Boris-Lawrie, K. and Huang, Y. (2011). Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. *Nucleic Acids Res* **39**, 3724-3734.

Jin, L., Hu, X. and Feng, L. (2005). NT3 inhibits FGF2-induced neural progenitor cell proliferation via the PI3K/GSK3 pathway. *J Neurochem* **93**, 1251-1261.

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C. and Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biol* **2**, e363.

Johnson, D. E. and Williams, L. T. (1993). Structural and functional diversity in the FGF receptor multigene family. *Adv Cancer Res* **60**, 1-41.

Jones, M. R., Quinton, L. J., Blahna, M. T., Neilson, J. R., Fu, S., Ivanov, A. R., Wolf, D. A. and Mizgerd, J. P. (2009). Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat Cell Biol* **11**, 1157-1163.

Josephson, R., Ording, C. J., Liu, Y., Shin, S., Lakshmipathy, U., Toumadje, A., Love, B., Chesnut, J. D., Andrews, P. W., Rao, M. S., et al. (2007). Qualification of embryonal carcinoma 2102Ep as a reference for human embryonic stem cell research. *Stem Cells* **25**, 437-446.

Kane, R., Godson, C. and O'Brien, C. (2008). Chordin-like 1, a bone morphogenetic protein-4 antagonist, is upregulated by hypoxia in human retinal pericytes and plays a role in regulating angiogenesis. *Mol Vis* **14**, 1138-1148.

Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M. and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**, 489-501.

Kawahara, H., Okada, Y., Imai, T., Iwanami, A., Mischel, P. S. and Okano, H. (2011). Musashi1 cooperates in abnormal cell lineage protein 28 (Lin28)-mediated let-7 family microRNA biogenesis in early neural differentiation. *J Biol Chem* **286**, 16121-16130.

King, C. E., Wang, L., Winograd, R., Madison, B. B., Mongroo, P. S., Johnstone, C. N. and Rustgi, A. K. (2011). LIN28B fosters colon cancer migration, invasion and transformation through let-7-dependent and -independent mechanisms. *Oncogene* **30**, 4185-4193.

Kong, Y. W., Cannell, I. G., de Moor, C. H., Hill, K., Garside, P. G., Hamilton, T. L., Meijer, H. A., Dobbyn, H. C., Stoneley, M., Spriggs, K. A., et al. (2008). The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci U S A* **105**, 8866-8871.

Kouhara, H., Hadari, Y. R., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I. and Schlessinger, J. (1997). A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell* **89**, 693-702.

Kroll, K. L. and Amaya, E. (1996). Transgenic Xenopus embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development* **122**, 3173-3183.

Kumar, B. M., Maeng, G. H., Jeon, R. H., Lee, Y. M., Lee, W. J., Jeon, B. G., Ock, S. A. and Rho, G. J. (2012). Developmental expression of lineage specific genes in porcine embryos of different origins. *J Assist Reprod Genet* **29**, 723-733.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902.

Kühl, M. (2002). Non-canonical Wnt signaling in Xenopus: regulation of axis formation and gastrulation. *Semin Cell Dev Biol* **13**, 243-249.

Lai, W. T., Krishnappa, V. and Phinney, D. G. (2011). Fibroblast growth factor 2 (Fgf2) inhibits differentiation of mesenchymal stem cells by inducing Twist2 and Spry4, blocking extracellular regulated kinase activation, and altering Fgf receptor expression levels. *Stem Cells* **29**, 1102-1111.

Landais, S., Landry, S., Legault, P. and Rassart, E. (2007). Oncogenic potential of the miR-106-363 cluster and its implication in human T-cell leukemia. *Cancer Res* **67**, 5699-5707.

Laneve, P., Di Marcotullio, L., Gioia, U., Fiori, M., Ferretti, E., Gulino, A., Bozzoni, I. and Caffarelli., E. (2007). The interplay between microRNAs and the neurotrophin receptor tropomyosin-related kinase C controls proliferation of human neuroblastoma cells. *Proc Natl Acad Sci U S A* **104**, 7957-7962.

Lea, R., Papalopulu, N., Amaya, E. and Dorey, K. (2009). Temporal and spatial expression of FGF ligands and receptors during Xenopus development. *Dev Dyn* **238**, 1467-1479.

Learish, R. D., Bruss, M. D. and Haak-Frendscho, M. (2000). Inhibition of mitogen-activated protein kinase kinase blocks proliferation of neural progenitor cells. *Brain Res Dev Brain Res* **122**, 97-109.

Lee, I., Ajay, S. S., Yook, J. I., Kim, H. S., Hong, S. H., Kim, N. H., Dhanasekaran, S. M., Chinnaiyan, A. M. and Athey, B. D. (2009). New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Res* **19**, 1175-1183.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854.

Lee, S. H., Schloss, D. J., Jarvis, L., Krasnow, M. A. and Swain, J. L. (2001). Inhibition of angiogenesis by a mouse sprouty protein. *J Biol Chem* **276**, 4128-4133.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419.

Lee, Y., Jeon, K., Lee, J. T., Kim, S. and Kim, V. N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-4670.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-4060.

Lehrbach, N. J., Armisen, J., Lightfoot, H. L., Murfitt, K. J., Bugaut, A., Balasubramanian, S. and Miska, E. A. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. *Nat Struct Mol Biol* **16**, 1016-1020.

Lei, X. X., Xu, J., Ma, W., Qiao, C., Newman, M. A., Hammond, S. M. and Huang, Y. (2012). Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Res* **40**, 3574-3584.

Levenstein, M. E., Ludwig, T. E., Xu, R. H., Llanas, R. A., VanDenHeuvel-Kramer, K., Manning, D. and Thomson, J. A. (2006). Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem Cells* **24**, 568-574.

Lewis, B. P., Burge, C. B. and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell* **115**, 787-798.

Li, C., Scott, D. A., Hatch, E., Tian, X. and Mansour, S. L. (2007). Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development. *Development* **134**, 167-176.

Li, N., Zhong, X., Lin, X., Guo, J., Zou, L., Tanyi, J. L., Shao, Z., Liang, S., Wang, L. P., Hwang, W. T., et al. (2012). Lin-28 homologue A (LIN28A) promotes cell cycle progression via regulation of cyclin-dependent kinase 2 (CDK2), cyclin D1 (CCND1), and cell division cycle 25 homolog A (CDC25A) expression in cancer. *J Biol Chem* **287**, 17386-17397.

Lightfoot, H. L., Bugaut, A., Armisen, J., Lehrbach, N. J., Miska, E. A. and Balasubramanian, S. (2011). A LIN28-dependent structural change in pre-let-7g directly inhibits dicer processing. *Biochemistry* **50**, 7514-7521.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S. and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L. and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-1441.

Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42**, 1113-1117.

Lombardo, A., Isaacs, H. V. and Slack, J. M. (1998). Expression and functions of FGF-3 in Xenopus development. *Int J Dev Biol* **42**, 1101-1107.

Loughlin, F. E., Gebert, L. F., Towbin, H., Brunschweiger, A., Hall, J. and Allain, F. H. (2012). Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat Struct Mol Biol* **19**, 84-89.

Lum, A. M., Wang, B. B., Li, L., Channa, N., Bartha, G. and Wabl, M. (2007). Retroviral activation of the mir-106a microRNA cistron in T lymphoma. *Retrovirology* **4**, 5.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* **303**, 95-98.

Lytle, J. R., Yario, T. A. and Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**, 9667-9672.

Ma, J. B., Ye, K. and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318-322.

Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D. and Doudna, J. A. (2006). Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195-198.

Maragkakis, M., Reczko, M., Simossis, V. A., Alexiou, P., Papadopoulos, G. L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., et al. (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res* **37**, W273-276.

Mariani, F. V., Ahn, C. P. and Martin, G. R. (2008). Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature* **453**, 401-405.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-7638.

Martin, I., Muraglia, A., Campanile, G., Cancedda, R. and Quarto, R. (1997). Fibroblast growth factor-2 supports ex vivo expansion and maintenance of osteogenic precursors from human bone marrow. *Endocrinology* **138**, 4456-4462.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563-574.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-635.

Mathonnet, G., Fabian, M. R., Svitkin, Y. V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W. C., Darzynkiewicz, E., Pillai, R. S., et al. (2007). MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317**, 1764-1767.

Mayr, F., Schutz, A., Doge, N. and Heinemann, U. (2012). The Lin28 cold-shock domain remodels pre-let-7 microRNA. *Nucleic Acids Res* **40**, 7492-7506.

Melton, C., Judson, R. L. and Blelloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621-626.

Mendell, J. T. (2008). miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133**, 217-222.

Mestdagh, P., Bostrom, A. K., Impens, F., Fredlund, E., Van Peer, G., De Antonellis, P., von Stedingk, K., Ghesquiere, B., Schulte, S., Dews, M., et al. (2010). The miR-17-92 microRNA cluster regulates multiple components of the TGF-beta pathway in neuroblastoma. *Mol Cell* **40**, 762-773.

Michlewski, G. and Caceres, J. F. (2010). Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* **17**, 1011-1018.

Michlewski, G., Guil, S., Semple, C. A. and Caceres, J. F. (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* **32**, 383-393.

Minoda, Y., Saeki, K., Aki, D., Takaki, H., Sanada, T., Koga, K., Kobayashi, T., Takaesu, G. and Yoshimura, A. (2006). A novel Zinc finger protein, ZCCHC11, interacts with TIFA and modulates TLR signaling. *Biochem Biophys Res Commun* **344**, 1023-1030.

Miraoui, H., Oudina, K., Petite, H., Tanimoto, Y., Moriyama, K. and Marie, P. J. (2009). Fibroblast growth factor receptor 2 promotes osteogenic differentiation in mesenchymal cells via ERK1/2 and protein kinase C signaling. *J Biol Chem* **284**, 4897-4904.

Mitchell, T. S. and Sheets, M. D. (2001). The FGFR pathway is required for the trunk-inducing functions of Spemann's organizer. *Dev Biol* **237**, 295-305.

Mohammadi, M., Honegger, A. M., Rotin, D., Fischer, R., Bellot, F., Li, W., Dionne, C. A., Jaye, M., Rubinstein, M. and Schlessinger, J. (1991). A tyrosine-phosphorylated carboxy-terminal peptide of the fibroblast growth factor receptor (Flg) is a binding site for the SH2 domain of phospholipase C-gamma 1. *Mol Cell Biol* **11**, 5068-5078.

Montcouquiol, M., Crenshaw, E. B., 3rd and Kelley, M. W. (2006). Noncanonical Wnt signaling and neural polarity. *Annu Rev Neurosci* **29**, 363-386.

Moss, E. G., Lee, R. C. and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**, 637-646.

Moss, E. G. and Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Biol* **258**, 432-442.

Mu, P., Han, Y. C., Betel, D., Yao, E., Squatrito, M., Ogrodowski, P., de Stanchina, E., D'Andrea, A., Sander, C. and Ventura, A. (2009). Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev* **23**, 2806-2811.

Murphree, A. L. and Benedict, W. F. (1984). Retinoblastoma: clues to human oncogenesis. *Science* **223**, 1028-1033.

Myeroff, L. L., Parsons, R., Kim, S. J., Hedrick, L., Cho, K. R., Orth, K., Mathis, M., Kinzler, K. W., Lutterbaugh, J., Park, K., et al. (1995). A transforming growth factor beta receptor type II gene mutation common in colon and gastric but rare in endometrial cancers with microsatellite instability. *Cancer Res* **55**, 5545-5547.

Nadiminty, N., Tummala, R., Lou, W., Zhu, Y., Zhang, J., Chen, X., eVere White, R. W., Kung, H. J., Evans, C. P. and Gao, A. C. (2012). MicroRNA let-7c suppresses androgen receptor expression and activity via regulation of Myc expression in prostate cancer cells. *J Biol Chem* **287**, 1527-1537.

Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080-1091.

Newman, M. A., Thomson, J. M. and Hammond, S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539-1549.

Newport, J. and Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675-686.

Nichols, J. and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487-492.

Nieuwkoop, P. D. and Faber, J. (1994). *Normal Table of Xenopus laevis (Daudin)*. New York: Garland Publishing Inc.

Niswander, L. and Martin, G. R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* **114**, 755-768.

Nutt, S. L., Dingwell, K. S., Holt, C. E. and Amaya, E. (2001). Xenopus Sprouty2 inhibits FGF-mediated gastrulation movements but does not affect mesoderm induction and patterning. *Genes Dev* **15**, 1152-1166.

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. and Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839-843.

O'Rourke, J. R., Georges, S. A., Seay, H. R., Tapscott, S. J., McManus, M. T., Goldhamer, D. J., Swanson, M. S. and Harfe, B. D. (2007). Essential role for Dicer during skeletal muscle development. *Dev Biol* **311**, 359-368.

Okamura, K., Phillips, M. D., Tyler, D. M., Duan, H., Chou, Y. T. and Lai, E. C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* **15**, 354-363.

