A neuro-specific hedgehog-responsive enhancer from intron 1 of the murine laminin alpha 1 gene

Thesis presented for the degree of PhD by
Kalin Dimitrov Narov

Department of Biomedical Science
University of Sheffield

September 2013

Acknowledgements

I thank my supervisors Anne-Gaëlle Borycki and Philip Ingham for all expert guidance, patience, and for the opportunity to study vertebrate development in their laboratories.

I also thank my advisors Pen Rashbass and Vincent Cunliffe for the helpful advices and their critical analyses on my work, and our collaborator Norris Ray Dunn for advising me on mouse transgenics and providing me with mouse embryos.

Many thanks also to Shantisree Rayagiri, Joseph B. Pickering, Claire Anderson, Ashish Maurya, Weixin Niah, Harriet Jackson, Raymond Lee, Xingang Wang, Yogavali Pooblan for helping me with text formatting and embryo harvesting, and for providing me with reagents and advices.

I am also grateful to the whole D-floor community, as well as to Martin Zeidler's and Marcelo Rivolta's labs for letting me use their equipment.

Last but not least, I thank my family for the constant support and encouragement, and especially my parents for nurturing in me the love to nature and knowledge.

Therefore, I dedicate this work to the memory of my father.

Abstract

Laminin alpha 1 (LAMA1) is a major component of the earliest basement membranes in the mammalian embryo. Disruption of the murine *Lama1* gene result in lethal failure of germ layer differentiation and extraembryonic membrane formation at gastrulation stages, while conditional deletion of *Lama1* leads to aberrant organization of retinal neurons and vasculature, and defects in cerebellar glia and granule cell precursors later in development. Similarly, inactivation of *lama1* in zebrafish affects lens, retina and anterior notochord development. This diverse range of phenotypes in *Lama1*-deficient animals reflects the complexity of its expression pattern during embryogenesis, which is largely conserved among vertebrates. Major sites of *Lama1* transcription in the mouse embryo are the neural tube, presomitic mesoderm, somites, nephrogenic mesoderm, head mesenchyme and the lens. However, little is known about the signaling mechanisms governing the spatio-temporal control of *Lama1* transcription. Previous studies in our lab revealed a requirement for SHH signaling in the transcription of *Lama1* in the somites and neural tube of mouse embryos. Therefore, I hypothesized that SHH might directly modulate *Lama1* expression via the binding of GLI transcription factors to regulatory regions in the *Lama1* locus.

In this study, I identified a *cis*-regulatory element that may be involved in the SHH-dependent control of *Lama1* expression in the murine embryo. I began my study with a phylogenetic footprinting approach that uncovered 25 conserved non-coding elements upstream of the murine *Lama1* locus, some of which contained GLI binding motifs. Subsequent luciferase reporter-based analysis in cell culture with a subset of the CNEs did not provide convincing evidence for enhancer- and/or silencer-like properties of the elements, except for CNE7. The CNEs were further characterised using an *in vivo* transgenesis reporter screen in zebrafish, which uncovered a skeletal-muscle specific regulatory region. In parallel, a detailed survey of the existing literature revealed the presence of a non-conserved GLI-occupied region in intron 1 of the murine *Lama1* gene. Subsequently, I showed that this element behaves as a tissue-specific enhancer driving reporter expression in the neural tube of mouse and zebrafish embryos. I provided evidence that active Hh signaling is required and sufficient for the activity of this enhancer. Finally, I demonstrated that the GLI binding motifs within the element are essential for its function. Altogether, these results suggest that SHH may directly control *Lama1* transcription in the mouse neural tube via an intronic enhancer, and also provide further insight in the relationship between cell signaling and the regulated expression of extracellular matrix components in development and disease.

Table of Contents

Figures

Tables

Abbreviations

BBE   billboard-like enhancer

BM   basement membrane

CNE   conserved non-coding element

EE   enhanceosome-like enhancer

HH   hedgehog protein

HH8   Hamburger-Hamilton stage 8

LCR   locus control region

MBM   myotomal basement membrane

TF   transcription factor

TFBS   transcription factor binding site

TRE   transcription-regulatory element

	Mouse	Human	Chicken	Zebrafish	*Drosophila*
Gene (full name)	hedgehog	hedgehog	hedgehog	*hedgehog*	*hedgehog*
Gene (symbol)	*Hh*	*HH*	HH	*hh*	*hh*
Protein (full name)	hedgehog	hedgehog	hedgehog	hedgehog	Hedgehog
Protein (symbol)	HH	HH	HH	Hh	HH
reference	www.informatics.jax.org	www.genenames.org	www.birdgenenames.org	www.zfin.org	www.flybase.org

Table i. Convention of gene/protein names and symbols applied in this study, according to established nomenclatures ("reference" row). The hedgehog gene and protein is provided as an example.

Chapter 1

Introduction

1. Introduction

1.1. Basement membranes and animal development

Metazoan development depends on intimate interactions between cells and their extracellular matrices, which provide chemical, mechanical and electrical cues to guide cell behavior. Basement membranes (BM) are a major form of extracellular matrix. They are multi-component meshwork-like sheets, underlying all epithelia and surrounding muscle cells, adipocytes, Schwann cells, nerves and endothelia (Colognato and Yurchenco 2000). Major constituents of BMs are the laminins, collagen type IV, nidogens, and perlecan, whereas a variety of other glycoproteins and proteoglycans are incorporated to BMs in a tissue- and stage-specific manner (Scheele et al. 2007) (Figure 1.1). Basement membranes are highly dynamic structures, which assemble and dis-assemble repeatedly during embryonic development, conferring stability and compartmentalization, and thus are central to organ morphogenesis and maintenance (Yurchenco 2011). Cell-surface receptor-mediated contacts of cells with BMs are involved in myriads of developmental processes: early germ layer polarization and differentiation (Miner et al. 2004), radial sorting of axons by Schwann cells (Wallquist et al. 2004), β-islet cell proliferation (Nikolova et al. 2006), hair growth and morphogenesis (Li et al. 2003), digit separation and neural tube closure (Miner et al. 1998), to name a few. When the interaction between BMs and cells is disturbed, it often results in pathological conditions such as muscular dystrophy observed in patients with deficiency in laminin α2, or junctional epidermolysisin patients with mutations in the *LAMA3*, *LAMB2* and *ITGA6* genes (Mitsuhashi and Hashimoto 2003).

The first basement membrane in the mouse embryo forms within the inner cell mass of the blastocyst at ~ stage E5.5, and separates the prospective epiblast from the primitive endoderm (Miner 2008), as described later in more detail. One of its major components is the product of the laminin alpha 1 gene, subsequently expressed in other structures of the developing embryo, and subject of the current study.

1.2. Laminin alpha 1 – gene organization, protein structure and assembly

The mouse laminin alpha 1 protein (laminin α1) was first isolated as part of a trimeric glycoprotein (laminin-111) from the Engelbreth-Holm-Swarm (EHS) embryonal carcinoma, that was also enriched in the basement membranes of normal tissues (Timpl et al. 1979).
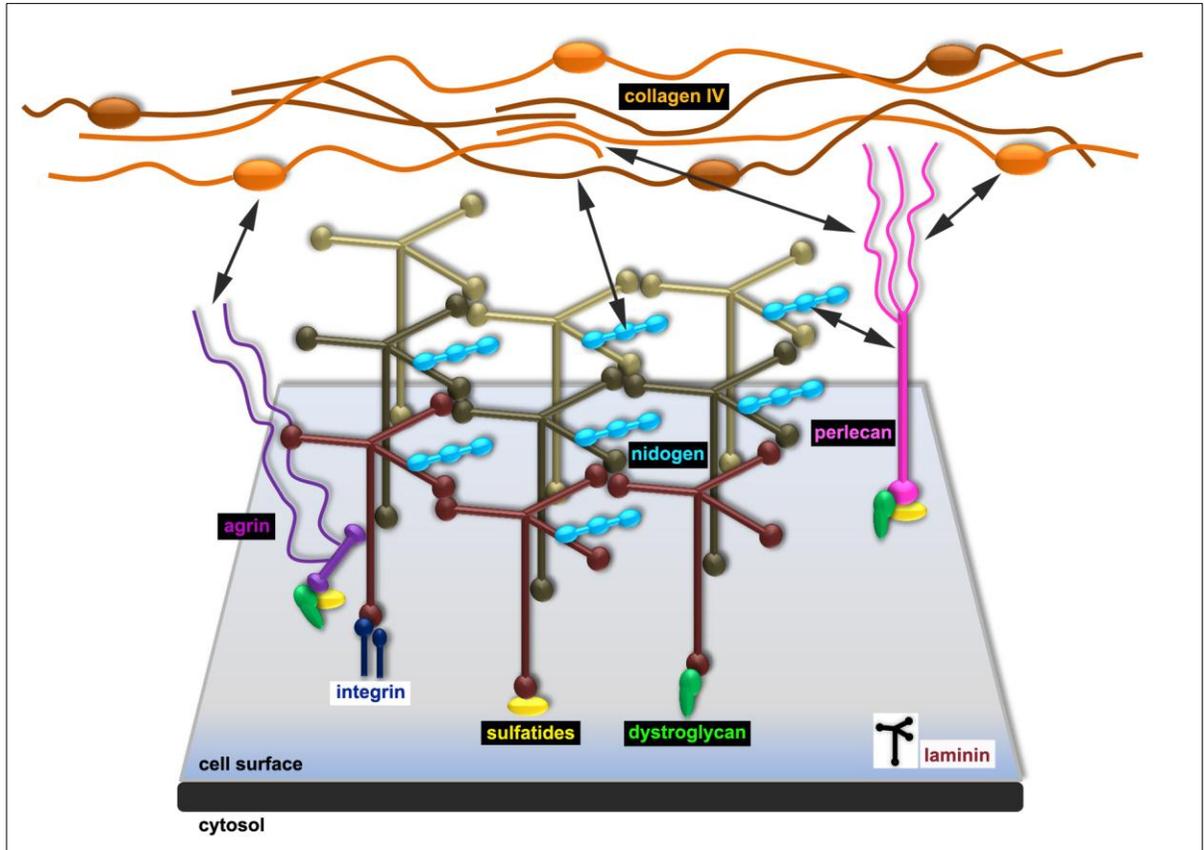
Figure 1.1. *Schematic representation of the structure of basement membranes.* Laminins interact with their LN domains to form a network, which is anchored to the cell surface via direct interactions of laminins with integrins, a-dystroglycan and sulfatides, and indirectly via agrin. Importantly, nidogen, perlecan and agrin connect the laminin network with the independently formed collagen IV network to establish the basement membrane (*double arrows* indicate physical interactions between basement membrane components). Adapted from Hohenester and Yurchenco (2013).

Subsequent molecular cloning studies revealed that the three subunits of laminin-111 are encoded by distinct but homologous genes – *Lama1*, *Lamb1* and *Lamc1*, for laminin α1, β1 and γ1 chains, respectively (Martin and Timpl 1987).

The *Lama1* gene is a member of a diverse family of genes encoding extracellular glycoproteins, with representatives in all Eumetazoans, including cnidarians, ecdysozoans (*Drosophila* and *Caenorhabditis*), and all vertebrates (Domogatskaya et al. 2013). Within vertebrate genomes, *Lama1* is located in a genomic region featuring conserved synteny where *Lama1* is always accompanied by the *Lrrc30*, *Ptprm* and *Arhgap28* protein-coding gene loci in a conserved order and orientation, hinting at the existence of constraints on gene regulation in the cluster, as discussed later in this study. Moreover, the exonic organization and amino acid sequence encoded by the murine *Lama1* are also highly similar to the

corresponding human, chicken and zebrafish orthologs, as summarised in Table 1.1. This suggests for conserved functions of the laminin α1 subunit across vertebrates.

	Gene length in bp[a]	Number of exons[a]	cDNA length in bp[a]	Number of aa in protein[b]	Predicted molecular mass of protein in kDa[c]	% Amino acid sequence identity with the murine ortholog[d]
Mouse	125, 380	63	9, 518	3, 084[e]	~ 338	-
Human	176, 070	63	9, 657	3, 075[e]	~ 337	~ 76%
Chicken	98, 854	63	9, 401	3, 097[f]	~ 340	~ 64%
Zebrafish	92, 360	63	8, 040	3, 075[f]	~ 336	~ 51%

Table 1.1. *Characteristics of laminin α1 gene and protein sequences.*
a, data obtained from the Ensembl database, www.ensembl.org
b, data obtained from the UniProt database, www.uniprot.org
c, data generated via the "Compute pI/mW" tool in ExPASy, www.web.expasy.org
d, data generated via the "Align" tool in UniProt, www.uniprot.org
e, evidence at protein level
f, evidence at transcript level

1.2.1. Laminin α1 is part of a heterotrimeric complex

Laminins constitute a diverse family of glycoproteins with critical roles in basement membrane composition and formation. Each laminin molecule is a heterotrimer composed of one α-, one β- and one γ-laminin subunits (Figure 1.2). The subunit composition determines the name of the laminin trimer where subunits α1, β1 and γ1 form laminin-111, for instance (Aumailley et al. 2005). In mammals, five alpha (including laminin α1), four beta and three gamma subunits have been described each encoded by an individual gene, which in different combinations participate in the formation of at least 16 distinct biochemically-recognised laminin trimer isoforms (Miner and Yurchenco 2004). Apart from laminin-111, the laminin α1 subunit is thought by some authors to be able to partner with laminin β2 and γ1 to form laminin-121 *in vivo*, although this idea is based only on coexpression studies and is currently controversial (Durbeej et al. 1996; Sasaki et al. 2010).

Laminin-111 is the first laminin described and the most extensively studied in terms of biochemical and biophysical properties (Timpl et al. 1979; Ekblom et al. 2003). Laminin-111 has a cross-shape architecture featuring three short arms and one long arm (Figure 1.2) (Beck et al. 1990). The three short arms are represented by the N-terminal parts of the α1, β1 and γ1 subunits, while the long arm is formed by the collective contribution of all three subunits (Colognato and Yurchenco 2000).

Figure 1.2. *Schematic drawing of the domain structure of laminin-111.* The laminin-111 heterotrimer is cross shaped and all three subunits (α1, β1 and γ1) feature common domain organization with globular LN and L4 domains, rod-like LE domains, and coiled coil domains. Uniquely, α1 (and all other α subunits) carries five globular LG domains at its carboxy terminus. Adapted from Hohenester and Yurchenco (2013).

The short arms consist of several globular domains - one LN domain and two L4 domains in the α1 subunit, and one LN domain and a single L4 domain in the β1 and γ1 chains. The globular domains in all three subunits are separated by tandem repeats of laminin-type epidermal growth factor, or LE, rod-like domains (Durbeej 2010). The long arm of each subunit consists of an extensive α-helical coiled-coil domain, which facilitates trimer formation (Beck et al. 1993). In addition, the alpha subunits, including α1, are unique in the possession of 5 globular domains (LG1-5) at the C-terminus (Tzu and Marinkovich 2008) (Figure 1.2). However, there are some exceptions to the so described laminin structure. For instance, laminin α3A and α4 are truncated – they lack the short arm, as is the case for laminin β3 and γ2 subunits (Miner et al. 1995).
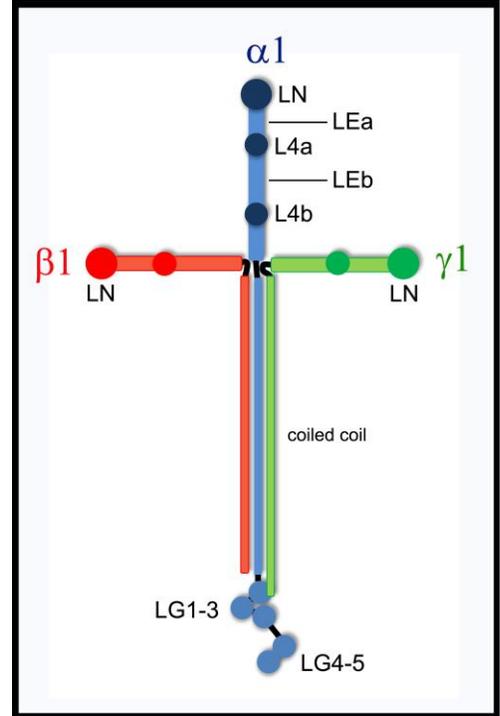
The LN domain of all three chains in laminin-111 features a beta-sandwich motif with several loops and is required for laminin network polymerisation (Schittny and Yurchenco 1990) as shown by electron microscopy where Ca^{2+}-dependent ternary interactions occur between the LN domains of one α, one β, and one γ subunit from adjacent laminin trimers, thus establishing the "three-arm interaction" model for laminin assembly (Cheng et al. 1997; Paulsson et al. 1988). The roles of the LE domains are poorly known except for LEβ3 in the γ1 subunit which binds to nidogen-1 and -2 and this interaction appears to be essential for kidney development, as discussed below (Willem et al. 2002). The LG1-5 domains of the alpha chains have a lectin-like beta-sandwich motif complexed with Ca^{2+} ions, and are especially important in establishing contacts with cell surface receptors (Hohenester et al. 1999).

Assembly of laminin heterotrimers occurs in the Golgi apparatus with the initial dimerization between the coiled-coil domains of the β and γ chains followed by incorporation of the α subunit, and the whole complex is stabilized by several disulphide bonds (Paulsson et al. 1985; Tokida et al. 1990). Addition of the α chain is required for secretion of the whole laminin trimer (Yurchenco et al. 1997), although studies in mouse and *Drosophila* have shown that α chains can be released from the secreting cells as single subunits (Kumagai et al. 1997; Yurchenco et al. 1997).

1.2.2. Interactions of laminin-111 with other ECM components and cellular receptors
Multiple interactions exist between laminins and other components of the extracellular matrix, as well as between laminin and cell surface receptors. In general, most interactions with the ECM occur on the short arms of all three laminin subunits, while interactions with the cell surface is mediated by the globular LN and LG1-5 domains of the alpha chains (Miner and Yurchenco 2004). For instance, the coiled-coil domain of γ1 interacts with the heparan sulphate proteoglycan (HSPG) agrin, and with nidogens, as mentioned earlier, which together with perlecan help to establish the linkage between laminin and the collagen IV networks (Bezakova and Ruegg 2003; Fox et al. 1991; Hohenester and Yurchenco 2013).

The major cell receptor partners for laminin-111 are integrins α1β1, α2β1, α6β1, α6β4, α7β1, and α9β1, and also α-dystroglycan and syndecans. Contacts with integrins are mainly established via the LG1-3 domains of the α subunit (Ido et al. 2004), while LG4-5 interact with α-dystroglycan, heparan sulfates and sulfated glycolipids (Talts et al. 1999). However, a glutamic residue in the C-terminus of the γ1 chain was shown to be required for complex formation between LG1-3 and the integrins (Ido et al. 2007).

The interaction of laminins with integrins and dystroglycan is thought to promote not only BM assembly and maintain stability, but also to serve signalling functions, which culminate in the activation of intracellular pathways. For instance, integrins activate the PI-3 kinase signaling pathway in mammary epithelial cells resulting in suppression of apoptosis (Boudreau et al. 1995). In another study, Li et al. (2002) examined the function of laminin-111, integrin and dystroglycan for cell differentiation in embryoid bodies (EBs) derived from mouse ES cells. They found that epiblast differentiation and cavitation are dependent on the presence of laminin-111 and that both integrin and dystroglycan are important for epiblast survival. Thus, laminin reception plays important roles in the control of cell differentiation,

survival, proliferation and migration in both embryonic and postnatal processes (Danen and Sonnenberg 2003; Yurchenco et al. 2004).

1.2.3. Laminin-111 and basement membrane assembly

Laminins, including laminin-111, are essential for BM assembly as demonstrated in culture and *in vivo*. For instance, laminin β1- and γ1-deficient mouse embryos are unable to assemble the earliest embryonic and Reichert's BMs caused by failure of laminin-111 and laminin-511 synthesis culminating in embryonic lethality at E5.5 (Smyth et al. 1999; Miner et al. 2004). Likewise, laminin α1 constitutive knockout mouse embryos completely lack Reichert's BM, while the embryonic basement membrane is present due to compensation by laminin α5, allowing epiblast polarization, cavitation and normal entry in gastrulation (Miner et al 2004). Nevertheless, this compensation is incomplete as *Lama1$^{-/-}$* embryos die shortly after E6.5 (Miner et al. 2004). Likewise, *Lamc1$^{-/-}$* embryoid bodies in culture fail to assemble a basement membrane between the endoderm and epiblast, while exogenous supplementation of laminin-111 restores BM assembly and epiblast polarization. Importantly, this restoration is prevented by blocking either polymerization or interactions between the LG domains of laminin α1 and cellular receptors (Li et al. 2002). Similarly, a mutant laminin α1 subunit that lacks all LG1-5 domains is unable to assemble a basement membrane on the surface of cultured Schwann cells (McKee et al. 2007).

In contrast to the requirement for laminins in early embryogenesis, genetic ablation of other BM components − nidogen-1/2, perlecan and collagen IV, reveal that they are not essential for initial assembly of BMs but are required later for maintenance of their integrity in the skeletal, vascular and respiratory systems, among others (Arikawa-Hirasawa et al. 1999: Murshed et al. 2000; Poschl et al. 2004).

1.3. Expression of Laminin α 1

Laminins are expressed in various tissues during embryonic and postnatal development. In general, the alpha subunits exhibit more restricted tissue-specific expression patterns as compared to the β and γ subunits. Thus, laminin α1, α3 and α5 are predominantly present in the basement membranes of epithelial tissues, while α2 and α4 are associated with endothelia and structures of mesenchymal origin (Tunggal et al. 2000).

1.3.1. *Lama1* expression in the mouse and human

In the mouse, *Lama1* mRNA is first detected at the 2-4 cell stage (Dziadek and Timpl 1985) suggesting the possibility of maternal deposition, while Laminin α1 protein is first observed in the intercellular space at morula stage (Cooper and McQueen 1983), most likely in the form of laminin-111. At early post-implantation stages (E5.5-7.5), Laminin α1 appears prominently in the extra-embryonic Reichert's basement membrane, as part of laminin-111, as well as in the embryonic basement membrane, which separates the epiblast from the visceral endoderm (Miner et al. 2004). At later embryonic stages (from E9.0 onwards), the major site of *Lama1* expression is the ventricular/subventricular zone of the central nervous system (in both the future spinal cord and in all major divisions of the brain), but also the presomitic mesoderm, newly-formed somites, sclerotome, head mesenchyme, lens of the eye, and in nephric structures (Chapter 3 of this study; Anderson et al. 2009; Lentz et al. 1997; Miner et al. 2004). Laminin α1 expressed in these tissues is observed in the pial basement membrane surrounding the CNS; in the somitic, dermomyotomal and myotomal BMs, and in the BM of the proximal tubules and ureteric buds of metanephric kidneys (Anderson et al. 2009; Lentz et al. 1997). With progression of development, *Lama1* expression is down-regulated in anterior somites (Anderson et al. 2009) (Table 1.2).

Analysis of laminin α1 protein distribution in the adult mouse revealed strong activity in only few structures in the nervous and genito-urinary systems, pia matter, brain blood vessels, the ECMs that cover the vitreous chamber and lens of the eye, adrenal gland cortex, proximal tubules of the kidney, testis; epididymis, prostate, and ovary (Ekblom et al. 2003; Falk et al. 1999). No staining was observed in the trachea, lung, adipose tissues, Schwann cells, endothelia, myocardium, smooth muscle, thymus, and spleen (Falk et al. 1999).

In humans, laminin α1 is similarly detected in both fetal and adult proximal tubules of the kidneys, in the BMs of the seminiferous tubules of the testes and intestinal mucosa glands, and associated with the capillary vasculature in the adult brain (Virtanen et al. 2000).

1.3.2. *Lama1* expression in zebrafish

The expression pattern of laminin α1 is largely conserved between mammals and zebrafish, where in the latter *lama1* is first detected as maternally deposited transcripts as early as the 2-cell stage and later zygotic *lama1* is expressed in the neural tube, presomitic mesoderm, somites, eyes, otic vesicles, gut and pronephros (Table 1.2; Figures 1.3 and 1.4) (Joseph B.

Pickering's Thesis 2012; Zinkevich et al. 2006). However, by 49 hpf, *lama1* expression is dramatically down-regulated in the anterior CNS and completely extinguished in the spinal cord, somites and presomitic mesoderm (Joseph Pickering's Thesis 2012), similar to the situation in the mouse (Miner et al. 2004).

1.4. Regulation of *Lama1* expression

The conserved pattern of *Lama1* mRNA expression between mouse and zebrafish suggests conservation in the mechanisms (signaling and transcriptional) that regulate *Lama1* gene activity (Anderson et al. 2009; Joseph Pickering's Thesis 2012; Zinkevich et al. 2006). However, the signalling pathways controlling *Lama1* expression remain largely unknown.

