Characterization of Transcripts Expressed from the *Meis2* locus in the Natal Long-fingered Bat (*Miniopterus natalensis*).

Lyle Curry

Department of Molecular and Cell Biology

University of Cape Town

Rondebosch

Supervisors:

Prof. Nicola Illing

Prof. David Jacobs

February 2014

This dissertation was conducted in fulfilment of the requirements of a Master's of Science Degree in Molecular and Cell Biology, MSc (MCB), at the University of Cape Town.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Declaration

I, ……, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: ……

Date: …….

Acknowledgements

I would like to thank my supervisors Prof. Nicola Illing and Prof. David Jacobs who played vital roles in moulding, financing and supporting this study. Funding opportunity was provided by the University of Cape Postgraduate Funding Office and the National Research Fund. I would like to thank my wife Rubina Bunjun for being my support during this project and playing an integral role during my write-up. I must also thank my colleague Mandy Mason, and all those in my lab. I would like to thanks my family and friends especically my parents and in-laws.

Abstract

The Myeloid ecotropic insertion site 2 (*Meis2*) gene is an important transcriptional regulator involved in the pattering of the limb during vertebrate development. In the forelimbs (wings) of bats the expression *Meis2* was shown to be differentially expressed when compared to mice forelimbs. *Meis2* was found to be present in the interdigital webbing of the autopod (hand) of the developing bat forelimb. In both the mouse and the bat, two separate transcripts were discovered to be expressed from the 3'and 5' region of the *Meis2* locus. The 3' transcripts corresponded to annotated *Meis2* mRNA transcripts, the 5' transcript corresponded to an mRNA transcript that did not code for any known protein. The 5' transcript was thought to be a long non-coding RNA and was termed *lncMeis2*. In this study Random Amplification of cDNA Ends (RACE) analysis was used to identify the RNA transcripts expressed from the *Meis2* locus and to verify the presence of *lncMeis2.* RACE was performed on the heads and forelimbs of the Natal Long-fingered bat (*Miniopterus natalensis*) and the mouse (*Mus musculus*) and the products sequenced and aligned to the mouse genome for verification. Three transcript variants (isoforms) of *Meis2* were found to present in the bat, one in the forelimb, one in the head and one shared between the tissues. These transcripts, as well as other bat *Meis2* transcripts, were tested for positive selection. The bat *Meis2* transcripts showed no positive selection when compare to transcripts from non-bat vertebrate tertapods. In the mouse forelimb and head four isoforms were determined. Analysis of the *lncMeis2* RACE ascertained that the *lncMeis2* transcript is an artefact of cDNA synthesis and that *lncMeis2* is a truncated part of the 5' untranslated region of a *Meis2* transcript.

Table of Contents

List of Illustrations

Figures

Tables

Glossary & List of Abbreviations

AER	Apical ectodermal ridge
AP	Anterior-posterior
Bioinformatics	A scientific field that develops methods for organizing and analysing biological data
bp	basepair
cDNA	Complementary DNA
contig	Contiguous sequence. Consensus assembled from overlapping sequences
CRABP	Cellular retinoic acid binding proteins
CS17	*Carollia* stage 17
CYP26	Cytochrome P450 enzyme
cyp26b1	*Cytochrome P450 enzyme b 1*
DEPC	Diethylene pyrocarbonate
DNA	Deoxyridonucleic acid
DV	Dorsal-ventral
E13.5	Embryonic day 13.5
EDTA	Ethylenediaminetetraacetic acid
EMBOSS	European Molecular Biology Open Source Software
ENCODE	The Encyclopaedia of DNA Elements is a public research project which aims to map all the functional elements in the human genome
Ensembl	An online database of annotated genes and genomes that allows gene/genome alignments
EST	Expressed Sequence Tag
Evo-devo	Evolutionary Development
Exd	Extradenticle
Exons	Sequences within the mRNA transcripts that encode the amino acid sequence of the protein
Fgf	*Fibroblast growth factors*
Galaxy	An web-based platform for users create of bioinformatic workflow systems for data integration and data analysis
GSP	Gene specific primer
Homeodomain	A domain in a protein that is encoded for by a homeobox, that consists of about 60 amino acid residues which are usually similar from one such domain to another, and that recognizes and binds to specific DNA sequences in genes regulated by the homeotic gene
HoxA11	*Homeobox A11*
HoxA12	*Homeobox A12*
HoxA13	*Homeobox A13*
HoxD11	*Homeobox D11*
HoxD13	*Homeobox D13*
Hth	Homothorax

Introns	Sequences within the mRNA transcript that is non-coding and is spliced out during mRNA processing
Irx	*Iroquois homeobox*
Isoforms	mRNA transcripts which are expressed from the same gene but differ in exons
LICR	Ludwig Institute for Cancer Research
lncMeis2	*Long non-coding Myoloid ecotropic insertion site 2*
lncRNA	Long non-coding RNA
Meis1	*Myoloid ecotrpic insertion site 1*
Meis2	*Myoloid ecotrpic insertion site 2*
Meis3	*Myoloid ecotrpic insertion site 3*
mg	milligram
mL	millilitre
mRNA	messenger RNA
NGSP	Nested gene specific primer
NUP	nested universal primer
Oligo dT	A single strand of poly thymine nucleotides
Ortholog	Genes in different species that evolved from a common ancestral gene by speciation
Paralog	Genes which are related by gene duplication
Pbx	*Pre-B cell leukaemia homeobox*
PCR	Polymerase chain reaction
PD	Proximal-distal
Poly-A tail	A stretch of adenine nucleotides at the 3' end of mRNA
Polyadenylation consensus sequence	A sequence of nucleotide (AAUAAA/AATAAA) found approximately 30 nucleotides from the poly-adenine tail of nascent mRNA transcripts. This sequence signals the termination of transcription and the of the poly-A tail to the mRNA
RA	Retinoic acid
RACE	Random amplification of cDNA ends
Raldh2	*Retinaldehyde dehydrogenase 2*
RARα	Retinoic acid receptor alpha
RARβ	Retinoic acid receptor beta
RARγ	Retinoic acid receptor gamma
RARE	Retinoic acid response element
Rdh10	*Retinol dehydrogenase 10*
RL	recombinant limb
RNA	Ribonucleic acid
RT	Reverse transcriptase
RXRα	Retinoid X receptor alpha
RXRβ	Retinoid X receptor beta
RXRγ	Retinoid X receptor gamma
sRL	somite recombinant limb

TALE	Three Amino acid Loop Extension
Tgif	*Transforming growth factor β induced factor homeobox*
tracks	A section on the UCSC Genom Browser that visualises specific information graphically.
UCSC Genome Browser	An online genome browser curated by the University of California, Santa Cruz that allows for alignments of query sequences to a chosen genome and for comparative genomics
µL	microlitre
UP	universal primer
UTR	Untranslated region
zRL	zeugopod recombinant limb

Chapter 1. Introduction and Literature Review

Introduction

It was Darwin who first pondered ”What can be more curious than the hand of a man, formed for grasping, that of a mole for digging, the leg of the horse, the paddle of the porpoise and the wing of the bat, should all be constructed on the same pattern and should include the same bones in the same relative position?” (Darwin, 1859). In the late 1800s and the early half of the 1900s embryologists knew that morphology was a consequence of development, but the mechanism responsible for the formation of the body plan eluded them. They were aware of substances, from certain tissues and regions of the embryo, which affected the patterning of of limbs and the overall body plan (Carroll *et al.*, 2001). Only with the advent of molecular biology and genetic techniques, were these patterning molecules identified, and the genes controlling their expression were discovered (Carroll *et al.*, 2001). It was only with the advent of molecular biology and advanced genetic techniques that the molecules responsible for these patterns were discovered, along with the regulatory genes controlling their expression (Carroll *et al.*, 2001).

The first of these developmental genes discovered, called homeotic genes, were from the larval stage of the fruit-fly, *Drosophila melanogaster*, and they controlled the formation of the body and appendages of the fly (Lewis, 1978). Interestingly, it was found that the mouse contained the exact counterparts (orthologs) of the fruit-fly genes (homeobox genes) responsible for governing the formation of the body plan (Duboule and Dolle, 1989). More fruit-fly larva homeotic genes were later identified and many of these genes had orthologous counterparts in vertebrates (Nüsslein-Volhard and Wieschaus, 1980). These genes were part of a conserved “tool-kit” of master genes that shaped and formed the body parts of metazoa, regardless of their morphology, from flies to vertebrates (Carroll, 2005).

The comparison of these developmental genes between species became a new field of study: evolutionary developmental biology or "evo-devo". Evo-devo was a merger of two distinct yet complementary disciplines, developmental biology and evolutionary studies (Carroll, 2005; Goodman and Coughlin, 2000).

One aspect of evo-devo involves examining patterned segments of the metazoan body as separate subunits or modules of distinct morphologic structure and characterising the signalling pathways that shape them. A clear example of this developmental modularity is the tetrapod (four-limbed vertebrate) (Bolker, 2000). The development of the tetrapod limb relies on the interaction between two tissues: mesenchymal cells (originating from the lateral plate mesoderm) and the apical ectodermal ridge (AER) (Vogt and Duboule, 1999). The AER is a thickening of the ectoderm which is located just above the mesenchymal cells (Saunders, 1948). Signals from the limb mesenchyme maintain the AER which release signals that cause the proliferation of cells just under the AER and this initiates the outgrowth of the limb bud (Carroll *et al.*, 2001). There are three separate axes of patterning that shape the limb bud, the proximal-distal (PD) axis, the anterior-posterior (AP) axis and the dorsal-ventral (DV) axis (Capdevila and Izpisua-Belmonte, 2001). Each axis is formed by distinct genetic signalling pathways (Tanaka and Tickle, 2007).

Fibroblast growth factors (FGFs), secreted by the AER, and the transcription factors *Myeloid ecotropic viral intergration site* (*Meis*) *1* and *Meis2* which are expressed in the proximal region, form a system that regulates development along the PD axis of the limb (Mercader *et al.*, 2000). The outgrowth of the limb along the PD axis establishes the formation of the skeletal elements: stylopod, zeugopod and autopod (Fig. 1.1)

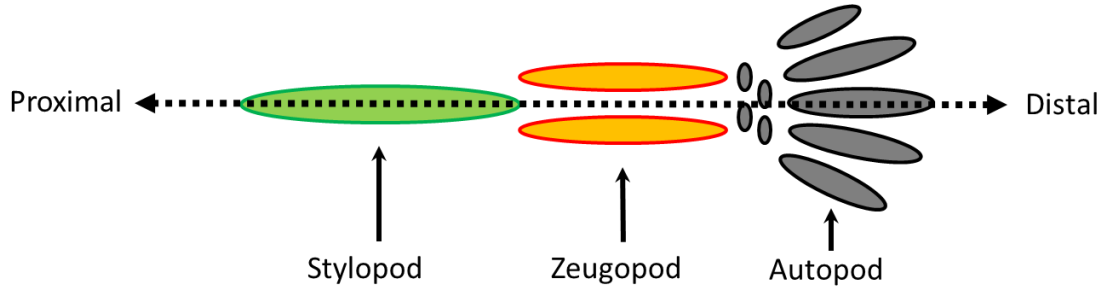

Figure 1.1 Diagram of limb skeletal elements along the proximal-distal axis: the proximal element, the stylopod (humerus or femur) (shown in green); the intermediate element, the zeugopod (ulna and radius, or fibula and tibia) (shown in yellow) and the distal element, the autopod (carpals and metacarpals, or tarsal and metatarsal, and phalanges) (shown in grey).

The skeletal domains are genetically defined along the PD axis (Fig. 1.1): the proximal stylopod is demarcated by expression of *Meis2* and *Meis1*, the intermediate zeugopod by *HoxA11* and *HoxD11* and the distal autopod by *HoxA13* and *HoxD13* (Davis *et al.*, 1995; Fromental-Ramain *et al.*, 1996; Mercader *et al.*, 1999).

1.1 *Meis* gene interactions and the role in limb patterning

MEIS proteins have been found to play an important role in the development and patterning of many vertebrate organs and structures, such as the limb, the mid-brain, the lens and retina of the eye and the female productive tract (Capdevila *et al.*, 1999; Heine *et al.*, 2008a; Marklund *et al.*, 2004; Mercader *et al.*, 1999; Sánchez-Guardado *et al.*, 2011; Williams *et al.*, 2005; Zhang *et al.*, 2002). The vertebrate *Meis* genes were first discovered as proto-oncogenes in the mouse cell line BXH-2 as sites where the murine leukaemia virus integrated into the host genome, causing aberrant gene expression (Moskow *et al.*, 1995; Nakamura *et al.*, 1996). There are three paralogs of the *Meis* genes in most vertebrates; *Meis1, Meis2* and *Meis3* (Nakamura *et al.*, 1996; Sánchez-Guardado *et al.*, 2011).The *Drosophila Homothorax* (*hth*) gene was subsequently identified as the ortholog of the *Meis* gene, which is required for nuclear translocation of the co-factor *extradenticle* (*exd*) during development of the *Drosophila* leg (Rieckhof *et al.*, 1997).

The *Meis* genes encode transcription factors that contain homeodomains; a conserved protein domain which recognizes specific *cis*-regulatory DNA sequences (Gehring *et al.*, 1994). MEIS proteins are a part of the Three Amino acid Loop Extension (TALE) superfamily of homeodomain proteins (Bürglin, 1997). The TALE superclass of homeodomain proteins are so named because they have an extension of three amino acids between α-helices 1 and 2 within the homeodomain (Yang *et al.*, 2000). Some of these TALE proteins, namely MEIS and PBX proteins (Table 1.1) are important cofactors of the homeobox (HOX) proteins, which are involved in anterior-posterior patterning of the body axis and the anterior-posterior patterning of the limb (Mann and Affolter, 1998; Shanmugam *et al.*, 1999).

Both *Meis1* and *Meis2* are expressed in the lateral plate mesoderm, and the proximal region of the developing limb bud in vertebrates (Fig 1.2 A and B) (Mercader *et al.*, 1999). MEIS1 and MEIS2 interact with PBX1, and regulate the translocation of PBX1 to the nucleus of the cell, where it acts as a transcription factor (Capdevila *et al.*, 1999; Mercader *et al.*, 1999).

Table 1.1: List of vertebrate TALE Homeobox Proteins

Family	**Name**	**No. of Mouse Paralogs**	**Hox gene interactions**
Meis	Myeoliod ecotropic viral insertion site homeobox	3	5' Hox proteins
Pbx	Pre-B cell leukaemia homeobox	4	3' Hox proteins
Irx	Iroquois homeobox	6	No interaction
Tgif	Tranforming growth factor β induced factor homeobox	6	No interaction

This interaction is conserved in *Drosophila*, HTH (which is the ortholog of MEIS2) and EXD which is a ortholog of PBX1) translocate to the nucleus of the leg imaginal disc (progenitor leg cells) after they interact (Rieckhof *et al.*, 1997). MEIS3 binds to PBX4, however, interactions with PBX1 have not yet been determined (Ravasi *et al.*, 2010). PBX1 is

produced in the entire limb, but is only present in the cell nuclei of the proximal region (Gonzalez-Crespo *et al.*, 1998). This suggests that PBX1 is found in the cytoplasm of distal cells because MEIS1 and MEIS2 are restricted to the proximal region of the limb bud during development (Capdevila *et al.*, 1999). Over-expression of both *Meis1* and *Meis2* in chick embryos induced the elongation of the stylopod and reductions, and sometimes deletions, of zeugopod and autopod elements of the limb during development (Capdevila *et al.*, 1999; Mercader *et al.*, 1999). The ectopic expression of *Meis1* in mouse limbs also induces a reduction in distal elements (zeugopod and autopod) and the expansion of the proximal element (stylopod) (Mercader *et al.*, 2009). The MEIS proteins have been shown to interact with the Hox proteins, for example, MEIS1 proteins have been shown to form stable heterodimers with 5' Hox proteins (HOXA11, HOXA12 and HOXA, B, C, D13) (Penkov *et al.*, 2013; Shen *et al.*, 1997). Additionally, HOXD13 and HOXD11 have been shown to bind MEIS1 and MEIS2 proteins *in vivo* and antagonize the expression of *Meis* gene (Salsi *et al.*, 2008). *Hoxd13* is expressed in the most distal part of the limb, the autopod, where it patterns the PD axis (Goff and Tabin, 1997; Kmita *et al.*, 2005). This suggests there is an antagonistic relationship between the *Meis* genes and HOXD13 and possibly other 5' Hox proteins to restrict MEIS1 and MEIS2 to the proximal region and promote distal limb patterning (Fig. 1.2) (Mercader *et al.*, 2009; Salsi *et al.*, 2008). In *Drosophila*, *hth* interacts with *Ultrabithorax* (*Ubx*) which is an ortholog of the Hox genes. This heterodimer binds to DNA and regulates the development of the limbs of the fly (Slattery *et al.*, 2011).

The intron-exon structure of vertebrate *Meis* genes is highly conserved, it consists of 13 coding exons (Irimia *et al.*, 2011; Sánchez-Guardado *et al.*, 2011). There are two alternative splice sites within *Meis2* (Fig. 1.3), a 3' alternate splice site in exon 11 and a cassette exon that contains a premature stop codon, exon 12a (Oulad-Abdelghani *et al.*, 1997; Sánchez-Guardado *et al.*, 2011). The alternative splice site is responsible for different transcripts being produced from the same gene (Hartwell *et al.*, 2008). The first 10 exons of *Meis2* are highly

conserved as they encode the Meis domain and the homeodomain which bind to DNA. The 3' region (from exon 11 to exon 12b) shows a diversity of structures and the different isoforms of the structures interact with different proteins (Irimia *et al.*, 2011).

The vertebrate *Meis2* locus has been shown to contain a large number of highly conserved non-coding regions (HCNRs) (Gómez-Skarmeta *et al.*, 2006). These HCNRs are mostly found within large introns of *Meis2* (Fig 1.3) and have been shown to induce expression during reporter assays in mice. This indicates that the HCNRs have enhancer function (Visel *et al.*, 2007).