Olejniczak, M., Galka, P. and Krzyzosiak, W. J. (2010). Sequence-non-specific effects of RNA interference triggers and microRNA regulators. *Nucleic Acids Res* **38**, 1-16.

Olive, V., Bennett, M. J., Walker, J. C., Ma, C., Jiang, I., Cordon-Cardo, C., Li, Q. J., Lowe, S. W., Hannon, G. J. and He, L. (2009). miR-19 is a key oncogenic component of mir-17-92. *Genes Dev* **23**, 2839-2849.

Olsen, P. H. and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**, 671-680.

Ong, K. K., Elks, C. E., Li, S., Zhao, J. H., Luan, J., Andersen, L. B., Bingham, S. A., Brage, S., Smith, G. D., Ekelund, U., et al. (2009). Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet* **41**, 729-733.

Ornitz, D. M. and Itoh, N. (2001). Fibroblast growth factors. *Genome Biol* **2**, REVIEWS3005.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y. and Seto, M. (2004). Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res* **64**, 3087-3095.

Pan, L., Gong, Z., Zhong, Z., Dong, Z., Liu, Q., Le, Y. and Guo, J. (2011). Lin-28 reactivation is required for let-7 repression and proliferation in human small cell lung cancer cells. *Mol Cell Biochem* **355**, 257-263.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-89.

Pawson, T., Olivier, P., Rozakis-Adcock, M., McGlade, J. and Henkemeyer, M. (1993). Proteins with SH2 and SH3 domains couple receptor tyrosine kinases to intracellular signalling pathways. *Philos Trans R Soc Lond B Biol Sci* **340**, 279-285.

Peng, S., Chen, L. L., Lei, X. X., Yang, L., Lin, H., Carmichael, G. G. and Huang, Y. (2011). Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells* **29**, 496-504.

Peng, S., Maihle, N. J. and Huang, Y. (2010). Pluripotency factors Lin28 and Oct4 identify a sub-population of stem cell-like cells in ovarian cancer. *Oncogene* **29**, 2153-2159.

Pera, E. M., Ikeda, A., Eivers, E. and De Robertis, E. M. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. *Genes Dev* **17**, 3023-3028.

Peter, M. E. (2009). Let-7 and miR-200 microRNAs: Guardians against pluripotency and cancer progression. *Cell Cycle* **8**, 843-852.

Petersen, C. P., Bordeleau, M. E., Pelletier, J. and Sharp, P. A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* **21**, 533-542.

Petrocca, F., Vecchione, A. and Croce, C. M. (2008a). Emerging role of miR-106b-25/miR-17-92 clusters in the control of transforming growth factor beta signaling. *Cancer Res* **68**, 8191-8194.

Petrocca, F., Visone, R., Onelli, M. R., Shah, M. H., Nicoloso, M. S., de Martino, I., Iliopoulos, D., Pilozzi, E., Liu, C. G., Negrini, M., et al. (2008b). E2F1-regulated microRNAs impair TGFbeta-dependent cell-cycle arrest and apoptosis in gastric cancer. *Cancer Cell* **13**, 272-286.

Piccolo, S., Sasai, Y., Lu, B. and De Robertis, E. M. (1996). Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589-598.

Pillai, R. S., Bhattacharyya, S. N., Artus, C. G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**, 1573-1576.

Piskounova, E., Polytarchou, C., Thornton, J. E., Lapierre, R. J., Pothoulakis, C., Hagan, J. P., Iliopoulos, D. and Gregory, R. I. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**, 1066-1079.

Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P. and Gregory, R. I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J Biol Chem* **283**, 21310-21314.

Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E. G. and Harel-Bellan, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev* **21**, 1125-1138.

Pownall, M. E., Isaacs, H. V. and Slack, J. M. (1998). Two phases of Hox gene regulation during early Xenopus development. *Curr Biol* **8**, 673-676.

Pri-Chen, S., Pitaru, S., Lokiec, F. and Savion, N. (1998). Basic fibroblast growth factor enhances the growth and expression of the osteogenic phenotype of dexamethasone-treated human bone marrow-derived bone-like cells in culture. *Bone* **23**, 111-117.

Qiu, C., Ma, Y., Wang, J., Peng, S. and Huang, Y. (2010). Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res* **38**, 1240-1248.

Rand, T. A., Petersen, S., Du, F. and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**, 621-629.

Ray, J., Peterson, D. A., Schinstine, M. and Gage, F. H. (1993). Proliferation, differentiation, and long-term culture of primary hippocampal neurons. *Proc Natl Acad Sci U S A* **90**, 3602-3606.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R. and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906.

Reynolds, B. A. and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255**, 1707-1710.

Richards, L. J., Kilpatrick, T. J. and Bartlett, P. F. (1992). De novo generation of neuronal cells from the adult mouse brain. *Proc Natl Acad Sci U S A* **89**, 8591-8595.

Richards, M., Tan, S. P., Tan, J. H., Chan, W. K. and Bongso, A. (2004). The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells* **22**, 51-64.

Rogers, C. D., Archer, T. C., Cunningham, D. D., Grammer, T. C. and Casey, E. M. (2008). Sox3 expression is maintained by FGF signaling and restricted to the neural plate by Vent proteins in the Xenopus embryo. *Dev Biol* **313**, 307-319.

Ruby, J. G., Jan, C. H. and Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83-86.

Saito, T. and Saetrom, P. (2010). MicroRNAs--targeting and target prediction. *N Biotechnol* **27**, 243-249.

Sakurai, M., Miki, Y., Masuda, M., Hata, S., Shibahara, Y., Hirakawa, H., Suzuki, T. and Sasano, H. (2012). LIN28: A regulator of tumor-suppressing activity of let-7 microRNA in human breast cancer. *J Steroid Biochem Mol Biol* **131**, 101-106.

Santiago, F. S., Lowe, H. C., Day, F. L., Chesterman, C. N. and Khachigian, L. M. (1999). Early growth response factor-1 induction by injury is triggered by release and paracrine activation by fibroblast growth factor-2. *Am J Pathol* **154**, 937-944.

Sasaki, A., Taketomi, T., Wakioka, T., Kato, R. and Yoshimura, A. (2001). Identification of a dominant negative mutant of Sprouty that potentiates fibroblast growth factor- but not epidermal growth factor-induced ERK activation. *J Biol Chem* **276**, 36804-36808.

Schenke-Layland, K., Angelis, E., Rhodes, K. E., Heydarkhan-Hagvall, S., Mikkola, H. K. and Maclellan, W. R. (2007). Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. In *Stem Cells*, pp. 1529-1538. United States.

Schnall-Levin, M., Rissland, O. S., Johnston, W. K., Perrimon, N., Bartel, D. P. and Berger, B. (2011). Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Res* **21**, 1395-1403.

Schulte, J. H., Horn, S., Otto, T., Samans, B., Heukamp, L. C., Eilers, U. C., Krause, M., Astrahantseff, K., Klein-Hitpass, L., Buettner, R., et al. (2008). MYCN regulates oncogenic MicroRNAs in neuroblastoma. *Int J Cancer* **122**, 699-704.

Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B. and Nusslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009-1015.

Sibley, C. R., Seow, Y., Saayman, S., Dijkstra, K. K., El Andaloussi, S., Weinberg, M. S. and Wood, M. J. (2011). The biogenesis and characterization of mammalian microRNAs of mirtron origin. *Nucleic Acids Res* **40**, 438-448.

Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737.

Sivak, J. M., Petersen, L. F. and Amaya, E. (2005). FGF signal interpretation is directed by Sprouty and Spred proteins during mesoderm formation. *Dev Cell* **8**, 689-701.

Skaper, S. D., Kee, W. J., Facci, L., Macdonald, G., Doherty, P. and Walsh, F. S. (2000). The FGFR1 inhibitor PD 173074 selectively and potently antagonizes FGF-2 neurotrophic and neurotropic effects. *J Neurochem* **75**, 1520-1527.

Slack, F. and Ruvkun, G. (1997). Temporal pattern formation by heterochronic genes. *Annu Rev Genet* **31**, 611-634.

Slack, J. M., Darlington, B. G., Heath, J. K. and Godsave, S. F. (1987). Mesoderm induction in early Xenopus embryos by heparin-binding growth factors. *Nature* **326**, 197-200.

Solchaga, L. A., Penick, K., Porter, J. D., Goldberg, V. M., Caplan, A. I. and Welter, J. F. (2005). FGF-2 enhances the mitotic and chondrogenic potentials of human adult bone marrow-derived mesenchymal stem cells. *J Cell Physiol* **203**, 398-409.

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J. and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**, 1026-1032.

Stavridis, M. P., Lunn, J. S., Collins, B. J. and Storey, K. G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development* **134**, 2889-2894.

Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. and Stern, C. D. (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature* **406**, 74-78.

Streit, A. and Stern, C. D. (1999). Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mech Dev* **82**, 51-66.

Strioga, M., Viswanathan, S., Darinskas, A., Slaby, O. and Michalek, J. (2012). Same or Not the Same? Comparison of Adipose Tissue-Derived Versus Bone Marrow-Derived Mesenchymal Stem and Stromal Cells. *Stem Cells Dev*.

Suarez, Y., Fernandez-Hernando, C., Yu, J., Gerber, S. A., Harrison, K. D., Pober, J. S., Iruela-Arispe, M. L., Merkenschlager, M. and Sessa, W. C. (2008). Dicer-dependent endothelial microRNAs are necessary for postnatal angiogenesis. *Proc Natl Acad Sci U S A* **105**, 14082-14087.

Suh, M. R., Lee, Y., Kim, J. Y., Kim, S. K., Moon, S. H., Lee, J. Y., Cha, K. Y., Chung, H. M., Yoon, H. S., Moon, S. Y., et al. (2004). Human embryonic stem cells express a unique set of microRNAs. *Dev Biol* **270**, 488-498.

Sweetman, D. (2011). In situ detection of microRNAs in animals. *Methods Mol Biol* **732**, 1-8.

Sylvestre, Y., De Guire, V., Querido, E., Mukhopadhyay, U. K., Bourdeau, V., Major, F., Ferbeyre, G. and Chartrand, P. (2007). An E2F/miR-20a autoregulatory feedback loop. *J Biol Chem* **282**, 2135-2143.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676.

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., et al. (2004). Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**, 3753-3756.

Tang, F., Hajkova, P., Barton, S. C., Lao, K. and Surani, M. A. (2006). MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res* **34**, e9.

Tanzer, A. and Stadler, P. F. (2004). Molecular evolution of a microRNA cluster. *J Mol Biol* **339**, 327-335.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199.

Tesařík, J., Kopečný, V., Plachot, M. and Mandelbaum, J. (1986). Activation of nucleolar and extranucleolar RNA synthesis and changes in the ribosomal content of human embryos developing in vitro. *Journal of Reproduction and Fertility* **78**, 463-470.

Thomas, J., Lea, K., Zucker-Aprison, E. and Blumenthal, T. (1990). The spliceosomal snRNAs of Caenorhabditis elegans. *Nucleic Acids Res* **18**, 2633-2642.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147.

Thomson, J. M., Newman, M., Parker, J. S., Morin-Kensicki, E. M., Wright, T. and Hammond, S. M. (2006). Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* **20**, 2202-2207.

Thomson, J. M., Parker, J., Perou, C. M. and Hammond, S. M. (2004). A custom microarray platform for analysis of microRNA gene expression. *Nat Methods* **1**, 47-53.

Tokumaru, S., Suzuki, M., Yamada, H., Nagino, M. and Takahashi, T. (2008). let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* **29**, 2073-2077.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., Gherzi, R. and Rosenfeld, M. G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010-1014.

Tsutsumi, S., Shimazu, A., Miyazaki, K., Pan, H., Koike, C., Yoshida, E., Takagishi, K. and Kato, Y. (2001). Retention of multilineage differentiation potential of mesenchymal cells during proliferation in response to FGF. *Biochem Biophys Res Commun* **288**, 413-419.

Uziel, T., Karginov, F. V., Xie, S., Parker, J. S., Wang, Y. D., Gajjar, A., He, L., Ellison, D., Gilbertson, R. J., Hannon, G., et al. (2009). The miR-17~92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. *Proc Natl Acad Sci U S A* **106**, 2812-2817.

Vadla, B., Kemper, K., Alaimo, J., Heine, C. and Moss, E. G. (2012). lin-28 controls the succession of cell fate choices via two distinct activities. *PLoS Genet* **8**, e1002588.

Vallier, L., Alexander, M. and Pedersen, R. A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495-4509.

Van Wynsberghe, P. M., Kai, Z. S., Massirer, K. B., Burton, V. H., Yeo, G. W. and Pasquinelli, A. E. (2011). LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. *Nat Struct Mol Biol* **18**, 302-308.

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., Newman, J., Bronson, R. T., Crowley, D., Stone, J. R., et al. (2008). Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132**, 875-886.

Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W. S., Karpilow, J. and Khvorova, A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11**, 674-682.

Villegas, S. N., Canham, M. and Brickman, J. M. (2010). FGF signalling as a mediator of lineage transitions--evidence from embryonic stem cell differentiation. *J Cell Biochem* **110**, 10-20.

Viswanathan, S. R., Daley, G. Q. and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science* **320**, 97-100.

Viswanathan, S. R., Powers, J. T., Einhorn, W., Hoshida, Y., Ng, T. L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L. A., Lockhart, V. L., et al. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet* **41**, 843-848.

Volinia, S., Calin, G. A., Liu, C. G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* **103**, 2257-2261.

Wang, C. L., Wang, B. B., Bartha, G., Li, L., Channa, N., Klinger, M., Killeen, N. and Wabl, M. (2006). Activation of an oncogenic microRNA cistron by provirus integration. *Proc Natl Acad Sci U S A* **103**, 18680-18684.

Wang, J., Cao, N., Yuan, M., Cui, H., Tang, Y., Qin, L., Huang, X., Shen, N. and Yang, H. T. (2012). MicroRNA-125b/Lin28 pathway contributes to the mesendodermal fate decision of embryonic stem cells. *Stem Cells Dev* **21**, 1524-1537.

Wasylyk, B., Hagman, J. and Gutierrez-Hartmann, A. (1998). Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. *Trends Biochem Sci* **23**, 213-216.

Watanabe, T., Takeda, A., Mise, K., Okuno, T., Suzuki, T., Minami, N. and Imai, H. (2005). Stage-specific expression of microRNAs during Xenopus development. *FEBS Lett* **579**, 318-324.

West, J. A., Viswanathan, S. R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I. H., Sero, J. E., Zhu, H., Perez-Atayde, A., Frazier, A. L., et al. (2009). A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909-913.

White, J. and Dalton, S. (2005). Cell cycle control of embryonic stem cells. *Stem Cell Rev* **1**, 131-138.

Wightman, B., Ha, I. and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855-862.

Wilder, P. J., Kelly, D., Brigman, K., Peterson, C. L., Nowling, T., Gao, Q. S., McComb, R. D., Capecchi, M. R. and Rizzino, A. (1997). Inactivation of the FGF-4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. *Dev Biol* **192**, 614-629.

Willems, E. and Leyns, L. (2008). Patterning of mouse embryonic stem cell-derived pan-mesoderm by Activin A/Nodal and Bmp4 signaling requires Fibroblast Growth Factor activity. *Differentiation* **76**, 745-759.

Wong, S. S., Ritner, C., Ramachandran, S., Aurigui, J., Pitt, C., Chandra, P., Ling, V. B., Yabut, O. and Bernstein, H. S. (2012). miR-125b promotes early germ layer specification through Lin28/let-7d and preferential differentiation of mesoderm in human embryonic stem cells. *PLoS One* **7**, e36121.

Woods, K., Thomson, J. M. and Hammond, S. M. (2007). Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *J Biol Chem* **282**, 2130-2134.

Wu, H., Ye, C., Ramirez, D. and Manjunath, N. (2009). Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA. *PLoS One* **4**, e7566.

Wylie, C. C. and Heasman, J. (1997). What my mother told me: Examining the roles of maternal gene products in a vertebrate. *Trends Cell Biol* **7**, 459-462.

Xu, B. and Huang, Y. (2009). Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res* **37**, 4256-4263.

Xu, B., Zhang, K. and Huang, Y. (2009). Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* **15**, 357-361.

Xu, R.-H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T. and Thomson, J. A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat Meth* **2**, 185-190.

Xu, X., Pantakani, D. V., Luhrig, S., Tan, X., Khromov, T., Nolte, J., Dressel, R., Zechner, U. and Engel, W. (2011). Stage-specific germ-cell marker genes are expressed in all mouse

pluripotent cell types and emerge early during induced pluripotency. *PLoS One* **6**, e22413.

Yang, D. H. and Moss, E. G. (2003). Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. *Gene Expr Patterns* **3**, 719-726.

Yang, J. S., Phillips, M. D., Betel, D., Mu, P., Ventura, A., Siepel, A. C., Chen, K. C. and Lai, E. C. (2011). Widespread regulatory activity of vertebrate microRNA* species. *RNA* **17**, 312-326.

Yang, W. J., Yang, D. D., Na, S., Sandusky, G. E., Zhang, Q. and Zhao, G. (2005). Dicer is required for embryonic angiogenesis during mouse development. *J Biol Chem* **280**, 9330-9335.

Yeoh, J. S., van Os, R., Weersing, E., Ausema, A., Dontje, B., Vellenga, E. and de Haan, G. (2006). Fibroblast growth factor-1 and -2 preserve long-term repopulating ability of hematopoietic stem cells in serum-free cultures. *Stem Cells* **24**, 1564-1572.

Yi, R., Qin, Y., Macara, I. G. and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016.

Ying, Q. L., Nichols, J., Chambers, I. and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523.

Yokoyama, S., Hashimoto, M., Shimizu, H., Ueno-Kudoh, H., Uchibe, K., Kimura, I. and Asahara, H. (2008). Dynamic gene expression of Lin-28 during embryonic development in mouse and chicken. *Gene Expr Patterns* **8**, 155-160.

Yoshida-Koide, U., Matsuda, T., Saikawa, K., Nakanuma, Y., Yokota, T., Asashima, M. and Koide, H. (2004). Involvement of Ras in extraembryonic endoderm differentiation of embryonic stem cells. *Biochem Biophys Res Commun* **313**, 475-481.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920.

Yuan, H., Corbi, N., Basilico, C. and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-2645.

Yuan, J., Nguyen, C. K., Liu, X., Kanellopoulou, C. and Muljo, S. A. (2012). Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis. *Science* **335**, 1195-1200.

Zaragosi, L. E., Ailhaud, G. and Dani, C. (2006). Autocrine fibroblast growth factor 2 signaling is critical for self-renewal of human multipotent adipose-derived stem cells. *Stem Cells* **24**, 2412-2419.

Zehir, A., Hua, L. L., Maska, E. L., Morikawa, Y. and Cserjesi, P. (2010). Dicer is required for survival of differentiating neural crest cells. *Dev Biol* **340**, 459-467.

Zeng, Y. and Cullen, B. R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* **32**, 4776-4785.

---- (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**, 27595-27603.

Zeng, Y., Yi, R. and Cullen, B. R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24**, 138-148.

Zhang, H., Kolb, F. A., Brondani, V., Billy, E. and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* **21**, 5875-5885.

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57-68.

Zhang, X., Ibrahimi, O. A., Olsen, S. K., Umemori, H., Mohammadi, M. and Ornitz, D. M. (2006). Receptor specificity of the fibroblast growth factor family. The complete mammalian FGF family. *J Biol Chem* **281**, 15694-15700.

Zhang, Z., O'Rourke, J. R., McManus, M. T., Lewandoski, M., Harfe, B. D. and Sun, X. (2011). The microRNA-processing enzyme Dicer is dispensable for somite segmentation but essential for limb bud positioning. *Dev Biol* **351**, 254-265.

Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W. S., Viswanathan, S. R., Takeuchi, A., Grasemann, C., Rinn, J. L., Lopez, M. F., et al. (2010). Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat Genet* **42**, 626-630.

Zimmerman, L. B., De Jesus-Escobar, J. M. and Harland, R. M. (1996). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599-606.

Development 140, 976-986 (2013) doi:10.1242/dev.089797

Lin28 proteins are required for germ layer specification in *Xenopus*

Laura Faas, Fiona C. Warrander, Richard Maguire, Simon Ramsbottom, Diana Quinn, Paul Genever and Harry V. Isaacs*

SUMMARY

Lin28 family proteins share a unique structure, with both zinc knuckle and cold shock RNA-binding domains, and were originally identified as regulators of developmental timing in *Caenorhabditis elegans*. They have since been implicated as regulators of pluripotency in mammalian stem cells in culture. Using *Xenopus tropicalis*, we have undertaken the first analysis of the effects on the early development of a vertebrate embryo resulting from global inhibition of the Lin28 family. The *Xenopus* genome contains two Lin28-related genes, *lin28a* and *lin28b*. *lin28a* is expressed zygotically, whereas *lin28b* is expressed both zygotically and maternally. Both *lin28a* and *lin28b* are expressed in pluripotent cells of the *Xenopus* embryo and are enriched in cells that respond to mesoderm-inducing signals. The development of axial and paraxial mesoderm is severely abnormal in lin28 knockdown (morphant) embryos. In culture, the ability of pluripotent cells from the embryo to respond to the FGF and activin/nodal-like mesoderm-inducing pathways is compromised following inhibition of lin28 function. Furthermore, there are complex effects on the temporal regulation of, and the responses to, mesoderm-inducing signals in lin28 morphant embryos. We provide evidence that *Xenopus* lin28 proteins play a key role in choreographing the responses of pluripotent cells in the early embryo to the signals that regulate germ layer specification, and that this early function is probably independent of the recognised role of Lin28 proteins in negatively regulating let-7 miRNA biogenesis.

KEY WORDS: *lin28a*, *lin28b*, *Xenopus*, Mesoderm, miRNA, let-7, Pluripotency, Germ layer, FGF, Activin, Nodal

INTRODUCTION

The RNA-binding protein LIN-28 was originally identified in *Caenorhabditis elegans* as a regulator of timing during development (Ambros and Horvitz, 1984; Moss et al., 1997). Lin28 family proteins contain a unique combination of both zinc knuckle and cold shock RNA-binding domains that has been conserved during evolution, and Lin28 proteins are present in a wide range of animal groups (Moss and Tang, 2003).

There is considerable interest in the function of Lin28 proteins as regulators of pluripotency in stem cells. LIN28A is expressed at high levels in proliferative, pluripotent embryonic stem (ES) cells, but is downregulated during human and murine ES cell differentiation (Darr and Benvenisty, 2008; Viswanathan et al., 2008). Furthermore, LIN28A, in combination with NANOG, OCT4 and SOX2, can reprogram somatic cells to a pluripotent ES cell phenotype (Yu et al., 2007). Lin28 proteins are involved in post-transcriptional regulation via direct association with target mRNAs (Polesskaya et al., 2007; Xu and Huang, 2009; Qiu et al., 2010; Jin et al., 2011; Peng et al., 2011; Lei et al., 2012) and recent studies show their role in negatively regulating the biogenesis of mature let-7 family miRNAs (Heo et al., 2008; Piskounova et al., 2008; Viswanathan et al., 2008; Hagan et al., 2009; Nam et al., 2011). let-7 miRNA expression levels generally correlate with the differentiated state of cells, and low let-7 expression has been linked with poor prognosis in several cancer types (Boyerinas et al., 2010). A prevailing model suggests that the balance between the levels of Lin28 proteins and let-7 miRNAs is important for the self-renewal and differentiation of ES cells (Viswanathan and Daley, 2010). However, a recent study indicates that LIN28A is not required for the self-renewal of ES cells (Darr and Benvenisty, 2008). At present, the exact role of Lin28 proteins in stem cell regulation is unclear.

Genome-wide association studies have also implicated *Lin28* genes as regulators of growth and timing of developmental events in humans (Lettre et al., 2008; Hartge, 2009; Viswanathan and Daley, 2010). Analysis of LIN28A-overexpressing mice indicates that these roles are conserved in other vertebrate species (Zhu et al., 2010). However, most of the evidence available at present is focused on processes in later development, and the role of Lin28 proteins in early development, at the time when the main body is established, remains elusive.

Here, we investigate the function of *lin28a* and *lin28b* in the early development of the amphibian *Xenopus*, a well-established model for understanding the role of growth factors in regulating lineage restriction during vertebrate development. We show that *lin28a* and *lin28b* have overlapping, but distinct, expression profiles. Notably, lin28b protein is present maternally in the early embryo, whereas lin28a protein is only present after the activation of the zygotic genome at the mid-blastula transition (MBT). *lin28a* and *lin28b* are expressed in the pluripotent cells of the animal hemisphere and are enriched in the cells of the early mesoderm. Our data suggest a role for fibroblast growth factors (FGFs) in regulating the expression of *lin28* genes within the mesoderm.

In addition, individual and compound knockdowns of the three *Xenopus* lin28 proteins lead to dramatic inhibition of dorsal development that is characterised by greatly reduced development of axial and paraxial mesoderm. We show that knockdown of lin28 function in pluripotent cells of the early embryo compromises their ability to respond appropriately to mesoderm-inducing growth factors. Furthermore, lin28 knockdown leads to shifts in the temporal expression profiles of a number of key genes involved in

Area 11, Department of Biology, University of York, York, YO10 5DD, UK.

*Author for correspondence (harry.isaacs@york.ac.uk)

Accepted 19 December 2012

mesoderm specification. This represents the first evidence that Lin28 family genes play a key role in regulating the timing of, and responses to, growth factor signalling in the early development of a vertebrate embryo.