1.4.1. Signalling pathways and control of *Lama1* expression *in vitro*

Some insights have been gained from studies in cell cultures. FGFR2-dominant-negative mutant ES cells feature abrogated FGF signaling causing reduced Akt/PKB activation, which correlates with decreased expression of laminin-111 and collagen type IV (Li et al. 2001). Interestingly, β1-integrin-null embryoid bodies fail to express laminin-111 (Li et al. 2002) due to down-regulation in laminin α1 expression, suggesting a feedback role for integrins in the control of laminin expression that might be mediated by, among others, the Akt/PKB signal transducers (Ekblom et al. 2003). Also, induction of endoderm differentiation in retinoic acid-treated F9 embryonal carcinoma cells in culture correlates with the simultaneous up-regulation of *Lama1*, *Lamb1* and *Lamc1* mRNA expression (Kleinman et al. 1987). The increase in *Lama1* expression during differentiation of F9 cells into parietal endoderm was shown to be dependent on SOX7, GATA4 and GATA6 inputs, where SOX7 is necessary for the activation of *Gata4* and *Gata6* (Futaki et al. 2004). Altogether, these studies reveal that *Lama1* expression is responsive to FGF, retinoic acid and β1 integrin signaling, although remains unclear whether these effects are direct or indirect.
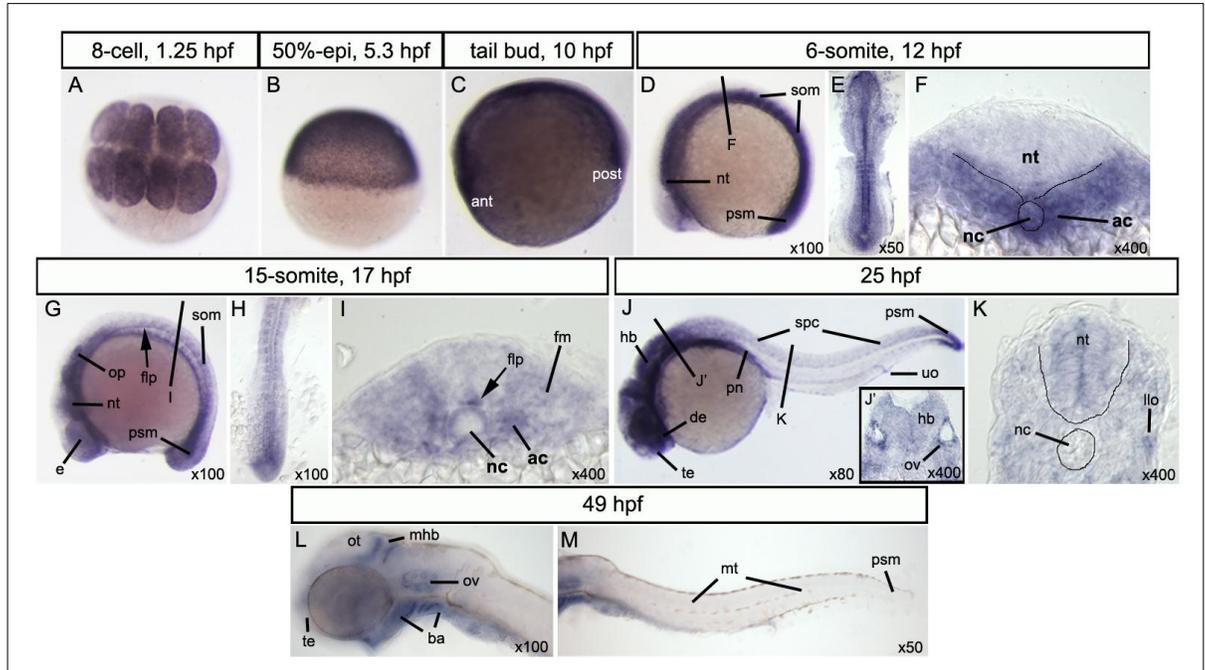
Figure 1.3. *Zebrafish lama1 mRNA expression.* Expression pattern of zebrafish *lama1* mRNA in wild type embryos. Maternal transcripts are detected at the 8-cell stage (A). (D-K) During somitogenesis, expression is strong in the neural tube, somites and presomitic mesoderm, but by 49 hpf (L, M) *lama1* transcripts are observed only in some brain structures, in the pouches associated with the branchial arches, and in the otic vesicles. (C, D, G, J, K, L, M) lateral views; (E, H) dorsal views; (I) a transverse section as indicated in G; (J', K) transverse sections as indicated in J. Abbreviations: ac, adaxial cells; ant, anterior end; de, diencephalon; flp, floor plate; fm, fast muscle progenitors; hb, hindbrain; mhb, midbrain-hindbrain boundary; mt, myotomes; nt, neural tube; nc, notochord; llo, lateral line organ; ot, optic tectum; ov, otic vesicle; som, somites; spc, spinal cord; pn, pronephros; post, posterior end; psm, presomitic mesoderm; te, telencephalon. (The images are kindly provided by Dr. Joseph B. Pickering, University of Sheffield).
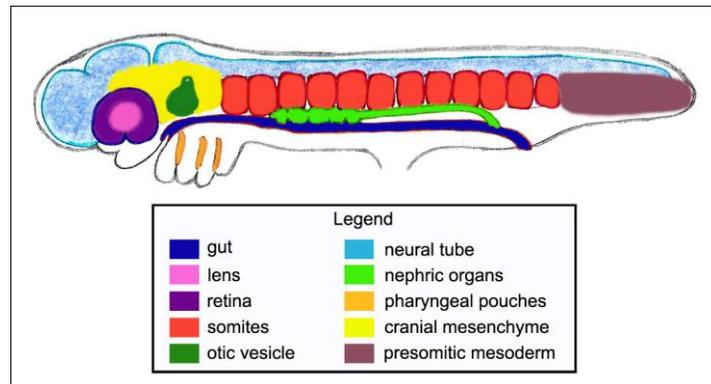
Figure 1.4. *Conserved domains of Lama1 mRNA expression pattern.* A schematic diagram of a generalized vertebrate embryo showing the major domains of *Lama1* mRNA expression that are shared between mouse and zebrafish embryos. Individual domains are highlighted in different colours. A more detailed description is provided in Table 1.2 and Section 1.3.

	mouse	zebrafish	chicken*
Lama1 expression pattern.	2-4 cell blastula (mat); morula; visceral and parietal endoderm; neural tube; presomitic mesoderm; somites; sclerotome; gut endoderm; nephrogenic mesoderm; head mesenchyme; lens; otic vesicle; optic cup; meninges; pharyngeal pouches.	2-cell stage (mat); neural tube; otic vesicle; optic cup; lens; presomitic mesoderm; somites; adaxial cells; fast muscle fibers progenitors; notochord; nephrogenic mesoderm; gut endoderm; hypochord. By 49 hpf, *lama1* transcripts remain detected only in some brain structures, in the pectoral fins and pharyngeal pouches.	Ingressing mesoderm during gastrulation; neural tube; foregut endoderm; head mesenchyme; somites; sclerotome; nephrogenic mesenchyme; optic cup; lens; posterior lateral plate mesoderm.
Involvement of HH signalling in regulation of *Lama1* expression.	*Lama1* transcription is absent from the neural tube and somites of **Shh$^{-/-}$, Gli2$^{-/-}$ and Shh$^{-/-}$;Gli3$^{-/-}$** embryos (Anderson et al. 2012).	**smo$^{-/-}$** mutants or **cyclopamine** treatment: early stage embryos (up to 15-somites) -> no defects in *lama1* expression. 25 hpf embryos -> reduced *lama1* mRNA levels in presomitic mesoderm and anterior neural tube; no change in somitic expression (Joseph B. Pickering's Thesis 2012) **ptch1$^{-/-}$;ptch2$^{-/-}$** or **dnPKA mRNA** treatment: in 15-somite stage embryos -> increased levels of *lama1* mRNA in the neural tube, somites and presomitic mesoderm; expansion of *lama1* expression along the D/V axis of the neural tube. 25 hpf embryos -> *lama1* mRNA levels in somites are similar to wt/untreated embryos, but levels in the neural tube and presomitic mesoderm remain increased (Joseph B. Pickering's Thesis 2012).	No data.
Other signalling molecules and TFs implicated in *Lama1* expression.	FGF, Akt/PKB signalling, GATA4, GATA6, retinoic acid, β1 integrins, SP1/SP3, YY1, SOX7/SOX17, KLF4/KLF5, DMRT2.	No data.	No data.

Table 1.2. *Summary of the Lama1 mRNA expression pattern together with relevant signaling molecules and transcription factors in the embryos of mouse, zebrafish and chicken.* Components and antagonists of the HH signaling pathway are highlighted in a bold font. The main text includes a more detailed account on the signaling pathways and transcription factors implicated in *Lama1* expression, together with the relevant references (see Section 1.3, 1.4 and 1.6 of this study). *data on *Lama1* expression in the chicken embryo is based on results from the current study (see Section 3.2.3). Abbreviations: mat, maternal; wt, wild type.

1.4.2. *cis*-**regulatory elements in** *Lama1* **transcription**

Although the majority of *cis*-regulatory sequences in the vicinity of *Lama1* are unknown, a few studies have reported the existence of regulatory sequences involved in the control of *Lama1* transcription in mouse and human cells in culture (Figure 1.5). For instance, a 435 bp

parietal endoderm enhancer (PEE) with binding sites for the ubiquitous transcription factors of the SP1/SP3 family, NFY and YY1 was found ~ 3 kb upstream of the murine *Lama1* gene, and was shown to be required for *Lama1* expression in the PYS-2 parietal yolk sac cell line (Niimi et al. 2003; Niimi et al. 2004). Subsequent investigation on the same enhancer revealed that SOX7 and SOX17 also bind to motifs within that region and suggested a synergistic mechanism between SP1/SP3, SOX7/SOX17 and YY1 in conferring parietal endoderm-specific expression of *Lama1* (Niimi et al. 2004). In addition, analyses of a 2 kb region upstream of the transcription start site of *Lama1*, in the murine Caco2-TC7 intestinal cell line, uncovered the existence of binding sites for the Krüppel-like transcription factors KLF4 and KLF5, as well as for JUN, CEBPA, FOXA2, NR3C1 and SP1. Mutations in all motifs revealed that they are essential for the activity of the 2 kb "promoter" region in Caco2 cells, while overexpression of exogenous KLF4 leads to repression of its activity (Piccinni et al. 2004). Based on this, the authors proposed a model where KLF4 expressed in post-mitotic cells of the intestinal epithelium competes for binding with the SP1 and/or KLF5, resulting in repression of *Lama1* transcription in those cells. In contrast, KLF5 is expressed in a complementary pattern in the proliferating zone of the adult gut epithelium - in cells of the intestinal crypts, where it might be involved in activation of *Lama1* expression (Piccinni et al. 2004).

Another study identified the basal promoter of the human *LAMA1* gene - a 237 bp region located between -206 and +31 relative to the transcription start site, that is highly conserved with the mouse *Lama1* promoter and both lack TATA and CCAAT elements (Niimi et al. 2006). It appears that the activity of the human promoter depends on the cumulative effect of six GC-rich binding sites for the SP1/3 and KLF4/6 transcription factors, as revealed in the Laminin-111- producing JAR choriocarcinoma cells (Niimi et al. 2006). Taken together, the observations in mouse and human cell lines suggest that conserved inputs by SP and KLF factors are involved in regulation of *Lama1* transcription.
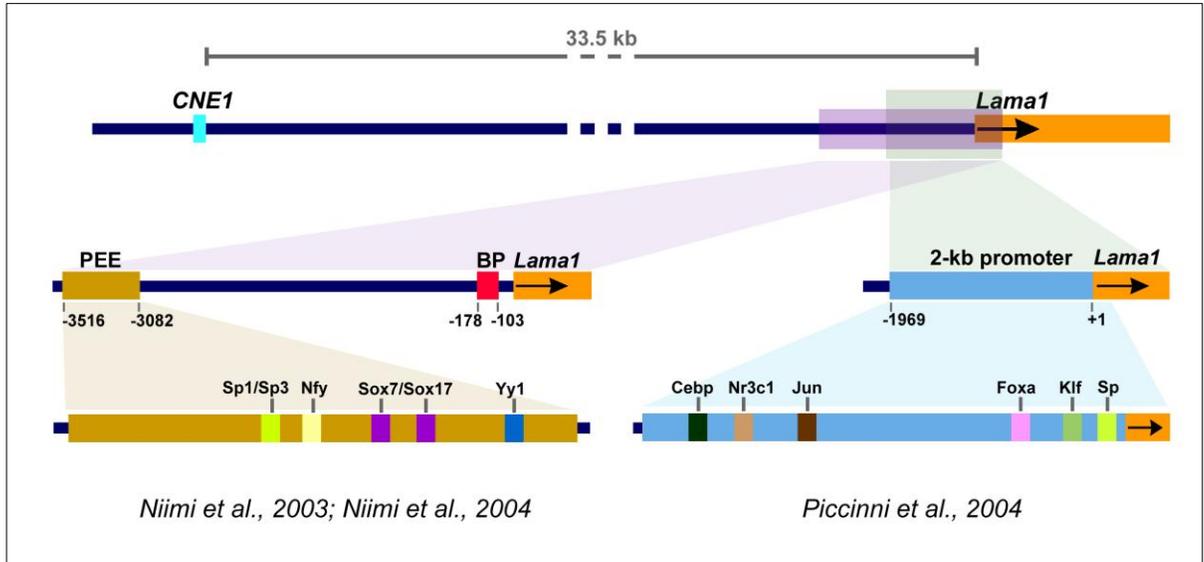
Figure 1.5. *Schematic representation of the known cis-regulatory elements of the murine Lama1 gene with experimentally validated transcription factor binding sites.* For more detailed description see section 1. 4. 2 in the main text. BP, basal promoter; PEE, parietal endoderm enhancer. Importantly, none of the previously identified regulatory elements of *Lama1* overlap with any of the conserved non-coding elements (CNEs) uncovered in this study (described in detail in Chapter 4).

1.4.3. Regulation of *Lama1* transcription in embryonic development

Knowledge on the regulation of *Lama1* expression during embryonic development *in vivo* is even scarcer. Intriguingly, E10.5 *Dmrt2$^{-/-}$* (doublesex and mab-3 related transcription factor 2) mouse embryos display lack of laminin α1 immunoreactivity in the somites, which correlates with morphological and molecular (such as reduced expression of *Pax3*, *MyoD* and *Pdgfa*) aberrations in the dermomyotome and myotome (Seo et al. 2006). However, it is unknown whether the absence of laminin α1 in this case is a result of defect in *Lama1* transcription.

1.5. Laminin expression in skeletal muscle development

Most muscles in the trunk of amniote embryos are derived from the myotome in two successive waves. First, the myotomes are built by delamination of MYF5$^+$ myogenic precursor cells from the dorso-medial edge of the dermomyotome followed shortly after by migration of cells from the other three lips of the dermomyotome – ventro-lateral, rostral and caudal (Figure 1.6) (Gros et al. 2004; Ordahl et al. 2001). During the second wave, the already formed myotomes increase in size by the addition of PAX3$^+$/PAX7$^+$ myogenic cells directly from the central dermomyotome, which contributes to the growth of the muscle masses in later embryogenesis and also provides progenitor cells for the stem cells of adult muscle – the satellite cells (Figure 1.6) (Gros et al. 2005; Relaix et al. 2005; Yusuf and Brand-Saberi 2006).

The distribution of laminins during myotome development is a highly dynamic process. Initially, a laminin-111- and laminin-511-containing BM surrounds the murine somite. Following de-epithelialisation of the sclerotome, laminin α1 is observed in the dermomyotomal and subsequently in the myotomal BMs (Anderson et al. 2009; Bajanca et al. 2006; Tosney et al. 1994), but also at the myotendinous junctions of developing intercostal muscles (Patton et al., 1997) and, as revealed by studies in our laboratory, at the sites of activated satellite cells in adult muscles (unpublished data, Shantisree Rayagiri's Thesis). Later in myotome development, laminin α2, α4 and α5 appear in the basement membranes enveloping individual myocytes (Patton et al. 1997). However, in postnatal muscle fibers α4- and α5-laminins are absent from the BM surrounding the myofibre, although they remain at the neuro-muscular junctions (Patton et al. 1997). Thus, laminin α2

is the main laminin α-chain present in the sarcolemmal BM of adult muscle (Gullberg et al., 1995; Schuler and Sorokin, 1995).
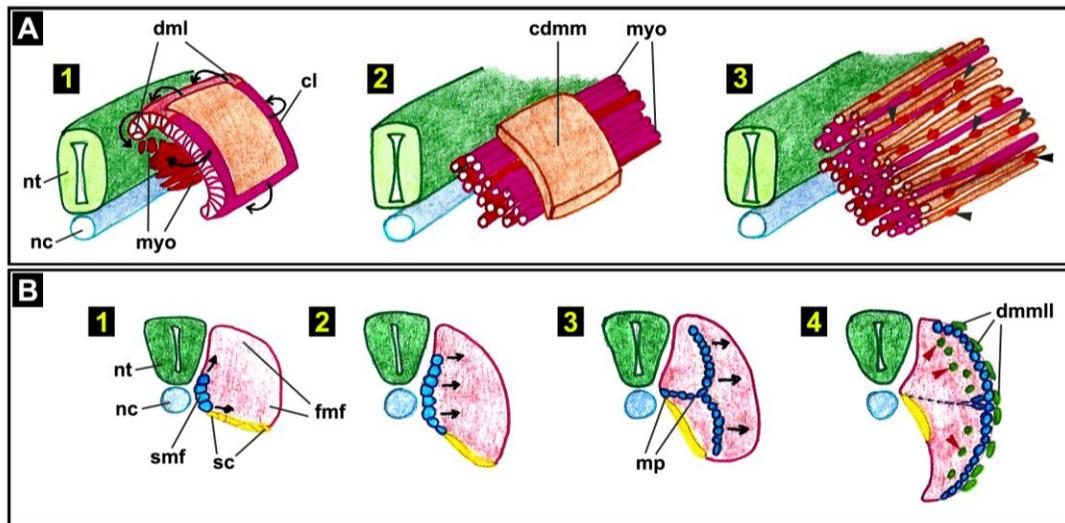
Figure 1.6. *Myotome morphogenesis in amniotes and zebrafish*. A schematic depiction of the morphological events during myotome development in (A) amniote (mouse and chicken) and (B) zebrafish embryos. In both groups the myotome is derived from the somites in a series of cell specification and translocation events regulated by the SHH, BMP and WNT signals emanating from the notochord, neural tube, surface ectoderm and the lateral plate mesoderm (Christ et al. 2007). However, the morphological rearrangements and the muscle progenitor compartments in the somites vary between amniotes and zebrafish (Bryson-Richardson and Currie 2008). (A1, A2) The myotome of amniotes is formed via myoblast delamination (curved *black* arrows in A1) from the four lips of the overlying dermomyotome (shown in *red* and *pink* at the edges of the dermomyotome in A1). Once in the myotome, the myoblasts become post-mitotic and elongate rostro-caudally (A2) (Bajanca et al. 2006; Christ et al. 2007). A second population of myotomal cells (*orange* tubes in A3) is provided via ingression of PAX3+/PAX7+ myoblast precursors from the central dermomyotome (an *orange* plate above the primary myotome in A2), which also establish the satellite cell population of adult muscles (*dark grey* arrowheads in A3) (Buckingham 2000). (B) Myotome development in zebrafish features the establishment of distinct populations of slow-twitch and fast-twitch muscle fibers (Jackson and Ingham 2013). The first cells committed to the trunk skeletal muscle lineage in zebrafish are the slow muscle fibers (SMFs). Their precursors are cuboidal cells located in the medial somite (*blue* circles in B1). Once specified to the myogenic lineage by Hh proteins from the notochord, these cells elongate, differentiate into mononucleate slow-twitch fibers and start migration – initially dorso-ventrally (*black* arrows in B1) and then laterally directly through the somite (*black arrows* in B2 and B3) and eventually populate the lateral myotome (B3) (Jackson and Ingham 2013; Ochi and Westerfield 2007; Wolff et al. 2003). Meanwhile, cells located in the posterior of somites start to elongate forming the medially positioned fast muscle cells (the *pink* area of somite in B3 and B4) (Stickney et al., 2000). Parallel to these events, an earlier cell population in the anterior of each somite relocates to the lateral surface of the myotome where it establishes the dermomyomal-like external layer of Pax3+/Pax7+ cells (*green* elipses on the lateral somitic surface in B4), which contribute to further growth of the myotome (*green* circles marked with *red* arrowheads in B4). Abbreviations: cdmm, central dermomyotome; cl, caudal lip of dermomyotome; dml, dorsomedial lip of dermomyotome; dmmll, dermomyotomal-like layer; fmf, fast muscle fibers; mp, muscle pioneers; myo, myotome; nc, notochord; nt, neural tube; sc, sclerotome; smf, slow muscle fibers. (Adapted from Bryson-Richardson and Currie 2008, and Stickney et al. 2000).

1.6. SHH is required for *Lama1* expression and myotomal basement membrane assembly

Important insight in the signaling pathways governing *Lama1* expression during myotome formation in the mouse embryo was provided by studies in our laboratory (Anderson et al. 2009).

1.6.1. *Shh*[-/-] mutant mouse embryos fail to express *Lama1* in the somites and neural tube

The formation of a laminin-rich myotomal basement membrane (MBM) is important for the achievement of correct myotome shape, for its separation from the sclerotome, for proper myogenic cell specification, and for neural crest migration (Anderson et al. 2009; Bajanca et al. 2006; Tajbakhsh et al. 1996; Tosney et al. 1994). The MBM contains laminin-511 and laminin-111 heterotrimers, as revealed by immunostaining (Anderson et al., 2009; Bajanca et al., 2006). Importantly, previous work in the lab uncovered an essential role of SHH in the formation of the MBM (Anderson et al. 2009). Indeed, in the absence of GLI-mediated SHH signalling, *Myf5*-expressing cells (which enter the myotome from the rostral, caudal and ventrolateral dermomyotomal lips) are abnormally located in the ventral somitic compartment. Further investigation showed absence of assembled myotomal BM in *Shh*-null and *Gli2*;*Gli3*-double null embryos, although fragments of laminin polymers and other BM components were detected associated with the myogenic precursor cells (Anderson et al. 2009).

Analysis of the expression of *Lama1* and *Lama5* genes revealed that laminin α1 was no longer expressed in the somites and neural tube of *Shh*-null embryos, whereas the somitic expression of laminin α5 was unaffected. Based on the experimental data, Anderson et al. (2009) hypothesised that SHH signalling is required for the activation of *Lama1*expression in the somite, and that the lack of Laminin α1 led to the failure of MBM assembly in *Shh*-deficient mouse embryos. Supporting this hypothesis, treatment of *Shh*[-/-];*Gli3*[-/-] mutant embryos with 30 μg/ml exogenous laminin-111 restored MBM assembly. Curiously, the efficiency of restoration appeared to depend on the dosage of endogenous laminin for the same treatment failed to restore the MBM in *Shh*[-/-];*Gli3*[+/-]embryos (Anderson et al., 2009) that express GLI3R protein, which presumably represses *Lama1* transcription.

Importantly, restoration of the myotomal basement membrane in *Shh*[-/-];*Gli3*[-/-]mutants after laminin-111 addition is a progressive process, in anterior-to-posterior direction, and

correlates with the recovery of *Lama1* expression, both in the sclerotome and neural tube (Anderson et al., 2009). However, the observed delay in the recovery suggests that *Lama1* expression also requires GLI2A function for timely activation in more posterior somites.

1.6.2. Shh is required for *lama1* expression in zebrafish

Interestingly, 25 hpf *smo*$^{-/-}$ mutant or cyclopamine-treated wild type fish embryos exhibit strong reduction in *lama1* expression in the presomitic mesoderm (PSM) and anterior neural tube, but not in somites, while *lama1* expression is unaffected in early (12-15 somites stage) embryos (Joseph B. Pickering's Thesis 2012), indicating that active Hh signaling is required for zebrafish *lama1* transcription in the central nervous system, similar to the mouse. Moreover, constitutive activation of the Hh pathway in *ptch1*;*ptch2* double mutant fish results in strong up-regulation of *lama1* expression in the neural tube and PSM, and also in the somites (Joseph B. Pickering's Thesis 2012) (Table 1.2).

Altogether, previous studies in our laboratory demonstrate a conserved requirement for SHH signaling in *Lama1* expression in the neural tube and paraxial mesoderm of mammals and teleosts.

1.7. Functions of laminin α1 in development

As part of laminin-111, laminin α1 is involved in the formation of the earliest basement membranes in the mouse embryo, while later tissue-restricted expression relates to its functions in the development of various structures, like the nervous system, eyes, kidneys and skeletal musculature. Moreover, several studies have provided insights into the roles of zebrafish laminin α1 in embryonic development, as discussed later.

1.7.1. Laminins and neural development

Laminins, including laminin-111, are known to participate in multiple aspects of nervous system development, from modulating the behavior of neural stem cells to cortical morphogenesis and peripheral axon path-finding.

1.7.1.1. Laminins and neural stem cells

Culturing human and mouse neural stem/precursor cells (NSPCs) isolated from the developing cerebral cortex on laminin-111 matrices enhances their migration, proliferation, survival, neurite elongation, arborisation and differentiation into neurons and astrocytes compared to cultures on fibronectin, Matrigel or poly-L-ornithine substrates (Flanagan et al.

2006). These effects of laminin-111, compared to other matrices, are consistent with the fact that human NSPCs express most of the Integrin receptors for laminin-111, namely integrins α1β1, α2β1, α3β1, α6β1, α6β4, and α7β1 (Flanagan et al. 2006).

Interestingly, when dissociated retinal neuroepithelial cells from 5-day old chicken embryos were grown on laminin-111 containing substrates, they proliferated, survived and differentiated into neurons in a cell-cell contact dependent manner. However, culturing these cells on laminin-211 also supported proliferation but not survival and failed to promote neuronal differentiation (Frade et al. 1996). Thus, although laminins containing the α2, α4 and β2 subunits are expressed in the neural stem cell-niche of the ventricular zone (Lathia et al. 2007), these laminins might have functions in modulation of neural stem cell behavior that are distinct from that of laminin α1.

1.7.1.2. Laminins and cortical morphogenesis

It appears that glial interactions with the pial basement membrane, which is rich in laminin α1, α2, α4, α5 and β1 (Anderson et al. 2009; Bajanca et al. 2006), are essential for proper CNS development. For instance, combined inactivation of laminin α2 and α4 leads to defects in radial glial cells (RGCs) attachment to the pial surface, apoptosis and cortex reduction (Radakovits et al. 2009), which are most likely caused by perturbation in β1 integrin signaling in RGCs. In fact, mice with conditional inactivation of β1 integrin in the RGCs exhibit reduced RGCs proliferation and increased apoptosis due to detachment of the basal processes of the RGCs from the pial basement membrane, which culminates in the reduced size of the telencephalic cortex (Radakovits et al. 2009).

Importantly, similar functions of laminin α1 were described in cerebellar morphogenesis. Heng et al. (2011) and Ichikawa-Tomikawa et al. (2012) reported that in adult *Lama1* conditional knockout mice generated by *Sox2-Cre* mediated deletion of *Lama1* in the epiblast, which circumvents the essential requirement for *Lama1* expression in the visceral and parietal endoderm during early development (Miner et al. 2004) there is an abnormal locomotor phenotype (Ichikawa-Tomikawa et al. 2012). This defect correlates with a severe disorganization of the cerebellar layers and with fusion of the folia, as well as with excessive proliferation of granule cell precursors in the external granular layer, and with defects in their migration, followed by massive cell death, causing a reduced cerebellum in *Lama1*-deficient animals (Heng et al. 2011). Interestingly, these perturbations are associated

with a discontinuous meningeal BM, disorganization of the Bergmann glial fibers and end-feet, and reduction in the number of dendritic processes in Purkinje cells (Heng et al. 2011; Ichikawa-Tomikawa et al. 2012). This demonstrates a role of the laminin α1-containing pial basement membrane in the control of GPCs proliferation at early stages and in GPCs survival later on, which is perhaps mediated by interactions of the radial glial cells with laminin-111 in the pial membrane.

A further insight into the function of laminins at the pial surface was provided by mutant mice with a deletion in the nidogen-binding site of the laminin γ1 subunit (Halfter et al. 2002). Although assembly of the pial BM initiates in the mutant embryos, it is unstable and subsequently degenerates, leading to the retraction of radial glial processes from the pial surface, and aberrant migration of Cajal-Retzius cells and cortical plate neurons (Halfter et al. 2002).

Taken together, these observations demonstrate a critical role of laminins (including laminin α1) in radial glial cell morphology, and in neuronal migration and differentiation, which is dependent on interactions with the radial glia.

1.7.1.3. Laminin α1 in axonal growth and migration

Laminins appear to be important for neurite outgrowth and axonal migration. Culturing avian and rodent retinal ganglion cells on laminin-111 or laminin-211 stimulates neurite outgrowth in an α6β1-integrin-dependent manner (Cohen and Johnson, 1991). Notably, distinct laminin isoforms exhibit different efficiency in promoting neurite outgrowth in cultured murine dorsal root ganglion (DRG) neurons, such that laminin-111 and laminin-511 are more potent at stimulating neurotrophin-independent neurite elongation compared to laminin-211 or laminin-411 (Plantman et al. 2008). Interestingly, antibody-blocking experiments revealed that the interactions of laminin-111 with α3β1 and α7β1 Integrins are essential in this process, but not the interactions with α6β1 Integrin, which is required for the effects of laminin-211 (Plantman et al. 2008). This observation suggests distinct roles for laminin-111 and laminin-211, mediated by interactions with different integrin receptors in the developing CNS.

Genetic ablation studies provide further insight into laminins and neuronal morphogenesis. Conditional inactivation of *Lamc1* in the murine neocortex leads to lamination abnormalities, failure of neurite ouugrowth and axonal path-finding defects and

these requirements are mediated by Integrin-dependent activation of the AKT/GSK-3β signaling pathway (Chen et al. 2009). Similarly, the *lama1*-deficient *bashful* (*bal*) mutant zebrafish displayed widespread defects in axonal path-finding and outgrowth in the CNS. For instance, the forebrain axons, hindbrain reticulospinal axons and retinal ganglion cell axons exhibited specific path-finding errors indicating that laminin α1 is required for the modulation of axon directionality (Paulus and Halloran 2006). In addition, some axon tracts were defasciculated and not fully extended, while other axons showed extensive branching (Paulus and Halloran 2006). This role of laminins in axonal migration seems to be conserved in evolution for mutational inactivation of either aA or aB laminin subunits in *Caenorhabditis elegans* is also manifested by nerve mis-positioning and defects in axonal outgrowth (Huang et al. 2003).

1.7.2. Laminins and neural crest cells

The involvement of laminins in neural crest (NC) cell development is well established (Perris and Perissinotto 2000). Trunk neural crest cells travelling between the neural tube and somites, along the ventro-lateral migration route, abruptly turn laterally upon reaching the level of the myotome, enter the anterior half of each sclerotome and preferentially use the myotomal basement membrane as a migratory substratum (Tosney et al. 1994).

Early studies of avian embryo explants in culture have shown that the interaction of NC cells with laminin is essential for their emigration and subsequent dispersal from the closing neural folds (Bilozur and Hay 1988). Interestingly, laminin and collagen IV proteins are down-regulated at the roof plate of the closing neural tube as neural crest cells undergo an epithelial-to-mesenchymal transition to leave the neuroepithelium (Duband et al. 1988). Notably, the basal surfaces of epithelia along the migratory routes of neural crest cells are lined with laminin and collagen IV, whereas the amount of these proteins within the NC populations is low (Duband et al. 1988; Krotoski et al. 1988). Curiously, termination of NC migration and aggregation into dorsal root and sympathetic ganglia correlates with an increase of interstitial laminin and collagen IV within the NC population, suggesting that these molecules promote stronger adhesion to the ECM which probably impedes further migration (Duband et al. 1988; Krotoski et al. 1988).