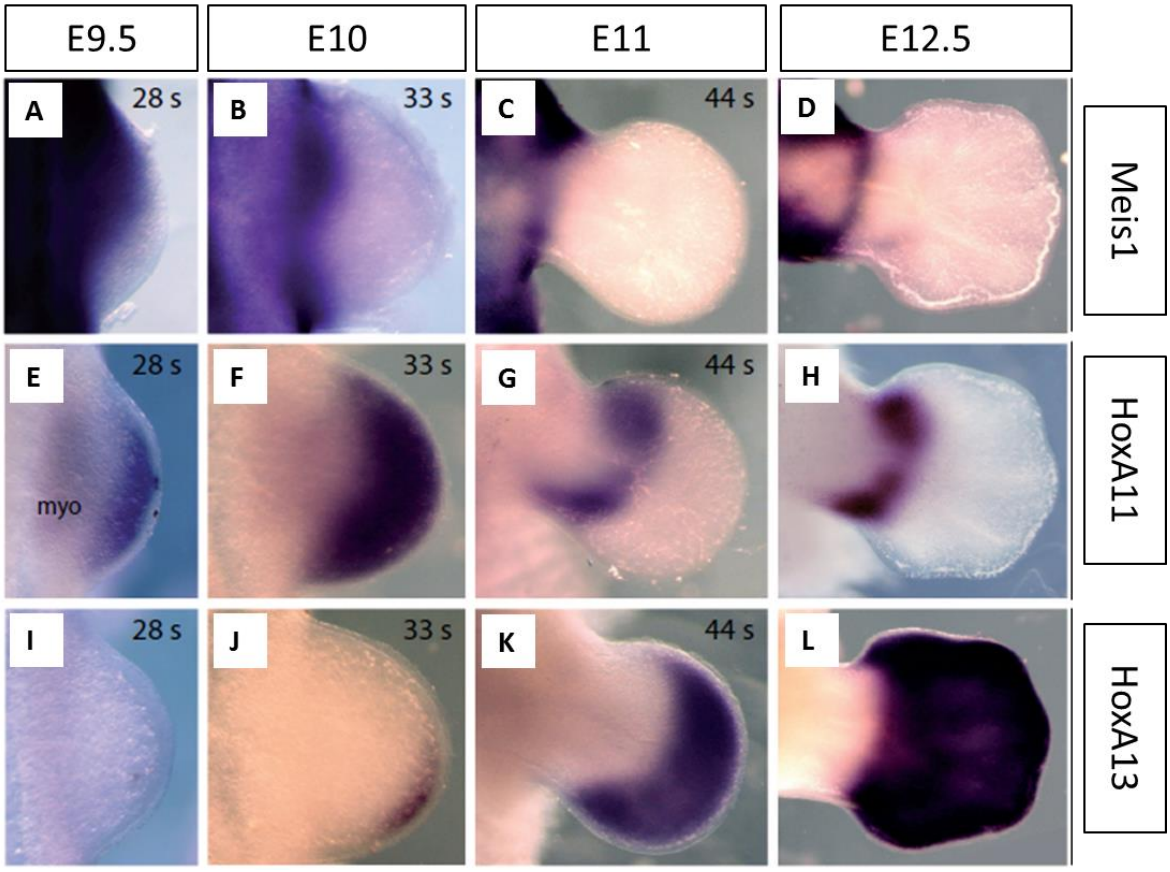

Figure 1.2: Wholemount *in situ* hybridisations of the developing forelimb of the mouse from developmental stages E9.5 (A, E and I), E10 (B, F and I), E11 (C, G and K) and E12.5 (D, H and L). The forelimbs were hybridised with probes for Meis1 (A-D), HoxA11 (E-H) or HoxA13 (I-L). The numbers on the top right of the photographs indicate the somite stage(s) of the embryos. A population of somite-derived myoblasts which are HoxA11-positive indicated by myo (E) are migrating into the developing limb bud (adapted from Mercader *et al.*, 2009)

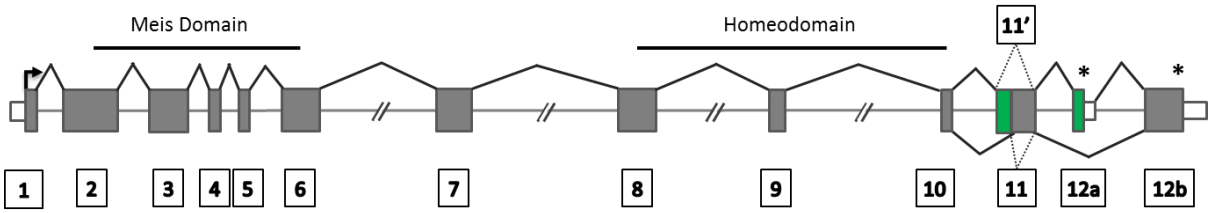

Figure 1.3: A representation of the intron-exon structure of vertebrate Meis genes. The grey blocks indicate the 13 exons (boxed numbers), the dashed lines indicate large introns and the alternative splice regions are shown in green. The regions encoding the Meis and homeodomain are indicated. Asterisks show the alternate in frame stop codons and the white boxes indicate the untranslated regions (UTR) (adapted from Sánchez-Guardado *et al.*, 2011).

Retinoic acid (RA) has been shown to induce a notable increase in *Meis2* and *Meis1* transcription in mouse cell lines (Oulad-Abdelghani *et al.*, 1997; Su and Gudas, 2008) .*In silico* analysis of the five kilobases upstream of the transcription start site of *Meis2*, on the mouse genome, has uncovered three retinoic acid response elements (RAREs). The RAREs induce gene expression when bound by retinoic acid receptor/retinoic acid complexes (Chambon, 1996) Subsequent *in vitro* experiments on a mouse embryonic cell line stimulated with RA, have shown two of the *Meis2* RAREs are occupied by RAR/RXR heterodimers however, this has not been characterised in limbs (Anno *et al.*, 2011).

1.2 Retinoic acid (RA) signalling in limb development

The induction of *Meis1* and *Meis2* expression by RA treatment and repression of *Meis1* and *Meis2* by FGF signaling became the basis of the hypothesis for the control of limb PD patterning, the so-called two signal model (Capdevila *et al.*, 1999; Mercader *et al.*, 2000; Zeller *et al.*, 2009). *In vivo* studies of RA biosynthesis in the limbs (Fig. 1.4) have shown that RA is catalysed from retinol (vitamin A) by two tissue specific enzymes, retinol dehydrogenase 10 (RDH10), which is encoded by *Rdh*10 and retinaldehyde dehydrogenase 2 (RALDH2), encoded by *Raldh*2 (Niederreither *et al.* 2000; Duester 2008). It has also been shown that FGFs oppose the expression of *Meis1* and *Meis2* in the distal region. FGF8 which is secreted by the AER can repress the distal induction of *Meis*1 and *Meis*2 (Mercader *et al.*, 2000). The antagonism between FGF and RA drives the outgrowth and patterning of limb

according to the proposed two signal model (Capdevila *et al.*, 1999; Mercader *et al.*, 2000; Zeller *et al.*, 2009)

Recent experiments involving the developing chicken limb support the finding that RA is responsible for the patterning of the PD axis (Cooper *et al.*, 2011; Roselló-Díez *et al.*, 2011). Dissociated limb mesenchymal cells were re-aggregated and were housed inside a jacket of limb bud ectoderm to form "recombinant limbs" (RLs) and these were grafted to the host embryo (Roselló-Díez *et al.*, 2011). The limb mesenchymal cells were cultured under various conditions before being sheathed in the ectoderm. When cultured with the appropriate stimuli, (ie FGF8, Wnt3a or RA) RLs have the ability to form identifiable limb structures. Additionally, when continuously exposed to RA, these cells specifically formed proximal structures (Cooper *et al.*, 2011). Somites express endogenous RA but there is little to no expression of RA in the zeugopod, as the concentration of RA diminishes along the PD axis (Vermot *et al.* 2005). RLs were grafted to the somites (sRLs) or the prospective zeugopod (zRLs) of the host embryos. The sRLs formed the three main skeletal element whereas the zRLs only formed the two distal elements (the zeugopod and autopod), suggesting that the formation of the complete PD axis requires signals from the graft region (Roselló-Díez *et al.*, 2011). The zRLs were treated with RA and all three elements were formed during outgrowth of the RL, meaning that the distally fated cells could be re-adjusted to form proximal elements from RA signalling (Mackem and Lewandoski, 2012; Roselló-Díez *et al.*, 2011).

However, results from the Duester group, which were based on knockout mouse experiments, do not support the proposed role of RA signalling in patterning the vertebrate limb along the PD axis (Cunningham et al., 2013, 2011; Zhao et al., 2010, 2009). Synthesis of RA was ablated in the mouse embryos through the deletion of limb specific RA synthesizing enzymes, retinol dehydrogenase-10 (RDH10) and retinaldehyde dehydrogenase-2 (RALHD2) (Cunningham *et al.*, 2011; Zhao *et al.*, 2010, 2009).

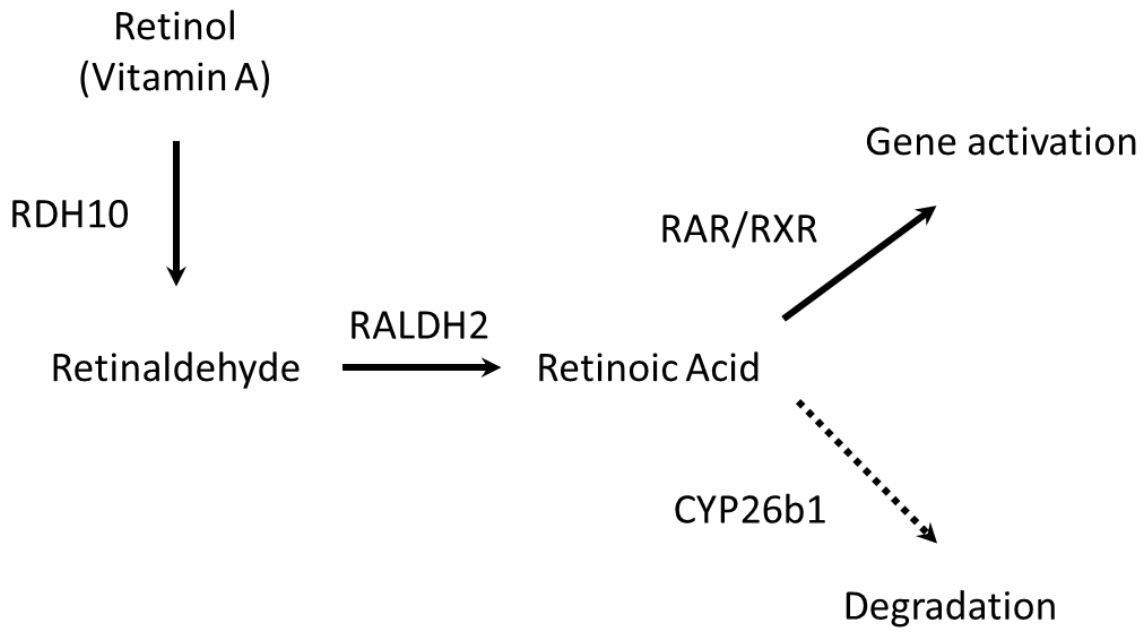

Figure 1.4: Retinoic acid limb biosynthetic pathway. Retinol is produced by RDH10 and RALDH2 to form RA. When RA enters target tissue it binds to RA/RX receptors which dimerize and bind to RAR elements on the DNA to induce or repress gene expression. If RA enters non-target tissue RA is enzymatically degraded by CYP26b1.

Mice with *Rdh10* null mutation have a specific phenotype termed t-rex, i.e. they have stunted forelimbs but normal hind limbs. The t-rex mice do not produce endogenous RA in the limb bud, although there was no reduction in the expression of *Meis*2 (Cunningham et al., 2013, 2011).This allowed normal patterning of the hind limb. However, these mice had stunted forelimbs, suggesting some degree of dependence on RA. Further investigation found that RA produced in the developing trunk of the embryo permits the induction of the forelimb when the limb bud is forming, but is not required for the patterning of the PD axis (Cunningham et al., 2013; Zhao et al., 2009). Mice with *Raldh2* or *Rdh*10 did not display PD patterning defects but they both display retention of the interdigital webbing; this could be because of a lack of RA signalling in the interdigital mesenchyme of the autopod (Cunningham *et al.*, 2011; Zhao *et al.*, 2010).

During development, the spatial and temporal expression of *raldh*2 correlates with the regions of RA production (Niederreither *et al.* 1997). RA signalling is induced through

nuclear hormone receptor complexes which consist of retinoic acid receptors (RAR-α, -β and –γ) and retinoid X receptors (RXR-α, -β and –γ) heterodimers. These complexes, once bound to ligand, bind to RA response elements (RARE) on the DNA and regulate their target gene's expression (Chambon 1996). The sequestering and degradation of cellular RA are facilitated by cellular retinoic acid binding proteins (CRABPs) (Donovan *et al.*, 1995) and by the RA-metabolizing CYP26 isozymes respectively (MacLean *et al.*, 2001). Different types of CYP26 enzymes are expressed in different regions of the developing embryo, and one type in particular, *cyp*26*b*1 is expressed in the developing limb bud (Abu-Abed *et al.*, 2002). The expression of *cyp26b1* is at the distal end of the developing limb. When the expression of *cyp26b1* is ablated in mice through null mutation, the phenotypes of the mice forelimbs resemble *Meis* gene overexpression. The limbs have reduced distal elements and the proximal region is enlarged (Yashiro *et al.*, 2004).

During vertebrate embryo development there are overlapping expression patterns of *Meis*1 and *Meis*2 (Capdevila *et al.*, 1999; Heine *et al.*, 2008a; Marklund *et al.*, 2004; Mercader *et al.*, 1999; Sánchez-Guardado *et al.*, 2011; Williams *et al.*, 2005; Zhang *et al.*, 2002). However, there has also been evidence of a significant difference in the spatial and temporal expression of *Meis1* and *Meis2*. In studies of whole chick development it was determined that *Meis1* is responsible for patterning the pretectum (early midbrain) and *Meis2* patterns certain areas of the gut (Sánchez-Guardado *et al.*, 2011).

1.3 Expression of Meis2 in bat wing formation

The differences seen in tetrapod limb types (i.e. fins, wings, paws etc.) can be attributed to modified regulation of developmental genes. In the bat, there is a stark difference in the morphology between the forelimbs and hind limbs. The bat forelimb, with its elongated skeletal elements and retained interdigital webbing, allows the bat the capability of powered flight. In comparison, the bat hind limb has shortened and uniform digits (Hockman *et al.*, 2009). A cross-species micro-array study was done on developing bat limbs to identify

differences in the mRNA transcript abundance between bat forelimbs and hind limbs, and also to determine which genes had differences in expression between the forelimb and hind limb that could be responsible for the differences in morphology (Mason MSc thesis, 2009). *Meis2* expression was significantly higher in the bat forelimb compared to the bat hind limb (Mason and Illing, unpublished data, Mason MSc thesis, 2009). There was no significant difference in signal determined for *Meis1* and *Meis3*. The *Meis2* gene was represented by two mouse probes (M400017713 and M400000987) on the Operon microarray slide. Blast analysis of the probes indicated that M400017713 (now referred to as 5'*Meis2*) corresponded to a mouse RIKEN clone (accession: AK043601.1) that in turn mapped to a region upstream of the *Meis2* 5' untranslated region (UTR) on the *M. musculus* genome (NCBI37/mm9). The RIKEN clone was isolated from a mouse embryo cortex cDNA library. The other OPERON probe M400000987 (now referred to as 3' *Meis2*) mapped to the 3' region of the *Meis*2 transcript (Mason MSc Thesis, 2009) The 5' *Meis2* probe exhibited a significantly higher signal than that of the 3' *Meis2* in the bat forelimb in the microarray dataset (Mason and Illing, unpublished data). Since AK043601.1 was annotated as a full length cDNA clone but did not show any overlap with existing mouse *Meis2* transcripts in the NCBI nr (non-redundant) and EST databases, it was suspected that it was an independently transcribed transcript. Sequence analysis of the RIKEN clone AK043601.1 determined that it contained an open reading frame of 36 amino acids but did not contain a start codon, and that its predicted amino acid sequence did not conform to any annotated protein domain (Mason and Illing, unpublished data). This suggested that bat mRNA transcripts that bound to the 5' Meis2 probe, and the RIKEN clone, were long non-coding RNAs (lncRNA), subsequently called *lncMeis*2 (Mason and Illing, unpublished data). The spatial expression of *Meis2* and *lncMeis2* were determined using *in situ* hybridisation analysis on *M. natalensis*. The analysis showed that *lncMeis2* and *Meis2* are co-expressed in the retained interdigital webbing of the forelimb (wing) of *M. natalensis* (Mason and Illing, unpublished data).

1.4 Long non-coding RNAs

LncRNAs are non-coding RNAs that are longer than 200 nucleotides in length (Kapranov *et al.*, 2007). Many of them have been shown to act as activators and repressors of gene expression, and have been shown to contribute to the development and differentiation of multicellular organisms (Amaral and Mattick, 2008). LncRNAs are defined according to their position and orientation in relation to nearby protein-coding genes (Fig. 1.5). Those that overlap the protein coding gene and are in the anti-sense orientation are termed "anti-sense lncRNAs". Those that are transcribed, in either orientation, within introns and terminate without overlapping an exon are called "intronic lncRNAs". Anti-sense lncRNAs that are transcribed in the opposite direction and orientation in relation to the nearby protein-coding gene from a bidirectional promoter are called "divergent lncRNAs". Intergenic lncRNAs are transcripts that are transcribed away from protein-coding genes with no overlap. These intergenic lncRNAs are required to be approximately 5 kilobases away from protein coding genes (Guttman *et al.*, 2009; Rinn and Chang, 2012). However, lncMeis2 does not conform to any of these definitions(Guttman *et al.*, 2009; Rinn and Chang, 2012).

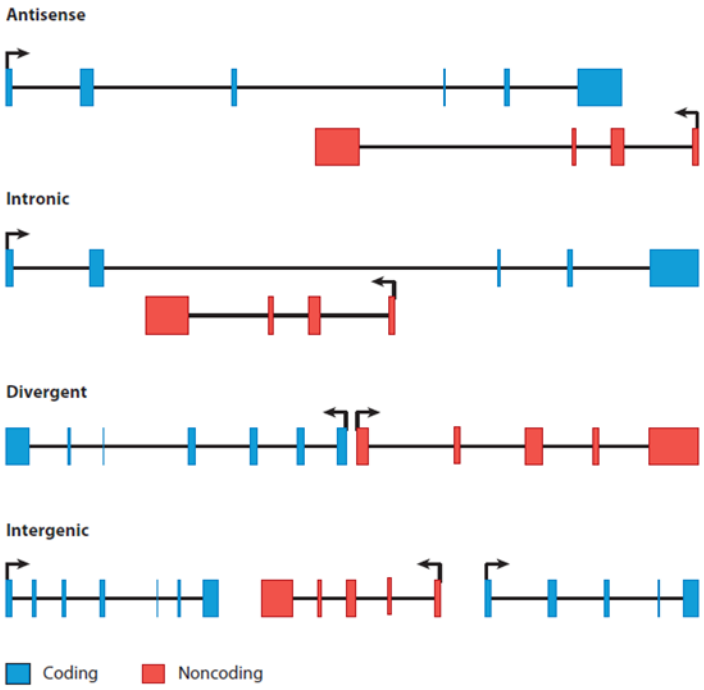

Figure 1.5: A representative diagram defining lncRNAs by their position relative to protein coding genes. Blue represents the coding transcripts and red the non-coding transcripts (taken from Rinn and Chang, 2012).

Additionally, lncRNAs can be classified based on the way they function, either in a *cis*-acting or *trans*-acting manner. *Cis*-acting lncRNAs influence neighbouring genes (Fig. 1.6A) but *trans*-acting lncRNAs influence genes on a different chromosome (Fig. 1.6B) (Amaral and Mattick, 2008; Guttman and Rinn, 2012). The most well-characterized examples of these type of lncRNAs are *HOTAIR* (HOX antisense intergenic RNA) and *HOTTIP* (*HoxA* transcript at distal tip) (Rinn *et al.*, 2007; Tsai *et al.*, 2010; Wang *et al.*, 2011). *HOTAIR* is transcribed from within the *HoxC* locus and acts as a scaffold for chromatin remodelling proteins, PRC2 (Polycomb repressor complex 2) and LSD1 (flavin contain monoamine oxidase). *HOTAIR* was first identified in a human cell line (Rinn *et al.*, 2007). *HOTAIR* bound by the histone modification complexes acts in *trans* and silences the expression of genes of the *HoxD* cluster (Tsai *et al.*, 2010). *HOTTIP* is expressed from the 5' region of the *HoxA* cluster. It has been shown to interact with WDR5 (WD repeat-containing protein 5). *HOTTIP* acts in *cis* with WDR5 to activate expression of the 5' HoxA genes. *HOTTIP* was

shown to be expressed in the mouse in the distal region of the developing limb and is required for patterning of the limb (Wang *et al.* 2011).

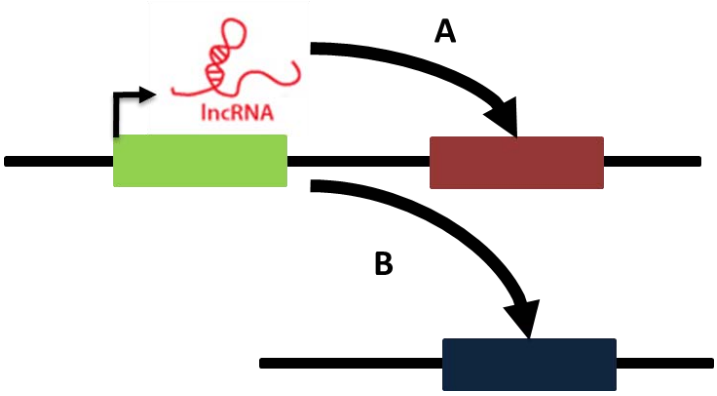

Figure 1.6: A diagram illustrating *cis*- (A) and *trans*-acting (B) lncRNAs.