A key finding of this study is that, although amphibian lin28 proteins exhibit let-7 miRNA-binding activity, no significant effects on the abundance of let-7a, let-7f and let-7g miRNAs are detected in lin28 knockdown embryos, at the stage when germ layer specification occurs. Our data indicate that the requirement for lin28 function in the initial specification of the mesoderm is likely to be independent of its role in let-7 family biogenesis. By contrast, lin28 function seems to be required to regulate let-7 levels at later stages, after germ layer specification, indicating differential roles for lin28 in amphibian development: a let-7-independent early role and a let-7-dependent later role.

MATERIALS AND METHODS

Embryo methods

X. tropicalis embryos were produced as previously described (Khokha et al., 2005; Winterbottom et al., 2010). Embryos were injected at the 2- or 4-cell stage and cultured at 22°C. Animal cap explants were dissected at Nieuwkoop and Faber (NF) stage 8 (Nieuwkoop and Faber, 1994) and treated with 10 U/ml recombinant FGF4 protein (Isaacs et al., 1992) or 5 U/ml recombinant murine activin (Sigma).

mRNAs

mRNA was synthesised as described previously (Branney et al., 2009). The dominant-negative *X. laevis* FGFR4a (dnFGFR4) plasmid was a gift from Harumasa Okamoto (Hongo et al., 1999). Constructs coding for C-terminal HA epitope-tagged lin28 proteins were generated by PCR using reverse primers including a sequence coding for the YPYDVPDYA peptide. Subclones of the coding region of each lin28 isoform lacking both 5′ and 3′ UTRs were generated by PCR.

Morpholino antisense oligonucleotides (MOs)

MOs were synthesized by Gene Tools. Standard control MO, 5′-CCTCCTACCTCAGTTACAATTTATA-3′; lin28a1 MO, 5′-GGTCTGCC-TTGAAGTTGTCCCAGCT-3′; lin28a2 MO, 5′-GAGTCTCTTCTA-TCTGAGGTCAGGC-3′; lin28b MO, 5′-TCCTTCGGCCATGATGCC-TCCTGCT-3′.

miRNA duplexes

The let-7 duplex and mutant let-7 duplex sequences injected have been reported previously (Kloosterman et al., 2004).

In situ hybridisation

DIG-labelled probes were synthesized using DIG Labeling Mix (Roche Diagnostics). *In situ* hybridisation was performed as previously described (Harland, 1991; Reece-Hoyes et al., 2002).

Western blots

Western blots were carried out as described previously (Winterbottom et al., 2010). Affinity-purified *X. tropicalis* anti-lin28 antisera were produced by Enzo Life Sciences (Exeter, UK) in rabbits by inoculation of peptides corresponding to the C-terminal sequences of *X. tropicalis* lin28a1/a2 (EEQPISEEQELIPETME) or lin28b (SRKGPSVQKRKKT) proteins. Antibody dilutions were as follows: anti-lin28a and anti-lin28b, 1:1000; anti-HA (Sigma), 1:5000; anti-ERK1/2 (Sigma), 1:100,000; anti-dpERK (Sigma), 1:4000; anti-p-Smad2 (Millipore), 1:1000; anti-GAPDH (Santa Cruz), 1:2000; anti-mouse POD (Amersham Biosciences), 1:5000; anti-rabbit POD (Amersham Biosciences), 1:10,000. POD detection was carried out using the BM Chemiluminescence Blotting Substrate (Roche) and ECL Hyperfilm (Amersham).

Immunofluorescence

Cryosectioned embryos were processed for immunohistochemistry as described (Roth et al., 2010), with the following modifications: prior to sectioning, embryos fixed in MEMFA for 1 hour at room temperature were transferred to 15% fish gelatine/15% sucrose for 16 hours followed by a second 16-hour incubation in 25% fish gelatine/15% sucrose. Embryos were then sectioned at 15 μm. Lin28a and lin28b rabbit polyclonal antibodies were used at 1:200. Goat anti-rabbit Alexa Fluor 488 (Invitrogen, A11034) was used at 1:250. Once processed, slides were mounted in Vectashield mounting medium with DAPI (Vector Labs, H-1200). Peptide competition assays were carried out using the relevant epitope peptides at 1 μg/ml for 45 minutes at room temperature. Fluorescence was imaged on a Zeiss LSM 710 confocal microscope mounted on an Axio Observer.Z1 inverted stage.

Quantitative PCR (Q-PCR) for mRNA expression

Total RNA was isolated using TRI Reagent (Sigma). cDNA was synthesised using SuperScript II reverse transcriptase and oligo(dT) primers (Invitrogen). Q-PCR reactions were carried out in triplicate on an ABI Prism 7300 detection system (Applied Biosystems) using SYBR Green I PCR Master Mix (Applied Biosystems). Relative expression levels of each gene were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) and normalised to *ornithine decarboxylase* (*ODC*).

Q-PCR for miRNA expression

Total RNA was isolated using miRVana miRNA Isolation Kit (Applied Biosystems). cDNA was synthesized with miRNA-specific primers for TaqMan assays (Applied Biosystems) using the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems). Q-PCR reactions were carried out in quadruplicate using TaqMan Universal Master Mix II (Applied Biosystems) with TaqMan miRNA probes (Applied Biosystems). miRNA expression levels were normalised to *U6*.

RNA electrophoretic mobility shift assay (EMSA)

RNA oligonucleotides (2 μM) were labelled using the KinaseMax Kit (Applied Biosystems). Binding reactions were performed at room temperature in 10 μl total volume (Piskounova et al., 2008), with 0.5 μl labelled probe and diluted embryo extract in lysis buffer comprising 50 mM Tris pH8, 25% glycerol, 50 mM KCl, 2 mM DTT, 0.1 mM EDTA and Proteinase Inhibitor Cocktail III (Calbiochem). Binding buffer comprised 60 mM KCl, 10 mM HEPES pH 7.6, 3 mM $MgCl_2$, 5% glycerol, 1 mM DTT, 5 μg/μl heparin and 150 ng yeast tRNA/reaction. Twenty units of RNasin (Promega) were added per binding reaction with embryo lysates. Binding reactions were resolved on 8-10% polyacrylamide native gels. Supershifts were carried out using a 1:20 dilution of custom lin28a or lin28b antiserum (Enzo Life Sciences), or the corresponding pre-immune bleed (serum control) per binding reaction. Antiserum was pre-incubated with the protein sample and binding buffer for 20 minutes on ice, before labelled probe was added for 20 minutes at room temperature. The sequence of the let-7g terminal loop is: 5′-UUUGAGGGUCUAUGAUACCACCCGGUAC-AGGAGAU-3′.

RESULTS

Identification of *X. tropicalis lin28a* and *lin28b*

A previous study showed that a *lin28a* orthologue is present in *Xenopus laevis* (Moss and Tang, 2003). Analysis of Lin28 family cDNAs in the closely related diploid species *X. tropicalis* identified one *lin28b* orthologue and two sequences closely related to *lin28a*. Alignments of the *X. tropicalis* lin28a and lin28b protein sequences are shown in supplementary material Fig. S1A. Analysis of the *lin28a* genomic locus indicates that the two *lin28a*-related sequences, which we have designated *lin28a1* and *lin28a2*, result from alternative splicing of two small exons coding for different N-termini (supplementary material Fig. S1B).

Expression of *X. tropicalis lin28a* and *lin28b*

Previous studies showed that *Lin28a* mRNA and protein are expressed widely in all three germ layers during early development of mouse and chicken (Yang and Moss, 2003; Yokoyama et al., 2008). In *Xenopus*, most zygotic gene expression commences after the MBT.

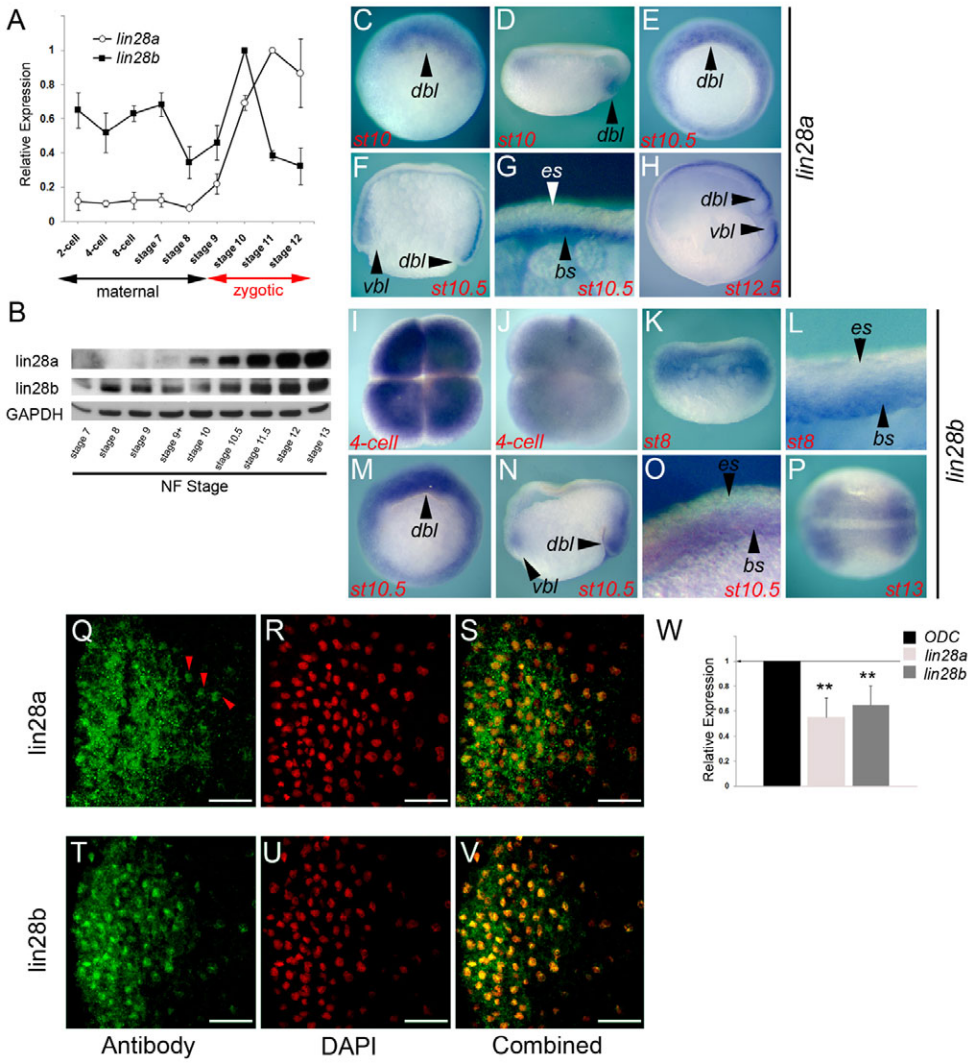


Fig. 1. Expression of *lin28a* and *lin28b* in the *X. tropicalis* embryo. (**A**,**B**) Temporal expression profiles of *lin28a* and *lin28b* mRNA (A) and protein (B) in early cleavage to late gastrula stage embryos as assessed by Q-PCR and western blot, respectively. (**C-P**) Spatial expression of *lin28a* (C-H) and *lin28b* (I-P) as assessed by *in situ* hybridisation. (C,E,M) Vegetal view; (D,F,H,K,N) transverse sections, lateral view; (G,L,O) higher magnification of F, K and N, respectively; (J) vegetal view of I. Stages are indicated: stage 8, mid-blastula; stage 10 to 10.5, early gastrula; stage 12.5 to 13, late gastrula. *dbl*, dorsal blastopore lip; *vbl*, ventral blastopore lip; *es*, ectodermal superficial layer; *bs*, basal sensorial layer. (**Q-V**) Fluorescence microscopy of cryosections from the circumblastoporal region of early gastrula stage 10.5 embryos. (Q) Immunolocalisation of lin28a expression. Red arrows indicate nuclear enrichment. (R) Nuclear DAPI staining in same section as Q. (S) Merge of Q and R. (T) Immunolocalisation of lin28b expression. (U) Nuclear DAPI staining in same section as T. (V) Merge of T and U. (**W**) Expression levels of *lin28a*, *lin28b* and the ubiquitously expressed housekeeping gene *ornithine decarboxylase* (*ODC*) in dnFGR4-injected embryos as compared with control embryos by Q-PCR. Solid line represents mRNA levels in corresponding control embryos (set at 1). $^{**}P<0.005$ (Student's *t*-test), for expression in experimental embryos compared with control embryos. Error bars indicate s.d. Scale bars: 50 μm.