1.7.3. Laminin α1 in ocular development

Laminin α1 is present in the two BMs around the retina – in the inner limiting membrane (ILM) separating the retina from the vitreous, and in the Bruch's membrane between the retina and choroid (Byström et al. 2006; Libby et al. 2000). Homozygous *Lama1^{nmf223}* mutant mice carrying the non-synonymous mutation Y265C in the short arm of laminin α1 (which affects LN domain protein-protein interactions) lack primary retinal vascular plexus and exhibit persistent vasculature in the vitreous due to ectopic entry of retinal vessels and astrocytes through the inner limiting membrane of the eye (Edwards et al. 2010; Edwards et al. 2011). These abnormalities are most likely caused by defects in the organization of the ILM which affect its interaction with Müller cells end-feet– the main glial cells in the retina. Notably, *Lama1$^{Δ/Δ}$* mutants with *Sox2-Cre*-mediated conditionally-deleted *Lama1* display similar but more severe defects in retinal cell number and vasculature correlating with a complete absence of the ILM and disorganized Muller cells end-feet. Interestingly, at 1 year of age, *Lama1^{Δ}* homozygotes feature 11% increase in eye diameter compared to wild type littermates (Edwards et al. 2010; Edwards and Lefebvre 2013). This is very intriguing, as it relates to a study by Zhao et al. (2011) who identified a significant association between high myopia in Chinese patients and the rs2089760 SNP located ~ 1.2 kb upstream of the transcription start site of *LAMA1*, presumably in a putative promoter-proximal regulatory element of *LAMA1* (Zhao et al. 2011). Thus, one can envisage a scenario where reduced expression of laminin α1 in the Bruch's membrane or in the sclera results in abnormal eyeball size and shape.

Interestingly, the larval-lethal zebrafish *bashful a69* (*bal^{a69}*) mutants, which carries a C56S point mutation in the N-terminal region of laminin α1, feature a plethora of ocular defects such as lens degeneration, corneal dysplasia, dismorphic hyaloid vasculature, ectopic photoreceptors in the inner retina, and defects in axonal projections of retinal ganglion cells (Zinkevich et al. 2006; Semina et al. 2006). Notably, the defects in the anterior chamber appear result from reduced focal adhesion kinase (Fak) activation and are not a consequence of the degenerating lens (Semina et al. 2006).

1.7.4. Laminin α1 and kidney development

Lama1 mRNA is expressed in the developing mouse mesonephros (Anderson et al. 2009; Miner et al. 2004) and metanephros (Miner et al. 1997; Sorokin et al. 1997), where it is transiently expressed by both the nephric duct and the nephrogenic mesenchyme, and is exclusively included in the basement membranes of the epithelial ureteric tree and nephric tubules but not in the BMs of kidney endothelia (Sorokin et al. 1997). Formation of the mammalian definitive kidney, the metanephric kidney, begins at stage E11, by evagination of the ureteric bud from the caudal end of the nephric (or Wolffian) duct, which invades the nearby metanephric mesenchyme and induces it to condense into the tubular epithelium of nephrons (Sorokin and Ekblom 1992). Interestingly, this mesenchyme-to-epithelium transition is accompanied by the formation of a new basement membrane around the forming tubules that predominantly contains laminin-111 but also laminin-411 (Sorokin et al. 1997). Importantly, incubation of mouse metanephric explants with antibodies against the laminin α1 subunit blocked the mesenchyme-to-epithelium transition of the nephrogenic mesenchyme, either by interfering with laminin network assembly or perturbing the interactions of laminin α1 with cellular receptors (Klein et al. 1988).

Interestingly, mutant mouse embryos harbouring a deletion in the nidogen-binding site on laminin γ1 display perturbation in the anterior-to-posterior elongation of nephric ducts, resulting in renal agenesis in 80% of the mutants (Willem et al. 2002). Other laminin isoforms are also prominently expressed in developing kidneys and have demonstrated roles at different stages of nephrogenesis. For instance, laminin-511 appears to be critical for glomerulogenesis (Miner and Li 2000), and together with laminin-3A32 are both required for ureteric bud branching (Miner and Li 2002; Zent et al. 2001). Thus, various laminin isoforms participate at distinct steps of kidney development.

1.7.5. Laminin α1 in skeletal muscle development

Laminin α1 is essential for early myotomal development in the mouse (Anderson et al. 2009). However, expression data suggest that it may also have some hitherto unknown functions in later skeletal muscle morphogenesis (Patton et al. 1997), whereas current research in our laboratory is uncovering unexpected roles of laminin α1 in modulating satellite cell behavior (Shantisree Rayagiri's Thesis), thus challenging former studies reporting its absence in adult murine muscles (Falk et al. 1999).

Although detailed characterization of laminin α1 functions in mammalian skeletal muscle development awaits future studies, *bashful* mutants and *lama1*-morpholino-treated wild type zebrafish embryos already provide some insights in this respect. Zebrafish *bal^{uwl}* mutants carry a 100 bp insertion in the *lama1* locus resulting in a frame-shift and predicted truncation of the laminin α1 chain at amino acid residue 1424 (Semina et al. 2006). Similar to other *bashful* mutants, 24 hpf *bal^{uwl}* homozygous embryos exhibit defects in cell differentiation in the anterior notochord, followed by curved tail and shortening of the body axis in 5 dpf larvae (Paulus and Halloran 2006). Interestingly, the overall gross morphology of skeletal muscles in *bal^{uwl/uwl}* fish appears normal, further evidenced by the normal expression of *gli1*, *gli2*, *myoD* and slow myosin gene expression in the somites, although some individual fibers are mis-oriented. However, injection of morpholinos against *lama1* in wild type fish leads to significant detachment of muscle fibres from the myotendinous junctions, and the defects are even stronger on a *lama2^{-/-}* background, revealing overlapping functions between laminin-211 and laminin-111 in zebrafish myofibre attachment to the extracellular matrix (Sztal et al. 2012).

An interesting role for laminin γ1-containing laminins in skeletal muscle development was uncovered in the *sleepy* (*sly*) zebrafish mutants (Dolez et al. 2011). *Sly* mutant alleles encode severely truncated laminin γ1 chains that lack most of the amino acid sequence after the ~ 100th residue (the predicted total length of zebrafish laminin γ1 is 1593 aa), and result in notochord cell apoptosis and shortened body axis, that are more severe than in *bal* mutants (Parsons et al. 2002). In addition, *sly* mutants display failure of *engrailed* expression in muscle pioneers and medial fast fibers due to the ectopic activation of Bmp signalling in the central adaxial domain, and this appears to be mediated by promoting the incorporation of heparan sulphate proteoglycans (HSPGs) in the BM on the medial surface of the somite (Dolez et al. 2011). Thus, laminin γ1, most likely incorporated in laminin-111, -211 and -411 (Szstal et al. 2011), creates a permissive environment for Shh-dependent induction of *engrailed* expression in the zebrafish somite (Dolez et al. 2011; Maurya et al. 2011)

1.8. The SHH signaling pathway

As described earlier, Sonic hedgehog (SHH) was found to be an essential signal for activating *Lama1* transcription in the somites and neural tube of mouse embryos (Anderson et al. 2009). *Shh* belongs to the hedgehog (hh) family of genes whose origin lies at the dawn

of Eumetazoa (Adamska et al. 2007; Ingham et al. 2011). The genome of *Drosophila* encodes only one HH protein that was originally characterized as essential for establishment of segment polarity in the fly embryo (Hidalgo and Ingham 1990). The amniote HH family includes two more paralogs apart from SHH – indian hedgehog (IHH) and desert hedgehog (DHH) (Bitgood et al. 1996; Chang et al. 1994; Echelard et al. 1993; Marigo et al. 1995), while the zebrafish genome harbours five hh genes: *shha*, *shhb* (or *tiggywinkle hedgehog, twhh*), *ihha*, *ihhb* (or *echidna hedgehog*, *ehh*), and a *dhh* ortholog (Avaron et al. 2006; Currie et al. 1996; Ekker et al. 1995; Krauss et al. 1993).

The SHH signalling pathway controls, directly or indirectly, the development of most vertebrate organs, and de-regulation leads to congenital defects, as well as to tumour formation/progression in humans (Bale 2002; Ingham and McMahon 2001). Notable examples of SHH-regulated processes in vertebrate development include cell-fate specification in the somite (Borycki et al. 1998; Wolff et al. 2003), craniofacial morphogenesis (Hu and Helms 1999), dorso-ventral regionalisation of the spinal cord (Dessaud et al. 2008; Ericson et al. 1995), antero-posterior patterning of the limbs (Johnson et al. 1994; Lopez-Martinez et al. 1995), gut wall patterning (Ramalho-Santos et al. 2000), ventral midline patterning of the prosencephalon (Ekker et al. 1995), tooth formation (Dassule et al. 2000), growth of external genitalia (Seifert et al. 2010), and many others (Huangfu and Anderson 2005).

1.8.1. Expression, secretion and diffusion of SHH

During vertebrate embryogenesis, SHH is expressed at various sites with important organizer-like properties, such as the notochord, floor plate, the zone of polarising activity (ZPA) in the limb bud, and the frontonasal ectodermal zone (FEZ) (Ericson et al. 1995; Hu and Marcucio 2009; Johnson et al. 1994). SHH usually acts as a morphogen at a distance from its secretion source, and this role is particularly well documented in dorso-ventral patterning of the neural tube, as discussed later (Ingham and McMahon 2001).

SHH is translated as a ~ 45 KDa pro-protein, that undergoes autocleavage to produce two polypeptides - a carboxy-terminal (SHH-C) and an amino-terminal one (SHH-N) (Lee et al. 1994). SHH-N carries all signalling activity of SHH and is palmitoylated in the N-terminus by the HHAT acyl transferase, whereas its C-terminus is covalently-linked to a cholesterol moiety (Beachy et al. 1997; Chen et al. 2004). Studies in mutant mouse embryos

show that release of SHH-N depends on the transmembrane protein dispatched 1 (DISP1), and studies in *Drosophila* demonstrate that the lipid modifications on HH are essential for its secretion and formation of diffusible multimeric HH complexes, and for its signaling potency and range of action (Caspary et al. 2002; Chen et al. 2004; Kawakami et al. 2002; Lewis et al. 2001). Once released from its source, SHH distribution is modulated by binding to its receptor – the twelve-transmembrane-pass protein Patched (PTCH), and to the vertebrate Hedgehog-interaction protein HHIP (which is absent in *Drosophila*), both restricting hedgehog's diffusion range (Bishop et al. 2009).

Mechanistically, hedgehog ligands exert their effects by blocking the generation of repressor forms of the transcription factor cubitus interruptus (CI) in *Drosophila*, and its homologs – the glioblastoma-associated (GLI) factors – GLI2 and GLI3, in amniotes, as described in more detail below (Alexandre et al. 1996; Aza-Blanc et al. 1997; Sasaki et al. 1999).

1.8.2. Transduction of the HH signal

The core of the canonical SHH signalling cascade is similar in flies and vertebrates. Below is presented a description of the *Drosophila* HH pathway with comments on its differences in vertebrates. In the absence of hh ligand, the transmembrane receptor patched (PTCH) inhibits the heptahelical transmembrane protein smoothened (SMO and prevents its localization to the plasma membrane (Beachy et al. 2010), perhaps by modulating the levels of phosphatidylinositol-4-phosphate (PI4P) (Yavari et al. 2010). Smoothened is essential for activation of hh signaling in both flies and vertebrates (van den Heuvel and Ingham 1996; Varga et al. 2001; Zhang et al. 2001). In these conditions, the serine/threonine kinases protein kinase A (PKA), glycogen synthase kinase 3 (GSK3) and casein kinase I (CKI) phosphorylate and target Cubitus Interruptus, or GLI in vertebrates, for proteasomal processing into a repressor form – CI-R, or GLI-R, respectively (Jia et al. 2005; Price and Kalderon 2002; Tempe et al. 2006). These repressor forms enter the nucleus where they silence the expression of SHH-target genes.

A key role in recruiting the kinases to CI/GLI is played by the motor protein Costal 2, COS2, (or KIF7 in vertebrates), which forms a complex with SMO, CI/GLI, PKA, GSK3 and CKI in the absence of hh (Endoh-Yamagami et al. 2009; Sisson et al. 1997). However, binding of HH to PTCH, which is assisted by the transmembrane proteins IHOG and BOI

(and their vertebrate orthologs - CDO and BOC, respectively) (Tenzen et al. 2006; Zheng et al. 2010), allows SMO to be translocated to the cell surface (Liu et al. 2007). This translocation appears to depend on the activity of another serine/threonine kinase – fused, FU, that phosphorylates COS2 at Ser572 resulting in the dissociation of CI/GLI from the COS2-PKA-GSK3-CKI complex and preventing the generation of CI-R/GLI-R (Liu et al. 2007; Ruel et al. 2007). Upon HH reception, Fused also phosphorylates Suppressor-of-fused, SUFU, permitting the release of full length CI/GLI from an inhibitory CI/GLI-sufu complex in the cytoplasm, and its subsequent translocation into the nucleus where it activates HH-responsive genes (Methot and Basler 2000; Wang et al. 2000).

Despite the described similarities with *Drosophila*, vertebrate HH signaling features some differences. For instance, instead of a single *Ptch* gene, vertebrate genomes encode two paralogs – *Ptch1* and *Ptch2* (Carpenter et al. 1998; Lewis et al. 1999; Motoyama et al. 1998). Also, vertebrate SUFU is required for GLI processing into a repressor form via recruitment of GSK3β to the PKA-CKI-KIF7 complex, in addition to inhibiting the release and nuclear translocation of full-length GLI (Cooper et al. 2005; Humke et al. 2010; Kise et al. 2009). Importantly, vertebrate HH signaling depends on the primary cilium (Corbit et al. 2005; Kim et al. 2010). In the absence of SHH, patched is present in the cilium where it blocks the ciliary accumulation of smoothened, while when present SHH causes eviction of patched from the cilium and accumulation of SMO and the GLI-SUFU complex, followed by release of full-length GLI (Rohatgi et al. 2007; Tukachinsky et al. 2010). The SHH-triggered translocation of GLI to the cilium requires KIF7, whereas in the absence of SHH, KIF7 localises to the basal body of the cilium which is enriched in PKA and proteasomal complexes (Ingham et al. 2011; Liem et al. 2009). In summary, the organisation of the HH signaling cascade is largely conserved between flies and vertebrates, where HH reception by patched blocks its inhibitory effect on smoothened which prevents the formation of CI-R/GLI-R forms, hence allowing the generation of CI-A/GLI-A and HH-target gene activation (Huangfu and Anderson 2005).

1.8.3. The GLI transcription factors are mediators of HH signaling in vertebrates

The mammalian genome encodes three paralogs of the GLI family – GLI1, GLI2 and GLI3 which are Zn-finger domain-containing DNA-binding transcription factors (Matise and Joyner 1999). GLI2 and GLI3 are bifunctional as they exist as either activator or repressor

forms depending on the presence or absence of HH ligands, whereas GLI1 possesses only activator properties (Hui and Angers 2011). The repressor ability of GLI2 and GLI3 is determined by an N-terminal repressor domain, which is absent in GLI1, while all three proteins have a C-terminal trans-activation domain, with the Zn-finger DNA-binding domain located in between (Ruiz i Altaba et al. 1999; Sasaki et al. 1999). In addition, GLI2 and GLI3 harbour, like CI, a processing determinant domain (PDD) which is necessary for limited proteolysis and generation of GLI-R forms (Smelkinson et al. 2007; Wang and Li 2006).

Genetic studies in mice have demonstrated that GLI2 and GLI3 are the major mediators of SHH signaling in mammals and are essential for normal embryogenesis, with GLI2 acting primarily as an activator, and GLI3 acting primarily as a repressor of HH target genes (Ding et al. 1998; McDermot et al. 2005; Motoyama et al. 1998; Persson et al. 2002). The cause for the predominantly activator function of GLI2 appears to be a less potent PDD than the one in GLI3, resulting in complete proteasomal degradation upon phosphorylation by PKA, GSK3-beta and CKI, such that very small amount of GLI2-R is produced (Pan and Wang 2007). *Gli1* gene expression is induced upon HH pathway activation, and is thought to amplify the target cell response to HH signaling, but appears to be dispensable for murine development (Bai et al. 2002; Bai et al. 2004; Park et al. 2000). However, in vertebrates like zebrafish and *Xenopus*, Gli1 rather than Gli2a/b is the main activator of Hh signaling in early development (Karlstrom et al. 2003; Lee et al. 1997). Interestingly, other post-translational modifications also modulate GLI factor activity. Deacetylation of acetylated GLI1 and GLI2 in NIH3T3 cells by HDAC1 promotes their trans-activatory potential (Canettieri et al. 2010), analogous to the effects of SUMO-ylation of GLI1/2/3 in the chicken neural tube catalyzed by the SUMO E3 ligase PIAS1 (Cox et al. 2010).

All three GLI transcription factors bind to DNA with similar affinity at the consensus motif 5'-TGGGTGGTC-3' using Cys_2His_2 Zn-fingers 4 and 5 of their DNA-binding domain (Kinzler and Vogelstein 1990; Hallikas et al. 2006; Pavletich and Pabo 1993). However, genome-wide studies of GLI occupancy in cell culture and *in vivo* revealed the presence of non-canonical GLI binding motifs, and that indirect binding of GLI to DNA mediated by protein-protein interactions is not unlikely (Vokes et al. 2007; Vokes et al. 2008). GLI target genes comprise both universal targets that are regulated in all HH-responsive tissues and these include *Ptch1*, *Ptch2*, *Gli1*, *Hhip, Boc* and *Cdo* (Chuang and McMahon 1999; Lee et al.

2010; Tenzen et al. 2006), and specific targets that are regulated in particular tissues only, such as the *Foxa2* gene in the floor plate, *Gremlin* and *Hand2* in the limb buds, and *Myf5* and *Lama1* in somites (Anderson et al. 2009; Borycki et al. 1999; Sasaki et al. 1997; Vokes et al. 2008).

1.8.4. SHH and dorso-ventral patterning of the neural tube

Initially secreted from the notochord, and then from the floor plate, the lipidated and multimeric SHH-N establishes a dynamic gradient directed along the dorso-ventral (DV) axis of the neural tube, where the spread and stability of SHH is negatively influenced by binding to atched and HHIP1, and positively affected by interactions with HSPGs, CDO, BOC and GAS1 (Echerald et al. 1993; Incardona et al. 2002; Martinelli and Fan 2007; Rubin et al. 2002). The ventral-to-dorsal gradient of SHH establishes six distinct progenitor domains in the ventral neural tube – floor plate, p3, pMN, p2, p1 and p0, by regulating the spatial expression of multiple homeodomain (HD) and basic helix-loop-helix (bHLH) transcription factors (Briscoe et al. 2000). The target genes of SHH signaling in the neural tube belong to two classes: Class I genes are repressed by SHH and these include *Pax3, Pax6, Pax7, Irx3* among others (Ericson et al. 1997; Mansouri et al. 1998), while Class II targets - like *Foxa2, Olig2, Nkx2.2, Nkx6.1, Dbx1, Dbx2* (Briscoe et al. 1999; Sander et al. 2000; Sasaki et al. 1997), are activated by SHH (Jessell 2000). Thus, graded SHH generates distinct expression domains of HD and bHLH factors with sharp boundaries, which are maintained by cross-regulatory interactions between the two classes of targets, and the combinatorial expression of different factors spatially defines the progenitor cell domains (Dessaud et al. 2008).

SHH patterns the neural tube via a gradient of GLI activity

Graded SHH signalling patterns the ventral neural tube by generating a parallel gradient of GLI activity, where GLI repressor potential is progressively attenuated, while GLI activator function is promoted in ventral-to-dorsal direction (Jacob and Briscoe 2003). Each GLI protein contributes to the sum gradient of GLI activity and this combined input determines the gene expression response of target cells to SHH signaling (Ribes and Briscoe 2013). Thus, the specific role of individual GLI factors is determined by their expression pattern and levels in the neuroepithelium, by their inherent transcriptional regulatory activity and by their post-translational regulation by SHH (Dessaud et al. 2008).

Expression of GLI 1, 2 and 3 in the murine neuroectoderm initiates slightly before E7.5 when all three genes are expressed uniformly along the dorso-ventral (DV) axis of the neural tube (Hui et al. 1994). At E9.5 GLI1 is restricted to the ventral half of the tube, GLI2 is expressed ubiquitously (but is excluded from the floor plate) with slightly higher expression in the dorsal neural tube, whereas GLI3 is restricted to intermediate and dorsal levels, and this pattern is maintained at later stages but limited to the ventricular zone (Hui et al. 1994; Lee et al. 1997; Sasaki et al. 1997). Zebrafish *gli* genes are similarly expressed along the DV axis of the neural tube. At early stages, *gli2a* and *gli3* mRNA transcription initiates throughout the neuroectoderm, but by 15 hpf they become restricted to the dorsal neural tube, while *gli1* expression displays a complementary pattern with high mRNA levels in the ventral neural tube and absence of expression in the floor plate and dorsal neural tube (Karlstrom et al., 1999; Karlstrom et al., 2003; Tyurina et al., 2005; Vanderlaan et al., 2005).

Interestingly, there are distinct requirements for GLI proteins during dorso-ventral patterning of the neural tube. In amniotes, GLI2 predominantly functions as a transcriptional activator and appears to be essential for generating the highest SHH signalling response for $Gli2^{-/-}$ mutant mouse embryos failed to specify the floor plate cells and V3 interneurons in the ventral-most domain of the tube (Ding et al. 1998). GLI1 appears more or less dispensable for neural tube patterning by SHH for *Gli1*-defficient mice are viable and display no behavioral abnormalities, indicating that its absence is compensated by another GLI factor (Park et al. 2000). In fact, *Gli1*;*Gli2* double mutants die perinatally and have multiple defects including loss of ventral spinal cord fates (Park et al. 2000), whereas replacement of *Gli2* with *Gli1* rescues the *Gli2*-null phenotype suggesting that it is the GLI activator function that is required for ventral NT patterning (Bai and Joyner 2001). In contrast, GLI3 operates mostly as a repressor for $Gli3^{-/-}$ embryos display dorsal expansion of intermediate neural progenitors (Persson et al. 2002), while combined inactivation of *Gli3* and *Shh* leads to normal specification of ventral progenitor domains except for the floor plate an V3 interneuron identities (Litingtung and Chiang 2000).

Thus, a model is proposed according to which SHH is required for motor neuron and V0-2 interneuron specification by removal of GLI3 repressor function resulting in de-repression of target genes, while patterning of V3 interneuron and floor plate identities by SHH depends on GLI2 activator inputs (Ribes and Briscoe 2009).

Dorso-ventral patterning of the zebrafish neural tube is also effected by Shh-regulated Gli transcription factor activity, although the exact role of each Gli factor is slightly distinct than in amniotes. In *gli1*-deficient fish (in the *detour* (*dtr*) mutants), induction of the cranial motor neurons fails, but spinal cord motor neurons are unaffected (Chandrasekhar et al. 1999; Karlstrom et al. 1996). The *yot* (*you-too*) mutants express a dominant repressor form of *gli2a* and feature severe reduction of motor neurons in the brain and spinal cord (Karlstrom et al. 1999). Application of *gli3* morpholinos leads to profound decrease in the number of cranial and spinal motor neurons, as well as retinal ganglional cells, and to ectopic activation of *gli1* and *fkd4* in the dorsal neural tube at 24 hpf (Tyurina et al. 2005; Vanderlaan et al. 2005). Taken together, these observations reveal that both Gli1 and Gli3 activator are required for ventral neural tube patterning in zebrafish, while Gli2a and Gli3 repressor forms are needed to oppose the activation of Shh target genes in the dorsal neural tube (Karlstrom et al. 2003; Tyurina e t al. 2005; Vanderlaan et al. 2005).

1.9. Transcription-regulatory elements

The precise control of developmental gene expression in time and space is ensured by the combinatorial binding of transcription factors to multiple, relatively short (100 bp - 1 kb) genomic regions called transcription-regulatory elements (TREs) (Lagha et al. 2012) which, apart from promoters, also include enhancers, silencers, insulators, locus control regions (LCRs), tethering elements, etc (Davidson 2006). Below is provided a brief description of the main classes of TREs, namely the core promoter, enhancers, silencers and insulators.

1.9.1. The core promoter

By definition, the core promoter is the minimal set of sequence motifs required for the initiation of gene transcription by RNA pol II (Watson et al. 2004). Core promoters can be focused or dispersed, differing in the number and distribution of transcription start sites: focused promoters initiate transcription from a single site or from a small cluster of start sites spread over few nucleotides; dispersed promoters harbour several start sites dispersed over a region of 50 – 100 base pairs (Muller and Tora 2013). Focused promoters are composed of a combination of sequence motifs including the TATA-box, Inr (initiator), BRE (TFIIB recognition element), DPE (downstream ptomoter element), MTE (motif ten element), and others, while dispersed promoters often correlate with CpG islands (Juven-Gershon et al.

2008). Importantly, core promoters serve as binding sites for the general transcription factors (GTFs) which recruit RNA polymerase II (RNA pol II), and together establish the pre-initiation complex poised for transcriptional initiation (Muller and Tora 2013).

However, efficient gene transcription *in vivo* often requires additional inputs from transcription factors bound to other regulatory elements, often located hundreds or thousands of base pairs away from the core promoter, which modulate the rate of transcription initiation through a variety of mechanisms (Watson et al. 2004).

1.9.2. Transcriptional enhancers

Transcriptional enhancers are classically defined as genomic regions that enhance transcription initiation from a promoter in a distance- and orientation-independent manner when tested on plasmids constructs (Bulger and Groudine 2011), although there are examples of enhancer-like elements whose activity is orientation-dependent or diminishes with increasing the distance to the target promoter (Hozumi et al. 2013; Swanson et al. 2010). Enhancers are located in the introns of their target genes, or upstream or downstream of the target promoters, often at huge distances – up to 1 Mb in the case of the ZPA-enhancer of the mammalian *Shh* gene (Lettice et al. 2003). Enhancer elements have diverse sequences, which in many cases feature high substitution rates during evolution (Taher et al. 2012), and serve as combinatorial-binding platforms for tissue-specific transcription factors (Laga et al. 2012). As such, transcriptional enhancers are central components of developmental gene regulatory networks and play important role in the evolution of organismal complexity (Davidson 2006; Levine 2010; Wittkop and Kalay 2011).

Mechanistically, enhancers modulate the transcription rate of their target genes through direct or indirect interactions of the enhancer-bound transcription factors with the transcription machinery at the core promoter (Faro-Trindade and Cook 2006; Fuda et al. 2009). There is also evidence that transcriptional activators at enhancers can recruit histone modifying and chromatin remodelling complexes that create a transcription-permissive environment by changing nucleosome conformation and distribution along DNA (Bulger and Groudine 2011), and that many enhancers are themselves transcribed as non-coding RNAs (Natoli and Andrau 2012). Further details on enhancer function and evolution are provided in the discussions of the result chapters of this study.

1.9.3. Transcriptional silencers

Transcriptional silencers are regulatory elements bound by transcriptional repressor proteins that negatively affect expression of their target genes by either directly interfering with the transcription machinery at a promoter, or by suppressing the activity of some enhancers (Ogbourne and Antalis 1998), which is often achieved by the induction of a repressive chromatin state at the target gene (Gaston and Jayaraman 2003; Talbert and Henikoff 2006). Similarly to enhancers, silencers can be located up- or downstream, or in the introns of their target genes, and some exhibit orientation-dependent behaviour (Shei and Broach 1995; Ogbourne and Antalis 1998). Well studied silencer elements are the PREs (Polycomb Response Elements), associated with the repressive function of the polycomb group proteins (Bantignies and Cavalli 2011), and the NRSE (Neuron-Restrictive Silencer Element) that binds the REST repressor and is involved in the silencing of neuronal gene transcription (Mori et al. 1992) (discussed in Section 7.4.5)

1.9.4. Insulators

For precise regulation of developmental gene expression it is essential that the enhancers of a particular gene do not interfere with the expression of other genes nearby. This is ensured by the CTCF (CCCTC-binding factor)-bound insulator elements which are a class of regulatory sequences that are able to insulate the promoter of a gene from the activity of nearby enhancers when located between the promoter and enhancers (Gaszner and Felsenfeld 2006), while other insulators act as boundary elements that prevent the spreading of repressive or activating chromatin states in neighbouring genomic regions with opposite chromatin states (Essafi et al. 2011). One of the models of enhancer-blocking insulator function postulates that insulators prevent the activation of neighbouring genes by the enhancers of another gene through the formation of chromosome loops where several insulator elements within a genomic region interact with each other (Herold 2012). There is evidence that the change in chromosome conformation is facilitated by the cohesin complex (Wendt et al. 2008), which can physically interact with CTCF (CCCTC-binding factor) and is essential for long-distance chromatin contacts (Nativio et al. 2009). According to this model, one of the insulators must be positioned in between the enhancer and the insulated genes (Gaszner and Felsenfeld 2006), which brings the enhancer in a chromosome loop that, by some unclear mechanism, prevents it from acting on genes located outside of the loop. This is the mechanism thought to

block expression of the maternal copy of the *Igf2* gene in mice (Kanduri et al. 2000; Kurukuti et al. 2006) and the process thought to mediate the function of the Fab insulators in the *Drosophila* bithorax complex of homeotic genes (Kyrchanova et al. 2011; Maeda and Karch 2006). In addition, there is also evidence that CTCF-bound insulators can physically associate with enhancers (Handoko et al. 2011) presenting another possible mechanism of enhancer-blocking activity by insulators.