1.5 Previous work done on bat species and limb development

Studies have been performed on bats to understand the morphologic and molecular evolution of their elongated limbs and retained interdigital webbing. The bat wing is a unique structure composed of flexible bones and membranes that allow bat the exclusive ability, among mammals, of powered flight (Swartz *et al.*, 2006). The forelimb skeletal elements of the bat wing are significantly elongated than those of non-flying mammals (Swartz, 1997). In the fossil record for the early Eocene period (49-53 million years ago), winged bats appeared fully developed without a record of any transitional species (Speakman, 2001).

To date there are five bat genomes that have been sequenced, the Little-brown bat also known as the Microbat (*Myotis lucifus*), David's Bat (*Myotis davidii*), Brandt's Bat (*Myotis brandtii*) the Large flying fox also known as the Megabat (*Pteropus vampyrus*) and the Black flying fox (*Pteropus alecto*) (Lindblad-Toh *et al.*, 2011; Seim *et al.*, 2013; Zhang *et al.*, 2013). The genomes of the Microbat and Megabat are available on USCS and Ensembl genome for comparative analyses (Fujita *et al.*, 2011; Hubbard *et al.*, 2002).Currently the genome for the

Natal long-fingered bat (*Miniopterus natalensis*) is being assembled (Schlebusch and Illing unpublished data).

Studies on many bat species to uncover the molecular manner in which flight has evolved, have either taken a transcriptome approach (Mason, 2009; Wang *et al.*, 2010) or candidate gene approach (Chen *et al.*, 2005; Cretekos *et al.*, 2008; Hockman *et al.*, 2008), to better understand the mechanisms involved in the development of wings.

In this study, the Natal long-fingered bat, *Miniopterus natalensis,* was used. This study was based on previous work by Cretekos *et al.* which staged a captive colony of bats, *Carollia perspicillata* (Cretekos *et al.*, 2005) but unlike those bats, *M. natalensis* could not be bred in captivity. Instead, pregnant *M. natalensis* females were caught in the wild, and the embryos were harvested. The bats were capture at the De Hoop Guano Cave which contains a breeding colony of 250,000 individuals (McDonald *et al*., 1990). The conservation status of *M. natalensis* is "least concerned" according to the IUCN Red list of Threaten species (Jacobs *et al.*, 2008) The harvested embryos were staged using an established staging system for *M .natalensis* embryos (Hockman *et al*., 2009). The conservation status, as well as the size of the breeding colonies make *M. natalensis* a suitable candidate for this study as removing a small number of individuals would have a small impact on the greater population.

For my MSc, I characterized the mRNAs transcribed from the *Meis2* gene locus in the autopods of *M. natalensis*, and compared it to the abundance of these mRNAs in the autopods of the mouse, *Mus musculus*.

1.6 Objectives

Specific Objective: To identify and characterise mRNA transcripts that are expressed from the *Meis2* and *lncMeis2* loci in *M. natalensis* and *Mus musculus.*

Rationale: The *lncMeis2* transcripts have been shown to be highly expressed the developing forelimb of *M. natalensis* when compared to the bat hind limb and to *M. musculus* autopod. As a first step in validating this observation, the transcriptional start sites, and exon structure of the *M. natalensis* and *M. musculus* transcripts need to be characterized.

Hypothesis: The transcriptional starts sites and exon structure of *lncMeis2* and *Meis2* are conserved in both *M. natalensis* and *M. musculus*, and differences in expression of *lncMeis2* accounts for the differences in limb morphology.

Chapter 2. Materials and Methods

2.1 Ethical clearance

Ethical clearance (application number 006/040) was acquired from The Science Faculty Animal Research Ethics Committee of the University of Cape Town to mate and sacrifice the mice.

Ethical clearance of the hunting and capture of Natal long-fingered bats (*Miniopterus natalensis*) (permit number AAA-064-000170-0035) was obtained from Cape Nature with the intent to harvest embryos. The bats were caught in the De Hoop Nature reserve in the Western Cape.

The experimentation on these animals was ethically cleared by The Science Faculty Animal Ethic Committee (ethics clearance number 2008/V16/DJ).

2.1.1 Mice

Laboratory-bred mice (outbred C57BL/6J) embryos were obtained from the Animal Unit of the University of Cape Town. Timed matings were performed and on embryological day 13.5 (E13.5) the pregnant female mice were euthanized by exposing them to halothane (Safe Line Pharmaceuticals Pty. Ltd., RSA) vapours for five minutes. The embryos were then dissected out. The embryos were stored in RNA*later*® (QIAGEN, Germany) at -20°C, for later dissection. The forelimbs and head were dissected from the whole embryos in phosphate buffered saline (PBS).

2.1.2 Bats

The bats were caught using a 3-bank harp trap (Faunatech, Australia) as they emerged from the cave mouth. Pregnant female *M. natalensis* bats were identified and the females with the largest abdomens were chosen as specimens. The bats that were not chosen were released. The specimens were euthanized with halothane (Safe Line Pharmaceuticals Pty. Ltd., RSA) and by decapitation as per the permit. The embryos were dissected out and the staged

according to developmental characteristics (Cretekos *et al.*, 2005). The embryos were store in RNA*later*® (QIAGEN, Germany) at -20°C and subsequently at -80°C for long term storage. Bat embyros staged at *Carollia* stage 17 (CS17) were dissected and the heads and forelimbs were used in this study.

2.2 Extraction and Amplification of RNA for RACE

2.2.1 RNA extraction

RNA was extracted from the forelimbs and heads of *M. natalensis* embryos at stages CS17 and the forelimbs and heads for the *M. musculus* at E13.5 using an RNeasy® Mini Lipid Tissue Kit (QIAGEN, Germany) as per the manufacturer's protocol (see supplementary information). The extracted RNA concentrations were as follows: *M. natalensis* forelimb 173.8 ng/µL and head 1772.5 ng/µ; and *M. musculus* forelimb (which was a pooled sample) 560.3 ng/µL and head 2649.8 ng/µL.

2.2.2 RACE cDNA synthesis and RACE PCR

The 3' and 5' RACE ready complementary DNA (cDNA) was synthesized from the total RNA from each tissue using the SMARTer™ cDNA Amplification kit (Clontech, USA) following manufacturer's instructions (see supplementary information. The RACE ready cDNA synthesized by using oligo dT which anneals to the poly-adenine tail of nascent mRNA transcripts (Fig 2.1). The RACE ready cDNA reactions were aliquoted into 20 µL aliquots and stored at -80°C. Gene specific primers (GSPs) (listed in Table A1) were designed to generate full-length cDNA transcripts expressed from 5' and 3'regions in the mouse and bat *Meis2* locus, in combination with the universal primer (UP). The nested gene specific primers (NGSPs) (Table A1) and the nested universal primer (NUP) are used in the second round RACE PCR to increase the specificity of the amplification of the transcripts (Fig. 2.1).

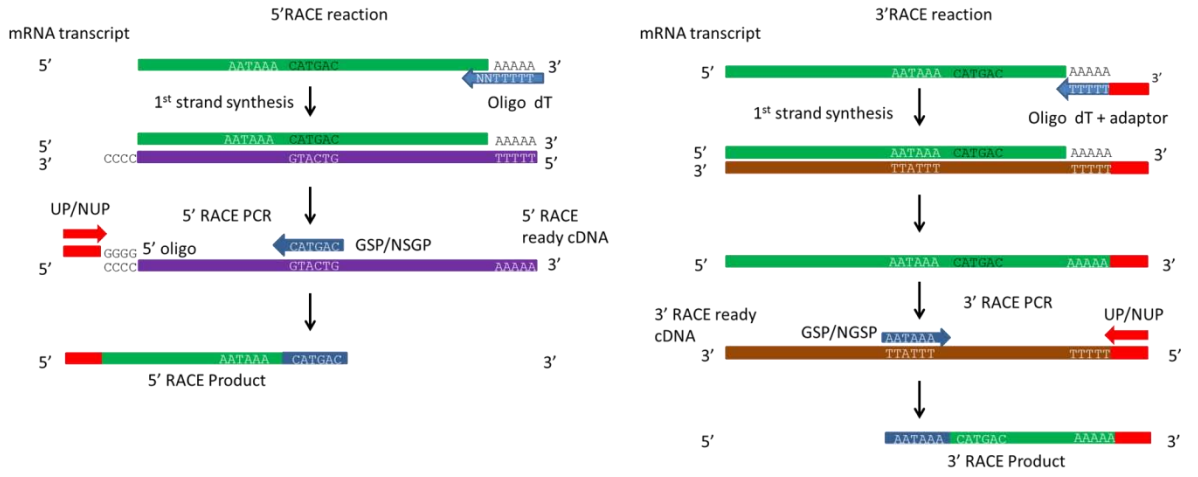

Figure 2.1: A diagram illustrating the generation of RACE ready cDNA and the amplification of RACE products during RACE PCR. The green bars indicate mRNA; the purple bars indicate the 5' RACE ready cDNA and the brown bar indicates the 3' RACE ready cDNA. The blue arrows indicate the primers: oligo dT used in the synthesis of the RACE ready cDNA and the gene specific primers/nested gene specific primers (GSP/NGSP) used in the RACE PCR. The red arrows indicate the universal primer/nested universal primer (UP/NUP) which anneals to the adaptor (red box) which is added to the RACE ready cDNA via oligo dT in the 3' reaction and via the 5' oligo in the 5' reaction.

The putative binding sites of the RACE primers are shown in (Fig. 2.2). The primers GSP1/NGSP1 and GSP3/NGSP3 were designed to amplify transcripts specifically from the *lncMeis*2 region, and to map the 5' and 3' boundaries of the *lncMeis2* RACE cDNA. The primers GSP2/NGSP2 and GSP4/NGSP4 amplified any cDNA transcripts, from the total mRNA, that were expressed from 3' region of *Meis*2. The GSP5/NGSP5 primers were designed to confirm the 5' transcription start of the Meis2 RACE cDNA. The primers GSP7/NGSP7, in combination with GSP5/NGSP5 were designed to validate whether *lncMeis2* was a separate mRNA transcript or if it was a part of the *Meis2* mRNA transcript.

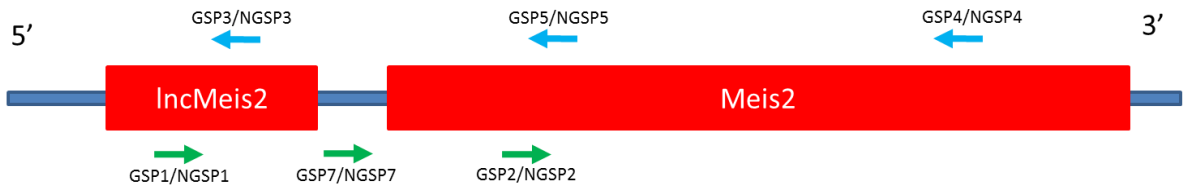

Figure 2.2: Schematic of binding regions of each RACE primers to the *Meis2* locus. The red blocks indicate the mouse EST clone boundaries corresponding to *lncMeis*2 and *Meis*2. The green arrows represent the 3' RACE primers and the blue arrows represent the 5' RACE primers.

The first round of gene specific RACE reactions on the RACE ready cDNA was carried out using Advantage™ 2 PCR kit (Clontech, USA), the gene specific primers (GSPs) and the Universal Primer (UP) (ClonTech, USA) (Table A1). The first round RACE reaction tube mix was setup as follows: 2.5µL of RACE ready cDNA, 5µL of universal primer mix (10 times concentration), 1µL of GSP (10µM), 5µL of 10XAdvantage 2 PCR buffer, 1µL of dNTPs (10µM), 1µL of 50X Advantage 2 Polymerase Mix and 34.5µL of PCR-grade water to make up the reaction volume to 50µL. These PCRs were carried out in a Techne Genius Thermal Cycler (Krackeler Scientific Inc., USA). The PCR program was 5 cycles of 94°C for 30 seconds and 72°C for 3 minutes; then 5 cycles of 94°C for 30 seconds, 70°C for 30 seconds and 72°C for 3 minutes; and concluded with 27 cycles of 94°C for 30 seconds, 68°C for 30 seconds and finally 72°C for 3 minutes.

The products of the first round RACE PCRs were then diluted 1:50 in Tricine-EDTA buffer (ClonTech, USA) and used as the templates for a second round of nested RACE PCRs. The products from GSP 1 to 5 were amplified using the Advantage™ 2 PCR kit, nested gene specific primers 1 to 4 (NGSPs) and the Nested Universal Primer (NUP) (ClonTech, USA) (Table A1). The nested RACE reaction tube was setup as stated for the first round RACE reaction tube but 2.5µL of diluted first round RACE PCR products was used as the template and 1µL of NGSP was used instead of GSP. The thermal cycling for the nested RACE was 20 cycles of 94°C for 30 seconds, 68°C for 30 seconds and 72°C for 3 minutes. The products from GSP7 were amplified using the Advantage™ GC 2 PCR kit (ClonTech, USA) with the NGSP7 and the NUP. The Advantage™ GC 2 PCR kit (ClonTech, USA) was used as it allows for amplification across regions that may have secondary structures due to high concentrations of guanine and cytosine. The PCR protocol was followed as per the manufacturer's instructions. The PCR program for the GSP/NGSP7 was as follows: 94°C for

1 minute; then 30 cycles of 94°C for 30 seconds, 68°C for 3 minutes and finally 68°C for 3 min.

2.3 Subcloning of RACE products

2.3.1 Gel extraction and subcloning of RACE products

The nested RACE reactions were electrophoresed on 2% (w/v) agarose gels stained with ethidium bromide (10mg/mL). The gels were viewed under UV radiation at 365 nm and prominent single bands were excised from the gel. DNA from these gel fragments was purified using Wizards® SV Gel and PCR clean-up kit (Promega, USA) as specified by the manufacturer. The purified PCR products were subsequently ligated using T4 ligase (Promega, USA) into the TA-vector pGEM®-T Easy using the pGEM®-T Easy vector system (Promega, USA). The ligation reaction was setup as per the manufacturer's protocol. The ligated RACE products were then transformed via heat shock into the *Escherichia. coli* strain XL-1 Blue (Sambrook *et al.*, 1989). The transformed *E. coli* (transformants) were spread onto petri dishes containing Luria Agar (LA) that was supplemented with the following: ampicillin (100mg/mL); X-gal (2% w/v) and IPTG (0.5mM). These cultures were incubated at 37°C for 16 hours.

2.3.2 Colony PCR

Up to ten white colonies (designated a-j) were picked and screened using colony PCR with M13 forward and reverse primers and KAPA*Taq* Ready Mix (with loading dye) (KAPAbiosystems, USA). The PCR mix (20µL) consisted of 10µL of 2 x KAPA*Taq* Ready Mix (KAPAbiosystems, USA), 0.8µL of each 10µM primer and 8.4µL of sterile nuclease free water. Each colony was inoculated separately into a PCR reaction and then subsequently streaked onto a reference master LA plates containing ampicillin (100mg/mL). Reactions were carried out in Techne Genius Thermal Cycler (Krackeler Scientific Inc., USA) for 2 minutes at 95°C followed by 30 cycles of 30 seconds at 95°C, 30 seconds at 45°C and 1

minute at 72°C, and a final elongation step of 2 minutes at 72°C. The PCR reactions were electrophoresed on a 2% (w/v) agarose gels stained with ethidium bromide (10 mg/mL). Products that had a fragment size greater than 200 base pairs (bp) (which is the size of the insert generated from a non-recombinant plasmid) were considered to contain inserts.

2.3.3 Sequencing of RACE products

A minimum of five colonies corresponding to the positive PCR product for each RACE product was inoculated in 75µL Luria broth and incubated for 16 hours at 37°C. Following incubation, 75 µL of 50% glycerol was added to each cultured broth to create 25% glycerol stock of the transformants. These stocks were sent to the University of Washington Highthroughput Genomic Unit, USA (UWHTGU) to be sequenced using M13 forward and reverse primers. Sequencing was done using the Sanger method (Sanger *et al.* 1977; Smith *et al.* 1986) on an ABI 3073xl DNA analyser.

The cDNA clones that were sent for sequencing were named according to their defining elements, including the species and tissue that they were generated from and the primer used in their PCR amplification (Table 2.1), to keep track of their origins.

An example using this nomenclature is the read name:

MF3ssN2C_b (see Table 2.1).

This transcript was generated from mouse (M) forelimb tissue (F), in a 3' RACE reaction (3) with sense strand cDNA (ss) and using NGSP2 (N2). It was the third largest band (C) extracted from the agarose gel and was found in the second colony (b) chosen to be screened in the colony PCR.

Table 2.1: Table of codes used for transcript labelling

Category	Type	Code
Species	Mouse	M
	Bat	B
Tissue	Forelimb	F
	Head	H
RACE	5' RACE	5'
	3' RACE	3'
Transcript	Sense	ss
Primer	NGSP#	N#
Gel Band	Largest	A
	Smallest	E
Colony	Colony 1	a
	Colony 5	e

2.4 Bioinformatics

2.4.1 RACE Sequence Assembly

The reads obtained from the UWHTGU were visualized using BioEdit program version 7.0.5.3 (Hall, 1999). Custom Galaxy pipelines (Blankenberg *et al.*, 2010) was created to process the sequences. The first pipeline was designed to reverse complement the reads that were in the opposite orientation and remove the vector sequence from the reads. The European Molecular Biology Open Software Suite (EMBOSS) tool vectorstrip (Rice *et al.*, 2000) was used to trim the vector sequence which was identified using the 5' and 3' flanking

sequence of the pGEM®-T Easy multiple cloning site (5'-TTCGATT-3' and 5'-AATCACTAG-3', respectively), leaving the insert sequence and associated primer sequences intact. Another pipeline was created to concatenate the trimmed forward and reverse reads, which were from the corresponding RACE reaction, using the labelling system (Table 2.1) into a matching file. The corresponding trimmed forward and reverse reads from each colony were aligned using the ClustalW (Thompson *et al.*, 1994) alignment feature. A contiguous consensus sequence (contig) was made by assembling the 5'RACE and 3'RACE reads using the Contig Assembly Program (CAP) (Huang, 1991) in BioEdit (Hall, 1999) (Fig. 2.3). The NGSP and NUP were used to determine the orientation of the transcript: the NGSPs orientated as the forward primer and the NUP at the reverse primer for 3' RACE transcripts, and *vice versa* for the 5' RACE transcripts.