Transcription of *lin28a* begins shortly after the MBT (Fig. 1A); before the MBT there is only a low level of maternally deposited *lin28a* mRNA present. The level of *lin28b* mRNA also begins to rise after the MBT, but in contrast to *lin28a* there is a relatively high level of maternal *lin28b* mRNA. In keeping with these data, western blot analyses show that low levels of lin28a protein are detected just after the MBT at blastula stage 9+. By the start of gastrulation at stage 10, the level of lin28a protein has risen significantly, and continues to rise through the gastrula stages (Fig. 1B). Maternally deposited lin28b is detected in pre-MBT stage embryos and protein levels remain constant through late blastula and gastrula stages.

Analysis of *lin28* mRNA localisation reveals that maternal *lin28b* mRNA is enriched in the animal hemisphere relative to the vegetal hemisphere (Fig. 1I,J). Animal localisation of *lin28b* persists in the deep sensorial layer of the presumptive ectoderm into blastula stages (Fig. 1K,L). The initial zygotic expression of *lin28a* and *lin28b* is enriched in the presumptive mesoderm of the circumblastoporal region of the embryo (Fig. 1C,E,M). Both genes show a marked dorsal to ventral expression gradient within the mesoderm (Fig. 1C,D,M,N). Expression of both genes persists in the deep sensorial cell layer of the animal hemisphere during gastrula stages (Fig. 1F,G,O).

By the end of gastrulation *lin28a* is expressed around the closed blastopore, with dorsal expression in the ectoderm extending further toward the anterior of the embryo than on the ventral side (Fig. 1H). *lin28b* expression is localised to the dorsal side of the embryo and is enriched at both the anterior and posterior ends of the neural plate (Fig. 1P).

Differential subcellular localisation of lin28a and lin28b proteins

It has recently been shown that mammalian Lin28a and Lin28b localise to different subcellular compartments and that this differential localisation in part accounts for the dissimilar modes of action of the two proteins; Lin28a protein is found mainly in the cytoplasm whereas Lin28b is highly enriched in the nucleus of cells in culture (Piskounova et al., 2011). We investigated whether amphibian lin28 proteins are also differentially localised within the cells of the early embryo. Immunofluorescence shows that lin28a and lin28b are localised in cells of the circumblastoporal region of the *Xenopus* embryo at the start of gastrulation (Fig. 1Q-V). Moreover, control experiments indicate that the observed immunofluorescence is specific for each of the proteins; the immunoreactivity can be effectively competed out by the relevant peptide immunogens (supplementary material Fig. S2).

Lin28a immunoreactivity is present throughout the cytoplasm of most cells and is also commonly seen in bright cytoplasmic puncta (Fig. 1Q-S). However, lin28a protein does not appear to be excluded from the nucleus, and in some cells (red arrows, Fig. 1Q) immunoreactivity is enriched in the nucleus. By contrast, immunoreactivity for lin28b is enriched in the nuclei of most expressing cells, but is also found at lower levels in the cytoplasm (Fig. 1T-V). We note that the bright, punctate, cytoplasmic fluorescence observed with lin28a is less common with lin28b.

FGF signalling and the regulation of *lin28* expression

Our previous study showed that *X. laevis lin28a* is downregulated in response to global inhibition of FGF signalling (Branney et al., 2009), highlighting lin28a as a component of a putative FGF-regulated pathway operating during the process of germ layer specification. Here we show that the levels of both *lin28a* and *lin28b* mRNAs are significantly downregulated in *X. tropicalis* gastrula stage embryos in response to FGF inhibition with a dominant-negative FGF receptor (Fig. 1W), indicating that FGF signalling is required for normal *lin28* expression during *Xenopus* development. Furthermore, our observation that *lin28* expression is enriched in the early mesoderm is in keeping with a role for FGF in regulating normal *lin28* expression because the early mesoderm represents a key domain of FGF activity in the early embryo (Christen and Slack, 1999; Branney et al., 2009).

Effects of lin28a and lin28b inhibition

To gain insight into lin28 function during development, we investigated the effects of overexpressing or knocking down lin28 proteins in the early embryo. Injection of up to 1 ng of synthetic *lin28a* or *lin28b* mRNAs has no gross phenotypic effects on the development of *X. tropicalis* embryos (data not shown). By contrast, single or compound knockdowns of lin28a1, lin28a2 and lin28b proteins using translation-blocking MOs produce a spectrum of phenotypic effects ranging from mild inhibition of dorso-anterior development, including loss of eyes and head structures, to a severe shortening of the dorsal axis (Fig. 2A,B; supplementary material Fig. S3A,C-F). The highest proportion of severe phenotypes is observed in the lin28a1, lin28a2 + lin28b (all lin28 MOs) compound knockdown embryos (Fig. 2B). Lin28 MOs used in this study effectively inhibit the translation of epitope-tagged lin28 proteins encoded by synthetic mRNAs containing the respective MO target sequences (supplementary material Fig. S3B), as well as the endogenous lin28a and lin28b proteins (Fig. 2C,D). Moreover, the observed effects are specific, as the phenotypic effects of lin28 knockdown are rescued by co-injection of non-targetable *lin28* mRNAs (supplementary material Fig. S3A,C-I).

Histological analysis of lin28 compound knockdown (morphant) embryos shows that the organisation of tissues in the dorsal axis is disturbed compared with controls (Fig. 2E,G), and that these effects are not restricted to a single germ layer. Thus, the patterning of the ectodermally derived brain and neural tube is highly abnormal in morphants (Fig. 2E,F), and, in keeping with the observed inhibited axial elongation, notochord and skeletal muscle tissue is reduced, indicating that the development of the dorsal axial and paraxial/somitic mesoderm is also abnormal (Fig. 2E-H).

Lin28 and the expression of genes in the early mesoderm

Given the observed effects of lin28 knockdown on the differentiation and patterning of mesodermal structures at larval stages, we were interested to see whether these late effects might result from prior effects on lin28 function in the early mesoderm, which, as discussed, is a site of zygotic *lin28a* and *lin28b* expression. Effects on the levels of a number of key genes expressed in, and required for, the specification of the mesoderm were analysed in lin28 morphant embryos at early gastrula stages.

Xenopus brachyury (Xbra) is a T-box transcription factor that is widely expressed in the early mesoderm (Smith et al., 1991). Chordin is a secreted BMP inhibitor expressed in the presumptive dorsal axial mesoderm (Sasai et al., 1994; Holley et al., 1995). MyoD is a bHLH transcription factor expressed in the myogenic lineage within the paraxial mesoderm from early gastrula stages (Hopwood et al., 1992). *In situ* hybridisation analysis shows that the levels of expression and size of the expression domains of the *Xbra*, *chordin* and *myoD* genes are greatly reduced in lin28 morphants (Fig. 3A). By contrast, no effects on the expression domain of the early endoderm marker *Sox17b* are observed in morphants (Fig. 3A).

Further quantitative analysis of gene expression levels in morphant embryos reveals significant inhibition of several genes expressed in the very early mesoderm, ranging from a 35% reduction for *mesogenin* to 80% reduction for *Tbox6* (Fig. 3B).

Lin28 knockdown alters the response of pluripotent cells to mesoderm-inducing signals

The FGF pathway is required for the normal development of the vertebrate mesoderm (Dorey and Amaya, 2010; Pownall and Isaacs, 2010). Our observation that *lin28* expression is reduced in response to FGF inhibition and that *lin28* genes are expressed in tissues that respond to FGF prompted us to investigate the effects of lin28 knockdown on FGF-mediated mesoderm induction.

The cells of the amphibian animal hemisphere constitute a pluripotent stem cell population that, upon exposure to appropriate growth factor signals, can be induced to form all three germ layers (De Robertis, 2006; Heasman, 2006). Blastula stage explants of animal hemisphere cells, known as animal caps, cultured in isolation differentiate as masses of atypical epidermis. However, when cultured in the presence of FGF these same cells respond by differentiating as a range of mesodermal tissue types (Slack et al., 1987; Isaacs, 1997). Fig. 4A shows control animal caps after 3 days in culture with or without FGF4 protein. The untreated control animal caps typically form rounded masses of tissue, which histology reveals as consisting of unlayered, atypical epidermis (Fig. 4B). By contrast, culture in the presence of FGF induces the formation of fluid-filled vesicles that are surrounded by epidermis but contain a range of mesodermal tissues, including mesenchyme and mesothelium (smooth muscle) (Fig. 4A,B). When the same

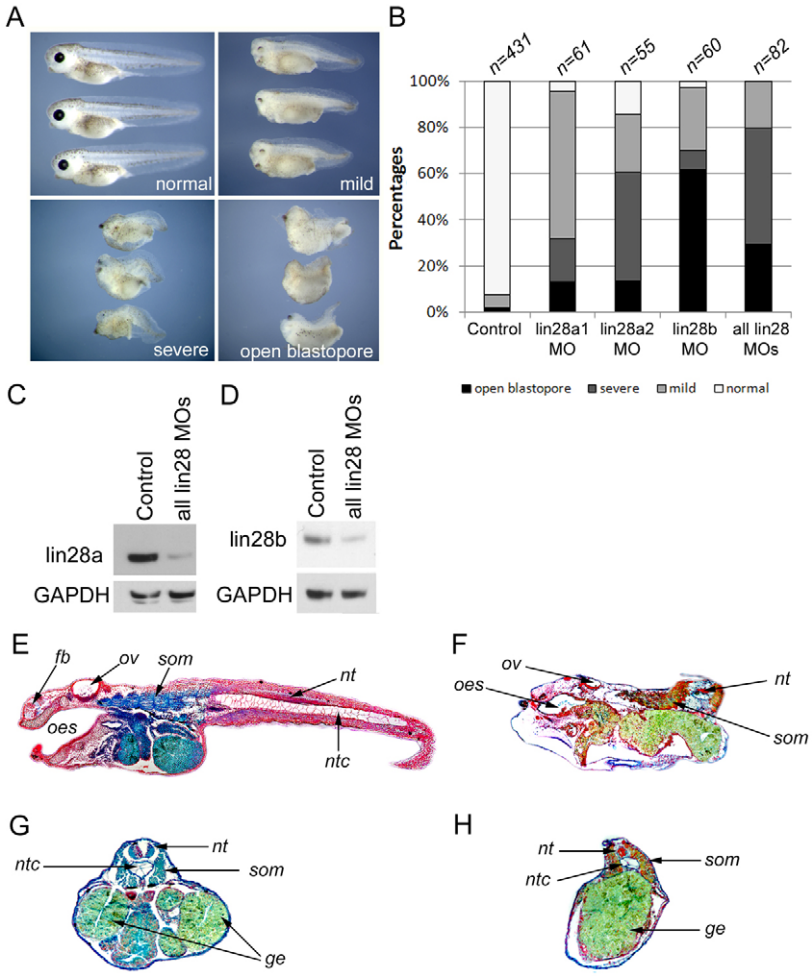


Fig. 2. Phenotype of lin28 morphants. (**A-D**) Range of axis defects in *X. tropicalis* lin28 morphants, from unaffected ('normal') to open blastopore. (B) Percentage of phenotypes in lin28a1, lin28a2, lin28b and all lin28 (i.e. in28a1 + lin28a2 + lin28b) morphant embryos. (C,D) Depletion of lin28a (C) and lin28b (D) proteins in lin28 (lin28a1 + lin28a2 + lin28b) morphant embryos as assessed by western blot at stage 20. GAPDH, loading control. (**E-H**) Histology on sagittal (E,F) and transverse (G,H) sections of lin28 morphants (F,H) and controls (E,G). *fb*, forebrain; *ge*, gut endoderm; *nt*, neural tube; *ntc*, notochord; *oes*, oesophagus; *ov*, otic vesicle; *som*, somites.

experiment is undertaken with animal caps from lin28 morphant embryos, the appearance of explant cultures in isolation is broadly unchanged. However, lin28 knockdown greatly inhibits the formation of fluid-filled vesicles and the differentiation of mesoderm following culture in the presence of FGF (Fig. 4A,B), indicating that lin28 function is required for animal caps to respond appropriately to FGF signalling.

Current models suggest that FGFs are not the primary signals for mesoderm specification in amphibians, but are secondary signals produced in the nascent mesoderm in response to the primary signals, such as nodal and activin (Isaacs et al., 1994; Schulte-Merker and Smith, 1995; Isaacs, 1997). We therefore investigated whether lin28 knockdown also interferes with the ability of activin to induce mesoderm and gastrulation-like elongation in animal cap explants. Similarly, we found that lin28 knockdown blocks the elongation of animal hemisphere explants and the differentiation of mesoderm induced by activin (Fig. 4C,D).

Thus far, we have shown that lin28 knockdown inhibits the ability of pluripotent cells from the embryo to differentiate into mesoderm in response to growth factor signals and that the expression levels of genes activated by these signals are reduced in lin28 morphant embryos *in vivo*, when germ layer specification is occurring. We next investigated whether lin28 knockdown interferes with ERK/MAP kinase phosphorylation, which is a characteristic, direct response to FGF signalling (Umbhauer et al., 1995; Christen and Slack, 1999; Branney et al., 2009). Fig. 4E is a western blot showing that activin or FGF treatment of animal cap explants upregulates the levels of ERK phosphorylation at early gastrula stages. However, ERK phosphorylation in response to FGF is reduced following lin28 knockdown, indicating that lin28 function is involved in the very earliest responses of pluripotent cells to FGF signalling.