1.10. Methods for the prediction of transcription-regulatory elements

The prediction of transcription-regulatory elements is generally more difficult compared to prediction of the coding portions of the genome due to the relatively small size, lack of ATG/stop codon marks and exon-intron boundaries, and the low level of sequence conservation in regulatory elements (Elgar 2009). However, several approaches exist, grouped in two categories – direct and indirect, which allow the prediction of putative TREs (Nelson and Wardle 2013).

1.10.1. Direct (biochemical) methods

Currently, the most frequently employed direct approaches include chromatin immunoprecipitation (ChIP) and DNaseI hypersensitivity (DHS) assays which, when combined with next-generation DNA sequencing methods, enable genome-wide prediction of TREs. Both methods are described in more detail below.

1.10.1.1. Prediction of TREs based on transcription factor occupancy and/or histone modifications as revealed by ChIP-seq

Mechanistically, TREs serve as scaffolds for the combinatorial binding of transcription factors to short binding motifs within the nucleotide sequence of the element (Davidson, 2006). The detection of such motifs in a sequence of interest forms the basis of the computational prediction of TREs. However, such prediction is not an indication of biochemical potential or the ability of the motif to be recognised and bound by its cognate transcription factor *in vivo*. Moreover, the binding of a particular TF to its motif depends on the tissue type and/or developmental stage (Levine, 2010). Chromatin immunoprecipitation allows for the identification of transcription factor binding sites of interest through-out the whole genome (ChIP-seq) in different tissues, stages and conditions (Park 2009). However, this method requires the availability of tested antibodies against the transcription factor of

interest. Due to the limited availability of GLI antibodies, ChIP is therefore not a preferred method for the prediction of *Lama1*'s TREs, in regard to this study.

However, the ChIP-seq method can be used to catalogue the genome-wide distribution of specific post-translational histone modifications like lysine mono-, di- or trimethylation on histone 3 and histone 4 (Rea et al. 2000; Martin et al. 2005), lysine acetylation on on all four core histones (Hebbes et al. 2000), serine phosphorylation on histone 3 (Mahadevan et al. 1991; Nowak et al. 2004), mono-ubiquitylation of H2A and H2B (Wang et al. 2004), etc. Importantly, many of these modifications are inherently plastic, or reversible, with specifically dedicated enzymes performing the opposite reactions (Shi et al. 2004; Kurdistani et al. 2003). Moreover, the same locus can harbour different histone marks in different tissues or developmental stages (Rada-Iglesias et al. 2011; Cotney et al. 2012). In effect, the combinatorial pattern of diverse histone marks is thought to substantially affect the regulation of gene transcription and chromosome organisation (Stock et al. 2007; Munshi et al. 2009).

Several studies have demonstrated that particular genomic regions like promoters, coding sequences and enhancers are enriched for certain types of histone modifications (Kimura 2013). For instance, the promoters of activated genes are generally marked by H3K27ac (acetylation of lysine 27 in histone 3) and H3K4me3 (tri-methylation of lysine 4 in histone 3) (Heintzman et al. 2007); the bodies of actively transcribed gene loci associate with H3K36me3 (Wagner 2012); active enhancers are enriched for H3K27ac and H3K4me1 (Rada-Iglesias et al. 2011; Spicuglia et al. 2012), while silenced regions are marked by H3K9me3 or H3K27me3 (Kim et al. 2012; Young et al. 2011). Thus, these modifications can serve as markers for the prediction of un-annotated regulatory elements, which is clearly demonstrated by the identification of 2489 putative melanocyte enhancer elements in a ChIP-seq assay for H3K4me1 enrichment in murine melanocytes (Gorkin et al. 2012).

In addition, the ChIP-seq method can also uncover enhancers characterised by the recruitment of the p300/CBP histone acetyltransferase co-activator to DNA-bound transcriptional activators (Imhof et al. 1997; Holmqvist et al. 2013). Consistent with this, a ChIP-seq assay of p300 binding in E11.5 mouse embryo heart tissue identified more than 3, 000 putative enhancers which show shallow evolutionary conservation and could not be isolated by phylogenetic footprinting (Blow et al. 2010). Remarkably, *lacZ* reporter

transgenesis revealed that 81 of the 130 tested sequences were active in the developing heart, clearly indicating the utility of using p300 as a marker of putative enhancer regions, especially divergent ones (Blow et al. 2010). Thus, unexplored genomic regions enriched for p300 binding are good candidates for enhancer function (Visel et al. 2009).

In summary, the ChIP-seq approach is a powerful method that is not only restricted to the confirmation of direct binding of transcription factors to their targets *in vivo*, but can also be used to identify previously unknown tissue-specific regulatory elements.

1.10.1.2. Prediction of TREs in DNaseI hypersensitive domains

A complementary approach for whole-genome discovery of putative regulatory elements is the DNaseI hypersensitivity (DHS) assay combined with next-generation sequencing platforms − DNase-seq (Boyle 2008). The DNaseI hypersensitivity assay relies on the identification of genomic regions with less compact nucleosome configuration which renders DNA more accessible (hypersensitive) to digestion by DNaseI (Ballare et al. 2013) as compared to DNA in tightly packed chromatin. It is believed that such "open chromatin" regions mark functionally active sites in the genome involved in the regulation of gene transcription (Thurman et al. 2012). In fact, multiple studies have shown that DNaseI hypersensitive sites correlate with various functional types of TREs. For instance, an extensive survey for hypersensitive sites in the intergenic region upstream of the *IL-3* gene in humans, identified a complex cluster of both constitutive and inducible hypersensitive sites which corresponded to hitherto unknown enhancers of *IL-3* that are required for precise *in vivo* expression of this gene in T-cells, myeloid progenitors and mast cells (Baxter et al. 2012). Other studies have revealed that DNaseI hypersensitivity marks not only enhancer and promoter elements but also locus control regions (Fu et al. 2002; Kim et al. 2010; Kim et al. 2013), silencers (Feng et al. 2005; Zarnegar et al. 2010) and insulators (Chen et al. 2001; Sultana et al. 2011; Follows et al. 2012).

Taken together, these observations illustrate the utility of DHS assays for uncovering transcriptional regulatory elements. Moreover, meta-analysis of genome-wide DNaseI hypersensitivity and ChIP studies have revealed significant overlap between hypersensitive sites and various histone modifications (Shu et al. 2011). This demonstrates that the combination of data from both DHS and ChIP assays is a powerful strategy for the prediction of developmentally relevant transcriptional regulatory elements.

1.10.2. Indirect methods

The indirect methods for prediction of transcription-regulatory elements rely on computational analyses *in silico* using genomic sequence information alone. There are two main approaches to *in silico* analyses. The most popular and straightforward strategy, phylogenetic footprinting, consists in performing multi-species genomic comparisons with the aim of identifying discrete evolutionary-conserved non-coding regions, which might possess characteristics of TREs (Blanchette and Tompa 2002). The other strategy employs a relatively novel technique which does not require sequence conservation, but predicts putative TREs using a model of the pattern of transcription factor binding sites (TFBSs), or motif grammar, that defines a particular functional class of regulatory elements (Spitz and Furlong 2012), as described below.

1.10.2.1. Computational prediction of TREs based on sequence organisation/pattern, irrespective of overall conservation

An increasing number of studies report on TREs with divergent sequence yet retaining similar or identical function across species (Fisher et al. 2006; Nelson and Wardle 2013). Such elements evade prediction by the methods based on evolutionary conservation of sequence, like phylogenetic footprinting (Blow et al. 2010), described later. In recent years however, new computational developments were designed to address this problem building on the idea that each TRE features a specific organisational pattern or architecture represented by a particular composition, number, order and spacing of TFBSs (Davidson 2006; Evans et al. 2012). In the cases where functionally conserved TREs lack overt sequence conservation, they may still retain corresponding individual TFBSs or clusters of TFBSs. This allows one to derive a picture of the pattern of TFBSs or "regulatory grammar" that characterises a particular functional class of TREs (Senger et al. 2004), which then can be used to predict novel TREs by probing the genome for the existence of regions with similar organisation. This approach has been successfully deployed in a small number of studies. For instance, Senger et al. (2004) performed SELEX (**S**ystematic **E**volution of **L**igands for **Ex**ponential **E**nrichment) assays to characterise the TFBSs of REL and GATA factors (two classes of transcription factors involved in the activation of innate immunity genes in *Drosophila*) followed by computational analysis of the 5'-flanking regions of 50 of the innate immunity genes. Interestingly, it was revealed that more than half of these genes

harbour a shared REL-GATA module with fixed/constrained organisation (distance and orientation between the REL and GATA motifs), features that appear to be essential for the regulatory activity (Senger et al. 2004).

In another genome-wide comparison between human and zebrafish combined with TFBSs-pattern identification analysis, Taher et al. (2011) demonstrated the existence of hundreds of divergent but corresponding non-coding regions in the human/zebrafish genome. Importantly, despite the lack of overt sequence conservation, the pairs of human/zebrafish elements displayed shared pattern of TFBSs architecture, suggestive of a conserved function. Consistent with this idea, many of the human sequences overlapped with sites where the transcriptional coactivator p300 is enriched, strongly hinting for roles in transcriptional regulation. In fact, 8 out of 18 human sequences displayed tissue-specific enhancer activity in transgenic zebrafish embryos, which was remarkably similar to the activity displayed by their orthologous sequences from zebrafish. Thus, a computational approach, based on the presence of common regulatory encryption, is able to successfully uncover functionally-conserved cryptic TREs (called "covert" elements) despite the lack of overall sequence similarity (Taher et al, 2011).

However, it is also true that many functionally-similar transcription-regulatory elements lack recognizable motif grammar; instead, they display considerable variation in the content, order, spacing and orientation of the TFBSs, yet generating the same output, as is the case with the regulation of the *RET* gene in human and zebrafish (Fisher 2006), the cardiac TREs in *Drosophila* (Junion 2012) and the promoters of ribosomal protein (RP) genes in yeasts (Hogues 2008). The latter case is particularly striking for *S. cerevisiae* uses a completely different set of transcription factors to regulate its ribosomal gene promoters compared to *C. albicans*. In the former species, the promoters of the RP genes are controlled by Rap1, Fhl1 and Ifh1, while in the latter the same function is performed by Tbf1 and Cbf1 (Hogues 2008).

Altogether, this flexibility in TFBSs composition and configuration renders many TREs difficult to predict computationally, especially in the absence of sequence conservation (Bery 2013).

1.10.2.2. Comparative genomics and TRE prediction

Comparative genomics consists in the alignment of genomic sequences from two or several different species, and relies on the conservation of functionally-relevant regions. This strategy is widely applied in the annotation of coding genes in newly-sequenced genomes, and in the reconstruction of organismal phylogenies (Ureta-Vida et al. 2003). Notably, comparative genomics can be used for the prediction of transcription-regulatory elements, in a procedure known as "phylogenetic footprinting". The latter method consists in aligning orthologous genomic regions from different species which results in the identification of evolutionary conserved non-coding elements (or CNEs) in the sequences. The phylogenetic footprinting rests on the classic molecular evolution paradigm that mutations (base-pair substitutions, indels, translocations) in functionally-important genomic regions are likely to be deleterious and subject to negative (purifying) selection leading to lower rates of change compared to functionally-neutral sequences (Blanchette and Tompa 2002). In effect, the conserved elements appear as "footprints" in the neutrally-evolving background (Zhang and Gerstein 2003).

The output of a phylogenetic footprinting analysis strongly depends on the choice of species for comparison (comparators), or the scope. Comparisons between highly divergent species (for instance, between mouse and zebrafish, whose last common ancestor lived more than 419 million years ago (Zhu et al. 2009)) would yield elements that are shared across deep phylogenies. Conversely, in a variant approach called "phylogenetic shadowing" (Berezikov et al. 2005), genomic alignments between closely related species could be used to uncover short species-specific differences (instead of similarities) embedded in long stretches of highly homologous sequence. Importantly, a phylogenetic footprinting study which combines both phylogenetically close and disparate comparators allows for the identification of clade-specific conserved non-coding elements (CNEs) (Müller et al. 2002).

Phylogenetic footprinting analyses can be deployed at different scales - genome-wide (Woolfe et al. 2005; Pennachio et al. 2006), chromosome-specific (Royo et al. 2011) or gene-centric (Uchikawa et al. 2003; Navratilova et al. 2009; Sato et al. 2012). The number of CNEs that is usually retrieved in a phylogenetic footprinting study depends on the scale of analysis, the phylogenetic distance between the compared species and the stringency parameters of the alignment algorithm and can vary from tens – in gene-centric alignments,

to several hundred thousand – in whole-genome alignments of placental mammals (Haeussler 2011; Bejerano et al. 2004; Visel et al. 2007).

The phylogenetic footprinting approach is advantageous over other methods for TRE prediction (such as ChIP-seq, DNaseI HS and the motif grammar-based methods) as it does not require prior knowledge on transcription-factor binding or the design of complicated motif-grammar algorithms, and can be used with pre-computed whole-genome alignments (Ureta-Vidal et al. 2003), thus enabling rapid identification of conserved elements. A major limitation of the phylogenetic footprinting is the difficulty or inability to identify functionally-conserved TREs that are highly divergent in sequence (Nelson and Wardle 2013). Consequently, phylogenetic footprinting may fail to detect rapidly-evolving, but still genuine TREs. Nevertheless, this method has been effectively used in several model species for the prediction of enhancers and other regulatory elements that were subsequently functionally validated *in vivo*, as described in Section 4.2.

1.11. Functional validation of candidate transcription-regulatory elements

Once a candidate transcription-regulatory element has been predicted either biochemically or computationally, the next step is validation of its putative activity. This can be achieved using reporter gene constructs in cell culture and/or in transgenic animals (*in vivo*) (Carey and Smale 2000). The cell culture approach, together with its advantages and limitations, is discussed in Chapter 5.

The best environment for functional validation of a predicted TRE is the whole embryo, ideally from the same biological species as the species-of-origin of the candidate element. This is so, for the whole embryo could potentially provide all signalling molecules and transcription factors necessary for the activity of the candidate TRE, and thus to enable the examination of the element's function in specific tissues and developmental stages (Haeussler and Joly 2011).

1.11.1. Zebrafish as a tool for the validation of candidate TREs

In order to examine *in vivo* the putative regulatory function of a genomic region, a CNE for instance, the tested element is cloned upstream of a minimal promoter driving a reporter gene, usually the bacterial *lacZ* gene (McGregor et al. 1991), or the *GFP* gene from the hydrozoan *Aequorea victoria* (Amsterdam et al. 1995; Chalfie et al. 1994), followed by

micro-injection the *CNE::reporter* constructs in 1-cell stage embryos. The ideal strategy for candidate mammalian TREs is to inject a *CNE::lacZ* construct in mouse oocytes and then assay for *lacZ* activity in the transgenic embryos (Nagy et al. 2003).

However, the latter approach is both laborious and expensive especially for screening a large number of elements. Moreover, the analyses can only be performed at one temporal stage at a time due to the intra-uterine nature of mouse development. Therefore, a faster and financially less demanding strategy is necessary for the initial screening of mammalian elements. Such an alternative is the analysis of these elements in transiently transgenic zebrafish embryos where a tested element regulates the expression of a fluorescent protein reporter gene, such as GFP (Amsterdam et al. 1995). Although this strategy leads to mosaicism, the reproducibility in reporter gene expression, combined with the large number of fertilized eggs, the extra-maternal nature of early zebrafish development, and the optical transparency of the embryos, renders monitoring the activity of putative regulatory elements possible in real time and across multiple stages (Higashijima et al. 1997; Nüsslein-Volhard and Dahm 2002).

Numerous studies have exploited this approach to uncover enhancers of developmental genes in mice and humans (Sacilotto et al. 2013; Oksenberg et al. 2013; Tamplin et al 2011; Hernandez-Vega et al. 2011; Ghiasvand et al. 2011; Amigo et al. 2011; Shin et al. 2005; Abbasi et al. 2007). For instance, a GFP-based reporter screen in transient transgenic zebrafish and mouse embryos revealed neural tube-specific enhancer activity of conserved intronic regions in the autism-associated human *AUTS2* gene (Oksenberg et al. 2013). Another investigation into the notochord-specific targets of the FOXA2 transcription factor in mice combined cell sorting, microarray-based gene expression screening and ChIP to identify the *cis*-regulatory modules that bind FOXA2 (Tamplin et al. 2011). The putative mouse regulatory regions were subsequently tested for their ability to direct notochord-specific GFP reporter expression in zebrafish embryos which resulted in the identification of 7 elements, whose activity requires intact FOXA2 sites. Importantly, these mouse enhancers are not conserved in zebrafish and could not have been detected on the basis of phylogenetic sequence comparisons (Tamplin et al. 2011). Thus, despite the lack of sequence conservation, murine regulatory elements were able to respond to the notochord-specific *trans*-environment in the zebrafish embryo and directed reporter expression according to the

40

expected pattern. This study, together with the discovery of functionally conserved but sequence-divergent enhancers of the *RET* gene in human and zebrafish (Fisher et al. 2006), illustrates well the power of the zebrafish model system to analyse the putative activity of non-conserved mammalian elements.

Importantly, the preliminary validation of a tissue-specific transcriptional enhancer in transiently transgenic zebrafish must be supported by a demonstration of its activity in stable lines (Ishibashi et al. 2013). Usually, several lines are generated and analysed for a common issue in enhancer screens is the "variability of position effects" (Roberts et al. 2014), that is the variation in reporter's expression pattern among transgenic lines due to different integration sites of the enhancer::reporter construct (Roberts et al. 2014; Ishibashi et al. 2013). The latter phenomenon is a combined result of: 1) the potential influence of cryptic TREs nearby the integration site that can enhance, attenuate or ectopically modify the reporter's gene expression; and 2) the random integration of the enhancer::reporter construct in the host genome (Ishibashi et al. 2013). Fortunately, Roberts et al. (2014) have recently provided an effective solution to this problem showing that stable lines generated via PhiC31 integrase-mediated targeted transgene integration exhibited nearly identical tissue-specific enhancer-driven reporter expression.

In conclusion, the zebrafish embryo is a well suited vertebrate model for the functional validation of candidate transcription-regulatory elements, including those predicted in mammalian genomes.

1.12. Objectives of the current study

Regulation of the tissue-specific expression pattern of the murine *Lama1* gene is poorly known. As described earlier, previous studies in our laboratory found that SHH is essential for *Lama1* expression in the somites and neural tube of mouse embryos, but did not inform about whether the control is direct, via the GLI transcription factors, or indirect – via intermediary factors in the gene network. Therefore, the main objective of the current study is to elucidate the link between SHH and *Lama1* transcription, by testing the hypothesis of a direct role of SHH in *Lama1* expression in the neural tube and somites, mediated by GLI binding to transcription-regulatory elements (TREs) in the vicinity of the *Lama1* locus.

To address the proposed hypothesis, I aim at the following specific objectives:

1) Identification of candidate TREs of the murine *Lama1* gene using comparative genomics approaches, based on the assumption that the conserved expression of *Lama1* mRNA across vertebrates is due to conserved regulatory mechanisms.

2) Functional screening of the set of identified candidate TREs in mammalian cell culture.

3) Functional screening of a subset of the identified candidate TREs in transgenic zebrafish embryos, relying on the assumption that the tested murine elements could be functional in zebrafish, which is supported by the similarities in *Lama1* expression pattern and the role of HH signaling between mouse and zebrafish.

5) In case none of the tested elements display transcriptional-enhancer properties, an alternative approach could be used to predict candidate TREs of *Lama1*, based on available data for transcription factor occupancy and enhancer-enriched histone marks at the murine *Lama1* locus.

4) Detailed analyses of the tissue-specific activity of promising candidate TREs in transgenic zebrafish embryos, including examination of the responsiveness of the tested elements to perturbations in Hh signaling.

\

Chapter 2

Materials and methods

2.1. Mouse embryo techniques

2.1.1. Mouse strains and embryo collection

Mouse embryos for whole-mount *in situ* hybridisation were harvested from the C57BL/6J and *Shh* (Anderson et al. 2009) strains. Pregnant females were euthanized by an overdose of anaesthetic followed by cervical dislocation. The embryos were harvested via hysterectomy and immediately placed in RNase-free qPBS solution, followed by removal of the yolk sacs and overnight fixation at 4°C in 4% HCHO + 2 mM EDTA in qPBS. Embryos were staged according to Kauffman (1992).

2.1.2. Embryo genotyping

DNA from embryos from the *Shh* strain was isolated via lysis of the yolk sacs in 100 uL tail mix solution at 95°C for 20 minutes, followed by the addition of 100 µL 40 mM Tris HCl (Sigma Aldrich). 1 µL of the lysis mix was used for a PCR at 58°C annealing temperature using Shh-P5, Shh-P6 and Shh-PGK primers (see Table 2.2) (Anderson et al. 2009). PCR products were analasyed with gel electrophoresis using 1.2% agarose/TAE gel.

a1-NSE::lacZ transient transgenic mouse embryos were genotyped from yolk sac material by our collaborator Dr. Norris Ray Dunn (IMB, Singapore) using the Generic LacZ F and LacZ R primers (Table 2.2) to generate a 315 bp *lacZ* transgene fragment (Figure 2.1). The *Mcc* gene was used as a positive internal control (Young et al. 2011).

2.1.3. Whole-mount *in situ* hybridization (WMISH)

Overnight-fixed embryos were rinsed and washed 3 x 15 minutes with 1 x PTW buffer and treated with 20 µg/mL Proteinase K in 1mL Proteinase K buffer at 37°C according to their stage, followed by 3 x rinses in 1 mL PTW buffer. Then, the embryos were post-fixed in 1 mL 4% HCHO + 0.1% glutaraldehyde in PTW for 20 minutes at room temperature, followed by 3x rinses and 3x washes with PTW for 5 minutes. Next, the embryos were rinsed in 1 mL 1:1 PTW:hybridisation buffer solution, replaced by 1 mL hybridisation buffer and stored at −20°C. On the day of WMISH, the embryos were incubated with 1 mL hybridization buffer for 2 hours at 69°C. Then, the old buffer was replaced with 1 mL fresh hybridization buffer,
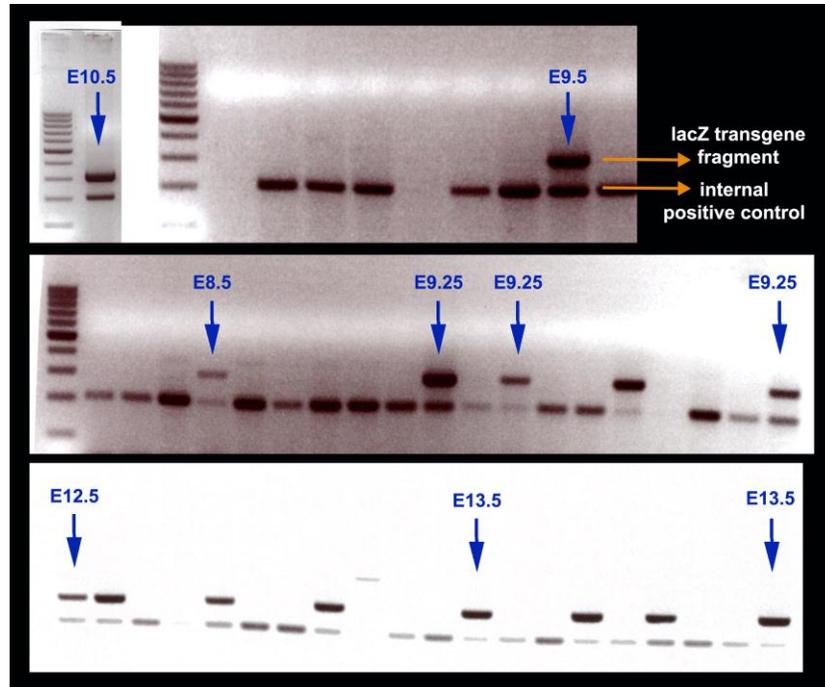
Figure 2.1. *Genotyping results for the nine b-Gal-positive embryos, as obtained by Dr. Norris Ray Dunn (IMB, Singapore).* All nine embryos carry the *lacZ* transgene represented by the upper of the two fragments on the electrophoretogram.

immediately followed by the addition of 3 uL 1 ug/mL DIG-labelled RNA probe, and the embryos were incubated at 69°C overnight. On the next morning, the old hybridization buffer + RNA probe were discarded and the embryos were quickly rinsed twice with 1 mL fresh hybridization buffer at 69°C, followed by two washes 30 minutes each with new 1 mL of hybridization buffer at 69°C. Then, the embryos were washed once with 1:1 hybridisation buffer:TBST buffer for 20 minutes at 69°C, followed by two 30 minutes washes in TBST at room temperature. Next, the embryos were rinsed 2 x in MABT buffer and incubated in 1 mL blocking solution for 2 hours at room temperature. Then, the old blocking solution was replaced with fresh one and 0.5 μL of anti-digoxigenin alkaline-phosphatase antibody (Roche) was added. The embryos were incubated at 4°C overnight. The following day, embryos were extensively washed with MABT, followed by 2 x 30 minutes washes with NTMT at room temperature. Staining was developed by the addition of 3.5 μl 5-bromo-4-chloro-3-indolyl phosphate (BCIP; Roche) and 4.5 μl 4-Nitro blue tetrazolium chloride (NBT; Roche) to 1 ml NTMT and incubating the embryos in the dark at room

temperature/4°C. The staining was terminated by rinsing embryos in PTW, followed by 20 minutes post-fixation at room temperature and 3x 15 mines washes in PTW.

2.1.4. β-galactosidase staining

Embryos were stained following a protocol kindly provided by our collaborator Dr. Ray Dunn (IMB, Singapore). Newly-harvested embryos were cleaned from the extra-embryonic membranes and fixed with 4% paraformaldehyde in PBS for 2 hours at 4°C, and then rinsed 3 x and washed 3 x 10 minutes in PBS at room temperature. Embryos were incubated in 1 mL staining solution (5 mM $K_3Fe(CN)_6$ + 5 mM $K_4Fe(CN)_6$ + 2 mM MgCl2 + 1 mg/mL X-gal + 0.02% NP-40 in PBS) at 37°C overnight. Stained embryos were extensively washed with PBS at room temperature. The E9.5 and E10.5 transgenic embryos were sectioned by the Histology and Pathology facility in IMCB, Singapore, and were counter-stained with eosin. The E12.5 and E13.5 transgenics were sectioned via a cryostat as described in 2.1.6.

2.1.5. Vibratome sectioning

Embryos stained according to the WMISH protocol were embedded in 2% agarose in PBS and kept at 4°C overnight. Blocks were cut out of the gel and glued to a chuck using SuperGlue (Bostik), and 80-100 μm sections were obtained using Vibratome 1500 sectioning apparatus (Vibratome®). The sections were collected in 80% glycerol and mounted to a glass slide using Glycergel mounting medium (Dako).

2.1.6. Cryostat sectioning

Stained embryos were incubated in 15% sucrose solution and embedded in OCT medium (BDH) on plastic mounting boats, immediately followed by transfer of the boats on dry ice-chilled 100% ethanol until freezing of the OCT medium. The embedded embryos were stored at -80°C. For sectioning, the embedded embryos were mounted on pre-chilled chucks via OCT medium, and cut into 30 μm sections using a cryostat (Bright Instruments). Sections were collected on superfrost slides (Menzel-Glaser) and left to dry for 1 hour at room temperature, and then kept at -20°C before use. Before use, sections were left at room temperature for 30 minutes, followed by 15 minutes rehydration with PBS, and mounted using Glycergel mouning medium (Dako).