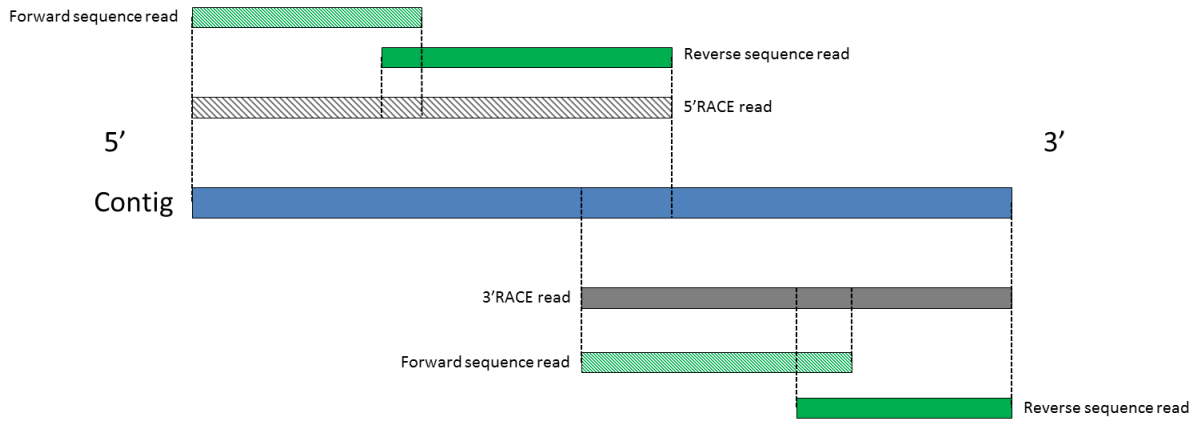

Figure 2.3: A graphical representation of the assembling a full length contig. First the forward and reverse reads are assembled to form either the 5'RACE or 3'RACE reads. The 3'RACE and 5'RACE reads are then assemble to form the full length contig

2.4.2 Analysis of Contiguous Sequences

The contigs were visualized using BioEdit and investigated to determine whether they contained the poly-adenylation site (AATAAA), a conserved non-coding region of nascent mRNA transcripts which is situated 30 nucleotides upstream of the poly-A tail. The analysis of the transcripts was performed using the UCSC Genome browser (Fujita *et al.*, 2011). The transcripts were aligned to the *M. musculus* genome assembly (NCBI37/mm9) using the BLAST-like Alignment Tool (BLAT) algorithm (Kent, 2002) on the USCS Genome Browser

(Fujita *et al.*, 2011). To validate the BLAT output, the matched reads were assessed by using the 'fitness score' algorithm (Wu *et al.*, 2008).

When visualised on the UCSC Genome browser, the best matched alignments were those which mapped to the correct chromosome and had a fitness score above 90%. Custom tracks were created for the RACE primers, *Meis2* overlap primers and the *lncMeis2* OPERON probe. Custom tracks were also created for the mouse ortholog *lncMeis2* (accession number: AK043601.1) and for three other expressed sequence tags ESTs that mapped to the 5' UTR of *Meis2* (accession numbers: BY106568, BY719373 and BY740861). For the *M. musculus* genome alignment, the uploaded tracks were compared with tracks from: UCSC Genes (Hsu *et al.*, 2006), Ensembl (Hubbard *et al.*, 2002) and NCBI (RefSeq) (Pruitt *et al.*, 2005). The tracks from the Ludwig Institute for Cancer Research (LICR)/ENCODE transcription factor binding site ChIP-seq experiments (Rosenbloom *et al.*, 2013) were used to investigate the transcription factors that bind in and around the *Meis2* locus.

2.5 *Meis2* Overlap Analysis

RNA was extracted from forelimb and head tissues of *M. natalensis* at CS17 and *M. musculus* at E13.5 as previously described in Section 2.2.1.

2.5.1 DNase Treatment and RNA Amplification

The RNA was quantified using a Nanodrop ND1000 (Thermoscientific, USA). A DNA free DNase I treatment kit (Ambion, USA) was used to digest any genomic DNA contaminating the RNA samples. The digestion recipe and protocol for 2 µg of total RNA in 10µl of nuclease free water was as follows: 1µl of 10x DNase I buffer and 1µl of rDNase I (enzyme) to the RNA and incubate for 20-30 minutes at 37°C. Incubations were carried out in Techne Genius Thermal Cycler (Krackeler Scientific Inc., USA). After the incubation, 2 µl of re-suspended DNAse I Inactivation Reagent was added and mixed well on a vortex for 10 seconds. The reaction was then incubated at room temperature for 2 minutes with occasional

mixing. The reaction tubes were then centrifuged at 10,000x g for 1.5 minutes and the supernatant containing treated RNA was collected.

This treated RNA was used to synthesize cDNA using SuperScript III reverse transcriptase (RT) (Invitrogen, USA). The reaction mix was setup using 1µg of the DNase-treated RNA in 5µl of solution. To the RNA, 1µl of 10mM dNTPs and 1 µl of 50mM random hexamers was added and the volume was topped up with DEPC treated water to 13 µl. The reaction tubes were then incubated for 10min at 25°C and then 65°C for 5 minutes; this was done to allow the primers to anneal to the mRNA transcripts. The reaction tubes were placed on ice and 1µl RNAseOUT ® (Invitrogen, USA) was added as well as 1µl of 1M dithiothreitol and 4µl of 5X FS buffer and 1µl SuperScript III RT. The reaction tubes were mixed on a vortex for a few seconds and incubated for 10 minutes at 25°C, 50°C for 60 min and 85°C for 5 minutes. Reactions were setup with the remaining 1µg of each DNase treated RNA, the reaction mixes contained the components as per above cDNA synthesis but excluded RT. These reactions were treated under the same conditions as the reaction containing RT. The "no RT" reaction was used as a control to check that there was no genomic DNA contamination (- RT reaction).

2.5.2 Amplification of *Meis2* Overlap

The cDNA synthesis and - RT reactions were amplified using PCR primer sets listed in Table 2.4.1. The *Meis2* and *lncMeis2* primer sets were used as positive controls to determine if the transcript was present in the tissue. The PCR reactions per reaction tube were setup as follows: 10µl of 5X KAPA HiFi GC buffer, 1.5µl of 10mM KAPA dNTP mix 1.5µl for each 10µM primer, 2µl of cDNA, 1µl of KAPA HiFi HotStart DNA Polymerase. The PCR cycling conditions were setup as follows: initial denaturing step 95°C for 3 minutes, then another denaturing step of 98°C for 20 seconds, an annealing step of 60°C for 15 seconds and an extension step of 72°C for 1min this was done for 30 cycles, and a final extension step of 72°C for 1min.

The reactions were visualized on a 2% w/v agarose gel stained with ethidium bromide (10 mg/mL) to confirm the amplification of single bands. The PCR reactions were cleaned-up using the PureYield® SV gel and PCR clean-up kit (Promega, USA) as per the manufacturer's instructions. The cleaned up PCR products were ligated into pGEM-T Easy® vectors for 16 hours at 4°C and these ligation mixes were transformed and cultured as previously described in Section 2.1.3. Up to five clones were tested for positive insertion of the PCR product using M13 primers via colony PCR, and one recombinant clone for each tissue was sent for sequencing to The Central Analytical Facilities at the University of Stellenbosch. The sequenced reads were visualized and edited in BioEdit (Hall, 1999) and were subsequently aligned to the mouse genome by means of the UCSC Genome Browser (Fujita *et al.*, 2010).

2.6 Positive selection analysis of coding region of *Meis2* in bats

The assembled RACE contig for *M. natalensis*: BFMeis2_1, BFMeis2_2 and BHMeis2_3 were analysed with the NCBI online tool ORFinder (Wheeler *et al.*, 2003) to determine the coding domain sequence (CDS) of the transcripts. Once the start and stop codons were found on each transcript, the 5' and 3' UTR were trimmed away using BioEdit (Hall, 1999). The *M. natalensis Meis2* transcripts were renamed for this analysis: BFMeis2_1 to mnFMeis2_1, BFMeis2_2 to mnFMeis2_2 and BHMeis3_3 to mnHMeis2_3. The *Meis2* transcripts for *Myotis lucifus* and *Myotis brandtii* were obtained from the NCBI GenBank database. Predicted *Meis2* transcripts were also obtained for *Equus caballus* (horse) and annotated *Meis2* CDS transcripts were obtained for: *Homo sapiens* (human) and *M. musculus* (house mouse) from GenBank as well (Benson *et al.*, 2009). The *Meis2* CDS transcripts were aligned using the ClustalW (Thompson *et al.*, 1994) and the transcripts were grouped according to intron/exon structure and each group of transcripts was analysed independently.

The phylogenetic program Molecular Evolutionary Genentic Analysis (MEGA) version 5.0.5 was used (Tamura *et al.*, 2011) to test for positive selection. The CDS sequences were

uploaded onto MEGA and the selection analysis Codon-based Z-test for Selection was chosen. To test whether positive selection was operating on *Meis2*, this method provided an estimate of the average of synonymous substitutions per synonymous site (d_S) and the average of non-synonymous substitutions per non-synonymous site (d_N) and their variances Var(d_S) and Var(d_N) respectively is calculated. Therefore, the null hypothesis H_0: $d_N = d_S$ was tested using a Z-test:

$$Z = (d_N - d_S) / SQRT(Var(d_S) + Var(d_N))$$

Where SQRT is the square root.

The alternative hypothesis HA: $d_N > d_S$.

The analysis was bootstrapped for 500 replicates using the Nei-Gojobori model, where the transition/transversion ratio was fixed (Nei and Gojobori, 1986). A one-tailed test was performed to calculate the level of significance. The null hypothesis, H_0 was rejected if the p-value was less than 0.05.

Chapter 3. Results

Rapid amplification of complementary DNA (cDNA) ends (RACE) polymerase chain reaction (PCR) is a technique used to amplify full-length cDNA transcripts when only a partial sequence is known (Frohman *et al.*, 1988). This technique has been used to map the 5' and 3' ends of transcripts, to identify conserved transcripts between organisms, and define the coding regions of mRNA transcripts (Casella *et al.*, 1989; Radicella *et al.*, 1997; Wu *et al.*, 2008b). Another form of mRNA analysis is the production of expressed sequence tags (ESTs), which are cDNA products of total mRNA. ESTs are not target amplified, the mRNA is reversed transcribed using oligo dT and the entire synthesis is partially sequenced to identify the genes present (Adams *et al.*, 1991).

Transcription is the process by which messenger RNA (mRNA) is synthesized from the gene through complementary base pairing. The starting pointing of transcription on the genome initiates at the transcription start site (TSS). The nascent mRNA undergoes RNA processing to produce a mature mRNA transcript. The 5' end of the mRNA transcript obtains a methylated cap and in most cases the 3' end receives a polyadenylated (poly-A) tail, the exception are some lncRNA (Hartwell et al., 2008; Kiyosawa et al., 2005; Ravasi et al., 2006). The poly-A tail is added to the mRNA transcript by a protein complex that detects the polyadenylation consensus sequence (AAUAAA) which is 11-30 nucleotides upstream of the poly-A tail (Hartwell *et al.*, 2008). This consensus sequence signals the RNA polymerase to terminate synthesis and cleave the 3' end of the mRNA to form a new 3' end, and this recruits a poly-A polymerase that adds adenines to the exposed 3' end.The methylated cap on the 5' end and the poly-A tail on the 3' end of the nascent mRNA provide stability to the transcript.

The nascent mRNA is made up of coding and non-coding regions, called exons and introns respectively. A protein complex known as a spliceosome "cuts out" the introns and joins the

exons to piece together the mature mRNA transcript. Alternative splicing can also occur, this is a process where one or more of the exons can be "cut out" of the mature transcript or different version of the exons can be present. This allows for different mRNAs to be produced from the same primary transcript (Hartwell *et al.*, 2008).

During cDNA synthesis there are conditions which could lead to artefactual transcripts. Messenger RNA transcripts which are transcribed from regions of genomic DNA that contain high concentration of guanine and cytosine have the tendency to form stable secondary structures (Galtier *et al.*, 2001). Secondary structures such as internal RNA hairpins or stem-loops can stall the elongation of cDNA transcripts by hindering the reverse transcriptase (Klasens *et al.*, 1999). The presence of internal adenine rich regions also hinders the correct synthesis of cDNA transcripts. The polyadenine regions bind oligo dT, mispriming the cDNA synthesis reaction and thereby creating truncated transcripts (Fig. 3.1).

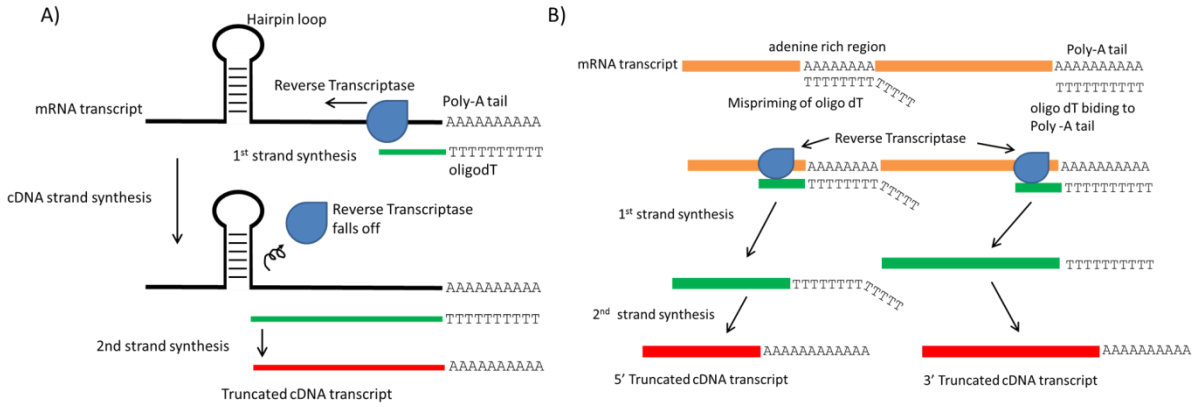

Figure 3.1: Diagram depicting the ways in which cDNA synthesis artefacts can arise. A) Illustrates the way in which trauncated cDNA transcripts can be synthesized when there is a hairpin loop downstream of the reverse transcriptase. B) depicts the manner in which artefactual cDNA transcripts are generated when there is a internal adenine rich region within the mRNA transcript that is being amplified using oligo dT

Initial analysis has determined that *lncMeis2* and *Meis2* mRNA transcripts are expressed from a region of DNA on the genome of *M. musculus* with a high concentration of guanine and cytosine. This means there could be a chance stable hairpin structures could form after the nascent mRNA is synthesised and processed.

3.1 RACE analysis of the *Meis2* locus

3.1.1 RACE PCR products

To characterize the mRNA transcripts which are transcribed from the *Meis2* locus Random Amplification of cDNA Ends (RACE) polymerase reaction was utilized. Primers were designed to highly conserved regions within *Meis2* to amplify transcripts from the forelimb and head tissue of *M. natalensis* and *M. musuclus.* *Meis2* transcripts of *M. natalensis* and *M. musuclus* were compared in order to determine if there was any sequence variation within the coding region. RACE was also used to determine the presence of *lncMeis2* mRNA transcripts in the tissues for both organisms.

The inserts from the 16 nested RACE reactions were electrophoresed in 1% (w/v) agarose gels stained with ethidium bromide and visualized using UV light (Fig. 3.2.). Figure 3.2A shows the 5' and 3' RACE inserts generated from *M. natalensis* forelimb total RNA. There were five prominent insert bands: one insert for each nested gene specific primer (NGSP) 1, 3 and 4 and two for NGSP 2 (Fig 3.2.1A). RACE inserts were amplified from *M. musculus* forelimb total RNA (Fig. 3.2B). There were four distinct bands visible with two dark smears occurring in the lanes for NGSP 1, and two each having a single insert (Fig. 3.2.B). NGSP4 has two insert bands, with NGSP3 having a single band (Fig. 3.2.B). In Figure 3.2.C, there are six separate bands visible: one band each for reactions using NGSP 1, 3 and 4 and three bands for NGSP2. The total RNA from the head of *M. natalensis* was used to synthesize the RACE inserts in Figure 3.2.C. Figure 3.2.D shows eight RACE inserts generated from *M. musculus* head tissue. There is one insert for NGSP 3, amplified two inserts for NGSP1 and four bands for NGSP2. The inserts amplified for NGSP4 are very faint but there appears to be two bands.

Table 3.1 lists the estimated sizes of the insert bands seen in Figure 3.2. The inserts amplified with the NGSP1 and NGSP3 primers for both *M natalensis* and *M. musculus* have similar sizes, which range from 300 – 700 nucleotides. The sizes of the NGSP2 and NGSP4 differ in

size between tissue and organism. The NGSP2 inserts for *M. natalensis* are much larger in size compared to *M. musculus*. The NGSP4 inserts for *M. musculus* show variability in size between tissues, the inserts from the head range from 2500 – 600 nucleotides whereas in the forelimb there was a 700 nucleotide band and a non-discernable smear above 900 nucleotides. The RACE reactions were electrophoresed in a higher density gel to obtain better resolution between inserts of similar sizes. This gel was used to extract the single inserts for sub-cloning.

Table 3.2 shows all the clones sequenced and analyzed for each RACE insert extracted. The letters in parentheses illustrate which bands of the reaction were successfully sub-cloned. There were at least five clones for each RACE insert which was sufficient for transcript analysis. In certain RACE reactions there was more than one insert, which are not apparent in Figure 3.2. This was due to the lower density of the gel used. Reactions that contained these extra inserts were: *M. musculus* NGSP1 forelimb (1 extra insert) and head (1); NGSP2 forelimb (1) and head (1), in *M. natalensis* NGSP2 forelimb (2) and head (1); and NGSP4 forelimb (1). During the sub-cloning of the RACE inserts, there were some inserts which failed to be ligated into the sequencing vector or were not allowed through the quality control process during sequencing and editing protocols. These inserts were: *M. musculus* NGSP1 head insert B; NGSP1 head insert A; NGSP4 head inserts B, C and D; and *M. natalensis* NGSP2 inserts A and B.

The successfully sequenced insert reads were assembled into 5'RACE and 3'RACE reads and aligned to the *M. musculus* genome (NCBI37/mm9) and the intron-exon structure of the transcripts were determined.

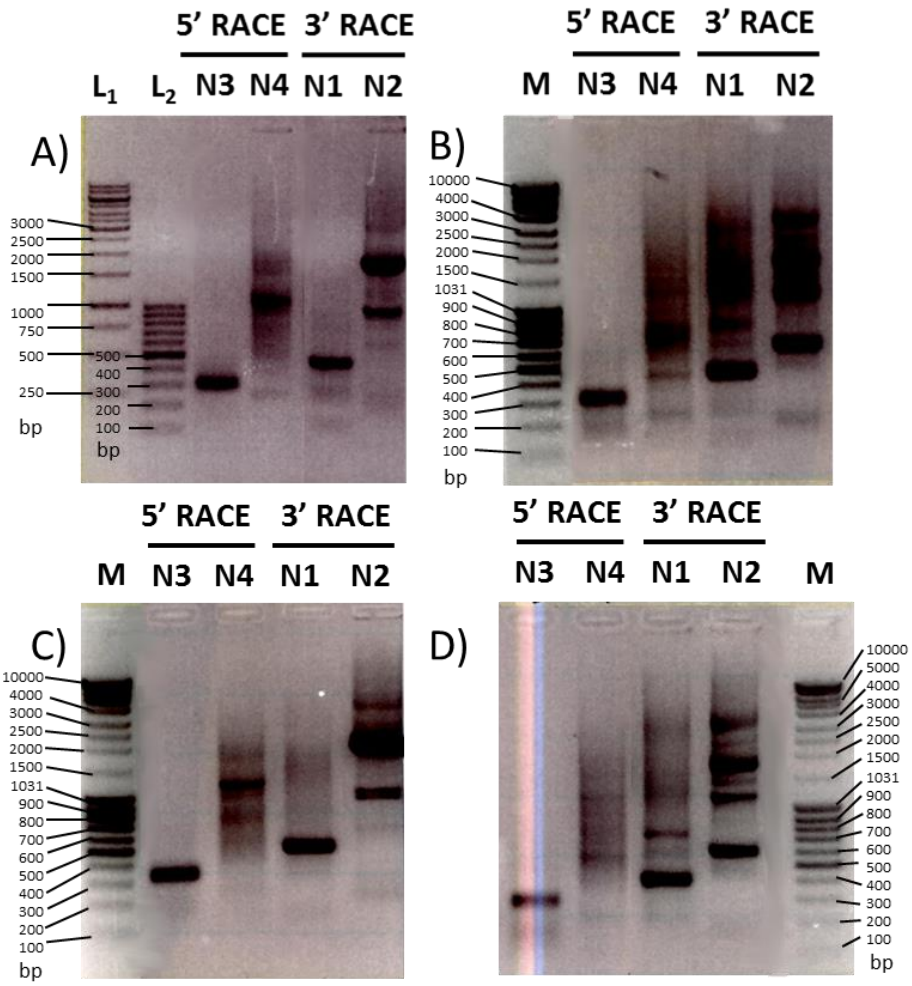

Figure 3.2: Agarose gel photographs showing the nested RACE inserts from A) *M. natalensis* forelimb; B) *M. musculus* forelimb; C) *M. natalensis* head and D) *M. musculus* head. N1 to N4 are the nested RACE primers used in the reaction. The molecular weight markers used were: M) MassRuler™ (Fermentas, USA) L_1) 1kb GeneRuler™ (Fermentas, USA) and L_2) 100bp GeneRuler™ (Fermentas, USA) with the sizes indicated in base pairs (bp).