Activin treatment of animal cap explants also induces ERK phosphorylation. Strikingly, lin28 knockdown greatly reduces ERK phosphorylation in response to activin. In contrast to ERK activation by FGF signalling, ERK phosphorylation in response to activin is a secondary, indirect response, requiring the downstream activation of transcription from FGF ligand-encoding genes (Schulte-Merker and Smith, 1995). As mentioned above, a pathway that involves primary mesoderm inducers, such as nodal/activin, activating the transcription of FGFs, including *FGF4* and *FGF8* as secondary mesoderm inducers/maintenance factors, is believed to operate during mesoderm formation *in vivo*. We investigated whether lin28 function is necessary for the normal activation of transcription of the endogenous *FGF4* and *FGF8* genes *in vivo*. The expression levels of *FGF4* and *FGF8* are significantly reduced in lin28 morphant embryos at early gastrula stages (Fig. 4F).

Heterochronic effects of lin28 inhibition during early amphibian development

lin-28 was originally identified as a heterochronic gene in *C. elegans*. *lin-28* mutations lead to the aberrant, precocious activation of later developmental events (Ambros and Horvitz, 1984; Moss et al., 1997). We examined whether the inhibition of lin28 function in *Xenopus* also leads to the misregulation of developmental events in

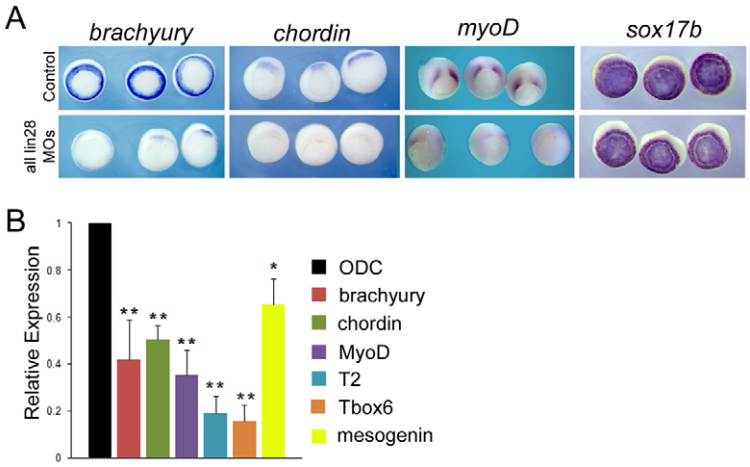


Fig. 3. Effect of lin28 knockdown on mesodermal gene expression. (**A**) Expression of *brachyury*, *chordin*, *myoD* and *sox17b* in control and lin28 morphant *X. tropicalis* embryos assessed by *in situ* hybridisation at gastrula stage 10.5. (**B**) Expression of the indicated mesodermal markers in lin28 morphants relative to expression in control embryos as assessed by Q-PCR at stage 10.5. **P*<0.05, ***P*<0.005 (Student's *t*-test), for morphant expression compared with control expression. *ODC* is a ubiquitously expressed housekeeping gene. Error bars indicate s.d.

the early amphibian embryo. To this end, we investigated the timing of activation of gene transcription and growth factor signalling in the early embryo during the process of germ layer specification.

Fig. 5 represents a developmental timecourse of gene expression in synchronously developing, sibling control and lin28 morphant embryos cultured at 27°C, with time points set at 30-minute intervals. This timecourse encompasses the pre-MBT to early gastrula stage of development, with the latter time points corresponding to the stages at which our analyses detect aberrant mesodermal gene expression in lin28 morphants. Transcription from *nodal-related 5* (*nr5*) and *nr6* is upregulated rapidly in control embryos following the MBT at mid-blastula stage 8. The upregulation of *nr1* expression in normal development is a slightly delayed response, relative to *nr5* and *nr6*. At the time resolution provided by our experiments, we see no evidence of precocious *nr5* or *nr6* transcription in response to lin28 knockdown; however, the expression of these genes at stage 8 is significantly higher in morphants (Fig. 5B,C). Subsequently, the shapes of the expression profiles from control and morphant embryos for *nr5* and *nr6* are very similar; however, the profiles in morphants are shifted to the

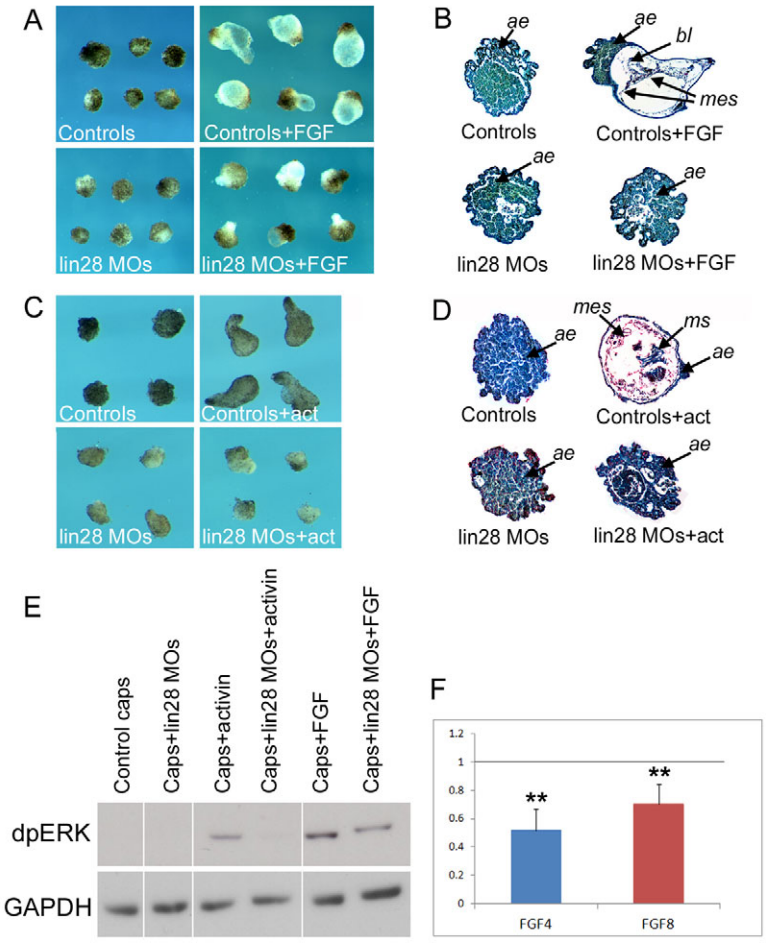


Fig. 4. Effect of lin28 knockdown on FGF and activin/nodal signalling. (**A**,**C**) Animal cap explants from control and lin28 morphant *X. tropicalis* embryos, with or without FGF (A) or activin (C) treatment. (**B**,**D**) Histology on sections of animal caps from A and C, respectively. *ae*, atypical epidermis; *bl*, blood; *mes*, mesenchyme; *ms*, muscle. (**E**) Western blots showing dpERK expression levels in animal cap explants from control and lin28 morphant embryos, with or without activin or FGF treatment. GAPDH, loading control. (**F**) Expression levels of *FGF4* and *FGF8* in lin28 morphants at stage 10.5 relative to controls as assessed by Q-PCR. Solid line represents mRNA levels in corresponding control embryos (set at 1). ***P*<0.005 (Student's *t*-test), for morphant expression compared with control expression. Error bars indicate s.d.

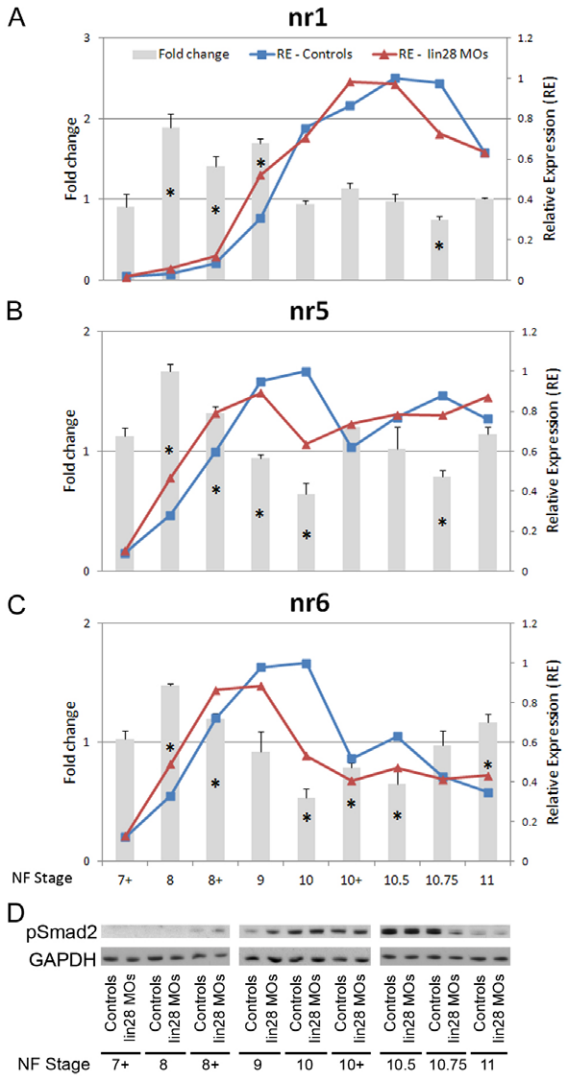


Fig. 5. Effects on the temporal regulation of nodal-related pathway components in lin28 morphants. (**A-C**) Temporal expression profiles of *nr1* (A), *nr5* (B) and *nr6* (C) from NF blastula stage 7+ to gastrula stage 11 as assessed by Q-PCR in control and lin28 morphant *X. tropicalis* embryos. Relative expression levels were calculated as the proportion of the maximum expression level in controls (set at 1) and fold change values are expressed relative to controls (set at 1). **P*<0.05 (Student's *t*-test), for morphant compared with control expression. (**D**) Western blot showing the expression levels of p-Smad2 at the NF stages analysed above. GAPDH, loading control. Error bars indicate s.d.

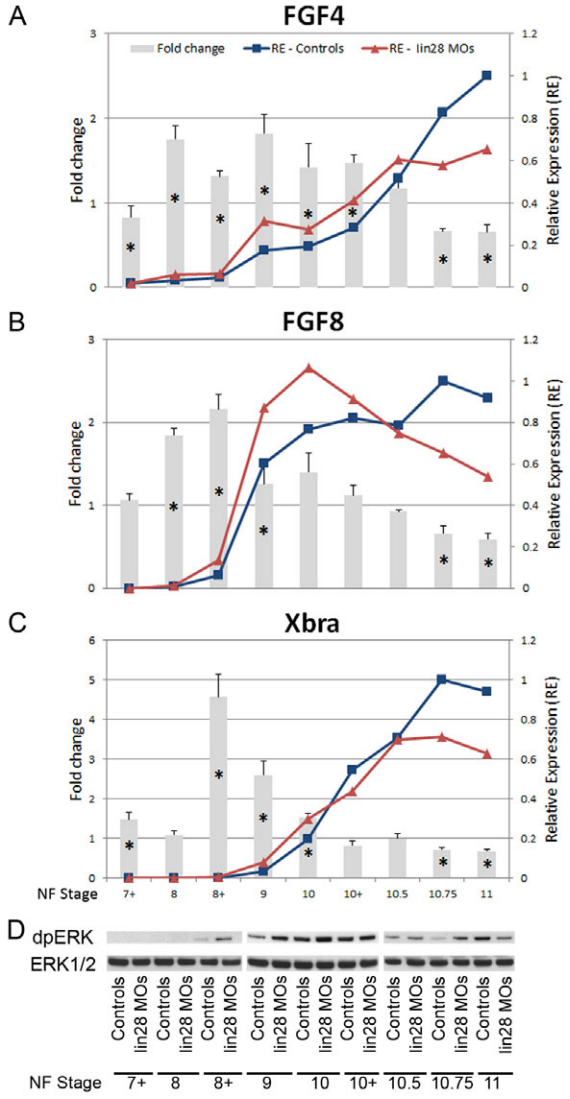


Fig. 6. Effects on the temporal regulation of FGF pathway components in lin28 morphants. (**A-C**) Temporal expression profiles of *FGF4* (A), *FGF8* (B) and *Xbra* (C) from NF blastula stage 7+ to gastrula stage 11 as assessed by Q-PCR in control and lin28 morphant *X. tropicalis* embryos. Relative expression levels and fold change values were calculated as in Fig. 5. **P*<0.05 (Student's *t*-test), for morphant compared with control expression. (**D**) Western blot showing the expression levels of dpERK at the NF stages analysed above. GAPDH, loading control. Error bars indicate s.d.

left. Thus, peak levels of *nr5* and *nr6* are reached earlier and the dip in expression levels, which is detected at the start of gastrulation in controls, occurs somewhat earlier in lin28 morphants. A similar shifted expression profile is also seen with *nr1*, with initial expression levels higher in morphants and the peak expression level being achieved earlier (Fig. 5A).