2.1.7. Imaging

Whole-mount embryos were visualized using a MZ12.5 steromicroscope (Leica) and images were captured by a SPOT® INSIGHT Colour camera using SPOT Advanced software (Diagnostic Instruments). Embryo sections were visualized using Leica DMR microscope (Leica) and images were captured via DMR DC300FX digital camera (Leica) using Leica IM50 Image software vl.20 (Leica). The obtained images were processed by Photoshop CS5 (Adobe).

2.1.8. Mouse oocyte injections

The *a1-NSE::lacZ* reporter plasmid was sent (dissolved in Milli-Q water) to the Mouse Transgenesis Group in IMCB, Singapore, where it was linearized with SacII releasing a 4.8 kb linear fragment containing the *a1-NSE::lacZ* construct, which was injected into oocytes from the FVB/N mouse strain. The day of injection was counted as E0 but embryos were harvested a day later than the desired stage to allow for delays in development as is often the case with transient transgenic mouse embryos (personal communication with Dr. Ray Dunn (IMB, Singapore)).

2.2. Chicken embryo techniques

2.2.1. Embryo incubation

Fertilised chicken eggs were obtained from local farms and incubated at 39°C until the desired embryonic stage, according to Hamburger and Hamilton (1951).

2.2.2. Embryo collection

At the desired embryonic stage, eggs were opened at the wider end and the embryos were dissected and temporarily kept in RNase-free qPBS on ice, immediately followed by fixation according to the protocol applied to mouse embryos and described earlier. Chicken embryos were staged according to Hamburger and Hamilton (1951).

2.2.3. WMISH, sectioning and imaging

Chicken embryos were processed for WMISH, vibratome sectioning and imaging following the protocols applied to mouse embryos, as described above.

2.3. Zebrafish husbandry and embryo techniques

2.3.1. Zebrafish lines and embryo collection

Zebrafish were kept at 28°C at 14 hours light/10 hours dark cycle. Fish embryos were collected from the wild type AB strain (IMCB, Singapore), from the *Tg(olig2:EGFP)* line kindly provided by Dr. Vladimir Korzh lab (IMCB, Singapore), and from the *smo*^{hi1640Tg} line provided by Dr. Sudipto Roy (IMCB, Singapore). Embryos were harvested from paired matings in the mornings and kept at 28°C in E3 embryo medium.

2.3.2. Generation of *CNE:: EGFP*-reporter constructs

CNEs were amplified from mouse genomic DNA (C57BL/6J) using the primer pairs in Table 2.2. The primers harboured HindIII linkers, which enabled cloning in the HindIII site of the *EGFP* reporter vector (McDonald et al. 2010; Yu et al. 2011). Successful *CNE::EGFP* reporter construct generation was confirmed via sequencing with the M13 forward and M13 reverse primers (See Table 2.2).

2.3.3. Microinjection and embryo maintenance

Embryos were obtained from paired matings in the mornings immediately before microinjection. Embryos were injected at the 1-cell stage using with ~ 1 nL plasmid (or plasmid + RNA) solution that also contained 0.05% Phenol Red (stock: 0.5% in DPBS, Sigma) in Milli-Q water. Following injection, the embryos were transferred at 28°C in E3 embryo medium and periodically monitored for reporter gene expression until the 96 hpf stage. All dead embryos were regularly discarded and the water was renewed every 24 hours.

2.3.4. Fixation and immunohistochemistry

Embryos with interesting reporter gene expression pattern were set aside and dechorionated with 0.5 mg/ml Pronase (stock 10mg/ml dissolved in water, Roche Applied Sciences) for 30 minutes at room temperature. Then, embryos were fixed with 4% PFA (Sigma-Aldrich) in PBS (Sigma-Aldrich) at 4°C on a shaker, overnight (~14 hours). On the following day, embryos were rinsed once and washed twice (10 minutes each) in PBTX buffer at room temperature, with agitation, followed by a single wash in 50% methanol (Merck) in PBTX. Finally, the embryos were kept in methanol and stored at -20°C.

Embryos for immunostaining were rehydrated by a rinse and a single wash in 50% methanol in PBTX, followed by two washes in PBTX. Then, embryos were incubated with

PBDT buffer for 1 hour at room temperature. Subsequently, the embryos were incubated with the primary antibody diluted in PBDT (1:1000) at 4°C, overnight, with agitation. On the next day, the embryos were rinsed once followed by two washes (30 minutes each) with fresh PBDT at room temperature, with agitation. Next, the embryos were incubated with the secondary antibody diluted in PBDT (1:5000), at 4°C overnight. On the following day, the embryos were rinsed once and washed two times in PBDT with DAPI (Invitrogen) added at one of the washing steps. Shortly before imaging, the PBDT buffer was replaced by 80% glycerol (Invitrogen) in PBTX.

Antibodies used

(1:800) anti-GFP rabbit polyclonal antibody, Alexa Fluor® 488 conjugate, Life Technologies (Invitrogen).

(1:200) anti-pax3/pax7 mouse monoclonal antibody, DSHB

(1:800) anti-mCherry, rabbit polyclonal DsRed antibody, Clontech

(1:1000) Alexa Fluor 546 goat anti-mouse IgG, Life Technologies (Invitrogen)

(1:1000) Alexa Fluor 568 goat anti-rabbit IgG, Life Technologies (Invitrogen)

(1:20) anti-Pax7 mouse monoclonal antibody (DSHB)

2.3.5. Cryosectioning

Whole-mount immunostained embryos were transferred from 80% glycerol to O.C.T. medium where they were embedded in the desired orientation, followed by rapid freezing in cooling bath of ethanol on dry ice. The O.C.T. blocks were stored at -80°C. For sectioning, the O.C.T. blocks were mounted to pre-chilled chucks and 12 μm sections were obtained using a cryostat (Bright Instruments). Sections were collected on Marienfeld-Superior glass slides, followed shortly after that by mounting using Vectashield medium (Vector Laboratories).

2.3.6. Imaging

Whole-mount embryos in 80% glycerol in PBTX were visualized via a Zeiss Axio Imager. M2 microscope equipped with Zeiss AxioCam HRc camera, and images were captured with the AxioVison v. 4.7.2 software. Confocal images were captured via Olympus BX61 microscope using the FV10-ASW software.

2.4. Luciferase reporter assays in cell culture

2.4.1. Reporter construct generation

CNEs were isolated from mouse genomic DNA using primer pairs in Table 2.2 (see section 2.5.1), and first cloned into the TOPO vector (see see section 2.5.2). CNEs cloned into the TOPO vector (Invitrogen) were excised with KpnI and XhoI and sub-cloned into the KpnI- and XhoI-sites of the multiple cloning region in the pGL3-Promoter vector (pGL3 Luciferase Reporter Vectors (Technical Manual, Promega 2007). Some CNEs were generated with primers harbouring KpnI-linkers, and were similarly cloned into the KpnI site of pGL3 (sections 2.5.3 – 2.5.5). Successful construct generation was confirmed via diagnostic digestion and sequencing using the pGL3 seq primer (see Table 2.2)

2.4.2. Cell culture maintenance

The C3H10T1/2 mouse embryonic fibroblast cell line was employed in all transient transfection assays of CNE function. Cells stored at liquid nitrogen were rapidly defrosted in water bath at 37°C and added to 5 ml pre-warmed DMEM medium (Gibco) containing 10% FBS (Gibco) and 1% PSF (Gibco), simply referred to as "medium" in this study. The cells were centrifuged at 1000 x g for 5 min and the pellet was re-suspended in 1 ml growth medium. Then, 0.1 ml cell suspension was added to 5 ml pre-warmed medium in 25 cm^3 Nunc flasks, which were kept in conditioned incubators at 37°C and 5% CO2. On the next day, the old medium was replaced with fresh one and the cells were subsequently passaged every 2 days after reaching 80-90% confluence. For passaging, the old medium was discarded and the cells in each 25 cm^3 flask were rinsed 3 times with 0.5 ml pre-warmed 1 x DPBS buffer without Ca^{2+} and Mg^{2+} (Life Technologies) followed by 5 minutes incubation with 1 x Trypsin-EDTA, Phenol Red (Life Technologies) at 37°C. The dissociated cells were collected in 5 ml growth pre-warmed growth medium and then pelleted by centrifugation at 1 x 10^3 for 5 min at RT. The supernatant was discarded and the cell pellet was re-suspended in 1ml growth medium. Finally, 0.1 μl cell suspension was added to a flask with fresh growth medium and cultured at the conditions described above until reaching 80-90% confluence when the passaging was repeated. Note that the cells were not maintained beyond the 20th passage to avoid the accumulation of aberrations. Thus all experiments were performed with cells between the 3rd (after reconstitution of frozen cells) and the 20th passage.

2.4.3. Transient cell transfection

Cells were transiently transfected via lipofection using Lipofectamine 2000 (Invitrogen) in Opti-MEM transfection medium (Life Technologies). The transfections were performed in 6 well-plates (Thermo Scientific) with 6×10^5 cells per well. On Day 1, the old growth medium from each well was discarded and the cell monolayer was washed 3 times with 1 ml of pre-warmed DPBS, followed by incubation of the cells with 1 ml pre-warmed Opti-MEM medium for 10-15 minutes at 37°C. Meanwhile, the transfection mix was prepared by the combination of two fractions. The first fraction was made by the addition of 3 μg DNA (containing the reporter plasmid, control plasmid and carrier DNA; see below*) to 0.3 ml Opti-MEM in 2 ml eppendorf tubes (one tube per well). For the second fraction, 4 μl Lipofectamine 2000 were added to 0.3 ml Opti-MEM in separate 2 ml eppendorf tubes (one tube per well) and incubated for 5 minutes at RT. Immediately after that, the two fractions were combined and left at RT for 30 minutes with gentle agitation every 10 minutes. In the meantime, the old Opti-MEM medium was discarded and 0.4 ml fresh pre-warmed Opti-MEM medium was put in each well with cells, followed by addition of the transfection mix (containing the DNA:Lipofectamine 2000 complex). The plate was gently rocked and left in the incubator. Three hours after transfection, the cells were supplemented with another 1ml/well pre-warmed Opti-MEM medium followed by overnight incubation. On Day 2, the old Opti-MEM transfection medium was removed and replaced with 3ml/well growth medium

*The total amount of DNA added to each well was 3 μg and it included 260 ng *CNE::pGL3* reporter plasmid, 60 ng pRL reporter vector and 2.68 μg pcDNA3 plasmid (used as carrier DNA).

2.4.4. Preparation of cell lysates and quantification of protein concentration

On Day 3, approximately 48 hours post-transfection, the old growth medium was removed and the cells were lysed by adding 0.5 ml/well 1x Passive Lysis Buffer (Promega) followed by incubation for 15 minutes at RT with agitation. In order to avoid errors in the luciferase assay due to unequal protein amount between a test CNE::pGL3 construct lysate and the control lysate, the total protein concentration in each lysate was estimated first using the Bradford protein method. For this, a standard curve was prepared using BSA (Promega) samples with known concentration: 0.125, 0.25, 0.5, 1, 2 and 5 mg/ml, where 20 μl of each protein sample was incubated with 980 μl 1 x Quick Start™ Bradford Dye Reagent (BioRad) for 10 minutes at RT, immediately followed by measurements of the light absorbance of the

samples at 595 nm on a spectrophotometre. The results were used to prepare a standard curve. Next, in analogous way, 20 µl of each cell lysate were mixed with 980 µl of the Bradford reagent and light absorbance was measured accordingly. The absorbance data from the lysates were plotted on the standard curve and the protein concentration was estimated for each lysate. Then, the concentrations of all lysates were standardised to 0.22 mg/ml (with Passive Lysis Buffer) and this was used in the subsequent luciferase reporter assays.

2.4.5. Dual Luciferase Reporter Assay

The activities of the two luciferases expressed in the transfected cells – the experimental luciferase (from the firefly *Photinus pyralis*) and the control luciferase (from the sea pansy *Renilla reniformis*), were measured using the Dual-Luciferase Reporter® (DLR™) Assay system (Promega) in luminometre at RT. At the onset, 10 µl of each lysate (after standardizing the concentrations to 0.22 mg/ml) were dispensed into the individual wells of Hard-Shell® Low-Profile Thin-Wall 96-well microplates (BioRad). The activities of the two luciferases were recorded sequentially beginning with the measurement of firefly luciferase activity. For this, 50 µl Luciferase Assay Reagent II (containing luciferin, the substrate of firefly luciferase) to each well followed by programmed agitation of the whole plate for 3 seconds. Then, another 3 seconds of programmed pre-read delay were included before the actual measurements of firefly luciferase activity. Upon completion of the recordings, the firefly luciferase was quickly quenched by the addition of 50 µl Stop & Glo® Reagent (which also includes coelenterazine – the substrate of the *Renilla* luciferase), followed by 3 seconds of plate agitation and 3 seconds of pre-read delay before the measurements of *Renilla* luciferase activity were taken (Promega).

2.5. Molecular biology techniques

2.5.1. Polymerase chain reaction (PCR)

2.5.1.1. Primer design

PCR primers were designed using the Primer Premier 5 or NetPrimer tools (www.premierbiosoft.com), to be between 19-26 nucleotides long, with similar A/T and G/C content, to possess at least one G/C nucleotide at the 3'-end, and to have Tm of 57-63°C. Primers harbouring linkers for restriction enzymes were designed similarly with reference to "Cleavage Activity Near DNA Termini" (Stratagene).

2.5.1.2. PCR reaction settings

The GoTaq® DNA Polymerase kit (Promega) was used for the amplification of sequences with size ≤ 1 kb and for standard genotyping procedures. The GoTaq polymerase requires 1 minute per kb during the extension phases. The reactions were performed in a 20 µl volume, including:

10 µl Green GoTaq® Reaction 2 x buffer

0.4 µl dNTP-mix (10 mM each)

0.8 µl F-primer (20 µM)

0.8 µl R-primer (20 µM)

0.1 µl GoTaq® DNA Polymerase (5 u/µl)

1 µl DNA template (30-100 ng/µl)

6.9 µl Milli-Q H_2O

Cycle conditions:

Intial denaturation 95°C 2 min

Denaturation 95°C 1 min

Annealing 1 min

Extension 72°C 1 min/kb

Repeat steps 2-4 34 times

Final extension 72°C 5 min

Keep 4°C indefinitely

The iProof™ High-Fidelity DNA Polymerase kit (Bio-Rad Laboratories) was used for highly accurate amplification of sequences with size ≤ 2 kb. iProof™ High-Fidelity DNA Polymerase requires 30 seconds per kb during the extension phases. The reactions were performed in a 50 µl volume, including.

10 µl iProof HF 5 x buffer

1 µl dNTP-mix (25 mM each)

1 µl F-primer (20 µM)

1 µl R-primer (20 µM)

1 µl MgCl2 (50 mM)

1 µl DNA template (100 ng/µl)

0.5 µl iProof DNA Polymerase (2 u/µl)

34.5 µl Milli-Q H_2O

Cycle conditions:

Initial Denaturation 98°C 30 sec

Denaturation 98°C 10 sec

Annealing at T°C + 3°C above the primer with the lowest T_m 30 sec

Extension 72°C 30 sec/kb

Repeat steps 2-4 34 times

Final extension 72°C 10 min

Keep 4°C indefinitely

Primer name	Sequence
Shh-P5	GTTGTTACTGCATCCCTTCCATC
Shh-P6	GGCTAGCTCAGTGCTTGCAAG
Shh-PGK-R	GGATGTGGAATGTGTGCGAG
CNE3 F	CATAGCGTGGAGGTGGAGAGAGC
CNE3 R	CCCTGGCTCTGGAAACCTAACTC
CNE4 F	CTGTCCCGAAGTCACTCTGTATTTG
CNE4 R	GCCGTGAGTGTCTCTGCTGTG
CNE5 F	CCAGTTGATGTCACAGCAGTAGC
CNE5 R	GCTCCACACTTGAGATGCTGCC
CNE6 F	CCTGGGAGGACCAAGATGAAG
CNE6 R	TCAGAGAGGGTGGGAAAGGAC
CNE7 F	CACTGGGAGCCATTAGGAGGG
CNE7 R	CGTGTGCTTTTCTGTCTCAGTAAC
CNE9 F	GACTCTGCTCAAGGGTATGTGTTCC
CNE9 R	GGTCAAGGAGCCTGGAAATCTGTC
CNE10 F	CCTTCGGAGACTTCTGGCTTTC
CNE10 R	GCTGTGACCCTGATTGTATCTGTATG
CNE13 F	CGATTTAGCCCTGCCCTGC
CNE13 R	GGGCAAAGCATCCAGTAGGC
CNE14 F	CACAGTGGAGACAAACACGAGGC
CNE14 R	CTGGTAGGGGTGATTTGGACGG
CNE15 F	GGAGACGCTGGGAGATTTGGAC
CNE15 R	GATTTGAAGGCACAGGCAGACC
CNE19 F	GTGACAGTCTTGCTTTCTGATAGGG
CNE19 R	GGTTTCGTTAGGTTCTTTCAGGG
CNE21 F	CTGCTCTGGCATTTCCGACC
CNE21 R	GCTCAGGTAGACACAGGAACGG
CNE22 F	CCAGAAGTCCCAGAAGAGAATGC
CNE22 R	GCGTTCCTTAATAGTATTAGTC
CNE23 F	CCACAATAGGTAAGAGACAGGTAGGG
CNE23 R	CCAGTGGCTTCCCAGTCAGG
CNE24 F	CAGCATTCTCCCTCTGAACATACAC
CNE24 R	GACAGAAACCTTGAGAGAAATCCC
a1-NSE F	TCTGACTCTAGGGGTCACCTGCTT
a1-NSE R	GGTCACTTTCAGCAACCTCACAG
np1230 seq	CTATGACCATGATTACGCCAAGC
pGL3 seq	CAAAATAGGCTGTCCCCAGTGC
M13 F	GTTTTCCCAGTCACGAC
M13 R	CAGGAAACAGCTATGAC
LacZ F	ATCCTCTGCATGGTCAGGTC
LacZ R	CGTGGCCTGATTCATTCC

Table 2.1. *Primer sequences used in this study.* All but Shh-P5, Shh-P6, Shh-PGK, M13 F, M13R, LacZ F and LacZ R were designed in this study.

2.5.2. TOPO cloning of PCR products

The TOPO® TA Cloning® Kit (Invitrogen, Life Sciences) was used for rapid cloning of Taq-polymerase generated PCR products. 1 µl PCR product was incubated with 1 µl Salt Solution, 3 µl Milli-Q H2O and 1 µl pCR™II-TOPO® vector for 5 minutes at RT, followed by transfer on ice for 20 minutes. Next, 2ul of the cloning reaction were added to 40 µl of TOP10F' chemically-competent cells (Invitrogen) and the mixture was incubated on ice for 30 minutes. The cells were heat-shocked for 30 seconds at 42°C and immediately placed on ice for 2 minutes, followed by the addition of 250 µl S.O.C medium and 1 hour incubation at 37°C with agitation (200 rpm). Finally, 40 µ-100 µl of each transformation were spread on a pre-warmed agar plate containing the appropriate antibiotic. Prior to spreading the cells, the plates had been coated with 40 µl of 40 mg/ml X-gal (Promega) and 40 µl of 100 mg/ml IPTG (Sigma-Aldrich) for blue/white screening of bacterial colonies. 10-15 promising clones – white colonies, were selected for further analysis of miniprep DNA.

2.5.3. DNA digestion with restriction enzymes

Digestion of DNA was performed in 100 µl reaction volume with the appropriate restriction endonuclease (RE, from New Engalnd Biolabs). For plasmid DNA, 5 µg were incubated with 10 µl 10 x Reaction Buffer (New Engalnd Biolabs) and 1.5-2 µl RE enzyme (1u/µl). 10 µl 10 x BSA (New England Biolabs) were also added where required. The reaction was incubated for 2 hours at 37°C.

Digestion of PCR amplicons with linkers was performed in 50 µl reaction volume. 5 µl 10 x Reaction Buffer and 1µl RE (1 u/µl) were added to 30 µl of purified PCR product in Milli-Q H2O. 5 µl 10 x BSA were added where required and the reaction was incubated overnight at 37°C.

2.5.4. Vector dephosphorylation

Prior to ligation with DNA fragments, the plasmid vector was dephosphorylated using the Calf Intestinal Alkaline Phosphatase (New England BioLabs), in order to prevent self-ligation of the vector DNA ends. The reaction volume was set up at 50 µl, where 5 µl NEB Buffer 3 and 1 µl phosphatase (10u/µl) were added to 30 µl solution of linearised plasmid in Milli-Q H2O. The reaction was incubated for 15 minutes at 37°C, followed by 15 minutes incubation at 50°C.

2.5.5. Ligation

DNA fragments and vector were ligated using the T4 DNA Ligase (Promega) in 1:3 or 1:1 molar ratio of vector to insert. The ligations were performed in 10 μl reaction volume. The appropriate volume of insert and vector solutions (both in Milli-Q H2O) were incubated with 5 ul of 2 x Ligase Buffer and 1ul of T4 DNA Ligase at 4°C overnight.

2.5.6. Bacterial transformation

2 μl of ligation reaction or 2 μl of plasmid (20-100 ng/μl) were added to 50 μl of chemically-competent *E.coli* cells (strain) and incubated on ice for 30 minutes. Then, the mixture was heat-shocked at 42°C for 30 seconds followed by immediate transfer on ice for 2 minutes. 400 μl of S.O.C medium were added and the cells were incubated with agitation (200 rpm) at 37°C for 1 hour. Finally, 80 μl of bacterial suspension were streaked on a pre-warmed LB agar plate containing the appropriate antibiotic (100 μg/ml Ampicillin, 25 μg/ml Kanamycin or 50 μg/ml Chloramphenicol (Sigma-Aldrich)) and incubated overnight at 37°C. On the following day, 10-20 bacterial colonies were picked up and used to prepare minicultures.

2.5.7. Plasmid miniprep

Preparation of plasmid DNA from bacterial minicultures (4 ml) was performed using the AxyPrep™ Plasmid Miniprep Kit (Axygen Biosciences). 3 ml of overnight LB culture were centrifuged at 12×10^3 rcf for 1 minute (repeated 2 times). The supernatant was discarded and the bacterial pellet resuspended in 250 μl of cold Buffer S1. Then, the cells were lysed with 250 μl Buffer S2 for 3 min at RT, followed by the addition of 350 μl Buffer S3, which neutralized the reaction. The sample was centrifuged at 12×10^3 rcf for 10 minutes to pellet the precipitate and the clear supernatant was transferred to an AxyPrep column. The loaded column was centrifuged for 1 minute at 12×10^3 rcf and then washed 2 x with Buffer W2 followed by another step of 1-minute centrifugation at 12×10^3 rcf to remove residual buffer. At the end, the DNA was eluted with 60 μl of Milli-Q H2O and stored at -20°C.

2.5.8. Plasmid midiprep

Preparation of plasmid DNA from bacterial midicultures (50 ml) was performed using the HiSpeed® Plasmid Midi Kit (QIAGEN). 50 ml of overnight LB culture were centrifuged at 6×10^3 rcf for 15 minutes at 4°C. The bacterial pellet was re-suspended in 6 ml Buffer P1 followed by the addition of 6 ml Buffer P2 (lysis buffer) and incubated at RT for 5 minutes.

Then, the reaction was neutralized with 6 ml chilled Buffer P3 after which the lysate was immediately added to a QIAfilter Cartridge and incubated for 10 minutes at RT. Next, the lysate was pressure-filtered from the cartridge and simultaneously loaded in a HiSpeed Midi Tip which has been equilibrated with 4 ml Buffer QBT. The tip was washed with 20 ml Buffer QC followed by elution of the DNA with 5 ml Buffer QF. The DNA was precipitated with 3.5 ml isopropanol (Merck) at RT for 5 minutes. The eluate/isopropanol mixture was pressure-filtered through a QIAprecipitator Midi Module and the flow-through was discarded. Next, the QIAprecipitator was washed with 2 ml 70% ethanol. Finally, the DNA was eluted with 1ml MilliQ-H2O and stored at -20°C.

2.5.9. Synthesis of digoxigenin-labelled RNA probes
Riboprobes were synthesized by in vitro transcription in 20µl reaction volume, containing:

1 µl linearised DNA (1 µg/µl)

4 µl 5 x TSC buffer (Promega)

2 µl 10 x DIG RNA Labeling Mix (Roche)

0.5 µl 50 mM DTT (Sigma Aldrich)

1 µl RNasin (Promega)

1.5 µl RNA polymerase (20 u/µl) (Promega)

10 µl DEPC-treated Milli-Q H2O

The reaction was incubated at 37°C for 2 hours. The DNA template was removed by the addition of 2 µl of RQ1 DNase I (Promega) and further incubation at 37°C for 30 minutes. To precipitate the RNA, 2 µl 200 mM EDTA, 2.5 ul 4 M LiCl and 70 µl 100% RNAse-free ethanol (Sigma Aldrich) were added and the mixture was kept at -80°C for 1 hour. Then, the RNA was pelleted by centriguation at 13×10^3 rpm for 30 minutes at 4°C; the supernatant was discarded and the pellet washed with 100 µl 70% RNAse-free ethanol followed by another centrifugation at 13×10^3 rpm for 10 minutes at at 4°C. The supernatant was discarded and the pellet was vacuum-dried at RT for 2-3 minutes. Finally, the RNA was resuspended in 50 µl DEPC-treated Milli-Q H2O, with the addition of 1.5 µl RNasin and 0.5 µl 100 mM DTT. The probes were stored at -20°C.

plasmid	restriction enzyme	RNA polymerase
Lama1-pCRII TOPO[a]	EcoRV	Sp6
Lamb1-pCRII TOPO[b]	EcoRV	Sp6
Lamc1-pCRII TOPO[a]	EcoRV	Sp6
cLama1-pBluescript II KS +	SacI	T3

Table 2.2. RNA probes used in this study. Indicated are the restriction enzyme used to linearise the plasmid template, and the RNA polymerase used for probe synthesis.

a plasmids were kindly provided by Claire Anderson (Anderson et al. 2009)
b plasmid was generated by Ms. Katherine Long

Chicken *Lama1* probe was synthesized form EST clone ChEST869c13 from the Source Bioscience (Life Sciences).

2.5.10. Capped-RNA synthesis

Capped RNA was synthesised using the mMESSAGE mMACHINE® Kit (Ambion), using appropriate RNA polymerase and linearised DNA template, depending on the vector and the insert orientation. The reaction was performed in 20 µl volume, where 1 µl of linearised DNA template (1 µg/µl), 10 µl 2 x NTP/CAP mix, 2 µl 10x Reaction Buffer, 2 µl appropriate RNA polymeraze (Enzyme Mix) and 5 µl Nuclease-free H2O were mixed together and incubated at 37°C for 2 hours. Then, DNA template was removed by the addition of 1 µl RQ1 DNase I (Promega) and further incubation at 37°C for 15 minutes. The reaction was stopped and RNA precipitated by the addition of 30 µl Nuclease-free H2O and 30 µl LiCl Precipitation Solution (7.5 M LiCl, 50 mM EDTA) followed by 1 hour incubation at -20°C. The precipitate was collected by centrifugation at 13×10^3 rpm for 15 minutes at 4°C; the supernatant was discarded and the pellet washed once with 1 ml of 70% RNase-free ethanol, and re-centrifuged at the same settings. Finally, the RNA was resuspended in Milli-Q H2O, aliquoted in small volumes and kept at -80°C.

RNA	Vector	RE for template linearisation	RNA polymerase
Tol2	pDB600	XbaI	T3
dnPKA	pCS2+	NotI	Sp6
H2B-mCherry	pCS2+	KpnI	Sp6
Shh	pSP64T	BamHI	SP6

Table 2.3. Capped RNAs used in this study. Shown are the encoding plasmids, the restriction enzyme (RE) used for plasmid linearization and the RNA polymerase used for transcription.

2.5.11. Estimation of DNA/RNA sequence size

DNA/RNA samples were loaded on 400 nM Ethidium Bromide 1% Agarose gel in TAE buffer, along with 3 µl of GeneRuler 1 kb Ladder (Fermentas Life Sciences). Size of DNA/RNA was determined by comparison of the distance travelled by the sample's band(s) with the bands on the ladder of known size.

2.5.12. Estimation of DNA/RNA concentration

2 µl of DNA/RNA sample (either in TE buffer or Milli-Q H2O) were loaded on NanoDrop ND-1000 Spectrophotometer (Thermo Scientific) and concentration was measured against a blank sample (TE buffer or Milli-Q H2O only).