Table.3.1: The estimated sizes of the RACE inserts

		M. natalensis		*M. musculus*	
		Forelimb	**Head**	**Forelimb**	**Head**
lncMeis2	**N3 5'RACE**	300	600	500	700,400
	N1 3'RACE	450	450	300	300
Meis2	**N4 5'RACE**	1000	1200	Smear, 700	2500, 1250 950, 600
	N2 3'RACE	1750	3000, 2000,1200	800, 450	600

Table 3.2: The number of clones sequenced and analysed for each RACE insert band

		M. natalensis		*M. musculus*	
		Forelimb	**Head**	**Forelimb**	**Head**
lncMeis2	**NGSP3 5'RACE**	6(A)	10(A)	9(A)	8(A)
	NGSP1 3'RACE	10(A)	10(A)	9(A), 5(B)	18(B)
Meis2	**NGSP4 5'RACE**	8(A),10(B)	10(A)	5(A)	9(A)
	NGSP2 3'RACE	5(A),7(B), 7(C)	9 (C), 10(D)	9(A), 6(B), 8(C)	9(A), 4(B),

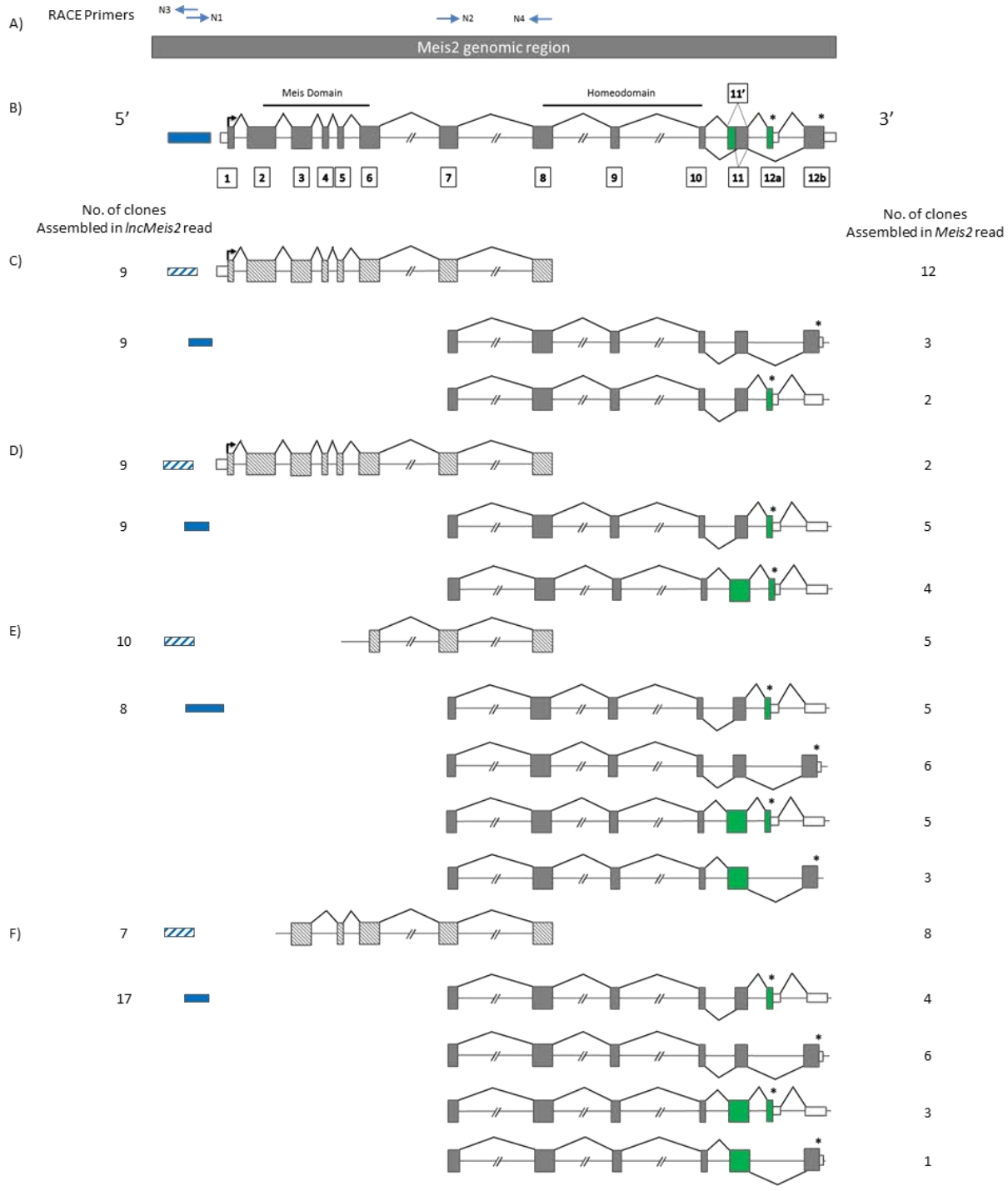

Figure 3.3: Schematic diagram showing transcripts expressed from the *Meis2* locus in *M. natalensis* and the *M. musculus* forelimb and head. A) The *Meis2* genomic region with arrows showing the putative binding sites of the RACE primers N1-N4. B) The position of possible *lncMeis2* transcripts (in blue) and the conserved *Meis2* (in gray) intron/exon structure containing 13 exons (boxed numbers).The regions encoding the *Meis* domain and homeodomain are indicated. There are alternatively spliced exons: 11' and 12a (shown in green), the asterisk (*) indicates stop codons and the white boxes represent the untranslated regions (UTRs). A representation of the assembled contigs for 5' RACE (in striped colour) and 3' RACE (in solid colour) for (C) *M. natalensis* forelimb and (D) head, and for (E) *M. musculus* forelimb and (F) head and the amount of clones used to assemble the read is shown.

3.1.2 *Meis2* contigs

The assembled RACE reads for the NGSP2 and NGSP 4 RACE reads showed that there are three different *Meis2* isoforms present in *M. natalensis* during development (Fig. 3.3 C & D in solid gray), one in the forelimb, one in the head and two shared between the two tissues. The isoforms present in *M. natalensis* were termed *BFMeis2_2* (assembled from four clones), *BHMeis2_3* (four clones), and *BFMeis2_1* (two clones) and *BHMeis2_1* (five clones), respectively. The transcripts *BFMeis2_1* and *BHMeis2_1* have exon 11 and exon 12a in their structural isoform whereas *BFMeis2_2* has exon 11 and exon 12b in its isoform. The *BHMeis2_3* isoform has exon 11' and exon 12a. The 5' RACE read for both the forelimb and head of *M. natalensis* (Fig. 3.3 C & D shown in striped grey) showed conserved structure with the vertebrate *Meis2* mRNA. Twelve clones were used to assemble the 5'RACE read for the limb and two were used to assemble the read from the head.

In *M. musculus*, the 5' Meis2 RACE reads (Fig. 3.3 E & F in striped grey) were truncated at the 5' end. The sequences that were assembled to construct the reads did not extend to exon 1 but halted at exon 6 and at exon 3 in *M. musculus* forelimb and head 5'RACE ready cDNA samples, respectively. The *M. musculus* 3' RACE reads (Fig. 3.3 E & F in solid grey) showed that there are four isoforms present in both the forelimb and the head.

The GSP5/NGSP5 5'RACE reactions were used to verify the integrity of the RACE ready cDNA. The amplification using NGSP5 RACE primer resulted in a in a smear with no prominent insert bands for both the forelimb and the head 5' RACE ready cDNA of *M. musculus* (Fig. 3.4A lanes 1 and 2, respectively) Conversely, the amplification of *M. natalensis* forelimb 5' RACE ready cDNA yielded a prominent band at the expected size (Fig. 3.4B shown with an arrow).

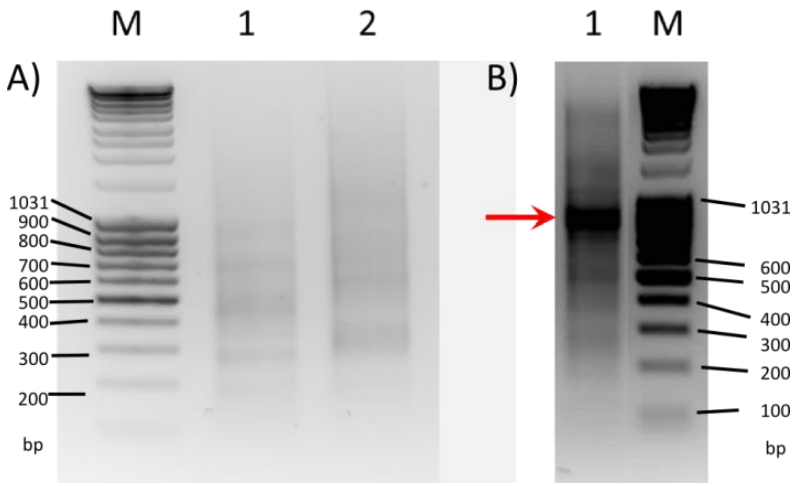

Figure 3.4: Agarose gel photographs of PCR reactions amplified using NGSP5. A) The gel photo of *M. musculus* reactions lane 1 is from forelimb and lane 2 is from the head. B) The gel photograph of the *M. natalensis* from the forelimb, a prominent insert is indicated with an arrow. Lanes M contain the DNA ladder MassRuler ™ (Fermentas, USA) with the sizes indicated in base-pairs (bp).

3.1.3 *lncMeis2* contigs

The RACE primer NGSP1 and NGSP3 were designed to determine if there were any transcripts expressed upstream of *Meis2* in *M. natalensis* and *M. musculus*. The NGSP1 and NGSP3 reads were assembled into *lncMeis2* RACE reads for each tissue (Fig. 3.3 C-F shown in striped and solid blue). There was a single read for each tissue type. Contigs amplified from *M. natalensis* forelimb and head were labelled as *BFlncMeis2* and *BHlncMeis2*, respectively and were approximately 580 nucleotides in length. There was 95% sequence similarity determined over 100% coverage between *BFlncMeis2* and *BHlncMeis2* (Fig A1). Contigs amplified from *M. musculus* forelimb and head were labelled *MFlncMeis2* (approximately 900 nucleotides) and *MHlncMeis2* (approximately 580 nucleotides), respectively. These had a sequence similarity of 98% over a 60% coverage between *MHlncMeis2* and *MFlncMeis2* (Fig A2). These contigs were aligned to the *M. musculus* genome (NCBI37/mm9) via USCS genome browser to determine their position relative to *Meis2* (Fig. 3.7). The transcription start sites (TSS) of the *lncMeis2* contigs (*M. natalensis* and *M. musculus*) initiate approximately 900 nucleotides away from the TSS of any *Meis2* transcripts. All the *lncMeis2* contigs terminated at a region 300 nucleotides from the

transcriptional start site of the UCSC transcripts of *Meis2* (Fig. 3.5 shown in blue) except for *MFlncMeis2* (Fig. 3.5 shown in black), which terminated within an overlapping region of about 130 nucleotides of the 5'UTR of *Meis2*. This was surprising, since extension of the *lncMeis2* transcripts into *Meis2* has not been previously observed.

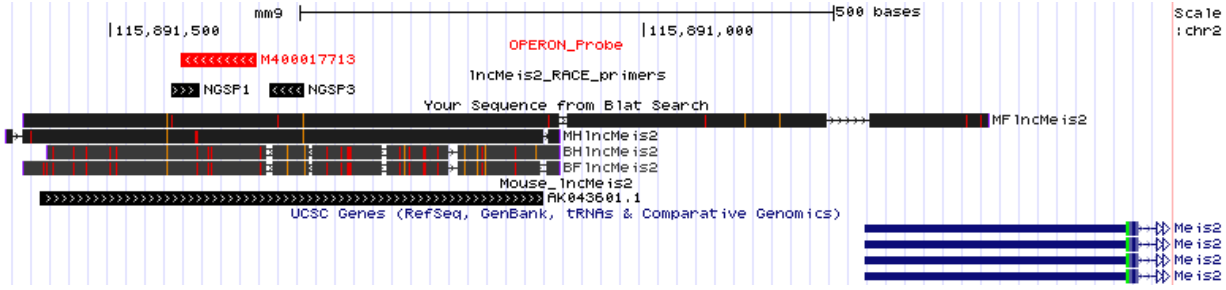

Figure 3.5: A graphical representation of the alignment of the *lncMeis2* contigs to the *M. musculus* genome via USCS genome browser. Custom tracks seen are of the RACE primers NGSP1 and NGSP3, OPERON probe M400017713 (Mason *et al.*, unpublished data) and a custom track for the EST AK043601.1 (mouse ortholog of *lncMeis2*). A custom track was designed for the aligned *lncMeis2* contigs: MFlncMeis2, MHlncMeis2. BHlncMeis2 and BFlncMeis2 (shown in black and grey).

The *lncMeis2* RACE contigs were inspected to determine if they contained the polyadenylation consensus sequence which should occur within 30 nucleotides upstream of the poly-A tail. Figure 3.6 shows the sequences of the *lncMeis2* contigs, *BFlncMeis2*, *BHlncMeis2* and *MHlncMeis2* (Fig. 3.6A) and *MFlncMeis2* (Fig. 3.6B). The poly-A tail and 30 nucleotides upstream of the poly-A tail is indicated (Fig 3.6A & B in brackets). From the sequence data of the *lncMeis2* contigs it was determined that none of the *lncMesi2* contigs contain the polyadenylation consensus sequence within the prescribed region (Fig 3.6).

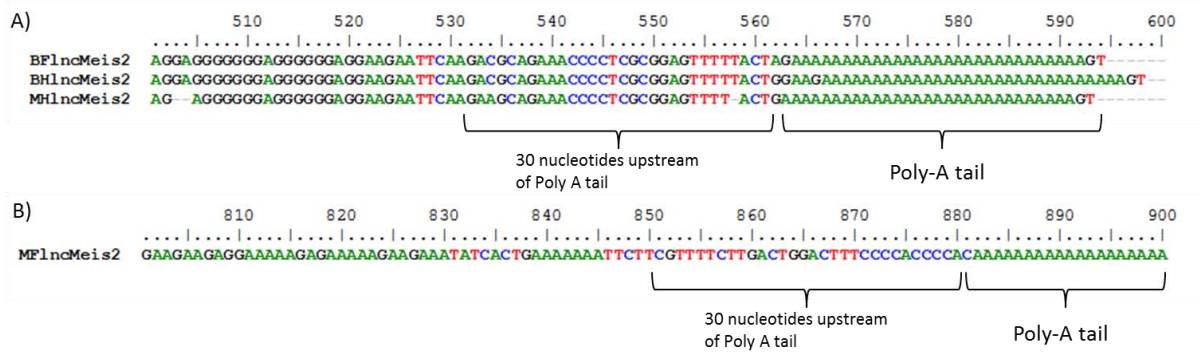

Figure 3.6: The 3' regions of the *lncMeis2* contigs from: A) *M. natalensis* forelimb and head, *BFlncMeis2* and *BHlncMeis2* respectively; *M. musculus* head, *MHlncMeis2* and B) M. musculus forelimb, *MFlncMeis2*. The brackets indicated the regions of the poly-A tail and 30 nucleotides upstream of the poly-A tail.

3.2 RACE Overlap analysis of *lncMeis2* and *Meis2*

A new RACE experiment was designed to amplify a region upstream of where the 3' ends of the *lncMeis2* transcripts terminate. This was done to identify any transcripts present in *M. natalensis* or *M. musculus* tissue that overlap with the 5' UTR of the *Meis2* and to determine if the transcripts extend further into the *Meis2* transcript. This was done as a result of the overlap evident in *MFlncMeis2*.

Single inserts were amplified in the nested 3' RACE PCR using the NGSP7, in *M. natalensis* and *M. musculus.* Both the head (Fig 3.7B) and forelimb tissue (Fig. 3.7A lane 1) of *M. musculus* yielded inserts but only forelimb tissue of *M. natalensis* yielded an insert (Fig. 3.7A lane 2). Table 3.3 lists the estimated sizes of the RACE insert amplified with NGSP7. Table 3.4 lists the number of inserts that were sub-cloned and the number of replicates that were sequenced and analysed.

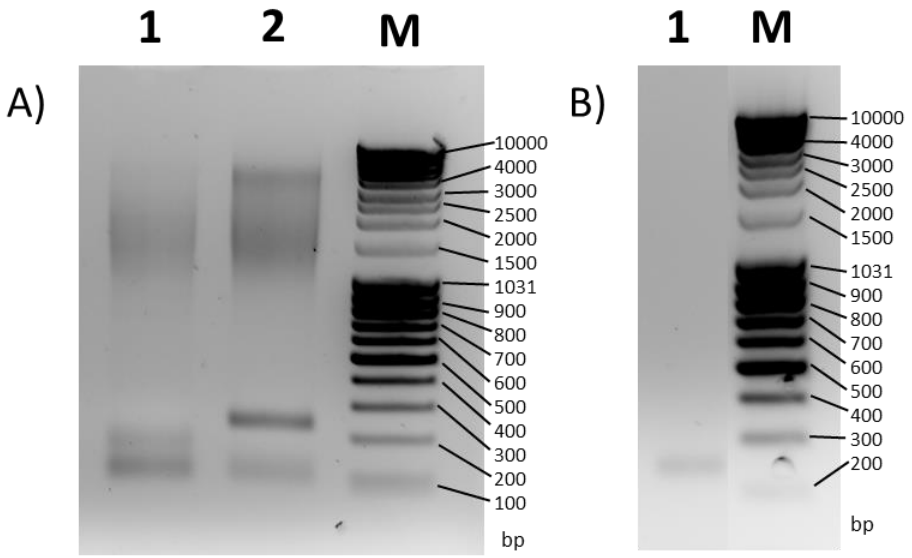

Figure 3.7: Agarose gel photographs of the inserts amplified in the nested RACE PCR using the NGSP7 primer. A) Lane 1contains the insert from mouse forelimb, and lane 2 the insert from *M. natalensis* forelimb. B) Lane 1 contains the insert from *M. musculus* head. Lanes M contain the MassRuler ™ DNA Ladder (Fermentas, USA) with the sizes indicated in base-pairs (bp)

Table 3.3: Estimated sizes of the RACE inserts

		M. natalensis		*M. musculus*	
		Forelimb	Head	Forelimb	Head
lncMeis2	NGSP7 3'RACE	250	250	150	-

Table 3.4: A summary of RACE inserts and replicates that were sub-cloned, sequenced and analysed

		M. natalensis		*M. musculus*	
		Forelimb	Head	Forelimb	Head
lncMeis2	NGSP7 3'RACE	5 (B)	5 (A)	5 (A)	-

The NGSP7 reads were aligned to the genome of M musculus with the *lncMeis2* contigs (Fig 3.8). As seen in figure 3.8 the *M. musculus* NGSP7 transcript, MF3ssN7A_1, aligned and terminated to the same region as *MFlncMeis2*. The *M. natalensis* transcripts BF3ssN7B_1 and BH3ssN7A_1 both aligned to the same region where *MFlncMeis2* is located but the two *M. natalensis* transcripts extended further into the 5' UTR of *Meis2* than *MFlncMeis2* (Fig. 3.8)

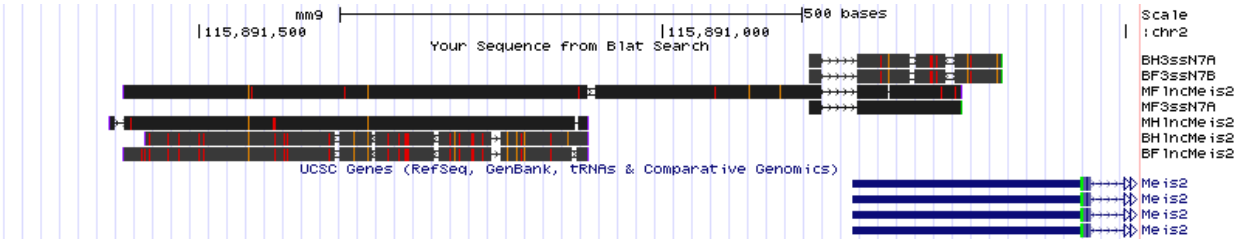

Figure 3.8: A UCSC genome alignment of the lncMeis2 contigs and the NGSP7 reads (shown in black) in relation to *Meis2* transcripts (shown in blue).