Smad2 is a key intracellular effector that is rapidly phosphorylated and activated in response to activin/nodal signalling (Baker and Harland, 1996; Bourillot et al., 2002; ten Dijke and Hill, 2004). The developmental timecourse of phospho-Smad2 (p-Smad2) levels was analysed and, in keeping with the expression profiles of the nodal-related genes, the initial level of p-Smad2 is higher and the reduction of p-Smad2 levels at early to mid-gastrula stages occurs somewhat earlier in morphants than in controls.

A similar timecourse analysis was also undertaken with the FGF ligand genes *FGF4* and *FGF8*, and with *Xbra*, which is activated as an immediate early response to both FGF and activin signalling. Again, the initial expression of all three genes is significantly elevated in morphants relative to control siblings (Fig. 6A-C), as are levels of activated ERK (Fig. 6D). However, in keeping with our previous results we find that gene expression levels are significantly reduced by early to mid-gastrula stages in lin28 morphant embryos.

let-7 miRNAs in *Xenopus* development

Recent evidence suggests that Lin28 proteins regulate the pluripotent state in stem cells *in vitro* through the negative regulation of let-7 miRNA biogenesis (Viswanathan et al., 2008;

Hagan et al., 2009; West et al., 2009). We examined whether the observed developmental effects in lin28 morphants correlate with altered let-7 expression levels. Fig. 7A shows the expression profiles in early *Xenopus* development of the mature forms of let-7a, let-7f and let-7g, which are regulated by Lin28 in mammalian cell culture (Viswanathan et al., 2008; Heo et al., 2009). We note that cycle threshold values (Cts) for all three let-7 miRNAs are relatively high, indicating that these miRNAs are present at low levels during early *Xenopus* development (data not shown).

Lin28 binding to the terminal loop region of immature let-7 miRNAs is crucial for regulating biogenesis of the mature miRNAs (Piskounova et al., 2008; Heo et al., 2009; Lightfoot et al., 2011). Cell lysates from *Xenopus* embryos overexpressing *lin28a1*, *lin28a2* or *lin28b* mRNAs contain activities that are able to bind to the terminal loop of let-7g (Fig. 7B-D). Moreover, we confirm that the binding activities correspond to the overexpressed lin28 proteins by showing that the addition of a relevant anti-lin28 antibody is able to supershift each riboprotein complex.

These data indicate that let-7 miRNAs are potential regulatory targets of lin28 proteins during amphibian development and it is possible that elevated levels of mature let-7 miRNAs contribute to the phenotype arising from lin28 knockdown. Indeed, overexpression of mature let-7a leads to a phenotype characterised by ventral bending of the main body axis, whereas overexpression of a mutant let-7 has less of an effect on axial development (supplementary material Fig. S4A,B). However, the phenotypic effects of let-7 overexpression are different from those produced by lin28 knockdown (compare with Fig. 2A).

We next analysed the levels of mature let-7a, let-7f and let-7g miRNAs in lin28 compound knockdown embryos (Fig. 7E). Surprisingly, we found no significant changes in the overall levels of let-7 miRNAs in lin28 morphant embryos at early gastrula stage 10.5. This is the stage when our analyses indicate that changes in mesodermal gene expression are already apparent in lin28 morphants. Therefore, the observed effects from lin28 knockdown on early mesoderm specification are unlikely to be mediated through a lin28/let-7 regulatory pathway. By contrast, by early tailbud stage 22, the overall level of let-7g is significantly higher in lin28 morphants compared with control embryos, indicating a role for lin28 proteins in regulating the biogenesis of at least some let-7 miRNAs during later development.

DISCUSSION

A model for lin28 function during germ layer specification

lin28 genes are expressed in pluripotent cells of the amphibian embryo. We propose that lin28 function is required in order for the pluripotent cells of the amphibian embryo to respond appropriately to mesoderm-inducing growth factors.

During amphibian development, both *lin28a* and *lin28b* are regulated by the FGF signalling pathway. In addition, knockdown of lin28 function inhibits the ability of pluripotent embryonic cells

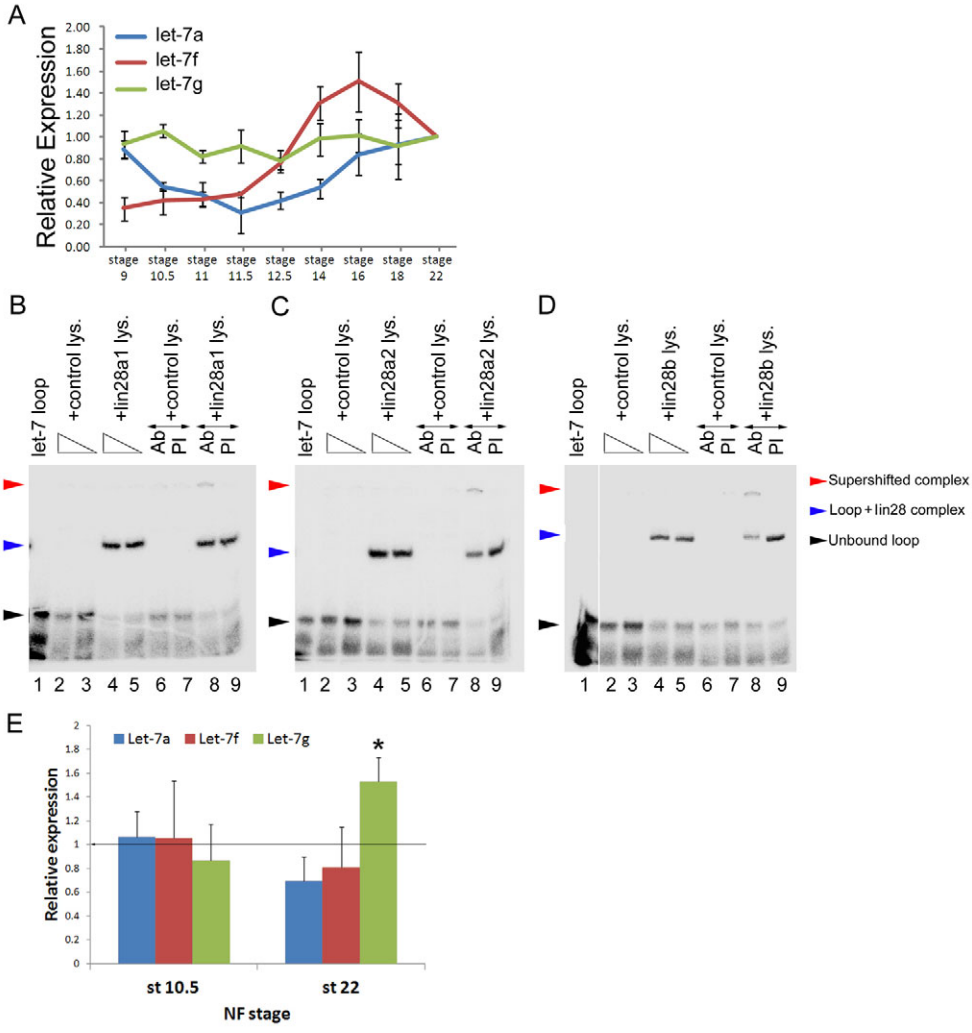


Fig. 7. let-7 in *Xenopus* development. (**A**) Q-PCR analysis of let-7a, let-7f and let-7g expression from stage 9 to stage 22. Expression is calibrated to stage 22 for each gene. (**B-D**) RNA EMSA supershifts showing let-7g terminal loop RNA-binding activity of dilutions of protein extracted from *Xenopus* embryos overexpressing lin28a1 (B), lin28a2 (C) or lin28b (D) proteins. Lane 1, labelled let-7g terminal loop alone; lanes 2 and 3, let-7g plus 1:4 and 1:8 dilutions of control embryo lysate, respectively; lanes 4 and 5, let-7g plus 1:4 and 1:8 dilutions, respectively, of lysate from lin28-overexpressing embryos; lane 6, let-7g plus 1:4 dilution of control embryo plus relevant lin28 antibody; lane 7, let-7g plus 1:4 dilution of control embryo lysate plus relevant pre-immune serum; lane 8, let-7g plus 1:4 dilution of lin28-overexpressing embryo lysate plus relevant lin28 antibody; lanes 9, let-7g plus 1:4 dilution of lin28-overexpressing embryo lysate plus relevant pre-immune serum. (**E**) Q-PCR analysis of relative expression of let-7 miRNAs in lin28 morphants compared with control embryos at stage 10.5 and 22. Solid line represents miRNA levels in corresponding control embryos (set at 1). *P<0.05 (Student's t-test), for morphant compared with control expression. Error bars indicate s.d.

to differentiate into mesoderm in response to FGF signals. Knockdown of lin28 function also compromises the ability of embryonic cells to respond to the mesoderm-inducing activity of activin, a TGFβ superfamily member that mimics the activity of endogenous mesoderm-inducing factors, such as nodal-related proteins. Furthermore, lin28 function is required for the appropriate early expression of FGF and nodal-related genes. Taken together, our data support a model in which *lin28* genes are components of the regulatory pathway involved in the specification and patterning of the amphibian mesoderm, modulating responses both upstream and downstream of FGF signalling.

Transcriptional regulation of *lin28* genes

Current understanding of the transcriptional regulation of vertebrate *Lin28* genes is limited. Our data suggest a role for FGF signalling in regulating the expression of *Xenopus lin28* genes. However, unlike some other known transcriptional targets of FGF signalling in amphibian development, such as *brachyury*, inhibiting FGF signalling does not completely block the expression of *Xenopus lin28a* or *lin28b* (Amaya et al., 1993). This suggests that additional pathways are likely to contribute to the transcriptional regulation of *lin28* genes in the amphibian embryo. We note that during the preparation of this manuscript it was reported that expression of *Lin28b* is regulated by FGF in the developing chick embryo (Bobbs et al., 2012), suggesting that an FGF-Lin28 regulatory axis might be a conserved feature of vertebrate embryo development.

Lin28 regulation of timing in early amphibian development

Lin28 knockdown leads to precocious elevation and subsequent precocious reduction in signalling by both the FGF and nodal pathways, indicating that lin28 function is involved in regulating the timing of signalling events in the early amphibian embryo. Recent evidence suggests that amphibian ventx transcription factors, which have Nanog-like activity, are also involved in regulating the timing of cell commitment to specific lineages in *Xenopus* (Scerbo et al., 2012). Both Lin28 and Nanog-related factors are implicated in the maintenance of the pluripotent state in mammalian stem cells. We propose that, similarly to ventx, lin28 participates in regulating the timing of responses to signalling pathways involved in cell lineage restriction in early amphibian embryogenesis.

lin-28 was originally identified as a heterochronic gene that encodes a factor promoting early developmental fates in *C. elegans* (Ambros and Horvitz, 1984; Moss et al., 1997). Subsequent evidence indicating that Lin28 proteins are involved in regulating the pluripotent state of mammalian stem cells in culture has further supported their role as factors promoting early stem cell phenotypes versus later differentiated cell fates (reviewed by Viswanathan and Daley, 2010). Our data provide the first evidence that *Lin28* genes are involved in regulating the timing of events in the early development of a vertebrate embryo, as opposed to stem cells in culture.

However, the role of lin28 in mesoderm development appears to be complex; although lin28 knockdown leads to precocious elevation of pathways involved in mesoderm specification, subsequent patterning and differentiation of the mesoderm are impaired. Thus, amphibian *lin28* genes appear to function as developmental 'gatekeepers' involved in regulating the transition from the pluripotent to mesodermal cell fate. Evidence indicates that this might represent a conserved function of *Lin28* genes, as it has recently been shown that Lin28a knockdown inhibits mesodermal gene expression during murine ES cell differentiation (Wang et al., 2012).

The proposed gatekeeper role of Lin28 proteins provides an intriguing link with the recently demonstrated function of FGFs as regulators of the transition from the pluripotent state to lineage restriction in ES cells. FGF signalling is required to allow ES cells to respond appropriately to signals involved in specifying particular lineages, including the mesoderm (Kunath et al., 2007). It will be interesting to determine whether an FGF-Lin28 axis, similar to that proposed in this study, operates in mammalian ES cells in culture.

let-7-independent function of lin28

Much recent work has focussed on Lin28 proteins as negative regulators of let-7 miRNA biogenesis. Although this is undoubtedly a key feature of Lin28 function, in some situations there is increasing evidence that Lin28 proteins have other, let-7-independent functions. A recent study indicates that LIN28A can function independently of regulating let-7 abundance during the differentiation of P19 embryonal carcinoma cells (Balzer et al., 2010). Furthermore, during *C. elegans* development LIN-28 has been shown to act through a two-step mechanism involving an initial let-7-independent phase and subsequent let-7-dependent phase (Vadla et al., 2012).