2.5.13. Gel Extraction

DNA fragments were recovered from agarose gels using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research). Agarose gel slices containing the desired fragment were excised from the rest of the gel and dissolved in ADB Buffer at 55°C for 10 minutes. The solution was loaded on Zymo-Spin™ I Columns and subsequently washed twice with DNA Wash Buffer. The DNA was eluted with Milli-Q H2O.

2.5.14. Phenol/chloroform DNA extraction

One volume of phenol (molecular biology grade, Sigma-Aldrich) was added to 1 volume of DNA solution and the mixtire was vortexed for 15 minutes, followed by centrifugation at for 10 minutes. The top aqueous phase containing the DNA was transferred to a new tube. 0.5 volume of phenol and 0.5 volume of chloroform/isoamyl alcohol (CIA, ratio 24:1) were added and the sample was vortexed for 15 minutes, followed by centrifugation at $13x10^3$ rpm for 10 minutes. The top layer was transferred to a new tube and 200 µl of CIA were added, followed by 1 minute of vortexing and 5 minutes of centrifugation. Finally, the DNA-enriched top phase was transferred to a clean tube and prepared for alcohol precipitation.

2.5.15. DNA precipitation

DNA was precipitated by adding 2 volumes of 100% ethanol (DNAase-free, Sigma Aldrich), 1/10 volume of 3 M sodium acetate (pH 5.3) (Sigma Aldrich) and 1 µl of glycogen (Sigma Aldrich). The sample was then incubated at -20°C overnight. Then, the sample was centrifuged at 13 x 10^3 rpm at 4°C for 30 minutes. After removal of the supernatant, the DNA pellet was washed with 100 µl of 70% DNAase-free ethanol and centrifuged at 13 x 10^3 rpm

at 4°C for 10 minutes. The supernatant was removed, the pellet was air-dried or vacuum-dried at 40°C for 3 minutes and re-suspended in Milli-Q H2O at the required concentration.

2.5.16. RNA precipitation

RNA was precipitated by adding 3 volumes of 100% ethanol (RNA grade, Sigma-Aldrich), 1/20 volume of 4 M LiCl (Sigma-Aldrich) and 1 µl of glycogen (Roche Applied Sciences). The sample was then incubated at -80°C for 1 hour. Then, the sample was centrifuged at 13×10^3 rpm at 4°C for 30 minutes. After removal of the supernatant, the RNA pellet was washed with 100 µl of 70% RNA-grade ethanol and centrifuged at 13×10^3 rpm at 4°C for 10 minutes. The supernatant was removed, the pellet was air-dried and re-suspended in DEPC-treated H2O at the required concentration.

2.6. *in silico* analyses

2.6.1. Genomic sequence retrieval

Genomic sequence information was retrieved from the Ensembl project database (www.ensembl.org), and from the UCSC Genome Bioinformatics database (www.genome.ucsc.edu).

2.6.2. Gene expression information retrieval

Information on the expression pattern of genes in the mouse, chicken and zebrafish was obtained from the Mouse Genome Informatics (MGI, www.informatics.jax.org), the Gallus Expression In Situ Hybridization Analysis (GEISHA, www.geisha.arizona.edu), and The Zebrafish Model Organism (ZFIN, www.zfin.org) databases.

2.6.3. Analysis of sequence conservation

Identification of conserved non-coding elements in the vicinity of the murine *Lama1* locus was performed using the ECR Browser (www.ecrbrowser.decode.org), the UCSC Genome Browser (www.genome.ucsc.edu), and the VISTA Browser (www.pipeline.lbl.gov/cgi-bin/gateway2).

The ECR Browser tool performs and visualises comparative multispecies whole-genome alignments in a dynamic way enabling the rapid identification of candidate TREs. In short, the ECR Browser aligns genomic sequences retrieved from the UCSC Genome Browser after masking of repetitive elements followed by construction of large-scale syntenic relationships. To facilitate synteny mapping, the browser initially aligns each species's genome to all others

in a pair-wise fashion (Ovcharenko 2004). By aligning the base genome to the genome sequences of selected species, the ECR generates a conservation profile, which is displayed in a 2D graphical format (Figure 4.1B; Figure 4.3B). There, the *x*-axis schematically represents the position in the base genome sequence while the *y*-axis represents the identity (in percentage, %) between the base and the aligned genome sequences at the specific position. Those segments of the alignment with significant length (in base pairs) which have equal to or higher percentage identity than a minimum custom-defined threshold appear as highlighted peaks, defined as evolutionary-conserved regions. All of the ECRs identified in this study are abbreviated as CNEs (**C**onserved **N**on-coding **E**lements) reflecting the fact that all of them are located in genomic regions with no known protein-coding potential. The rest of the aligned sequence which have bp length and percentage identity lower than the set-up thresholds does not appear as peaks and such regions are considered more or less neutrally evolving.

Customised multiple alignments in of the *Lama1*-containing genomic region from mouse, human, opossum, chicken and zebrafish were performed in mVISTA using the Shuffle-LAGAN algorithm. The aligned genomic sequences were manually retrieved from the Ensembl database; each sequence spans from the 3'-end of *Ptprm* locus to the 5'-end (transcription start site) of *Arhgap28* locus. Identification of peaks in the mouse/zebrafish alignment was achieved using the low stringency parametres of minimum 50% sequence identity over a 50 bp window, maximising the identification of conserved, albeit short, regions between the genomes of mouse and zebrafish.

2.6.4. Analysis of transcription factor binding motif presence and conservation

Detailed analyses of the transcription factor binding motifs content of CNEs and a1-NSE were performed using the rVISTA 2.0 (www.rvista.dcode.org), MatInspector and FrameWorker tools (Cartharius et al. 2005; www.genomatix.de). The rVISTA tool enables the computational prediction of TFBSs using PWMs (**P**ositional **W**eight **M**atrices) from the TRANSFAC database, using a combination of pattern recognition and comparative sequence analysis. Such approach significantly reduces the number of false positive matches by 95% while maintaining sensitivity of the search. After the TFBSs are localised in both sequences, rVISTA proceeds with the identification of pairs of aligned TFBSs which are further tested for sequence conservation. To be regarded as conserved, the aligned TFBSs must meet the

requirement of at least 80% sequence identity over a 20bp-sliding window encompassing the core of the binding site (Loots and Ovcharenko 2004). Despite the fact that a significant fraction of the predicted TFBSs are not functionally relevant, the rVISTA tool is nevertheless efficient in providing a set of limited number of putative TFBSs which can be further examined by functional studies. Each of the identified CNEs was individually submitted for analysis in the rVISTA 2.0 tool using default parameters: the "Vertebrates" set from the TRANSFAC professional V10.2 library and the "Optimized for function" option from the "Matrix similarity" parameter. In the next section, the programme was set up to screen the CNE sequence using the PWMs for all available TF families, including the Gli family.

2.6.5. Analyses of chromatin features
Information on histone marks, transcription factor occupancy and DNaseI hypersensitivity was obtained using the UCSC ENCODE Browser/database (Rosenbloom et al. 2012; www.genome.ucsc.edu).

2.6.6. Design of mutations in the Gli motifs within a1-NSE
Base pair substitutions in the Gli motifs of a1-NSE were designed using the SequenceShaper tool from Genomatix (www.genomatix.de).

2.6.7. Sequence visualization and manipulation
All sequence files were stored and manipulated in the ApE (A plasmid editor v1.10.4) software tool (www.biologylabs.utah.edu/jorgensen/wayned/ape/).

Solution or buffer	Composition
Blocking solution	2% Blocking reagent (Roche), 10% Horse serum, 0.1% Tween 20 in DEPC H2O
E3 embryo medium	15mM NaCl, 0.5mM KCl, 1mM CaCl2, 1.5mM KH2PO4, 0.05mM Na2HPO4, 1mM MgSO4, 0.7mM NaHCO3 and 3-4 drops per litre of methylene blue
Genius 3	100mM Tris HCl, 50mM MgCl2, 150mM NaCl
Hybridisation buffer	50% Formamide, 1.3x SSC, 5mM EDTA, 50ug/mL tRNA, 0.2% Tween 20, 0.5% CHAPS, 100ug/mL in DEPC H2O, pH7.5
LB agar	10g/L Tryptone, 5g/L yeast extract, 10g/L NaCl, 15g/L agar in deionised H2O, pH7.0
LB broth	10g/L Tryptone, 5g/L yeast extract, 10g/L NaCl in deionised H2O, pH7.0
MABT	100mM Tris HCl, 150mM NaCl, 0.1% Tween 20 in DEPC H2O
NTMT	1% Tween 20 in Genius 3 solution, pH7.5
PBDT	1% BSA (Sigma-Aldrich), 1% DMSO (Merck), 0.5% Triton X-100 in PBS
PBTX	0.1% Triton X-100 (BDH Chemicals) in PBS
Proteinase K buffer	100mM Tris HCl, 50mM EDTA in DEPC H2O, pH8.0
PTW	0.1% Tween 20 in qPBS
qPBS	137mM NaCl, 5.37mM KCl, 0.14mM CaCl2, 1.25mM MgSO4, 1.1mM KH2PO4, 1.1mM Na2HPO4 in DEPC H2O
Tail mix solution	25mM NaOH, 0.2mM EDTA
TBST (10x)	0.4M NaCl, 0.2mM KCl, 0.25M Tris HCl, 1% Tween 20 in DEPC H2O, pH7.5

Table 2.4. *Composition of solutions and buffers used in this study.*

reporter construct	"parent" vector	cloning site	sequencing primer[c]
CNE::pGL3	pGL3 Promoter	KpnI or KnpI/XhoI	pGL3 seq
CNE::EGFP	EGFP (-)[a]	HindIII	M13 F and M13 R
a1-NSE::EGFP	EGFP (-)[a]	HindIII	M13 F and M13 R
a1-NSE::lacZ	np1230[b]	KpnI/XhoI	np1230 seq

Table 2.5. *Reporter constructs generated in this study.*

a, the EGFP (-) vector was kindly provided by Harriet Jackson (University of Sheffield) and is based on McDonald et al. (2010) and Yu et al. (2011).

b, the np1230 vector was developed by Dr. Sarah Coy during her Thesis project (Coy et al. 2011) and is based on Yee and Rigby (1993).**c**, all sequencing primers are indicated in Table 2.2

Chapter 3

Laminin α1 expression in the amniote embryo

3.1. Hypothesis and Aims

As described earlier in Chapter 1, previous research in our laboratory revealed that SHH is required for *Lama1* expression in the neural tube and somites of E9.5 mouse embryos, where absence of *Lama1* mRNA correlates with failure to assemble the myotomal basement membrane culminating in defective morphogenesis and cell fate specification of the myotome (Anderson et al. 2009). However, it was unclear whether SHH acts as an activating or maintaining factor for *Lama1* expression in the neural tube and somites. Thus, I performed whole-mount *in situ* hybridization (WMISH) to examine *Lama1* mRNA expression in *Shh*-deficient embryos at an earlier stage, at E8.5. Another unanswered question related to the expression of the *Lamb1* and *Lamc1* genes in $Shh^{-/-}$ embryos. Was their expression disturbed similarly to that of *Lama1*? I addressed this question by studying the mRNA expression pattern of *Lamb1* and *Lamc1* in *Shh*-mutant embryos. Finally, I hypothesized that the expression pattern of *Lama1* is conserved across vertebrates, based on existing knowledge in the mouse and zebrafish models, and to demonstrate this I investigated *Lama1* expression in a third vertebrate species – the embryo of the domestic chicken.

3.2. Results

First, I confirmed our previous observations on the expression of *Lama1* in E9.5 wild type and $Shh^{-/-}$ embryos, using a 950 bp digoxigenin-labelled antisense RNA probe complementary to a region of the murine *Lama1* mRNA spanning from position 4, 193 to position 5, 142 relative to the transcription start site (Claire Anderson's Thesis 2009).

In wild type embryos (here labelled as $Shh^{+/+}$), the most prominent sites of *Lama1* expression were the neural tube, presomitic mesoderm and somites, the mesonephros, optic and otic vesicles, and the head mesenchyme (Figure 3.1A-F). Notably, *Lama1* expression in the neural tube and somites was completely obliterated in *Shh*-deficient embryos, while expression in the presomitic mesoderm, and to some extent in the head mesenchyme, remained unaffected (Figure 3.1G-L). This result is in agreement with earlier studies in our lab (Anderson et al. 2009).

Further examination of *Lama1* expression in E10.25-E10.5 wild type embryos showed little change from E9.5. Similarly, *Lama1* transcripts were abundant along the neural tube, in the head mesenchyme, in the presomitic mesoderm and differentiating sclerotomes,
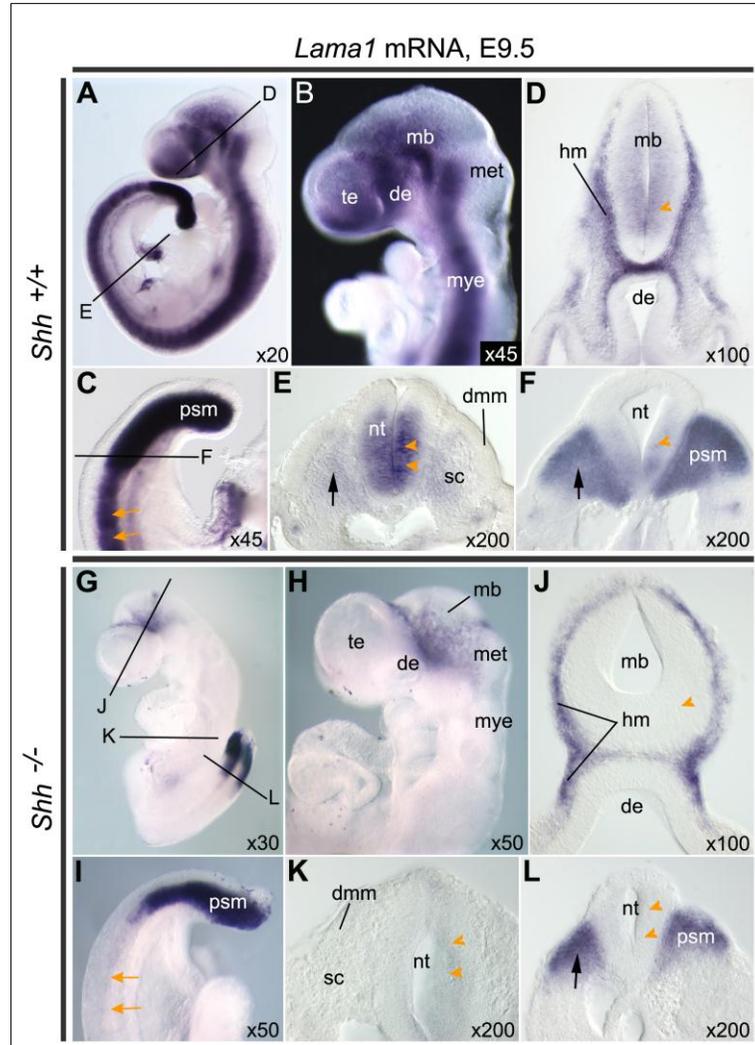
Figure 3.1. *Expression of Lama1 mRNA in E9.5 wild type and Shh mutant mouse embryos.* *Lama1* expression in E9.5 *Shh*[+/+] (A-F) (n=3), and *Shh*[-/-] (G-I) (n=2) mouse embryos was analysed by whole-mount *in situ* hybridisation. (A) Note that *Lama1* expression extends throughout the entire neural tube. (B) In the head, *Lama1* is detected in all divisions of the brain and in the surrounding head mesenchyme (section in D). (C) High magnification of the tail region showing high levels of *Lama1* expression in the presomitic mesoderm and posterior somites. (E) and (F) are sections through the flank and tail regions, respectively, showing strong expression in the neural tube (*orange arrowheads* in E) and PSM (*black arrow* in F), and weak expression in the sclerotome (*black arrow* in E). (G) Note the absence of expression in the neural tube and somites of the *Shh*[-/-] embryo. (H) Absence of *Lama1* expression in all brain divisions, but not in the head mesenchyme, is observed in the *Shh* mutant embryo (section in J). (I) High magnification of the tail region showing the complete absence of *Lama1* transcripts in all somites (*orange arrows*) but not in the presomitic mesoderm. (K) and (L) are sections through the trunk and tail regions of the *Shh*[-/-] embryo revealing the lack of *Lama1* expression in the neural tube (oranhe arrowheads) and somites (K), while expression in the presomitic mesoderm remains intact (*black arrow* in L). Abbreviations: de, diencephalon; dmm, dermomyotome; hm, head mesenchyme; mb, midbrain; met, metencephalon; mye, myelencephalon; nt, neural tube; psm, presomitic mesoderm; sc, sclerotome; te, telencephalon.

67
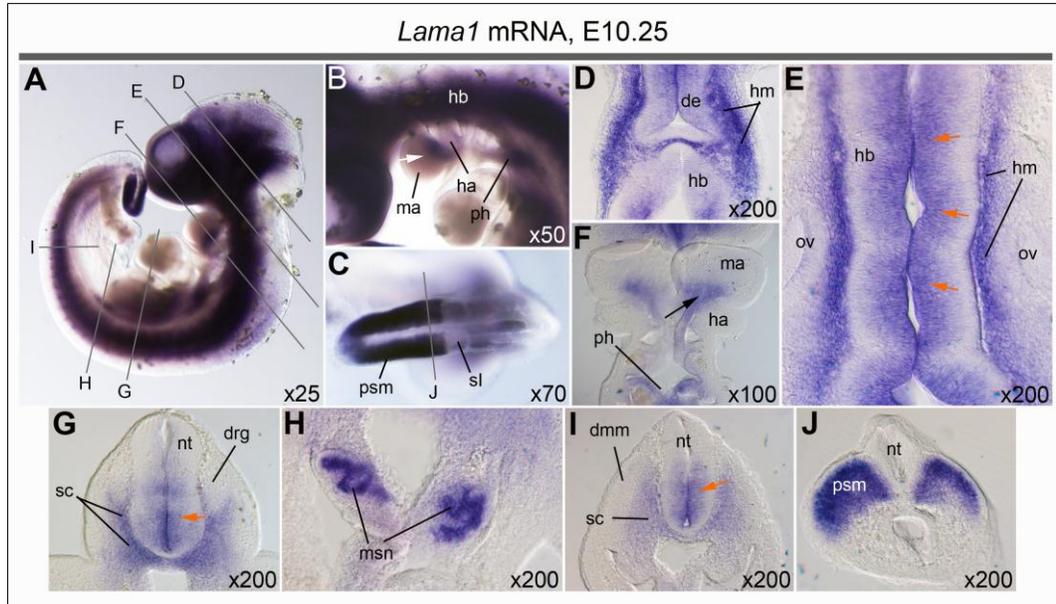
Figure 3.2. *Expression of Lama1 mRNA in E10.25 wild type mouse embryos. Lama1 expression in E10.25 wild type (A-F)* (n=3) *mouse embryos was analysed by whole mount* in situ *hybridisation. (A) Note the strong expression of* Lama1 *in the entire neural tube. (B) High magnification of the left side of the facial region showing* Lama1 *expression between the* 1st *and* 2nd*pharyngeal arches (*white arrow*). (C) Magnified dorsal view of the tail bud showing high* Lama1 *expression in the presomitic mesoderm and weaker expression in the somites. (D-F) Horizontal sections through the head where anterior is* up*, posterior is* down*. Lama1 is expressed in the brain and surrounding head mesenchyme (D, E) and in the* 1st *pharyngeal pouch (indicated by a* black arrow *in F). (G-J) Transverse sections through the trunk and tail regions. Lama1 is expressed in the condensing sclerotome and in the ventral neural tube, but not in the dermomyotome (G) and dorsal root ganglia (I). (H) Expression in the mesonephric tubules is very prominent. Note the higher levels of* Lama1 *expression in the ventricular zone of the neural tube (*orange arrows *at E, G and I). Abbreviations: de, diencephalon; dmm, dermomyotome; drg, dorsal root ganglion; ha, hyoid arch (*2nd *pharyngeal arch); hb, hindbrain; hm, head mesenchyme; ma, mandibular arch (*1st *pharyngeal arch); msn, mesonephros; nt, neural tube; ov, otic vesicle; ph, pharynx; psm, presomitic mesoderm; sI, somite I (designates a newly formed somite); sc, sclerotome.*

and in the mesonephros (Figure 3.2). Notably, *Lama1* expression in the neural tube was confined to the ventricular zone in the ventral half of the tube (Figure 3.2E, G and I). Interestingly, a new domain of expression has appeared in the endodermal lining of the 1st pharyngeal pouch (Figure 3.2B and F).

3.2.1. *Shh* is required for *Lama1* transcription in the somites, but not in the neural tube, of E8.5 embryos

Shh is essential for *Lama1*expression in the neural tube and somites of E9.5 mouse embryos. However, it is unclear whether SHH initiates or maintains *Lama1* transcription. To address this question, I studied *Lama1* expression at an earlier stage, in E8.5 *Shh*[+/+] and *Shh*[-/-] mouse embryos. *Lama1* is already expressed in the neural tube, presomitic mesoderm and somites of

E8.5 $Shh^{+/+}$ embryos (Figure 3.3A-D). In $Shh^{-/-}$ embryos, *Lama1* transcripts are absent from somites (Figure 3.3E, F, G), but maintained in the presomitic mesoderm (Figure 3.3H), as in E9.5 $Shh^{-/-}$ embryos. However, in contrast to E9.5 *Shh*-null embryos, *Lama1* expression in the neural tube of E8.5 $Shh^{-/-}$ embryos was largely unaffected (Figure 3.3E, F, G), except for the posterior-most region of the neural tube where it was absent. These observations suggest that SHH is required for the initiation of *Lama1* expression in the somites. In contrast, SHH appears to function in the maintenance of *Lama1* expression in the neural tube and not in its initiation. Altogether, these observations reveal distinct requirements for SHH in *Lama1* expression in the neural tube and somites, and raise interesting questions about the molecular mechanisms governing *Lama1* transcription in these tissues.
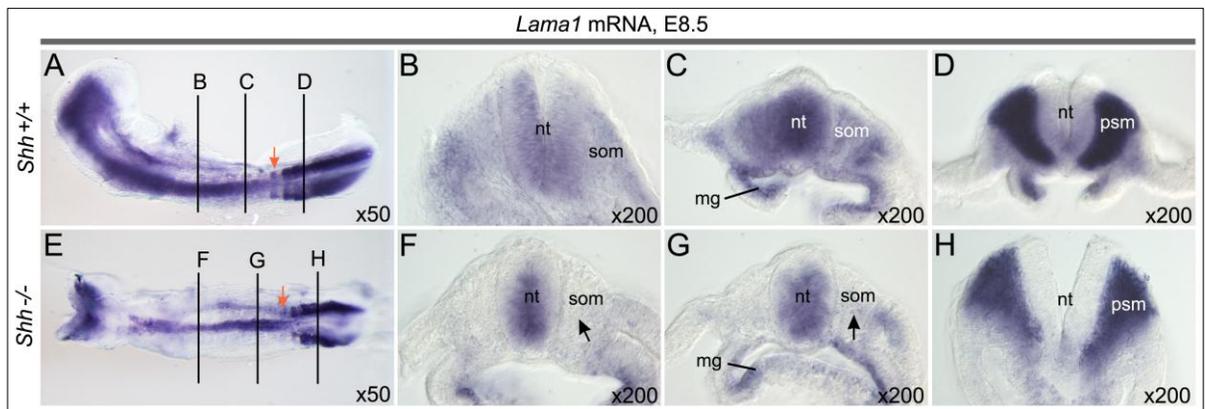
Figure 3.3. *Expression of Lama1 mRNA in E8.5 wild type and Shh mutant mouse embryos. Lama1* expression in E8.5 $Shh^{+/+}$ (A-D) (n=3), and $Shh^{-/-}$ (E-H) (n=3) mouse embryos was analysed by whole- mount *in situ* hybridisation. (A) and (E) present dorsal views of the whole embryo, while (B-D) and (F-H) are transverse sections. *Lama1* is expressed throughout the neural tube and somites (A, B, C, D), and presomitic mesoderm (D) of the wild-type embryo. In the $Shh^{-/-}$ embryo there is lack of expression in the somites (E; *black arrow* in F and G), but not in the neural tube (E, F, G) and presomitic mesoderm (H). The *orange arrow* in (A) and (E) signify the position of a newly-formed somite, sI. Abbreviations: nt, neural tube; psm, presomitic mesoderm; som, somite.

3.2.2. Expression of *Lamb1* and *Lamc1* is unaffected in *Shh*-deficient mouse embryos

According to the current model in our lab, the failure in myotomal basement membrane assembly in $Shh^{-/-}$ mouse embryos is caused by the loss of *Lama1* expression in the somites, which leads to inability to secrete the laminin-111 trimer that is critical to initiate the assembly of the myotomal basement membrane (Anderson et al. 2009). However, the question of whether expression of *Lamb1* and *Lamc1*, encoding the other two subunits of laminin-111, is also affected in *Shh*-null embryos, particularly in the neural tube and somites,

remained unanswered. Therefore, I investigated the expression of these laminin genes in wild type and *Shh* mutant embryos. Consistent with earlier reports (Tunggal et al. 2000), I observed ubiquitous expression of both *Lamb1* and *Lamc1* in *Shh*$^{+/+}$embryos (Figure 3.4A-C and Figure 3.5A, B, respectively). Transcripts were detected in the neural tube, somites, presomitic mesoderm, pharyngeal arches, etc. A notable difference between the expression pattern of *Lamb1* and *Lamc1* in *Shh*$^{+/+}$ mice was the absence of *Lamb1* transcripts from the heart, while *Lamc1* was expressed there (Figure 3.4A and Figure 3.5A). However, most importantly, the expression of both *Lamb1* and *Lamc1* was unaffected in the *Shh*$^{-/-}$ embryos, and the staining in the neural tube and somites, in particular, did not significantly differ from that observed in wild type embryos (Figure 3.4D-F and Figure 3.5C, D).
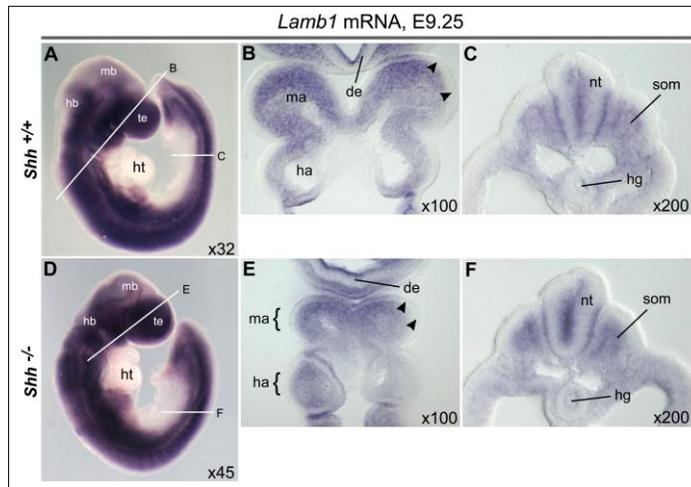
Figure 3.4. *Expression of Lamb1 mRNA in E9.25 wild type and Shh mutant mouse embryos. Lamb1* expression in E9.25 *Shh*$^{+/+}$ (A-C) (n=2), and *Shh*$^{-/-}$ (D-F) (n=2) mouse embryos was analysed by whole- mount *in situ* hybridisation. (A) and (D) are images of the whole embryo, (B) and (E) are horizontal sections through the head region, while (C) and (F) are transverse sections through the trunk. (A) *Lamb1* mRNA is expressed in most tissues, but not in the heart (A) and the ectoderm of the pharyngeal arches (*black arrowheads* in B, E). (C) Both the neural tube and somites express *Lamb1*. (D) The pattern of *Lamb1* expression is largely unchanged in *Shh*$^{-/-}$ embryos, and particularly in the head (E), as well as in the neural tube and somites (F). Abbreviations: de, diencephalon; ha, hyoid arch (2nd pharyngeal arch); hb, hindbrain; hg, hindgut; ht, heart; ma, mandibular arch (1st pharyngeal arch); mb, midbrain; nt, neural tube; som, somite.