An investigation of nucleotide sequences of the genomic regions where the NGSP1 and NGSP7 3' RACE inserts for *M. natalensis* and *M musculus* terminates revealed that all the 3' transcripts ended in an adenine rich region (Fig. 3.9 highlighted in red). These adenine rich regions are conserved within vertebrates as seen by the MultiZ alignments of the different vertebrate genomes (Fig. 3. 9) (Blanchette *et al.*, 2004).

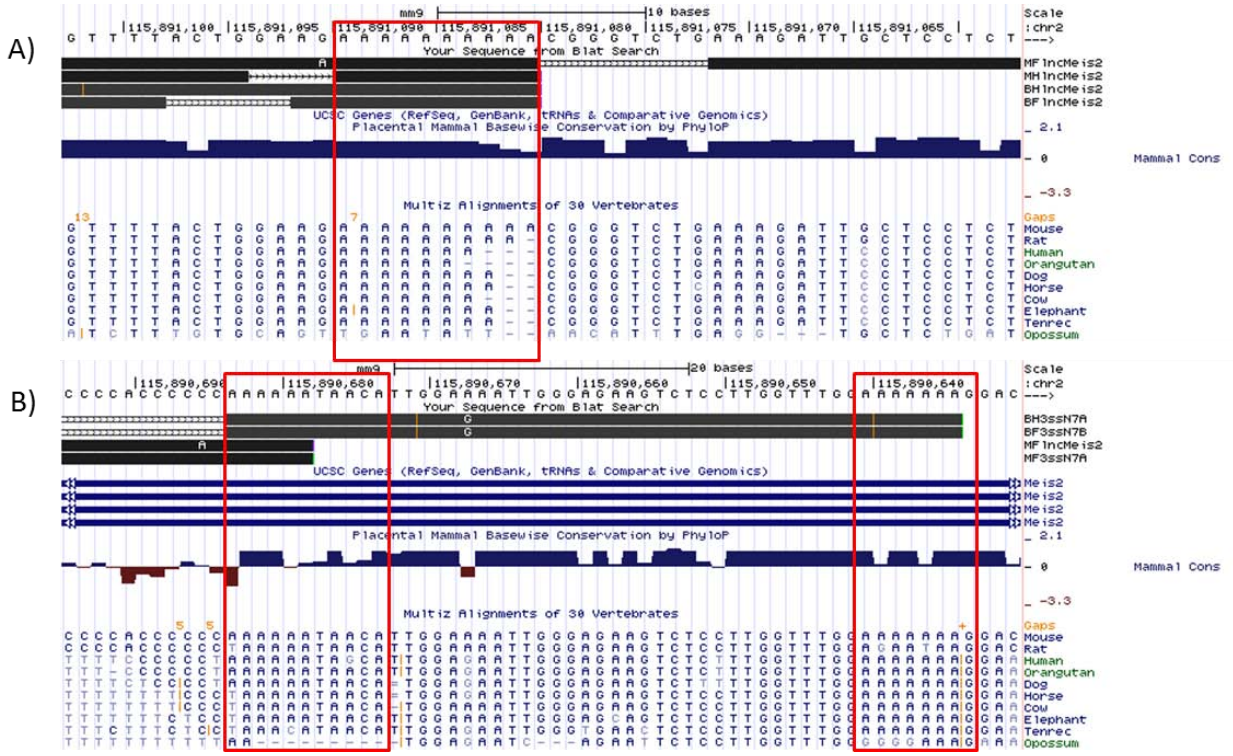

Figure 3.9: A USCS genome browser output showing the alignments of the A) *lncMeis2* reads, B) NGSP7 inserts and the MFlncMeis2 contig (shown in black) aligned to the mouse genome with the track for USCS genes present (*Meis2* transcripts shown in blue). The red boxes enclose the adenine rich regions in which the *lncMeis2* and the NGSP7 insert terminate. The blue bar graph indicates the conservation of the nucleotide sequences across the mammalian lineage. The alignments of other vertebrate sequences to the mouse are shown at the bottom of each alignment (A) and (B).

3.3 Ascertaining *Meis2* overlap via conventional PCR

To address the concern of artefactual sequences, a conventional PCR experiment was designed to amplify any mRNA transcripts that span from *lncMeis2* to within the 5' UTR of *Meis2* in both *M. natalensis* and *M. musculus*. During the cDNA synthesis random hexamers were used instead of oligo dT to amplify the mRNA. Random hexamers are oligonucleotides, six nucleotides in length and contain randomly generated sequences. These oligonucleotides are able to bind anywhere within the mRNA transcript in an unbiased manner. The cDNA was then amplified using Meis2 Overlap (MO) primers: the forward primer was designed to bind within *lncMeis2* and the reverse primer was designed to bind within the 5' UTR of *Meis2*. Primers that specifically amplified *Meis2* and *lncMeis2* only were used as controls for the experiment. Figure 3.10 shows *lncMeis2* and *Meis2* were present in all tissues. The MO

primers amplified an insert approximately 650 nucleotides in *M. natalensis* tissue and an insert approximately 550 nucleotides in *M. musculus* (Fig 3.10). The inserts were sequenced and aligned to the *M. musculus* genome (Fig. 3.11). The MO inserts mapped to the region overlapping *lncMeis2* and the 5'UTR of *Meis2* (Fig 3.11A). There are also three EST which map to the overlap region (Fig. 3.11B).

Figure 3.11C shows a graph indicating the abundance of RNA polymerase II (RNAP II) bound to the genome in the region of *lncMeis2* and *Meis2*. These data were obtained from the LICR/ENCODE transcription factor binding site track by using chromatin immunoprecipatation to identify binding regions of specific transcription factors. This particular track shows the prevalence of RNAP II on *M. musculus* genome using mouse embryonal fibroblast (MEF) cells. By using a gene annotation online application BioGPS (Wu *et al.*, 2009) and the MOE430 gene atlas data set (Lattin *et al.*, 2008), it was confirmed that the MEF cells express *Meis2*. The peak of RNAP II occupancy extends from the TSS of *lncMeis2* into the 5' UTR of *Meis2.* This suggests that the transcription of *Meis2* initiates at the TSS of *lncMeis2*. There are 2 Ensembl transcripts ENSMUST00000110907 and ENSMUST0000010908 (Fig. 3.13D) which have 5' UTRs that extend over the region onto which *lncMeis2* maps.

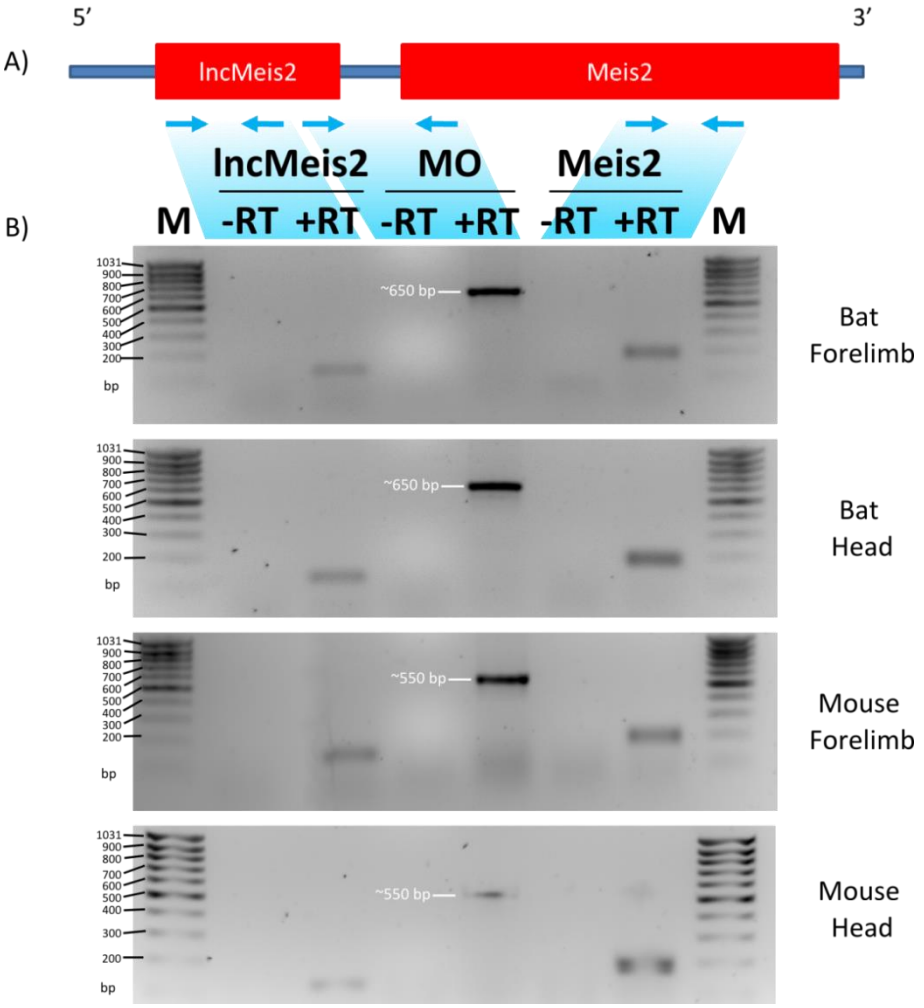

Figure 3.10: Gel photos from the *Meis2* overlap experiment. A) A schematic diagram indicating the putative binding sites of the primers. B) Gel photos of the PCR reactions from each tissue. Lane M contains the MassRuler ™ DNA Ladder (Fermentas, USA) with the sizes indicated in base-pairs (bp).

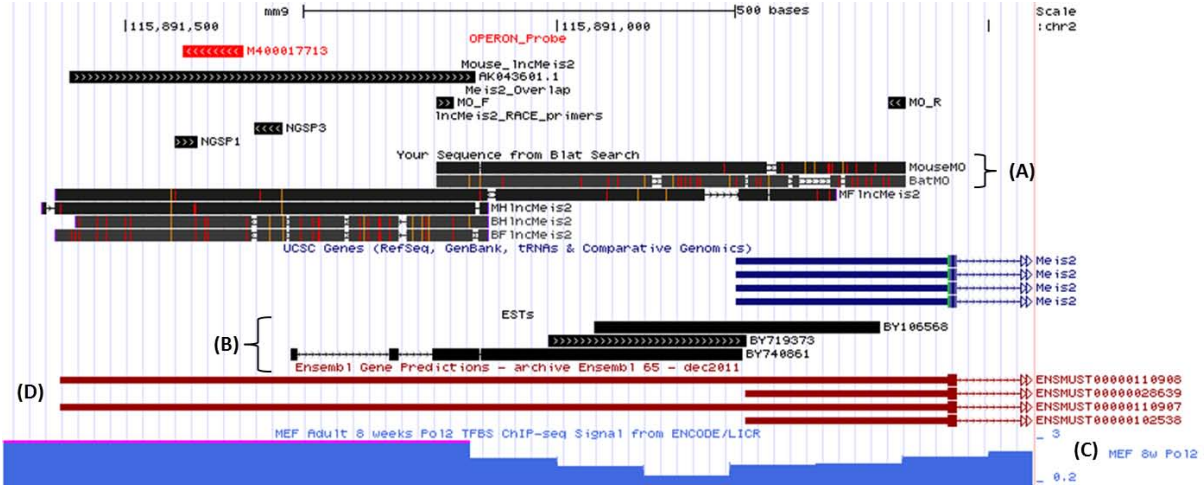

Figure 3.11: A UCSC genome alignment of *lncMeis2* contigs and *Meis2* overlap contigs with the *M. musculus* genome indicating (A) the *Meis2* overlap contigs, (B) the position of the three ESTs overlap ping *lncMeis2* and *Meis2*, (C) the track indicating the presence of RNA polymerase II bound to the genome in the MEF cell line and (D) Ensembl transcripts in red.

3.4 The evolution of *Meis2* in bats

One of the main facets of evo-devo is the comparison of developmental genes across species (Carroll, 2005). A simple method to test adaptive evolution is to determine the selection model acting upon a given locus between two or more species. This can be done by comparative analysis of the protein coding region of the gene and examining the rate of nucleotide substitutions that occur over evolutionary time (Kimura, 1980). There are two types of nucleotide substitutions within the coding domain of a gene: synonymous substitution, where the nucleotide substitution does not alter the codon sequence or amino acid, and non-synonymous substitution where the nucleotide change results in a different amino acid.

3.4.1 Selection analysis of *Meis2* in bats

The *M. natalensis Meis2* transcripts assembled in section 3.2 were used in this analysis. There are only two bat species which have *Meis2* sequences in GenBank, *Myotis lucifugus* and *Myotis brandtii*. The *Meis2* transcripts for these species are predicted sequences that were assembled computationally, not through *in vitro* studies (Birney *et al.*, 2004). The predicted CDS transcripts for each *Meis2* transcript was obtained from GenBank (Benson *et al.*, 2009) and used in this study. Other predicted CDS transcripts were also obtained for, *Equus caballus* (horse) and annotated *Meis2* CDS transcripts were obtained for: *Homo sapiens* (human) and *M. musculus* (house mouse).

The phylogenetic software package Molecular Evolutionary Genetics Analysis (version 5.0.2) (Tamura *et al.*, 2011) was used to perform selection analysis on the *Meis2* CDS transcripts. The Codon-based Z-test for positive selection and the Nei-gojobori were chosen and the bootstrap value was set to 500 replicates.

Table 3.5 contains the transcript name and accession number of the grouped Meis2 transcripts that was used in the analysis. The transcripts were divided into three groups according to

intron/exon structure: group 1 contained the transcripts which have exon 11 and exon 12a; group 2 has exon 11 and exon 12b; and group 3 transcripts have exon 11' and exon 12a. Table 3.6 contains the test statistic (*p*-values) of the Z-tests that were performed on the *Meis2* coding domain sequences. A *p*-value of less than 0.05 for the Z-test would mean the transcript is under positive selection. But as seen in table 3.6 all the *p*-values have a value of 1, the Z-test show in favour of the null hypothesis of neutral selection (the number of non-synonymous changes = the number of synonymous changes) and reject the alternative of positive selection.

Table 3.5: The names and accession numbers of the grouped Meis2 transcripts used in the selection analysis

Group	Meis2 Transcript Name	Accession Number
1	mnFMeis2_1*	
	mlMeis2_x4	XM_006087500
	Horse Meis2_x6	XM_005603163
	Human Meis2_b	NM_170674.4
	Mouse Meis2_v1	NM_001136072.2
2	mnFMeis2_2*	
	mlMeis2_x2	XM_006087498
	mbMeis2_x3	XM_005879353
	Horse Meis2_x2	XM_001503627
	Human Meis2_d	NM_170675.4
	Mouse Meis2_v2	NM_010825.3
3	mnMeis2_3*	
	mlMeis2_x3	XM_006087499
	mbMeis2_x5	XM_005879355
	Horse Meis2_x4	XM_005603162
	Human Meis2_a	NM_170677.4
	Mouse Meis2_v3	NM_001159567.1

mn – *M. natalensis,* ml – *M. lucifugus,* mb – *M. brandtii* *the *M.natalensis* transcripts do not have accession numbers

Table 3.6: The p-values of a codon based Z-test on positive selection of the *Meis2* coding region for two bat species and other vertebrates

Group 1

	Meis2 Transcript Name	1	2	3	4	5
1	mnFMeis2_1					
2	mlMeis2_x4	1.000				
3	Horse Meis2_6	1.000	1.000			
4	Human Meis2_b	1.000	1.000	1.000		
5	Mouse Meis2_v1	1.000	1.000	1.000	1.000	

Group 2

	Meis2 Transcript Name	1	2	3	4	5	6
1	mnFMeis2_2						
2	mlMeis2_x2	1.000					
3	mbMeis2_x3	1.000	1.000				
4	Horse Meis2_x2	1.000	1.000	1.000			
5	Human Meis2_d	1.000	1.000	1.000	1.000		
6	Mouse Meis2_v2	1.000	1.000	1.000	1.000	1.000	

Group 3

	Meis2 Transcript Name	1	2	3	4	5	6
1	mnMeis2_3						
2	mlMeis2_x3	1.000					
3	mbMeis2_x5	1.000	1.000				
4	Horse Meis2_x4	1.000	1.000	1.000			
5	Human Meis2_a	1.000	1.000	1.000	1.000		
6	Mouse Meis2_v3	1.000	1.000	1.000	1.000	1.000	

Chapter 4. Discussion

To date much is known about the diverse roles *Meis2* plays in embryonic development. The expression of *Meis2* is well documented in the vertebrate limb however, little is known of the role Meis2 plays in the formation of the bat wing. In this study, Random Amplification of cDNA Ends (RACE) polymerase chain reaction was used to determine the transcripts which are expressed from the *Meis2* locus in the Natal Long-fingered bat (*Miniopterus natalensis*) and the mouse (*Mus musculus*).

The results of the analysis identified three transcripts variants (isoforms) of *Meis2* in the bat, two isoforms were identified in the forelimb and two in the head, one isoform was shared between the tissues. The contigs amplified were labelled *BFMeis2_1*, *BFMeis2_2*, *BHMeis2_1* and *BHMeis2_3*. The variability of the isoforms comes from the different combinations of the alternative splice sites at the 3' end of the *Meis2* transcripts (Irimia *et al.*, 2011; Sánchez-Guardado *et al.*, 2011). The alternative splice sites are conserved among vertebrates, therefore the different isoforms are conserved as well (Irimia *et al.*, 2011). The 3'region of the *Meis* isoforms encode for different interacting domains, e.g. transcriptional activation domains and regulatory modules that alter protein transcriptional activity and response to cell signalling (Huang *et al.*, 2005; Irimia *et al.*, 2011; Yang *et al.*, 2000). Changing the combination of the isoforms could allow subtle changes in the control MEIS proteins have on transcriptional activity, e.g. MEIS proteins have the propensity to enhance the cell proliferation by activating cell cycle genes (Bessa *et al.*, 2008; Heine *et al.*, 2008a; Huang *et al.*, 2005; Irimia *et al.*, 2011; Sánchez-Guardado *et al.*, 2011). Although this hints at the function of Meis2 isoforms, these experiments were tested in the eyes and cerebral tissue of zebrafish and mice, and not in any limb tissue (Bessa *et al.*, 2008; Heine *et al.*, 2008a; Huang *et al.*, 2005).

Analysis of the amplification of *Meis2* transcripts from *M. musculus* showed four isoforms present in both the head and forelimb tissue. These results were determined from 3' RACE results as 5' ends of the 5'RACE reads were truncated. The truncations of the 5' RACE read were due to the degraded RNA from *M. musculus* which was used to synthesize the RACE ready cDNA.