Our data also suggest early let-7-independent and later let-7-dependent roles for lin28 during amphibian development. It has been proposed that the initial let-7-independent function in *C. elegans* involves LIN-28 directly interacting with, and promoting translation from, target mRNAs (Vadla et al., 2012). There is increasing evidence that such direct, miRNA-independent regulation of mRNA translation by Lin28 proteins is a key element of their function, and a wide range of mRNA targets have been identified (Polesskaya et al., 2007; Xu and Huang, 2009; Qiu et al., 2010; Jin et al., 2011; Peng et al., 2011; Lei et al., 2012). We speculate that lin28 function during early amphibian development might similarly involve direct regulation of mRNA translation. This would seem particularly relevant to lin28b, which is present before the onset of zygotic transcription and is therefore a potential regulator during a phase of development that is critically dependent upon the translational regulation of maternally deposited mRNAs. In this regard, we note that Y-box proteins, which like lin28 contain a cold shock domain, are involved in translational control during the maternal phase of *Xenopus* development (Bouvet and Wolffe, 1994; Matsumoto and Wolffe, 1998). At present, the target mRNAs with which lin28 proteins might interact during early amphibian development remain to be identified.

Acknowledgements

We thank Dr Betsy Pownall for critical reading of the manuscript and Dr Mark Coles for helpful discussions during the course of the work.

Funding

This work was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) project grant [BB/H000925/1] to H.V.I.; a University of York, Department of Biology Pump Priming Award to H.V.I. and P.G.; and a BBSRC studentship to F.C.W.

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.089797/-/DC1

References

Amaya, E., Stein, P. A., Musci, T. J. and Kirschner, M. W. (1993). FGF signalling in the early specification of mesoderm in Xenopus. *Development* **118**, 477-487.

Ambros, V. and Horvitz, H. R. (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* **226**, 409-416.

Baker, J. C. and Harland, R. M. (1996). A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway. *Genes Dev.* **10**, 1880-1889.

Balzer, E., Heine, C., Jiang, Q., Lee, V. M. and Moss, E. G. (2010). LIN28 alters cell fate succession and acts independently of the let-7 microRNA during neurogliogenesis in vitro. *Development* **137**, 891-900.

Bobbs, A. S., Saarela, A. V., Yatskievych, T. A. and Antin, P. B. (2012). FGF Signaling during gastrulation negatively modulates the abundance of microRNAs that regulate proteins required for cell migration and embryo patterning. *J. Biol. Chem.* **287**, 38505-38514.

Bourillot, P. Y., Garrett, N. and Gurdon, J. B. (2002). A changing morphogen gradient is interpreted by continuous transduction flow. *Development* **129**, 2167-2180.

Bouvet, P. and Wolffe, A. P. (1994). A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. *Cell* **77**, 931-941.

Boyerinas, B., Park, S. M., Hau, A., Murmann, A. E. and Peter, M. E. (2010). The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* **17**, F19-F36.

Branney, P. A., Faas, L., Steane, S. E., Pownall, M. E. and Isaacs, H. V. (2009). Characterisation of the fibroblast growth factor dependent transcriptome in early development. *PLoS ONE* **4**, e4951.

Christen, B. and Slack, J. M. (1999). Spatial response to fibroblast growth factor signalling in *Xenopus* embryos. *Development* **126**, 119-125.

Darr, H. and Benvenisty, N. (2008). Genetic analysis of the role of the reprogramming gene LIN-28 in human embryonic stem cells. *Stem Cells* **27**, 352-362.

De Robertis, E. M. (2006). Spemann's organizer and self-regulation in amphibian embryos. *Nat. Rev. Mol. Cell Biol.* **7**, 296-302.

Dorey, K. and Amaya, E. (2010). FGF signalling: diverse roles during early vertebrate embryogenesis. *Development* **137**, 3731-3742.

Hagan, J. P., Piskounova, E. and Gregory, R. I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **16**, 1021-1025.

Harland, R. M. (1991). In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol.* **36**, 685-695.

Hartge, P. (2009). Genetics of reproductive lifespan. *Nat. Genet.* **41**, 637-638.

Heasman, J. (2006). Patterning the early *Xenopus* embryo. *Development* **133**, 1205-1217.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J. and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor microRNA. *Mol. Cell* **32**, 276-284.

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., Yeom, K. H., Han, J. and Kim, V. N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708.

Holley, S. A., Jackson, P. D., Sasai, Y., Lu, B., De Robertis, E. M., Hoffmann, F. M. and Ferguson, E. L. (1995). A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. *Nature* **376**, 249-253.

Hongo, I., Kengaku, M. and Okamoto, H. (1999). FGF signaling and the anterior neural induction in Xenopus. *Dev. Biol.* **216**, 561-581.

Hopwood, N. D., Pluck, A., Gurdon, J. B. and Dilworth, S. M. (1992). Expression of XMyoD protein in early *Xenopus laevis* embryos. *Development* **114**, 31-38.

Isaacs, H. V. (1997). New perspectives on the role of the fibroblast growth factor family in amphibian development. *Cell. Mol. Life Sci.* **53**, 350-361.

Isaacs, H. V., Tannahill, D. and Slack, J. M. W. (1992). Expression of a novel FGF in the *Xenopus* embryo. A new candidate inducing factor for mesoderm formation and anteroposterior specification. *Development* **114**, 711-720.

Isaacs, H. V., Pownall, M. E. and Slack, J. M. (1994). eFGF regulates Xbra expression during Xenopus gastrulation. *EMBO J.* **13**, 4469-4481.

Jin, J., Jing, W., Lei, X. X., Feng, C., Peng, S., Boris-Lawrie, K. and Huang, Y. (2011). Evidence that Lin28 stimulates translation by recruiting RNA helicase A to polysomes. *Nucleic Acids Res.* **39**, 3724-3734.

Khokha, M. K., Yeh, J., Grammer, T. C. and Harland, R. M. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Dev. Cell* **8**, 401-411.

Kloosterman, W. P., Wienholds, E., Ketting, R. F. and Plasterk, R. H. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res.* **32**, 6284-6291.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S. and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* **134**, 2895-2902.

Lei, X. X., Xu, J., Ma, W., Qiao, C., Newman, M. A., Hammond, S. M. and Huang, Y. (2012). Determinants of mRNA recognition and translation regulation by Lin28. *Nucleic Acids Res.* **40**, 3574-3584.

Lettre, G., Jackson, A. U., Gieger, C., Schumacher, F. R., Berndt, S. I., Sanna, S., Eyheramendy, S., Voight, B. F., Butler, J. L., Guiducci, C. et al. (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40**, 584-591.

Lightfoot, H. L., Bugaut, A., Armisen, J., Lehrbach, N. J., Miska, E. A. and Balasubramanian, S. (2011). A LIN28-dependent structural change of pre-let-7g directly inhibits Dicer processing. *Biochemistry* **50**, 7514-7521

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402-408.

Matsumoto, K. and Wolffe, A. P. (1998). Gene regulation by Y-box proteins: coupling control of transcription and translation. *Trends Cell Biol.* **8**, 318-323.

Moss, E. G. and Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.* **258**, 432-442.

Moss, E. G., Lee, R. C. and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**, 637-646.

Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. and Sliz, P. (2011). Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080-1091.

Nieuwkoop, P. D. and Faber, J. (1994). *Normal Table of Xenopus laevis (Daudin)*. New York, USA: Garland Publishing.

Peng, S., Chen, L. L., Lei, X. X., Yang, L., Lin, H., Carmichael, G. G. and Huang, Y. (2011). Genome-wide studies reveal that Lin28 enhances the translation of genes important for growth and survival of human embryonic stem cells. *Stem Cells* **29**, 496-504.

Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P. and Gregory, R. I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* **283**, 21310-21314.

Piskounova, E., Polytarchou, C., Thornton, J. E., LaPierre, R. J., Pothoulakis, C., Hagan, J. P., Iliopoulos, D. and Gregory, R. I. (2011). Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**, 1066-1079.

Polesskaya, A., Cuvellier, S., Naguibneva, I., Duquet, A., Moss, E. G. and Harel-Bellan, A. (2007). Lin-28 binds IGF-2 mRNA and participates in skeletal myogenesis by increasing translation efficiency. *Genes Dev.* **21**, 1125-1138.

Pownall, M. E. and Isaacs, H. V. (2010) *FGF Signalling in Vertebrate Development*. San Rafael, CA: Morgan & Claypool Life Sciences.

Qiu, C., Ma, Y., Wang, J., Peng, S. and Huang, Y. (2010). Lin28-mediated post-transcriptional regulation of Oct4 expression in human embryonic stem cells. *Nucleic Acids Res.* **38,** 1240-1248.

Reece-Hoyes, John S., Keenan, Iain D. and Isaacs, Harry V. (2002). Cloning and expression of the Cdx family from the frog *Xenopus tropicalis*. *Dev. Dyn.* **223**, 134-140.

Roth, M., Bonev, B., Lindsay, J., Lea, R., Panagiotaki, N., Houart, C. and Papalopulu, N. (2010). FoxG1 and TLE2 act cooperatively to regulate ventral telencephalon formation. *Development* **137**, 1553-1562.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K. and De Robertis, E. M. (1994). *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779-790.

Scerbo, P., Girardot, F., Vivien, C., Markov, G. V., Luxardi, G., Demeneix, B., Kodjabachian, L. and Coen, L. (2012). Ventx factors function as Nanog-like guardians of developmental potential in Xenopus. *PLoS ONE* **7**, e36855.

Schulte-Merker, S. and Smith, J. C. (1995). Mesoderm formation in response to Brachyury requires FGF signalling, *Curr. Biol.* **5**, 62-67.

Slack, J. M., Darlington, B. G., Heath, J. K. and Godsave, S. F. (1987). Mesoderm induction in early *Xenopus* embryos by heparin-binding growth factors. *Nature* **326**, 197-200.

Smith, J. C., Price, B. M., Green, J. B., Weigel, D. and Herrmann, B. G. (1991). Expression of a *Xenopus* homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79-87.

ten Dijke, P. and Hill, C. S. (2004). New insights into TGF-beta-Smad signalling. *Trends Biochem. Sci.* **29**, 265-273.

Umbhauer, M., Marshall, C. J., Mason, C. S., Old, R. W. and Smith, J. C. (1995). Mesoderm induction in Xenopus caused by activation of MAP kinase. *Nature* **376**, 58-62.

Vadla, B., Kemper, K., Alaimo, J., Heine, C. and Moss, E. G. (2012). lin-28 controls the succession of cell fate choices via two distinct activities. *PLoS Genet.* **8**, e1002588.

Viswanathan, S. R. and Daley, G. Q. (2010). Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445-449.

Viswanathan, S. R., Daley, G. Q. and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science* **320**, 97-100.

Wang, J., Cao, N., Yuan, M., Cui, H., Tang, Y., Qin, L., Huang, X., Shen, N. and Yang, H. T. (2012). MicroRNA-125b/Lin28 pathway contributes to the mesendodermal fate decision of embryonic stem cells. *Stem Cells Dev.* **21**, 1524-1537.

West, J. A., Viswanathan, S. R., Yabuuchi, A., Cunniff, K., Takeuchi, A., Park, I. H., Sero, J. E., Zhu, H., Perez-Atayde, A., Frazier, A. L. et al. (2009.) A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909-913.

Winterbottom, E. F., Illes, J. C., Faas, L. and Isaacs, H. V. (2010). Conserved and novel roles for the Gsh2 transcription factor in primary neurogenesis. *Development* **137**, 2623-2631.

Xu, B. and Huang, Y. (2009). Histone H2a mRNA interacts with Lin28 and contains a Lin28-dependent posttranscriptional regulatory element. *Nucleic Acids Res.* **37**, 4256-4263.

Yang, D. H. and Moss, E. G. (2003). Temporally regulated expression of Lin-28 in diverse tissues of the developing mouse. *Gene Expr. Patterns* **3**, 719-726.

Yokoyama, S., Hashimoto, M., Shimizu, H., Ueno-Kudoh, H., Uchibe, K., Kimura, I. and Asahara, H. (2008). Dynamic gene expression of Lin-28 during embryonic development in mouse and chicken. *Gene Expr. Patterns* **8**, 155-160.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920.

Zhu, H., Shah, S., Shyh-Chang, N., Shinoda, G., Einhorn, W. S., Viswanathan, S. R., Takeuchi, A., Grasemann, C., Rinn, J. L., Lopez, M. F. et al. (2010). Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat. Genet.* **42**, 626-630.