This indicates that SHH is not required for the expression of *Lamb1* and *Lamc1* at the examined stages and further supports the model that it is the lack of *Lama1* transcription that is mainly responsible for the failure to assemble the myotomal basement membrane in *Shh*-deficient embryos.
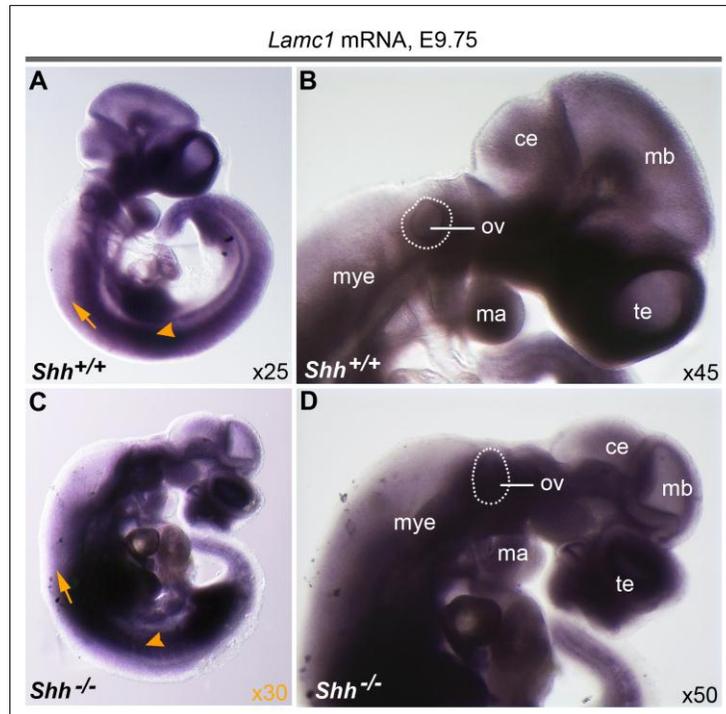
Figure 3.5. *Expression of Lamc1 mRNA in E9.75 wild type and Shh mutant mouse embryos. Lamc1* expression in E9.75 *Shh*$^{+/+}$ (A, B) (n=2), and *Shh*$^{-/-}$ (C, D) (n=2) mouse embryos was analysed by whole- mount *in situ* hybridisation. (A) and (C) images of the whole embryo; (B) and (D) lateral view of the right side of the head. (A) and (B) *Lamc1* is expressed throughout the body of the wild type embryo, including the heart, all brain divisions, the future spinal cord (*orange arrow*), the somites (*orange arrowhead*), pharyngeal arches and otic vesicle (outlined in *white*). (C) and (D) Notably, the expression of *Lamc1* is unaffected in *Shh* mutant embryos. Abbreviations: ce, cerebellum; ma, mandibular arch (1st pharyngeal arch); mb, midbrain; mye, myelencephalon; ov, otic vesicle; te, telencephalon.

3.2.3. *Lama1* transcription in the chicken embryo

Considering the conserved pattern of *Lama1* expression between mouse and zebrafish embryos, I was eager to learn whether this phenomenon is a general characteristic of vertebrates. Therefore, I decided to examine *Lama1* mRNA expression in a third vertebrate embryo – that of the domestic chicken (*Gallus gallus dom.*).

The expression pattern of the chicken *Lama1* gene (*cLama1*) is poorly documented. Zagris et al. (2000) performed *in situ* hybridization studies with radiolabelled anti-sense RNA probes to chicken *Lama1* mRNA in embryos from pre-gastrulation Hamburger-Hamilton stage 1 (HH1) until 10-somites stage (HH11). The authors reported strong expression in both the epiblast and hypoblast at HH1, and at the stage of primitive streak formation (HH3-4) *Lama1* was strongly expressed by mesenchymal cells ingressing through the primitive groove and more weakly by the ectoderm. During head fold development (HH5-6), intensive *Lama1*

71

staining appeared in the neuroectoderm and lateral mesoderm. Curiously, Zagris et al. (2000) failed to detect *Lama1* expression in the neural tube and sclerotome of 10-13-somites stage embryos (HH10-11). Instead, they reported high expression in the pronephros, weaker expression in the dermomyotome and absence of *Lama1* in the sclerotome. This lack of *Lama1* transcripts in the early avian neural tube contrasts with data from the rodent (Anderson et al. 2009; Miner et al. 2004) and zebrafish (Figure 1.3; Joseph Pickering's Thesis 2012; Zinkevich et al. 2006) models where, in the case of the mouse embryo, *Lama1* is highly expressed in the neural tube even at early stages of somitogenesis at E8.9 (this study). Moreover, during murine somite maturation *Lama1* transcripts disappear from the dermomyotome, while expression is maintained at low levels in the sclerotome (Anderson et al. 2009; this study), which is opposite to the report from Zagris et al. (2000).

However, there were several issues with the latter study. First, the authors used mouse embryo cDNA to generate their "anti-chicken *Lama1* mRNA" probes which could theoretically have an effect on the probe's ability to correctly recognize the avian *Lama1* mRNA, and hence generate trustful staining. Second, their data for the lack of neural tube and sclerotomal expression is derived from only a single HH10-11 chicken embryo. Thus, it could be argued that the lack of *Lama1* expression in these tissues is a random artifact of the assay and does not represent the true pattern.

Therefore, I decided to re-analyse the expression of the chicken *Lama1* and add data for later stages as well, which were not covered by Zagris et al. (2000). To achieve this, I generated a 819 bp digoxigenin-labelled antisense RNA probe that was complementary to a region of the chicken *Lama1* mRNA spanning from position 506 to position 1, 324 relative to the transcription start site. This probe was then used to examine the expression pattern of the *cLama1* gene.

cLama1 displays a tissue-specific but dynamic expression pattern. At the earliest stage investigated – Hamburger-Hamilton stage 4 (or HH4, 18 h of incubation), *cLama1* transcripts are restricted to ingressing mesodermal cells that have invaded the space between the ectoderm and endoderm (Figure 3.6A-D). Later, at HH8 (28 h of incubation), the pattern dramatically changes with high levels of *cLama1* expression in the forming neural tube anteriorly and the neural plate posteriorly, while expression in most of the mesoderm is diminished (Figure 3.7A-F). Interestingly, there is a domain of *cLama1* expression restricted
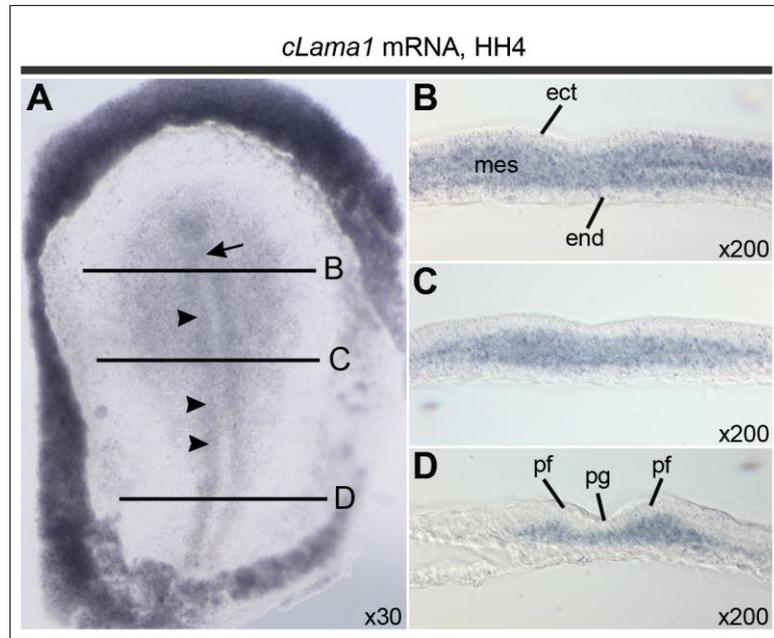
Figure 3.6. *Expression of cLama1 mRNA in HH4 chicken embryos. cLama1* expression in HH4 chicken embryos (n=2) was analysed by whole-mount *in situ* hybridization. (A) Dorsal view of the whole embryo showing the fully developed primitive streak (*black arrowheads*) with the Hensen's node (*black arrow*) at its anterior-most end. Expression of *cLama1* is evident in the ingressed mesoderm. (B-D) Transverse sections showing high expression in the mesoderm and low or absent expression in the ectoderm and endoderm. Abbreviations: ect, ectoderm; end, endoderm; mes, mesoderm; pf, primitive fold; pg, primitive groove.

to the floor of the foregut endoderm and in what appears to be the posterior-most edge of the unsegmented cranial paraxial mesoderm (anterior to somite 1) (Figure 3.7B and H). Most posteriorly, where gastrulation movements are still proceeding, there is only speckled expression in the mesoderm (Figure 3.7G).

By HH15 (51 h of incubation), new domains of *cLama1* transcription have appeared (Figure 3.8A-H). At this stage, prominent sites of *cLama1* expression are the somites, nephrogenic mesenchyme, posterior lateral plate mesoderm and the tail bud (Figure 3.8B, D, F, H). Similar to the mouse, *cLama1* was expressed throughout the epithelium in newly formed somites, but in more mature somites *cLama1* mRNA is restricted to the sclerotome, while expression in the dermomyotome is extinguished (Figure 3.8D, E). Surprisingly, in contrast to the situation in mouse and zebrafish embryos, there is a conspicuous lack of *cLama1* transcripts in the presomitic mesoderm (Figure 3.8B, F). Expression in the neural tube continues and spans the whole length of this structure. Interestingly, in the posterior neural tube *cLama1* expression extends through the whole medio-lateral width of the

neuroepithelium, while anteriorly it is restricted to the ventricular zone of the ventral half of the neural tube (Figure 3.8C, F). Another curious detail is the lack of *cLama1* transcripts in the nephric duct while high levels are detected in the nephrogenic mesenchyme that will subsequently generate the mesonephric tubules (Figure 3.8D).
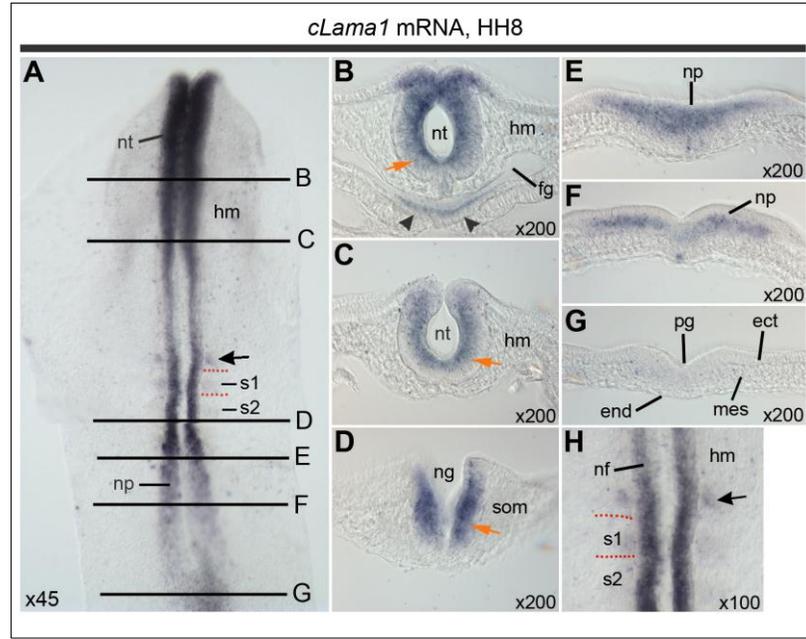
Figure 3.7. *Expression of cLama1 mRNA in HH8 chicken embryos. cLama1* expression in HH8 chicken embryos (n=3) was analysed by whole-mount *in situ* hybridization. (A, H) Dorsal views. (B-G) Transverse sections. (A) *cLama1* mRNA is abundant in the closing neural tube and in the neural plate at the posterior of the body (A, E, F), as well as in the posterior-most edge of the unsegmented cranial paraxial mesoderm (*black arrow* in A and H). Low levels of expression are observed in the 1st somte (H). The *dotted orange lines* delineate the intersomitic clefts. Note the higher expression of *cLama1* in the ventricular zone of the neural tube (orange arrows in B, C and D). The floor of the foregut also expresses *cLama1* (*black arrowheads* in B). Abbreviations, ect, ectoderm; end, endoderm; fg, foregut; hm, unsegmented head mesoderm; mes, mesoderm; nf, neural fold; ng, neural groove; np, neural plate; nt, neural tube; pg, primitive groove; s1, 1st somite; s2, 2nd somite; som, somite.

At the latest examined stage, at HH18 (67 h of incubation), *cLama1* continues to be expressed in the somites, neural tube, mesonephros and posterior lateral plate mesoderm with no expression in the presomitic mesoderm (Figure 3.9A-D). Similar to HH15, expression in mature somites is restricted to the sclerotome, while in the neural tube *cLama1* RNA is confined to the ventricular zone in the ventral half of the tube (Figure 3.9D). The nephric duct is devoid of *cLama1* expression in contrast to the nephrogenic mesenchyme, congruent to the pattern at HH15 (Figure 3.9D').

Figure 3.8. *Expression of cLama1 mRNA in HH15 chicken embryos. cLama1* expression in HH15 chicken embryos (n=3) was analysed by whole-mount *in situ* hybridization. (A) Dorso-lateral view of a whole embryo showing expression in the neural tube, somites, nephric mesoderm and lateral plate mesoderm. (B) High magnification of the posterior half of the embryo (dorsal view, anterior is to the top); a *green arrow* indicates *cLama1* expression in a newly-forming somite (s0) in the rostral-most edge of the presomitic mesoderm. (B, F) Note the absence of *cLama1* mRNA from most of the presomitic mesoderm. (H) High magnification of the tail bud region (dorsal view, anterior is to the top). (C-G) Transverse sections. (C, F) *cLama1* transcripts are enriched in the ventricular zone of the neural tube (*orange arrows*). (E, D, D') *cLama1* is expressed in the nephrogenic mesenchyme (*black arrow*), but not in the nephric duct (*black arrowhead*). (F, H') Note the presence of *cLama1* mRNA in the posterior lateral plate mesoderm. Abbreviations: dmm, dermomyotome; lpm, lateral plate mesoderm; nc, notochord; mt, myotome; nt, neural tube; sc, sclerotome; sI, somite I (designates a newly formed somite); som, somite; psm, presomitic mesoderm; tb, tail bud.
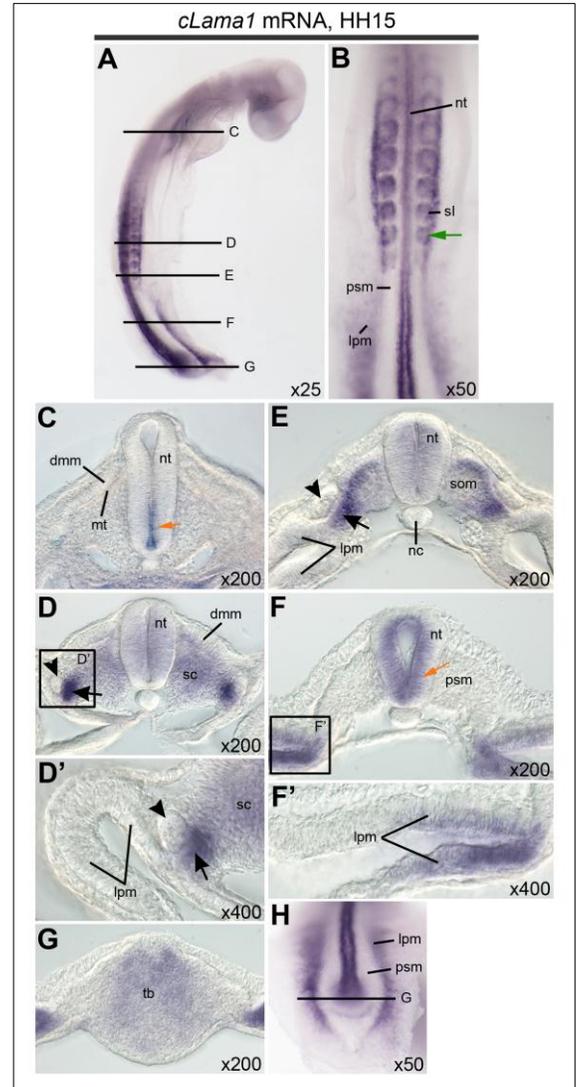
In summary, contrary to Zagris et al. (2000), I observed high expression of *Lama1* in the chicken neural tube, weaker signal in the sclerotome, absent from the dermomyotome, and curiously, also absent from the presomitic mesoderm, unlike the situation in mouse and zebrafish.

Overall, the expression pattern of the chicken *Lama1* gene is highly similar to that in the mouse and zebrafish embryos, suggesting conserved mechanisms of transcriptional control on *Lama1*. Importantly, this hints for the possibility that SHH signaling is similarly involved in the regulation of *Lama1* expression in the ventral neural tube and somites across vertebrates.
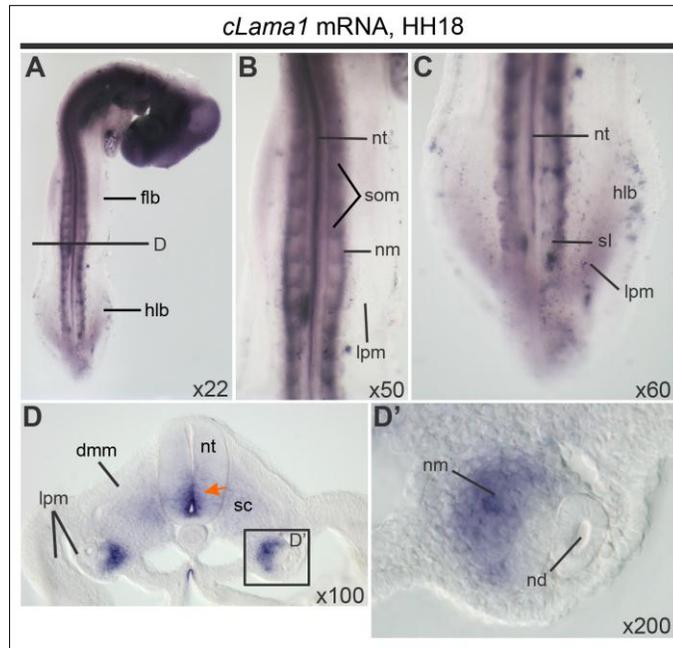
Figure 3.9. *Expression of cLama1 mRNA in HH18 chicken embryos. cLama1* expression in HH18 chicken embryos (n=2) was analysed by whole-mount *in situ* hybridization. (A) Dorsal view of the whole embryo showing intense *cLama1* expression in the future brain and spinal cord, somites and nephrogenic mesenchyme. (B) and (C) High magnification of the trunk at the forelimb level and of the tail bud region, respectively (dorsal views, anterior is to the top). (D, D') A transverse section through the interlimb region indicating strong *cLama1* expression in the ventricular zone of the ventral neural tube (*orange arrow*) and nephrogenic mesenchyme, lower levels of expression the sclerotome, and absence in the nepric duct. Abbreviations: dmm, dermomyotome; flb, forelimb bud; hlb, hindlimb bud; lpm, lateral plate mesoderm; nd, nephric duct; nm, nephrogenic mesenchyme; nt, neural tube; sc, sclerotome; sI, somite I (designates a newly formed somite); som, somites.

3.3. Discussion

I examined the expression pattern of the *Lama1* gene in rodent and avian embryos. Prominent sites of expression in the mouse embryo are the neural tube, somites, presomitic mesoderm, nephric structures and head mesenchyme, which is consistent with older reports. Confirming previous findings from our laboratory, I found that *Lama1* transcription in the neural tube and somites was lost in *Shh*-deficient E9.5 mouse embryos, while expression in the presomitic mesoderm was unaffected. Interestingly, SHH was not required for *Lama1* expression in the neural tube of E8.5 embryos, in contrast to such requirement in the somites. I also showed that expression of *Lamb1* and *Lamc1* genes is not affected in *Shh*-null mouse embryos. Finally, I studied the expression pattern of the chicken *Lama1* gene and found that it is strongly expressed in the neural tube, nephric mesoderm, cranial mesenchyme and somites,

but not in the presomitic mesoderm, and concluded that the expression pattern of the *Lama1* gene in vertebrates is largely conserved.

3.3.1. *Lama1* expression in the neural tube of mouse and chicken embryos.

In agreement with prior studies, I found that *Lama1* is characteristically expressed along the neural tube (from the ventral telencephalon to most posterior region of the spinal cord) in mouse embryos and I also showed that expression in this structure is conserved in the chicken.

3.3.1.1. SHH is not required for *Lama1* gene activation in the neural tube

Intriguingly, *Shh*-defficient E8.5 embryos displayed persistent *Lama1* expression in the neural tube but not in the somites (Figure 3.3), in contrast to the situation in E9.5 $Shh^{-/-}$ animals where both the CNS and somites were devoid of *Lama1* mRNA. Interestingly, a similar relationship exists in zebrafish, where expression of *lama1* in the neural tube is unaffected in early $smo^{-/-}$ or cyclopamine-treated wild type embryos (12-15 somites stage), while it is reduced in the brain of cyclopamine-treated zebrafish at 27 hpf (Table 1.2; Joseph Pickering's Thesis 2012).

The observations in the mouse embryo suggest that: 1) *Lama1* expression in the somites and neural tube is governed by distinct *cis*-regulatory elements, and 2) SHH plays a role in the maintenance of *Lama1* transcription in the CNS rather than in its initial activation there, similarly to the WNT3a-dependent maintenance, but not initiation, of brachyury gene transcription in the paraxial mesoderm (Galceran et al. 2001).

The latter conclusion also hints to the existence of separate enhancers controlling the activation and the maintenance of *Lama1* expression in the neural tube. Such scenario is consistent with studies reporting the presence of distinct enhancers dedicated to either activation or maintenance of developmental gene expression. For instance, the modulation of *Pax2* expression in the murine midbrain-hindbrain boundary (MHB) is governed by two enhancers (Pfeffer et al. 2002). A POU5f1-dependent 120 bp element, termed the "activation" enhancer, is sufficient and required to activate *Pax2* expression in the anterior neural plate at late gastrulation. Another 410 bp enhancer with functional PAX2/5/8 binding motifs is both sufficient and necessary to maintain *Pax2* expression in the MHB at later stages, indicating a positive feedback by PAX2 on its own gene's transcription (Pfeffer et al. 2002). Similarly, investigation of the complex transcriptional regulation of the *Myf5-Mrf4*

locus identified separate enhancers for activation and maintenance of *Myf5* expression in the pharyngeal arches, as well as in the axial muscles (Carvajal et al. 2001).

Alternatively, both activation and maintenance of *Lama1* expression within the mouse neural tube could be controlled by a single enhancer, which would require distinct inputs from different transcription factors and signaling pathways at early (activation) and later (maintenance) stages. This is the case of the *orthodenticle* (*otd*) gene in *Drosophila* that is essential for ocellar development. Expression of *otd* in the eye-antennal disc is controlled by a single enhancer – *ocelliless*, which is initially activated by inputs from the Wingless and Hedgehog signalling pathways, while its later activity is maintained by autoregulatory inputs by the Otd transcription factor itself (Blanco et al. 2009).

3.3.1.2. Laminins and neural progenitor cells

Interestingly, *Lama1* expression accumulates in the ventricular zone (VZ) of the neural tube, particularly in the anterior neural tube at E8.5, 9.5 and 10.25 in the mouse (Figure 3.1D; Figure 3.2E; Figure 3.3B, F) and at HH8/HH15 in the chicken (Figure 3.7B, C; Figure 3.8C), while in later stage embryos (Figure 3.2G and Figure 3.9D) this ventricular enrichment is also observed in the posterior neural tube. This suggests that *Lama1* is predominantly transcribed by the pool of proliferating neural progenitor cells located in the ventricular zone and is down-regulated in post-mitotic neural or glial cells. Thus, it is possible that laminin α1-containing laminins play a role in modulating the behavior of neural progenitor cells. Consistent with this possibility, several studies have reported the effects of laminins on the maintenance of neural stem cells (Flanagan et al. 2006; Hall et al. 2008), and led to the concept that laminins may be an essential component of the ECM in the neural stem cell niche (Lathia et al. 2007), as described in more detail in Chapter 1.

3.3.1.3. Laminins in cortical morphogenesis and axonal growth and migration.

The pattern of *Lama1* expression in both the mouse and chick embryos is also consistent with roles of this laminin subunit in glial development and cortical morphogenesis, and multiple studies have elucidated critical functions of the laminin α1, α2 and γ1 subunits in axonal migration, fasciculation, and glial development in the CNS, as described in Chapter 1. The *Lama1*-knockout studies by Heng et al. (2011) and Ichikawa-Tomikawa et al. (2012) were especially important for they demonstrated a critical role of the laminin α1-containing pial basement membrane in cerebellar development. These findings are consistent with my

observations that *Lama1* is transcribed in the embryonic neuroepithelium, and also in mesenchymal cells aggregated at the basal surface of the neural tube (Figure 3.2D, E, G). These cells form the meninges of the forebrain and spinal cord and are derived from the cranial neural crest and the sclerotome, respectively (Couly et al. 1993; Christ et al. 2007). Future studies using conditional inactivation of *Lama1* in specific neural cell populations at different temporal stages would be instrumental in elucidating the plethora of unknown functions of *Lama1* in the developing nervous system.

3.3.2. Laminins and neural crest cells

Laminin α1 expressed by the neural tube and somites may also participate in the migration of neural crest (NC) cells along basement membranes of the neural tube, dermomyotome and myotome, as described in Chapter 1 where it was highlighted that neural crest cells up-regulate laminin synthesis at the onset of ganglia formation (Duband and Thiery 1988). It would be interesting to determine the laminin isoform present within the condensing ganglia, although laminin α1 is a unlikely candidate as I found no *Lama1* expression of this subunit in the dorsal root ganglia (Figure 3.2G). However, laminin-111 synthesised by sclerotomal cells may be utilized by aggregating NC cells.

3.3.3. *Lama1* expression in somites and presomitic mesoderm

Lama1 is expressed in the somites of all three vertebrates analysed to date – in mouse, zebrafish and chicken. While expression is uniform in newly-formed somites of mouse and chicken embryos (Figure 3.3C), *Lama1* expression is down-regulated in the dermomyotome upon or before epithelial-to-mesenchymal transition of the ventral somitic half (Figure 3.8E), but remains expressed in the mesenchymal sclerotome (Figure 3.1E; Figure 3.2G, I; Figure 3.8D). A similar pattern is observed in zebrafish embryos (Figure 1.3; Joseph Pickering's Thesis 2012).

 Lama1 expression in the mouse somites is required for proper myotomal morphogenesis and differentiation and, directly or not, for trunk neural neural crest migration, through the involvement of laminin α1 in the formation of the myotomal basement membrane (Anderson et al. 2009). It is likely that chicken laminin α1 performs similar functions. This could be tested via antibody-mediated blocking of laminin α1 interactions with cellular receptors like integrin and/or dystroglycan, or by siRNA-mediated down-regulation of *cLama1* mRNA expression in the paraxial mesoderm.

3.3.3.1. The requirement for SHH in the presomitic mesoderm and somites is different in mouse and zebrafish

Here, I confirmed previous studies from our lab (Anderson et al. 2009) demonstrating the requirement for SHH in the expression of *Lama1* in the mouse somites and neural tube at E9.5 (Figure 3.1), but not in the presomitic mesoderm. I also provided evidence suggesting that *Shh* is required for the initiation of *Lama1* expression in somites, whereas expression in the presomitic mesoderm is independent of SHH signalling (Figure 3.3). Interestingly, the opposite is true in zebrafish, as *lama1* expression requires Hh signalling in the presomitic mesoderm, but not in the somites. Indeed, 27 hpf *smoothened*-deficient fish embryos displayed a strong reduction of *lama1* mRNA in the presomitic mesoderm, but not in the somites (Table 1.2; Joseph Pickering's Thesis 2012).

These observations suggest that distinct signalling mechanisms control *Lama1* expression in the somites and in the presomitic mesoderm. One may speculate that *Lama1* transcription in these structures is controlled 1) via a single "paraxial mesoderm" enhancer, or alternatively, 2) via separate "somite" and "PSM"-dedicated enhancers. The first scenario would assume that the availability of some transcription factors differs between the somites and presomitic mesoderm. For instance, GLI are not expressed in the murine PSM but are expressed in the somites, while in zebrafish they are expressed in both tissues (Hui et al. 1994; Thisse et al. 2004), suggesting that the mouse PSM employs an alternative GLI-independent mechanism to induce *Lama1*. According to the second scenario, the somite enhancer(s) of the murine *Lama1* gene would be expected to be SHH-responsive, while the PSM enhancer(s) would not. The opposite would be true for zebrafish *lama1*, although constitutive activation of Hh signaling is sufficient to up-regulate *lama1* in zebrafish somites (Table 1.2; Joseph Pickering's Thesis 2012), suggesting that a somite-specific regulatory element(s) which is Hh-responsive is present in the zebrafish *lama1* locus. In fact, what appears to be a lack of requirement for Hh in the somitic expression of fish *lama1* could be due to combinatorial regulation by other signaling pathways and their downstream transcription factors, whose combined inputs form an "OR" logic processing gate, instead of an "AND" one in the control of *lama1* transcription in the fish somites (Davidson 2006). The same could be true for the presomitic mesoderm in the mouse.