The analysis of the *lncMeis2* transcripts showed two contigs from *M. natalensis* (one from the head, *BHlncMeis2* and one from the forelimb, *BFlncMeis2*) but because they share almost 100% sequence similarity they could be considered the same transcript. Two contigs were also amplified from *M. musculus* tissue *MHllncMeis2* and *MFlncMeis2*. There was a sequence similarity of 98% over 60% coverage between *MHlncMeis2* and *MFlncMeis2*. All the *lncMeis2* contigs mapped to region upstream of annotated Meis2 transcripts when aligned to the M. musculus genome. The *M. natalensis lncMeis2* contigs and *MHlncMeis2* showed no overlap with the *Meis2* transcripts except *MFlncMeis2* which overlap with the 5'UTR of Meis2. Through further RACE and PCR analysis it was determined that the *lncMeis2* transcripts from both *M. natalensis* and *M. musculus* overlap with *Meis*. The overlap of *lncMeis2* and *Meis2* has previously not been observed. Further analysis of the *lncMeis2/Meis2* overlap has shown that *lncMeis2* could be an artefact transcript or a low-level upstream transcript of unknown significance.

All the *lncMeis2* contigs do not contain the polyadenylation consensus sequence, which is pivotal to the processing of nascent mRNA transcripts during transcription (Das *et al.*, 2001). The lack of the polyadenylation consensus in all the *lncMeis2* transcripts suggests that these transcripts are not valid mRNAs and the poly-A tails are artefacts from cDNA synthesis (Proudfoot and Brownlee, 1976).

The *lncMeis2* contigs from both *M. natalensis* and *M. musculus* terminated within conserved adenine rich regions upstream of the *Meis2* 5'UTR. The *Meis2* overlap analysis demonstrated

that when random hexamers are used to synthesize the cDNA, a transcript can be amplified that overlaps *lncMeis2* and *Meis2*. If this represents a legitimate, overlapping transcript then it contains internal adenine rich region, the same regions where the *lncMeis2* transcripts terminate. These poly-adenine regions allow for the mispriming of oligo dT, a primer which is used to amplify mRNA transcripts (Das *et al.*, 2001; Nam *et al.*, 2002). The mispriming of oligo dT at adenine rich regions causes the truncation of a full length transcript at the adenine region. These internal adenine rich regions mimic the poly-A tail of nascent mRNAs. The initiation of amplification from these regions, by the polymerase, results in truncation of the transcript (Nam *et al.*, 2002; Rampias *et al.*, 2012). The truncated transcript would not have a valid poly-A tail therefore the transcript would not have the polyadenylation sequence which would signal RNA polymerase to add the valid poly-A tail (Das *et al.*, 2001; Proudfoot and Brownlee, 1976).

For correct initiation of transcription, the initiation complex, which contains RNAP II, binds 30 to 50 nucleotides upstream of the TSS before scanning for the TSS to begin RNA synthesis (Mayer *et al.*, 2010). Here we see the presence of RNAP II approximately 950 nucleotides away from the TSS of *Meis2* and closer to the TSS of *lncMeis2*. This evidence suggests that initiation of *Meis2* starts at the TSS of *lncMeis2* and that *lncMeis2* is a part of the 5' UTR of *Meis2*, and therefore not a valid mRNA transcript.

There are two *Meis2* transcripts for *M. musculus* present in the Ensembl database which have overlapping 5'UTRs with *lncMeis.* However, these transcripts were assembled from annotated transcript data and ESTs from mouse and human databases (Birney *et al.*, 2004). There are no transcripts described that overlap with *Meis2* and *lncMeis2* that were transcribed *in vitro.* But evidence suggested here indicates there is a the presence of a transcript with an overlapping 5'UTR *in vitro.*

The implications of oligo dT mispriming due to internal adenine regions, as stated previously in Nam *et al,* 2002, can cause problems in gene identification as the synthesis of full length cDNAs will be interrupted. Truncation of the gene makes the gene assembly difficult. Mispriming of oligo dT can lead to problems in gene identification and gene expression analysis (Nam *et al.*, 2002). The truncated transcripts that are generated from the mispriming during cDNA can increase the background and give false positives; the truncated transcripts can disperse the signal from a single transcript, among multiple detected transcripts. (Nam *et al.*, 2002)

Selection analysis of the coding domain sequence (CDS) of *M. natalensis Meis2* transcripts and the predicted transcripts from *M. lucifugus* and *M. brandtii* has shown that *Meis2* in bats is not under positive selection. *Meis2* CDS is highly conserved, the first 10 exons contain the MEIS domain and the homeodomain, which are vital to MEIS2 proteins function as a transcription factor (Bürglin, 1998; Irimia *et al.*, 2011). The only variable region of *Meis2* transcript is the last three exons, 11, 12a and 12b, where the alternative splicing occurs. The 3' region contains several domains which have been involved in transcriptional activation and cell signalling (Bessa *et al.*, 2008; Heine *et al.*, 2008a; Huang *et al.*, 2005; Irimia *et al.*, 2011; Sánchez-Guardado *et al.*, 2011). If the coding sequence of *Meis2* is not responsible for the variation seen in bat limbs then it could the regulation of Meis2 expression that leads to the morphological differences (Carroll, 2005).

I acknowledge the limitations of this study: the number of clones that contained RACE products. These could have been increased to increase the chances of isolating more isoforms from *M. natalensis*. It is possible that there are more isoforms present in *M. natalensis* that I did not isolate. The truncated *M. musculus* 5' RACE reads was due to degraded RNA that was extracted from the head and forelimb. Due to limited embryos and prioritization of another study the extraction and 5'RACEready cDNA synthesis could not be repeated.

In the positive selection analysis, more bat *Meis2* transcripts could have been incorporated, but there have not been any recent predicted or annotated *Meis2* transcripts for any bat species.

Chapter 5. Concluding Remarks

Here, evidence is provided to suggest three isoforms of *Meis2* present in the developing *M. natalensis* embryo, and two of the isoforms are found in the developing limb bud. These isoforms are also under no positive selection and are highly conserved. Evidence is also provided to suggest that *lncMeis2* is not a valid lncRNA and is a truncated transcript artefact.

The Meis2 isoforms have been shown to illicit transcriptional activation especially in cell proliferation (Bessa *et al.*, 2008; Heine *et al.*, 2008b; Irimia *et al.*, 2011), this same mechanism could be responsible for the outgrowth of the limb bud. In terms of *M. natalensis*, *Meis2* was shown to be expression in the retained interdigital webbing of the developing autopod. *Meis2* could be aiding the proliferation of ectodermal cells of the webbing which would later be the membrane of the wing.

The *lncMeis2* transcript was determined to be a false positive. It was generated from the mispriming of oligo dT to internal adenines within the 5'UTR of *Meis2*. In the transcriptome analysis of the bat wing formation *lncMeis2* shown a significantly differential signal than *Meis2* (Mason, Msc Thesis 2009). Gene expression analysis was also used to confirm the expression of *lncMeis2* in the developing wing of *M. natalensis*. The result of the analysis showed lncMEis2 expression was significantly higher than *Meis2* (Mason and Illing, unpublished data). In both these studies oligo dT was used to generate the cDNA, which as suggested, created a false *lncMeis2* transcript at high levels which skewed the results in favour of *lncMeis2*.

Further work needs to be done to determine whether there is in fact a lncMeis2 transcript present in bat or mice. The use of Northern blot analysis is recommend as there is no need for the generation of cDNA or the use of oligo dT, total RNA is used. Targeted RACE can also be used where adaptor oligos can be ligated to the 3' ends of the total RNA and in conjunction with the gene specific primer, can amplify the target without the need for oligo dT.

In conclusion, I have shown that there are three isoforms present in the developing bat (*M. natalensis*) and that the once thought *lncMeis2* is not a valid mRNA transcript but an artefact of cDNA synthesis via oligo dT.

Chapter 6. References

Abu-Abed, S., MacLean, G., Fraulob, V., Chambon, P., Petkovich, M., Dolle, P., 2002. Differential expression of the retinoic acid- metabolizing enzymes CYP26A1 and CYP26B1 during murine organogenesis. Mech. Dev. 173–177.

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science (80-.). 252, 1651–6.

Amaral, P.P., Mattick, J.S., 2008. Noncoding RNA in development. Mamm. Genome 19, 454–92.

Anno, Y.N., Chatagnon, A., Samarut, E., Poch, O., Laudet, V., Benoit, G., Lecompte, O., 2011. Genome-wide in Silico Identification of New Conserved and Functional Retinoic Acid Receptor Response Elements (Direct Repeats Separated by 5 bp) * □. J. Biol. Chem. 286, 33322–33334.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2009. GenBank. Nucleic Acids Res. 37, D26–31.

Bessa, J., Tavares, M.J., Santos, J., Kikuta, H., Laplante, M., Becker, T.S., Gómez-Skarmeta, J.L., Casares, F., 2008. meis1 regulates cyclin D1 and c-myc expression, and controls the proliferation of the multipotent cells in the early developing zebrafish eye. Development 135, 799–803.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res. 14, 988–95.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., Haussler, D., Miller, W., 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14, 708–15.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. Chapter 19, Unit 19.10.1–21.

Bolker, J.A., 2000. Modularity in Development and Why It Matters to Evo-Devo. Integr. Comp. Biol. 40, 770–776.

Bürglin, T.R., 1997. Analysis of TALE superclass homeobox genes (MEIS , PBC , KNOX , Iroquois , TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 25, 4173–4180.

Bürglin, T.R., 1998. The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? Dev. Genes Evol. 208, 113–6.

Capdevila, J., Izpisua-Belmonte, J.C., 2001. Patterning Mechanisms Controlling Vertebrate Limb Development. Gene Expr. 87–132.

Capdevila, J., Tsukui, T., Zappavigna, V., Izpisu, J.C., Jolla, L., Expression, G., 1999. Control of Vertebrate Limb Outgrowth by the Proximal Factor Meis2 and Distal Antagonism of BMPs by Gremlin. Cell 4, 839–849.

Carroll, S.B., 2005. Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kindgdom, First. ed. W.W. Norton & Company, New York.

Carroll, S.B., Grenier, J.K., Weatherbee, S.D., 2001. From DNA to diversity: molecular genetics and the evolution of animal design, First. ed. Oxford: Blackwell Publishing, Malden, Mass.

Casella, J.F., Casella, S.J., Hollands, J.A., Caldwell, J.E., Cooper, J.A., 1989. Isolation and Characterization of cDNA Encoding the α Subunit of Cap Z(36/32), an Actin-Capping Protein from the Z Line of Skeletal Muscle. PNAS 86, 5800–5804.

Chambon, P., 1996. A decade of molecular biology of retinoic acid receptors. FASEB J. 10, 940–54.

Chen, C., Cretekos, C.J., Iv, J.R., Behringer, R.R., 2005. Hoxd13 expression in the developing limbs of the short-tailed fruit bat , Carollia perspicillata. Development 141, 130 –141.

Cooper, K.L., Hu, J.K., Berge, D., Fernandez-teran, M., Ros, M.A., Tabin, C.J., 2011. Initiation of Proximal-Distal by Signals and Growth. Science (80-.). 332, 1083–1086.

Cretekos, C.J., Wang, Y., Green, E.D., Cooper, K.L., Tabin, C.J., Comparative, N., Program, S., 2008. Regulatory divergence modifies limb length between mammals Regulatory divergence modifies limb length between mammals. Genes Dev. 141–151.

Cretekos, C.J., Weatherbee, S.D., Chen, C., Badwaik, N.K., Niswander, L., Behringer, R.R., Rasweiler, J.J., 2005. Embryonic Staging System for the Short-Tailed Fruit Bat , Carollia perspicillata , a Model Organism for the Mammalian Order Chiroptera , Based Upon Timed Pregnancies in Captive-Bred Animals. New York 721–738.

Cunningham, T.J., Chatzi, C., Sandell, L.L., Trainor, P.A., Duester, G., 2011. Rdh10 Mutants Deficient in Limb Field Retinoic Acid Signaling Exhibit Normal Limb Patterning but Display Interdigital Webbing. Dev. Dyn. 1142–1150.

Cunningham, T.J., Zhao, X., Sandell, L.L., Evans, S.M., Trainor, P. a, Duester, G., 2013. Antagonism between retinoic acid and fibroblast growth factor signaling during limb development. Cell Rep. 3, 1503–11.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life., First. ed, Notes and Records of the Royal Society. John Murray, Albemarle Street, London.

Das, M., Harvey, I., Chu, L.L., Sinha, M., Pelletier, J., Diemen, C.C. Van, Postma, D.S., Vonk, J.M., Bruinenberg, M., Schouten, J.P., Boezen, H.M., Hourfar, M.K., Roth, W.K.,

Seifried, E., Schmidt, M., 2001. Full-length cDNAs : more than just reaching the ends. Physiol Genomics 57–80.

Davis, A.P., Witte, D.P., Hsieh-Li, H.M., Potter, S.S., Capecchi, M.R., 1995. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. Nature 375, 791–5.

Donovan, M., Olofsson, B., Gustafson, A.L., Dencker, L., Eriksson, U., 1995. The cellular retinoic acid binding proteins. J. Steriod Biochem. Mol. Biol. 53, 459–465.

Duboule, D., Dolle, P., 1989. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J. 8, 1497–1505.

Duester, G., 2008. Retinoic Acid Synthesis and Signaling during Early Organogenesis. Cell 921–931.

Frohman, M.A., Dush, M.K., Martin, G.R., 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. PNAS 85, 8998–9002.

Fromental-Ramain, C., Warot, X., Messadecq, N., Lemeur, M., Dollé, P., Chambon, P., 1996. Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod 3011, 2997–3011.

Fujita, P. a., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T.R., Giardine, B.M., Harte, R.A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R.M., Learned, K., Li, C.H., Meyer, L.R., Pohl, A., Raney, B.J., Rosenbloom, K.R., Smith, K.E., Haussler, D., Kent, W.J., 2010. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 1–7.

Fujita, P., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T.R., Giardine, B.M., Harte, R. a, Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R.M., Learned, K., Li, C.H., Meyer, L.R., Pohl, A., Raney, B.J., Rosenbloom, K.R., Smith, K.E., Haussler, D., Kent, W.J., 2011. The UCSC Genome Browser database: update 2011. Nucleic Acids Res. 39, D876–82.

Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-Content Evolution in Mammalian Genomes : The Biased Gene Conversion Hypothesis. Genetics 907–911.

Gehring, W., Affolter, M., Burglin, T., 1994. Homeodomain proteins. Annu Rev Biochem 487–526.

Goff, D.J., Tabin, C.J., 1997. Analysis of Hoxd-13 and Hoxd-11 misexpression in chick limb buds reveals that Hox genes affect both bone condensation and growth. Analysis 636, 627–636.

Gómez-Skarmeta, J.L., Lenhard, B., Becker, T.S., 2006. New technologies, new findings, and new concepts in the study of vertebrate cis-regulatory sequences. Dev. Dyn. 235, 870–85.

Gonzalez-Crespo, S., Abu-Shaar, M., Torres, M., Martinez-A, C., Mann, R., Morata, G., 1998. Antagonism between extradenticle function and Hedgehog signalling in the developing limb. Nature 196–200.

Goodman, C.S., Coughlin, B.C., 2000. The evolution of evo-devo biology. PNAS 97, 4424–4425.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L., Lander, E.S., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–7.

Guttman, M., Rinn, J.L., 2012. Modular regulatory principles of large non-coding RNAs. Nature 482, 339–346.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleis Acid Symp. Ser. 41, 95–98.

Hartwell, L.H., Hood, L., Goldberg, M.L., Reynolds, A.E., Silver, L., Veres, R., 2008. Genetics : From Genes to Genomes, Thrid. ed. McGraw-Hill, New York.

Heine, P., Dohle, E., Brien, K.B., Engelkamp, D., Schulte, D., 2008a. Evidence for an evolutionary conserved role of homothorax / Meis1 / 2 during vertebrate retina development. Development 811, 805–811.

Heine, P., Dohle, E., Bumsted-O'Brien, K., Engelkamp, D., Schulte, D., 2008b. Evidence for an evolutionary conserved role of homothorax/Meis1/2 during vertebrate retina development. Development 135, 805–11.

Hockman, D., Cretekos, C.J., Mason, M.K., Behringer, R.R., Jacobs, D.S., Illing, N., 2008. A second wave of Sonic hedgehog expression during the development of the bat limb. Spring.

Hockman, D., Mason, M.K., Jacobs, D.S., Illing, N., 2009. The Role of Early Development in Mammalian Limb Diversification : A Descriptive Comparison of Early Limb Development Between the Natal Long-Fingered Bat (Miniopterus natalensis) and the Mouse (Mus musculus). Dev. Dyn. 965–979.

Hsu, F., Kent, W.J., Clawson, H., Kuhn, R.M., Diekhans, M., Haussler, D., 2006. The UCSC Known Genes. Bioinformatics 22, 1036–46.

Huang, H., Rastegar, M., Bodner, C., Goh, S.-L., Rambaldi, I., Featherstone, M., 2005. MEIS C termini harbor transcriptional activation domains that respond to cell signaling. J. Biol. Chem. 280, 10119–27.

Huang, X., 1991. A Contig Assembly Program Based on Sensitive Detection of Fragment Overlaps. Genomics.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R.,

Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., Clamp, M., 2002. The Ensembl genome database project. Nucleic Acids Res. 30, 38–41.

Irimia, M., Maeso, I., Burguera, D., Hidalgo-Sánchez, M., Puelles, L., Roy, S.W., Garcia-Fernàndez, J., Ferran, J.L., 2011. Contrasting 5' and 3' evolutionary histories and frequent evolutionary convergence in Meis/hth gene structures. Genome Biol. Evol. 3, 551–64.

Jacobs, D., Cotterill, F.P.D., Taylor, P.J., Griffin, M., 2008. Miniopterus natalensis. IUCN Red List of Threaten Species. URL http://www.iucnredlist.org/details/44862/0 (accessed 6.1.14).

Kapranov, P., Cheng, J., Dike, S., Nix, D. a, Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., Gingeras, T.R., 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484–8.

Kent, W.J., 2002. BLAT −− The BLAST-Like Alignment Tool BLAT — The BLAST-Like Alignment Tool. Genome Res. 656–664.

Kimura, M., 1980. A Simple Method for Estimating Evolutionary Rate of Base Substitutions Through Comparative Studies of Nucleotide Sequences. J. Mol. Evol. 16, 111–120.

Kiyosawa, H., Mise, N., Iwase, S., Hayashizaki, Y., Abe, K., 2005. Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res. 15, 463–74.

Klasens, B.I.F., Huthoff, H.T., Das, A.T., Jeeninga, R.E., Berkhout, B., 1999. The effect of template RNA structure on elongation by HIV-1 reverse transcriptase. Biochim. Biophys. Acta - Gene Struct. Expr. 1444, 355–370.

Kmita, M., Logan, M., Tabin, C.J., Duboule, D., Tarchini, B., Za, J., 2005. Early developmental arrest of mammalian limbs lacking HoxA / HoxD gene function. Nature 435, 1113–1116.

Lattin, J.E., Schroder, K., Su, A.I., Walker, J.R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C.K., Hume, D.A., Kellie, S., Sweet, M.J., 2008. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 4, 5.

Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. Nature 276, 565–570. doi:10.1038/276565a0

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L.D., Lowe, C.B., Holloway, A.K., Clamp, M., Gnerre, S., Alföldi, J., Beal, K., Chang, J., Clawson, H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M.J., Jaffe, D.B., Jungreis, I., Kent, W.J., Kostka, D., Lara, M., Martins, A.L., Massingham, T., Moltke, I., Raney, B.J., Rasmussen, M.D., Robinson, J., Stark, A., Vilella, A.J., Wen, J., Xie, X., Zody, M.C., Baldwin, J., Bloom, T., Chin, C.W., Heiman, D., Nicol, R., Nusbaum, C., Young, S., Wilkinson, J., Worley, K.C., Kovar, C.L., Muzny, D.M., Gibbs, R.A., Cree, A., Dihn,

H.H., Fowler, G., Jhangiani, S., Joshi, V., Lee, S., Lewis, L.R., Nazareth, L. V, Okwuonu, G., Santibanez, J., Warren, W.C., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Delehaunty, K., Dooling, D., Fronik, C., Fulton, L., Fulton, B., Graves, T., Minx, P., Sodergren, E., Birney, E., Margulies, E.H., Herrero, J., Green, E.D., Haussler, D., Siepel, A., Goldman, N., Pollard, K.S., Pedersen, J.S., Lander, E.S., Kellis, M., 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476–82.

Mackem, S., Lewandoski, M., 2012. Limb Cells Don ' t Tell Time. Science (80-.). 1038.

MacLean, G., Abu-Abed, S., Dollé, P., Tahayato, P., Chambon, P., Petkovich, M., 2001. Cloning of a novel retinoic-acid metabolizing cytochrome P450, Cyp26B1, and comparative expression analysis with Cyp26A1 during early murine development. Mech. Dev. 195–201.

Mann, R.S., Affolter, M., 1998. Hox proteins meet more partners. Curr. Opin. Gene Dev. 423–429.

Marklund, M., Sjödal, M., Beehler, B.C., Jessell, T.M., Edlund, T., Gunhaga, L., 2004. Retinoic acid signalling specifies intermediate character in the developing telencephalon. Development 131, 4323–32.

Mason, M.K., 2009. Cross-Species Microarray Analysis of Limb Development in the Bat, Miniopterus natalensis. University of Cape Town.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., Cramer, P., 2010. Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. 17, 1272–8.

McDonald, J.T., Rautenbach, I.L., Nel, J.A.J., 1990. Roosting requirements and behaviour of five bat species at De Hoop Guano Cave, southern Cape Province of South Africa. South African J. Wildl. Res. 157–161.

Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martinez, C., Torres, M., 1999. Conserved regulation of proximodistal limb axis development by Meis1/Hth. Nature 402, 425–429.

Mercader, N., Leonardo, E., Piedra, M.E., Martínez-a, C., Ros, M.A., Torres, M., 2000. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. Development 127, 3961–3970.

Mercader, N., Selleri, L., Criado, L.M., Pallares, P., Parras, C., Cleary, M.L., Torres, M., 2009. Ectopic Meis1 expression in the mouse limb bud alters P-D patterning in a Pbx1-independent manner. Int. J. Dev. Biol. 53, 1483–94.

Moskow, J.J., Bullrich, F., Huebner, K.A.Y., Daar, I.R.A.O., Buchberg, A.M., 1995. Meis1 , a PBX1-Related Homeobox Gene Involved in Myeloid Leukemia in BXH-2 Mice. Microbiology 15, 5434–5443.

Nakamura, T., Jenkins, N.A., Copeland, N.G., 1996. Identification of a new family of Pbx-related homeobox genes. Oncogene 13, 2235–42.

Nam, D.K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J.D., Wang, S.M., 2002. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. PNAS 99, 6152–6.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–26.

Niederreither, K., McCaffery, P., Dräger, U.C., Chambon, P., Dollé, P., 1997. Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. Mech. Dev. 62, 67–78.

Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P., Dollé, P., 2000. Retinoic acid synthesis and hindbrain patterning in the mouse embryo. Development 127, 75–85.

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801.

Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P., Dollé, P., 1997. Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. Dev. Dyn. 210, 173–83.

Penkov, D., Mateos San Martín, D., Fernandez-Díaz, L.C., Rosselló, C. a, Torroja, C., Sánchez-Cabo, F., Warnatz, H.J., Sultan, M., Yaspo, M.L., Gabrieli, A., Tkachuk, V., Brendolan, A., Blasi, F., Torres, M., 2013. Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. Cell Rep. 3, 1321–33.

Proudfoot, N.J., Brownlee, G.G., 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211–214.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 33, D501–4.

Radicella, J.P., Dherin, C., Desmaze, C., Fox, M.S., Boiteux, S., 1997. Cloning and characterization of hOGG1 , a human homolog of the OGG1 gene of Saccharomyces cerevisiae. PNAS 94, 8010–8015.

Rampias, T.N., Fragoulis, E.G., Sideris, D.C., 2012. Efficient cloning of alternatively polyadenylated transcripts via hybridization capture PCR. Curr. Issues Mol. Biol. 14, 1–8.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C.O., Forrest, A.R.R., Gough, J., Grimmond, S., Han, J.-H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van Nimwegen, E., MacPherson, C.R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R.D., Tegnér, J., Lenhard, B., Teichmann, S.A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D.A., Ideker, T., Hayashizaki, Y., 2010. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, 744–52.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., Grimmond, S.M., Hume, D.A., Hayashizaki, Y., Mattick, J.S., 2006. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. 16, 11–9.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-shaar, M., Mann, R.S., 1997. Nuclear Translocation of Extradenticle Requires homothorax , which Encodes an Extradenticle-Related Homeodomain Protein 91, 171–183.

Rinn, J.L., Chang, H.Y., 2012. Genome Regulation by Long Noncoding RNAs. Annu Rev Biochem 81, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S. a, Goodnough, L.H., Helms, J. a, Farnham, P.J., Segal, E., Chang, H.Y., 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–23.

Roselló-Díez, A., Ros, M.A., Torres, M., 2011. Diffusible signals, not autonomous mechanisms, determine the main proximodistal limb subdivision. Science 332, 1086–8.

Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., Lee, B.T., Barber, G.P., Harte, R.A., Diekhans, M., Long, J.C., Wilder, S.P., Zweig, A.S., Karolchik, D., Kuhn, R.M., Haussler, D., Kent, W.J., 2013. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 41, D56–63.

Salsi, V., Alessandra, M., Cocchiarella, F., Mantovani, R., Zappavigna, V., 2008. Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. Dev. Bio. 317, 497–507.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A laboratory Manual, Second. ed. Cold Spring Harbor Laboratory Press.

Sánchez-Guardado, L.Ó., Irimia, M., Sánchez-Arrones, L., Burguera, D., Rodríguez-Gallardo, L., Garcia-Fernández, J., Puelles, L., Ferran, J.L., Hidalgo-Sánchez, M., 2011. Distinct and redundant expression and transcriptional diversity of MEIS gene paralogs during chicken development. Dev. Dyn. 240, 1475–92.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. PNAS 74, 5463–7.

Saunders, J.W., 1948. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. J. Exp. Zool. 108, 363–403.

Seim, I., Fang, X., Xiong, Z., Lobanov, A. V, Huang, Z., Ma, S., Feng, Y., Turanov, A. a, Zhu, Y., Lenz, T.L., Gerashchenko, M. V, Fan, D., Hee Yim, S., Yao, X., Jordan, D., Xiong, Y., Ma, Y., Lyapunov, A.N., Chen, G., Kulakova, O.I., Sun, Y., Lee, S.-G., Bronson, R.T., Moskalev, A. a, Sunyaev, S.R., Zhang, G., Krogh, A., Wang, J.,

Gladyshev, V.N., 2013. Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat. Commun. 4, 2212.

Shanmugam, K., Green, N.C., Rambaldi, I., Saragovi, H.., Featherstone, M.., 1999. PBX and MEIS as Non-DNA-Binding Partners in Trimeric Complexes with HOX Proteins. Mol. Cell. Biol. 7577–7588.

Shen, W., Montgomery, J.C., Rozenfeld, S., Moskow, J.J., Lawrence, H.J., Buchberg, A.M., Largman, C., 1997. AbdB-Like Hox Proteins Stabilize DNA Binding by the Meis1 Homeodomain Proteins. Microbiology 17, 6448–6458.

Slattery, M., Ma, L., Négre, N., White, K.P., Mann, R.S., 2011. Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in Drosophila. PLoS One 6, e14686.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., Hood, L.E., 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–9.

Speakman, J.O.H.N.R., 2001. The evolution of flight and echolocation in bats : another leap in 31, 111–130.

Su, D., Gudas, L.J., 2008. Gene expression profiling elucidates a specific role for RARgamma in the retinoic acid-induced differentiation of F9 teratocarcinoma stem cells. Biochem. Pharmacol. 75, 1129–60.

Swartz, S.M., 1997. Allometric Patterning in the Limb Skeleton of Bats : Implications for the Mechanics and Energetics of Powered Flight. J. Morphol. 294, 277–294.

Swartz, S.M., Bishop, K., Aguirre, M.-F.I., 2006. Dynamic complexity of wing form in bats: Implications for flight performance, in: Zubaid, A., McCracken, G.F., H., K.T. (Eds.), Functional and Evolutionary Ecology of Bats. Oxford University Press, Oxford, pp. 110–130.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–9.

Tanaka, M., Tickle, C., 2007. The Development of Fins and Limbs, in: Hall, B.K. (Ed.), Fins into Limbs: Evolution, Development and Transformartion. The University of Chicago Press, Chicago, pp. 65–78.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., Chang, H.Y., 2010. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689–93.

Vermot, J., Llamas, J.G., Fraulob, V., Niederreither, K., Chambon, P., Dollé, P., 2005. Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. Science (80-.). 563–566.

Visel, A., Minovitsky, S., Dubchak, I., Pennacchio, L.A., 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 35, D88–92.

Vogt, T.F., Duboule, D., 1999. Antagonists Go out on a Limb. Cell 99, 563–566.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R. a, Gupta, R.A., Wysocka, J., Lei, M., Dekker, J., Helms, J. a, Chang, H.Y., 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120–4.

Wang, Z., Dong, D., Ru, B., Young, R.L., Han, N., Guo, T., Zhang, S., 2010. Digital gene expression tag profiling of bat digits provides robust candidates contributing to wing formation. BMC Genomics 11, 619.

Wheeler, D.L., Church, D.M., Federhen, S., Lash, A., Madden, T.L., Pontius, J.U., Schuler, G.D., Schrimi, L.M., Sequeira, E., Tatusova, T.A., Wagner, L., 2003. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 31, 28–33.

Williams, T.M., Williams, M.E., Innis, J.W., 2005. Range of HOX / TALE superclass associations and protein domain requirements for HOXA13 : MEIS interaction. Dev. Bio. 277, 457–471.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., Su, A.I., 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 10, R130.

Wu, J.Q., Du, J., Rozowsky, J., Zhang, Z., Urban, A.E., Euskirchen, G., Weissman, S., Gerstein, M., Snyder, M., 2008a. Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. Genome Biol. 9, R3.

Wu, J.Q., Du, J., Rozowsky, J., Zhang, Z., Urban, A.E., Euskirchen, G., Weissman, S., Gerstein, M., Snyder, M., 2008b. Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. Genome Biol. 9, R3.

Yang, Y., Hwang, C.K., D'Souza, U.M., Lee, S.H., Junn, E., Mouradian, M.M., 2000. Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription. J. Biol. Chem. 275, 20734–41.

Yashiro, K., Zhao, X., Uehara, M., Yamashita, K., Nishijima, M., Nishino, J., Saijoh, Y., Sakai, Y., Hamada, H., 2004. Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. Dev. Cell 6, 411–422.

Zeller, R., López-Ríos, J., Zuniga, A., 2009. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nat. Rev. Genet. 10, 845–58.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K. a, Fang, X., Wynne, J.W., Xiong, Z., Baker, M.L., Zhao, W., Tachedjian, M., Zhu, Y., Zhou, P., Jiang, X., Ng, J., Yang, L., Wu, L., Xiao, J., Feng, Y., Chen, Y., Sun, X., Zhang, Y., Marsh, G. a, Crameri, G., Broder, C.C., Frey, K.G., Wang, L.-F., Wang, J., 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science (80-.). 339, 456–60.

Zhang, X., Friedman, A., Heaney, S., Purcell, P., Maas, R.L., 2002. Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. Genes Dev. 16, 2097–2107.

Zhao, X., Brade, T., Cunningham, T.J., Duester, G., 2010. Retinoic acid controls expression of tissue remodeling genes Hmgn1 and Fgf18 at the digit-interdigit junction. Dev. Dyn. 665–671.

Zhao, X., Sirbu, I.O., Mic, F.A., Molotkova, N., Molotkov, A., Kumar, S., Duester, G., 2009. Retinoic acid promotes limb induction through effects on body axis extension but is unnecessary for limb patterning. Curr. Biol. 19, 1050–7.

Appendices and Supplementary Information

Table A 1: List of Primers used in this study

Experiment	Primer Name	Sequence (5'- 3')	No. Bases	Tm
3'RACE	GSP1	ACG GGC TAT GGC CAC CAC GAC TTC CGG GTT CC	32	71.4
3'RACE	GSP2	ATG ACG CAA CCT CCA CCC ACT CAG CAG GCA CC	32	70.4
5'RACE	GSP3	ATC CCT CCC TCT CTC GCT CGT TCT CAC TCG CGC T	34	70.4
5'RACE	GSP4	ATG CAG GCC GGA TTC CCA TGT GTT GCT GAC C	31	68.4
5'RACE	GSP5	ATG CTG TTG TCT CCA CTC TGG GAA GC	26	69.8
3'RACE	GSP7	CCG CTC GTA ACC TTC AGT TCG GG	23	69.1
3'&'5RACE	UP	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	46	76.9
Nested 3'RACE	NGSP1	ACG AAC CGC GCC GCC AAA CTG AGG CTC TTC TA	32	69.7
Nested 3'RACE	NGSP2	AGC TTC CCA GAG TGG AGA CAA CAG CAG TGA GCA AGG G	37	69.2
Nested 5'RACE	NGSP3	ATC GCT CTC TCG CGC TCG CTC TCT CTC GCT CTC T	34	70
Nested 5'RACE	NGSP4	ACG GAT GTG TGA GAT GCT GGA AGA GCC ACG C	31	67.6
Nested 5'RACE	NGSP5	ATT GAG GTT GCG TCA TCG TGG TCT C	25	69.5
Nested 3'RACE	NSGP7	AGG AGT AAA GGA GGA GGA GGA AGA TCA G	28	66.3
Nested 3'&'5RACE	NUP	AAG CAG TGG TAT CAA CGC AGA GT	23	62.4
Meis2 overlap	Meis2Overlap_F	GGG AGG AAG AAT TCA AGA AGC	21	59.3
Meis2 overlap	Meis2Overlap_R	CCA GTC CGG ATA AGA AAG TGA	21	59.2
Meis2 overlap	Meis2_F	GAA GAA ACA GTT AGC GCA AGA CA	23	60.9
Meis2 overlap	Meis2_R	ACC ATC CAA CAC AAA GCT CC	20	60
Meis2 overlap	lncMeis2_F	CTA TGG CCA CCA CGA CTT C	19	59.1
Meis2 overlap	lncMeis2_R	TGT CAG TAG GTG TTG GCA GG	20	59.7

Table A 2: List of Transcripts that failed to be sub -cloned or sequenced

Failed sub-cloning	Failed sequencing
BF5ssN1B	BH3ssN2A, BH3ssN2B
	MF3ssN2D
	MH3ssN1A
	MH5ssN4B, MH5ssN4C, MH5ssN4D
	MH3ssN2D, MH3ssN2C

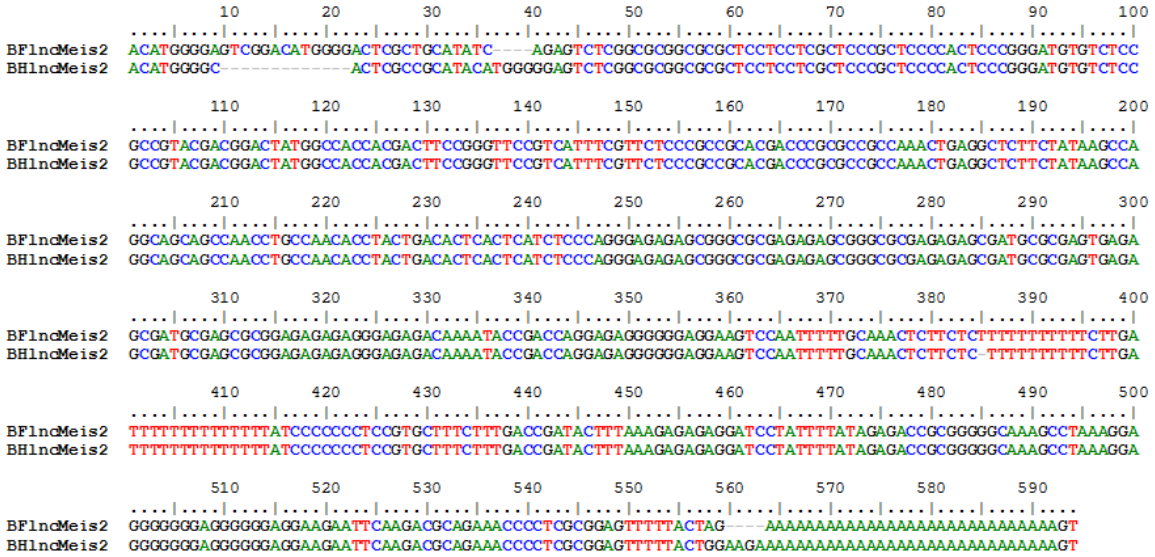

Figure A1: Multiple sequence alignment of *BFlncMeis2* and *BHlncMeis2*.

Figure A2: Multiple sequence alignment of *MFlncMeis2* and *MHlncMeis2*

Supplementary Methods:

1. RNA extraction (RNeasy Mini kit, QIAGEN)

The protocol for purification of total RNA from animal tissue was followed (RNeasy Mini kit, QIAGEN). The head and forelimb tissue put in in a 1.5ml sterile tubes containing 600µl of Buffer RLT and were homogenized using a pestle. The lysate was centrifuged for 3 minutes at 140,000 rpm in a microcentrifuge. One volume of 70% (v/v) ethanol was added and mixed by pipetting. A measure of 700µl of the lysate was added to the RNeasy spin column placed in a 2ml collection tube. The spin column lid was closed and centrifuged for 15 seconds at 10,000 rpm. The flow-through collected in the collection tube was discarded. Buffer RW1, at a volume of 700µl, was added to the spin column, the lid was closed and the tube was centrifuged for 15 seconds at 10,000 rpm. The flow-through was discarded form the collection tube. Buffer RPE was added to the spin column at a volume of 500µl. The spin column lid was closed and the assembled tube was centrifuged for 15 seconds at 10,000 rpm. The flow-through from the collection tube was discarded. The RPE buffer step was repeated

as stated above. The RNeasy spin column was placed in a sterile 1.5ml collection tube and 50µl of RNase-free water was added directly to the spin column membrane. The spin column lid was closed and the assembled tube was centrifuged for 1 minute at 10,000 rpm to elute the immobilised RNA on the membrane.

2. RACE-Ready cDNA Amplification

The generation of RACE-ready cDNA was followed as per the manual. A master mix (MM1) was made, containing: 2µl of 5x First-Strand Buffer, 1µl of 20mM dithiothreitol and 1µl of dNTPs for each reaction. tHe miaster mix was then placed on ice. Separate reactions were prepared for 5'RACE-Ready cDNA generation for each species and tissue type, and 3' RACE ready cDNA generation for each specie and tissue type. To each reaction tube 2µl of RNA was added and either 1µl 5'-CDS Primer A (for 5' RACE-Ready cDNA) or 3'-CDS Primer A (for 3' RACE-Ready cDNA) was added. Sterile water was added to the reactions for a final volume of 3.75µl for 5' RACE or 4.75µl for 3'RACE. Each reaction tube was mixed by vortexing and briefly centrifuge in a microcentrifuge. The reactions were incubated at 72°C for 3 minutes and then cooled to 42°C for 2 minutes. After the reactions have cooled the reaction tube was briefly centrifuged for 10 seconds at 140,000 rpm. To just the 5'RACE-Ready cDNA synthesis reactions 1µl of the SMARTer IIA Oligo was added. The 5'RACE reactions were briefly mixed by vortexing and centrifuged briefly. Another master mix (MM2) was made, containing: 4µl of MM1, 0.25µl of RNase Inhibitor (40 U/ml) and 1µl of SMARTScribe™ Reverse transcriptase (100U) for each reaction. To the cooled reaction tubes 5.25µl of MM2 was added, which will bring the total volume of each reaction tube to 10µl. The contents of each reaction tube were mixed by gentle pipetting and each tube was briefly centrifuged. The reaction tubes were incubated at 42°C for 90 minutes and then at 70°C for 10 minutes. The reactions were then diluted with 100µl Tricine-EDTA.