Based on this, one can envision a hypothetical scenario where somitic and presomitic expression of *Lama1* in the last common ancestor of zebrafish and mouse was modulated by two regulatory elements, one for the somite and one for the PSM, both of which receiving inputs from HH signaling, perhaps in the form of direct binding by GLI factors (Figure 3.10). Beside the GLI inputs, other inputs – from factor X in the somite enhancer and from factor Y in the PSM enhancer, acted in a combinatorial, quasi-redundant manner with the GLI factors to drive *Lama1* expression in these tissues. Upon divergence of the actinopterygian (leading to zebrafish) and sarcopterygian (leading to mouse) lineages nearly 419 million years ago (Zhu et al. 2009), one may hypothesise that the two enhancers underwent independent loss of transcription factor inputs, such that the fish lineage lost input Y from the PSM enhancer, while the mouse lineage lost input X from the somite enhancer (Figure 3.10). In effect, the losses rendered the zebrafish PSM enhancer and the mouse somite enhancer dependent on GLI inputs, and these dependencies were revealed upon experimental perturbation of the HH signaling pathway. In contrast, the murine PSM- and fish somite enhancers were insensitive to HH signaling deficiencies for they still harbored the quasi-redundant Y and X inputs, respectively (Figure 3.10).
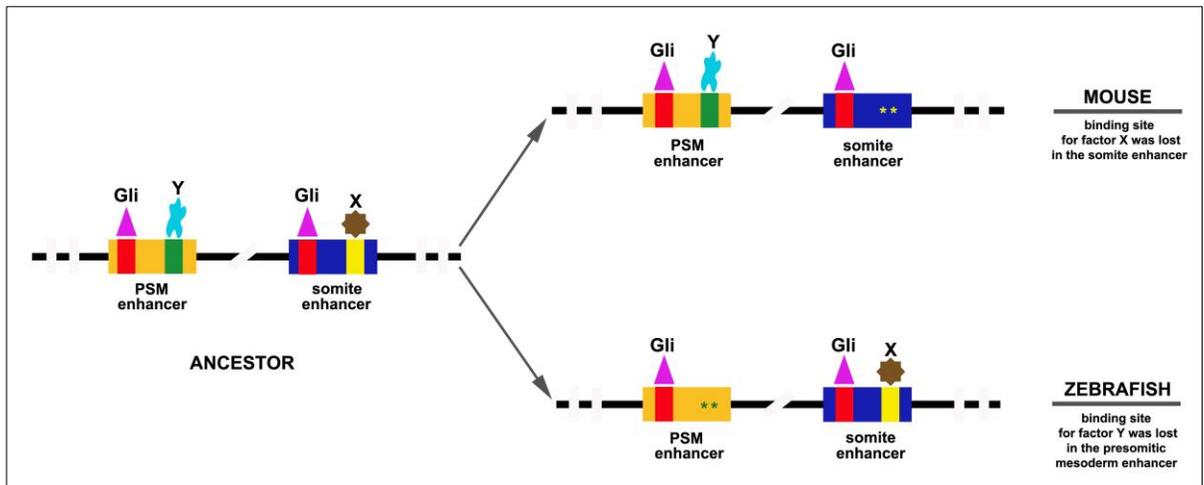
Figure 3.10. *A hypothetical scenario for the evolution of the putative somite and presomitic mesoderm enhancers of Lama1*. A schematic diagram illustrating the hypothetical presence of separate somite and presomitic mesoderm (PSM) enhancers of *Lama1* in the last common ancestor of teleosts and mammals, which were driven by semi- (quasi-) redundant inputs from GLI combined with X or Y factors in the somite and PSM enhancer, respectively. After divergence of the sarcopterygian and actinopterygian lineages, the somite enhancer lost the binding site for factor X in the lineage lieading to mouse, whereas the zebrafish lineage lost the binding site for factor Y in the PSM enhancer. Thus, the murine somite enhancer and the fish PSM enhancer were rendered strongly dependent on Hh signaling (as discussed in the text). Legend: *magenta triangles*, Gli factors; *cyan shape*, factor Y; *brown shape*, factor X; *red rectangles*, GLI binding sites; *green rectangles*, factor Y binding sites; *yellow rectangles*, factor X binding sites; *green asterisks*, lost factor Y binding site; *yellow asterisks*, lost factor X binding site.

81

A striking example is provided by the eve stripe 2 enhancer (es2E) in *Drosophila melanogaster*, which is regulated by inputs from products of the *hunchback*, *bicoid*, *kruppel* and *giant* genes, where some of the binding motifs appear to be redundant indicated by the lack of qualitative changes in the expression pattern driven by mutant variants of es2E (Arnosti et al. 1996; Lagha et al. 2012). Similarly, Drewell (2011) investigated the pattern of distribution of Kruppel and Hunchback transcription factor binding sites (TFBSs) within the BX-C (*bithorax complex*) in *D. melanogaster* and concluded that clustering of TFBSs allows for extensive functional redundancy such that losses of individual sites do not result in obvious phenotypic changes (Drewel 2011). This suggests that enhancers can tolerate extensive sequence turnover with gain and losses of TFBSs without overt effects on the regulatory output of the elements, perhaps because of quasi-redundancy between particular TFBSs (Weirauch and Hughes 2010).

3.3.3.2. Putative functions of laminin α1 in the presomitic mesoderm (PSM)

It is possible that laminin-111 synthesised in the murine PSM incorporates into the basement membrane surrounding the neural tube, which could explain the presence of laminin α1 protein on the pial surface of the neural tube even in $Shh^{-/-}$ embryos, as hypothesized by Anderson et al. (2009).

In regard to functions within the PSM, laminin-111 is in a position of hypothetically modulating various aspects of somitogenesis, from survival of PSM cells to clefting and budding of epithelial somites, similarly to the requirement for fibronectin in the formation of chicken somites (Rifes et al. 2007). The putative function of laminin α1 in the murine presomitic mesoderm could be addressed by soaking cultured mouse embryos in the presence of antibodies that prevent the interaction of laminin a1 with cell surface receptors, or by Cre-based conditional deletion of the *Lama1* gene within the cells of the presomitic mesoderm (Marinic et al. 2013).

However, the avian presomitic mesoderm was largely devoid of *cLama1* mRNA (Figure 3.8A, B, F), in striking contrast to the mouse and zebrafish embryos. This suggests that *cLama1* expression in the presomitic mesoderm was lost in the avian lineage at some point after the great sauropsid/synapsid dichotomy nearly 310 million years ago (Benton 2005; van Tuinen and Hadly 2004). An alternative, but less parsimonious scenario is that *Lama1* expression in the presomitic mesoderm was independently invented by teleosts and

mammals. Nevertheless, I observed *cLama1* in the anterior-most end of the chicken PSM where the newly formed somites bud off from (green arrowhead in Figure 3.8B). This pattern is very intriguing as it suggests some unique function for laminin α1 at this very rostral end of the PSM, as hypothesised above for the mouse. Strikingly, despite the lack of *cLama1* mRNA expression through most of the avian PSM, a recent study provided detailed 3-dimentional view of the laminin and fibronectin protein matrices within and around the avian presomitic mesoderm (Rifes and Thorsteinsdottir 2012). Interestingly, both matrices display progressive increase in structural organization in a caudal-to-rostral direction, where laminin in particular is presented by small patches within the less mature posterior PSM while in the anterior half of the PSM, where somitic epithelialization and budding occur, laminin is deposited as a fenestrated meshwork on the surface of the mesodermal cell (Rifes and Thorsteinsdottir 2012), again hinting to functions in PSM morphogenesis.

But if the chicken PSM does not transcribe *cLama1* then which cells provide laminin to the PSM? One possibility is that the neural tube provides laminin-111 for *cLama1* is expressed there. Alternatively, the absence of *cLama1* expression in the avian presomitic mesoderm may be compensated by the expression of another alpha laminin subunit, most likely laminin α5, which appears to be present in the chicken presomitic mesoderm (Coles et al. 2005) but not in the murine PSM (Anderson et al. 2009). Thus, the putative function of laminin in the development of the PSM might be performed by different isoforms in the mouse and chicken embryos.

Moreover, another unreported aspect of avian *cLama1* expression pattern is the domain at the very caudal edge of the unsegmented cranial paraxial mesoderm (black arrow in Figure 3.7A, H). It could be speculated that laminin α1 plays a role in the initial separation of the cranial from presomitic paraxial mesoderm. These scenarios can be addressed either by interfering with the stability of *cLama1* mRNA or by blocking antibodies against laminin α1, as mentioned previously.

3.3.4. *Lama1* expression in the nephrogenic mesoderm

Another conserved domain of *Lama1* expression is the nephrogenic mesoderm. The results from my studies on *Lama1* mRNA expression in the mouse embryo (Figure 3.1E; Figure 3.2H) are consistent with previous reports on *Lama1* expression in the developing nephric systems (Anderson et al. 2009; Miner et al. 1997; Miner et al. 2004; Sorokin et al. 1997), as

described in Chapter 1. Interestingly, I found that in the avian embryo *cLama1* mRNAs are present only in the nephrogenic mesenchyme but not in the nephric duct (Figure 3.8D, D'; Figure 3.9D, D'), at least at the examined stages. Perhaps, the minor difference in the expression pattern of the chicken and murine *Lama1* has species-specific functional consequences for nephric development. However, Ekblom et al. (1990) reported that, nevertheless, in the mouse embryo, laminin α1 disappears from the ureters and collecting ducts (which are derivatives of the nephric duct) by day 16 of gestation and remains only in the proximal tubules of the nephrons (derived from the nephrogenic mesenchyme) where it also down-regulated eventually (Ekblom et al. 1990). Despite these minor differences, the transient expression of *Lama1* in the developing kidneys suggests important functions of laminin α1 in kidney morphogenesis, as described in Chapter 1.

3.3.5. Other domains of *Lama1* expression

Intriguingly, I observed *Lama1* mRNA expression in the floor of the foregut of HH8 chicken embryos (*black arrowheads* Figure 3.7B) as well as in the endoderm of the 1st pharyngeal pouch and pharynx of mouse embryos at E10.25 (Figure 3.2B, F). It is possible that laminin α1, in the form of laminin-111, could perform a role in the complex processes of epithelial sheet folding during pharyngeal pouch and thyroid diverticulum evagination. In this respect, the observation that blocking the interaction between laminin γ1 and nidogen leads to a severe reduction in branching morphogenesis of the submandibular salivary gland in E13 mouse embryos (Kadoya et al. 1997), suggests that the presence of assembled Laminin network is required for branching/folding of epithelia. A study in *Caenorhabditis elegans* uncovered the involvement of laminin in the apical localization of the PAR-3 protein in pharyngeal precursors that is required for establishment of the apical-basal polarity of the nematode pharynx epithelium, independent from laminin's incorporation in basement membranes (Rasmussen et al. 2012). Thus, expression of *Lama1* in the pharyngeal endoderm of mouse and chicken embryos hints for hitherto unknown functions of laminins in this region of the embryo, which can be addressed by the application of blocking antibodies to laminin α1 and/or its receptors in cultured embryos or by RNAi-mediated down-regulation of *Lama1* translation.

Another interesting domain of *cLama1* mRNA expression is in the ingressing mesoderm in HH4 chick embryos (Figure 3.6), which is consistent with the observations

from Zagris et al. (2000). However, the functional significance of this expression by the newly formed mesoderm is unclear, especially in consideration of the requirement for basement membrane breakdown at the site of the primitive groove, where epiblast cells undergo epithelial-to-mesenchymal transition (EMT) and turn into mesoderm (Nakaya et al. 2008). In relation to this, the EMT marker *Snail2* was up-regulated and mesodermal differentiation accelerated in *Lamc1*[-/-] embryoid bodies which also, expectedly, lacked basement membrane (Fujiwara et al. 2007), again demonstrating the negative relationship between Laminins and mesodermal ingression in gastrulation.

Although I did not examine *cLama1* expression in pre-gastrulation chicken embryos (stages 1-2), it was reported to be expressesed by the epiblast and more intensively by the hypoblast (Zagris and Chung 1990; Zagris et al. 2000) at that stage, and culturing chicken embryos in solution containing laminin antibodies perturbed the adhesion and migration directionality of epiblast cells during formation of the primitive streak, which eventually culminated in disintegration of the whole area pelucida (Zagris and Chung 1990), thus revealing important roles for laminin-111 in early avian embryogenesis. This is reminiscent to the essential functions of laminin α1 in murine epiblast differentiation and extraembryonic membrane stability (Miner et al. 2004).

3.3.6. *Lamb1* and *Lamc1* mRNA expression is unaffected by the absence of *Shh*

The absence of *Shh* did not have an effect on the expression of *Lamb1* and *Lamc1* at E9.5. This further consolidates the model in our laboratory that the failure of myotomal basement membrane (MBM) assembly is a result of perturbed *Lama1*, but not other Laminin genes' transcription (Anderson et al. 2009). However, Anderson et al. (2009) detected patches of laminin β1 protein in the somites of *Shh*[-/-] embryos, which were not organized in a continuous sheet as in wild type animals. It is not implausible that laminin β1 in these patches could be part of residual laminin-111 produced before somitogenesis, in the presomitic mesoderm, that failed to polymerise due to perturbation of *Lama1* transcription in the somites of *Shh*-null embryos, illustrating that dosage of laminin α1, and hence – laminin-111, is important for effective assembly of the myotomal basement membrane (Anderson et al. 2009). In fact, synthesis of the alpha subunits and interaction with the beta-gamma dimer is considered essential for secretion of the whole laminin heterotrimer from laminin-producing cells (Yurchenco et al. 1997). However, a more likely explanation is that the laminin β1[+] patches

of unassembled laminin are formed by laminin-511 instead, as *Lama5* expression was unaffected in the somites of *Shh*-defficient embryos (Anderson et al. 2009), and therefore could have facilitated the secretion of a laminin β1-containing trimer.

In summary, I confirmed our previous finding that SHH is required for *Lama1* expression in the somites and neural tube of E9.5 mouse embryos. I reported a differential requirement for SHH signaling in the somites and neural tube for the initiation and maintenance of *Lama1* expression, respectively. Finally, I investigated the expression pattern of the chicken *Lama1* gene and concluded that its expression is largely conserved with that of its mouse and zebrafish orthologs, hinting to the conservation of signaling and transcription-regulatory mechanisms. This has led me to hypothesise that *Lama1* expression in the murine somites and neural tube is directly regulated by SHH via binding of GLI transcription factors to *cis*-regulatory elements in the vicinity of the *Lama1* locus. Moreover, based on the conserved expression of *Lama1* across vertebrates, particularly in the neural tube and somites, it is highly plausible that the SHH-responsive regulatory elements are conserved, which may aid their identification via *in silico* approaches, as described in the next chapter.

Chapter 4

in silico identification of conserved non-coding elements in the vicinity of the *Lama1* locus in mice

4.1. Hypothesis and Aims

Previous research in our laboratory implicated the SHH signalling pathway and GLI2/GLI3 transcription factors in the regulation of *Lama1* expression in the neural tube and somites of mouse embryos (Anderson et al. 2009). Based on this, I hypothesised that SHH exerts a direct control on *Lama1* transcription, mediated by binding of the GLI transcription factors to regulatory elements in the vicinity of the *Lama1* locus. Here, I attempted to identify putative GLI-binding-sites-containing regulatory elements in the mouse *Lama1* locus using the method of phylogenetic footprinting.

4.2. Introduction

4.2.1. Phylogenetic footprinting and regulation of developmental gene expression

An emerging picture from genome-wide studies is that most conserved-non coding elements (CNEs) predominantly map to regions flanking genes implicated in embryonic development (Dermitzakis et al. 2002; Bejerano et al. 2004; Woolfe et al. 2005; Pennacchio et al. 2006; Elgar 2009). This hints for a role of CNEs as modulators of gene transcription. For instance, such a role was suggested by Bejerano et al. (2004), who uncovered 481 ultraconserved elements (defined as displaying 100% sequence identity over a ≥ 200 bp window) in phylogenetic footprinting between human, rat and mouse whole genome sequences. Notably, 111 of these elements mapped to exonic portions of known human genes predominantly involved in RNA binding and splicing, but the majority were located in the introns or intergenic territories around genes with known roles in transcriptional control and development (Bejerano et al. 2004). Indeed, numerous *in vivo* studies at different scales and scopes have demonstrated that a substantial fraction of CNEs have gene regulatory potential (Dickmeis et al. 2004; Poulin et al. 2005), as revealed for the murine endothelial and neural enhancers of the *Scl* gene (Gottgens et al. 2000) for instance, illustrating the power of phylogenetic footprinting to detect putative TREs, as detailed below.

Despite the low level of sequence conservation of vertebrate CNEs with those identified in flies or nematodes, suggesting that these elements are perhaps phylum-specific and related to the establishment of the *bauplan* (Vavouri and Lehner 2009; Elgar 2009; Woolfe et al. 2005), a recent study by Clarke et al. (2012) provided the first example of functional pan-bilaterian CNEs that span the deuterostome-protostome dichotomy. Two such elements with enhancer activity were detected in alignments between human, zebrafish, sea

urchin, owl limpet, sea hare and tick genomic sequences and were named Bicores, for *Bi*laterian *Co*nserved *R*egulatory *E*lement*s*. Bicore 1 and 2 were mapped upstream of the orthologs of the *Id2* and *Zn503* genes, respectively. In all of the above metazoans, Bicore 1 and 2 displayed 60-65% sequence identity over ~100 bp in the vertebrate/protostome comparisons. Remarkably, despite the vast phylogenetic distance, the Bicores from different species drove similar spatio-temporal patterns of GFP reporter expression in the CNS of transgenic zebrafish, corresponding to the domains of endogenous *Id2* and *Zn503* gene activity (Clarke et al. 2012).

Importantly, mammal/teleost comparisons proved to be highly informative and effective in detecting functional CNEs as evident from a genome-wide comparison between human and pufferfish, which recovered 3,100 CNEs with at least 70% identity (Pennacchio et al. 2006). 167 of the top-ranking human fragments were further examined in a high-throughput mouse transgenesis approach, which revealed that 75 behaved as transcriptional enhancers driving *lacZ* reporter gene expression predominantly in the central and peripheral nervous system (Pennacchio et al. 2006). Curiously, more than 50% of the candidate regions did not exhibit enhancer activity which could be due to species-specific differences in the transcription-factor milieu, to the fact that some of these CNEs are not enhancers, and/or to technical limitations of the study, such as terminating reporter expression screening prior to the stage of enhancers' activity.

Similarly, another study with a narrower phylogenetic scope demonstrated the power of phylogenetic footprinting to uncover CNEs with pan-vertebrate conserved function. Navratilova et al. (2009) performed human/zebrafish alignments and catalogued several highly-conserved non-coding elements (HCNEs) with at least 70% identity over $\geq$ 100bp sequence window in the *SOX3* and *PAX6* loci of the human genome. Notably, upon reporter construct injection in zebrafish embryos, most of the HCNEs (80%) generated reproducible patterns of GFP fluorescence in tissues that correspond to domains of *sox3* and *pax6* mRNA expression. Furthermore, some of the human and fish HCNEs showed similar functional properties, i.e. the expression driven by the orthologous elements was remarkably congruent and consistent with previous functional analyses of some of the human regions around *PAX6* in transgenic mice (Navratilova et al. 2009; Kleinjan et al. 2001; Kleinjan et al. 2006). This

demonstrates the potency of zebrafish as a useful system to test the activity of mammalian non-coding elements.

In some exceptional, gene-centric investigations, phylogenetic footprinting has succeeded in the identification of the almost complete set of tissue-specific enhancers of a gene, as is the case of the murine *Six1* gene. There, Sato et al. (2012) compared the mouse genome with those of the opossum, chicken, *Xenopus* and teleosts and detected 15 CNEs flanking the *Six1* locus. The elements ranged in size from 0.1 to 0.7 kb in length and showed at least 50% identity over 100bp among mouse/chicken. Notably, subsequent *in vivo* analyses in chicken and mouse embryos revealed that 7/15 of the sequences exhibit distinct enhancer activities in the cranial placodes, dorsal root ganglia, somites, notochord and cranial mesoderm, which in sum covered the major domains of *Six1* expression (Sato 2012).

Nevertheless, although CNEs are frequently associated with enhancer properties *in vivo*, CNEs may also belong to other types of transcription-regulatory elements such as silencers, insulators (Royo et al. 2011), locus control regions, miRNA binding sites, or even non-coding RNAs (Birney et al. 2007).

Considering the efficiency of phylogenetic footprinting combined with the availability of web-based automated programmes allowing relatively rapid customised multispecies genome comparisons, I decided to use this approach for the detection of conserved non-coding elements in the murine *Lama1* locus, some of which I predicted would harbour Gli-binding sites if Shh directly controls *Lama1* (Anderson et al. 2009).

4.3. Results

4.3.1. *in silico* identification of CNEs in the *Lama1* locus region

There are few web-based bioinformatics tools which perform multi-species genome alignments combined with graphical display of sequence conservation, namely the UCSC Genome Browser (Kent et al. 2002), the VISTA Genome Browser (Frazer et al. 2004) and the ECR Browser (Ovcharenko et al. 2004). I chose to use the ECR (for *E*volutionary *C*onserved *R*egions) Browser over the other alternative tools for the following advantages (see section 2.3.3 for detailed description of the ECR Browser):

1. It affords dynamic and interactive visualization of conservation profiles across multiple species.
2. It allows the inclusion of phylogenetically critical species, such as the opossum *Monodelphis domestica*, which are not available in alternative programmes.
3. It automatically highlights regions meeting the stringency criteria, which enables faster screening of the alignments.
4. It provides a direct link to the rVISTA tool (Loots and Ovcharenko 2004) for rapid examination of TFBSs present in the CNE sequences.

The ECR Browser also allows the adjustment of conservation parameters according to the user's needs. In this study, two such parameters, minimum window length (in base pairs) and minimum sequence identity (in percentage, %) were customized at 100 bp and 60%, respectively. Such level of stringency, $\geq$ 60% identity over a 100 bp window, was chosen for it has been previously shown in comparisons between chicken and mammalian genomes that these stringency settings are effective in the identification of CNEs with important regulatory functions (Uchikawa et al. 2003, Coy et al. 2011), and they decrease the frequency of false positives, providing a manageable number of CNEs for subsequent experimental studies. The 100 bp window was chosen for, although the average size of transcriptional enhancers is 500 bp (Loot 2008), these elements can range in size from 100 bp (Krebsbach et al. 1996; Loot 2008) to few kilobases (Danielian et al. 1997), whereas the 60% stringency cut off would enable the detection of candidate enhancers that might be more weakly conserved than the reported > 95% and > 70% sequence identity of putative TREs derived from human/rodent and human/teleost sequence comparisons, respectively (Bejerano et al. 2004; Sandelin et al. 2004; Woolfe et al. 2005).

Based on this approach and parameter settings, I aligned the mouse genomic region harboring the *Lama1* locus with the corresponding genomic regions from several other vertebrate species: human, cow, opossum, chicken, *Xenopus*, zebrafish and fugu (Figure 4.1). The mouse/human and mouse/cow alignments generated too high number of peaks, due to close evolutionary relationship, most of which are perhaps functionally irrelevant. In contrast, the mouse/chicken, mouse/Xenopus and mouse/teleosts comparisons produced either too few or no conserved elements, suggesting too stringent conservation parametres.
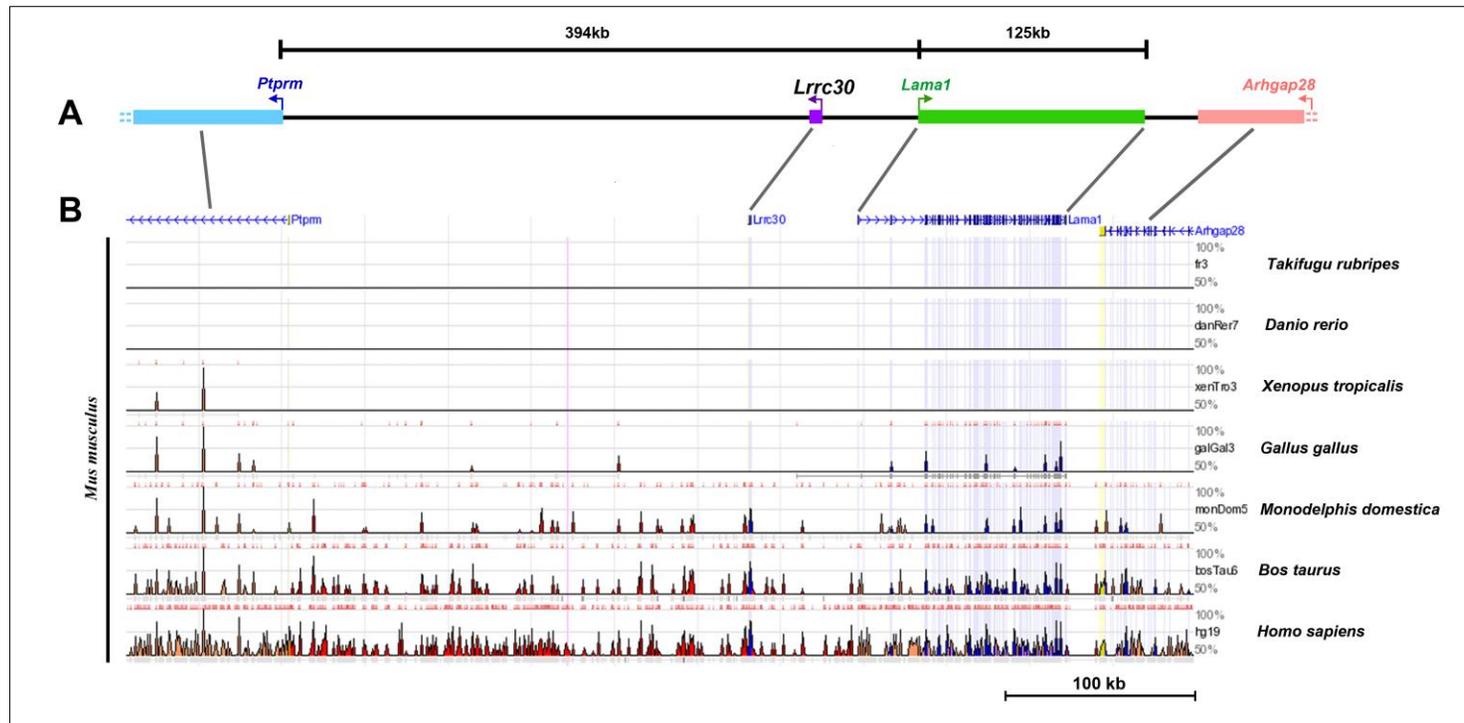
Figure 4.1. *Phylogenetic footprinting analysis of the Lama1 locus on mouse chromosome 17.* (A) Schematic representation of the murine *Lama1* locus (shown in *green*) and its nearest neighbouring genes on chromosome 17: *Arhgap28* (salmon), *Lrrc30* (purple) and *Ptprm* (blue). The centromere of chromosome 17 is further down on the left, while the scale bar above indicates the length of the mouse *Lama1* gene (125kb) in kilobases and the distance (394kb) between *Lama1* and the *Ptprm* gene. Curved arrows represent the transcriptional start site of the gene and the direction of transcription. Straight grey lines connect genes from the schematic representation in (A) to their corresponding positions in the ECR Browser output in (B). (B) Conservation plot of the genomic region shown in (A), generated by the ECR Genome Browser. Each row of the plot represents the alignment of the mouse sequence with the genome of a comparator species with decreasing phylogenetic relatedness moving from bottom (mouse/human alignment) to top (mouse/fugu alignment). Conserved genomic regions meeting or exceeding the stringency criteria (≥60% identity over at least 100bp) appear as colour-coded peaks: blue (exons), yellow (UTRs), salmon (introns), green (repeats) and red (intergenic non-coding sequence).Note the decrease in peaks number or absence thereof in the alignments with *Xenopus* and teleots fish, respectively. A scale bar of 100 kb is shown below the graph.

92

Therefore, I repeated the same ECR Browser alignment but decreased the stringency down to 30% identity over a 30 bp window. However, even under these relaxed conditions, the ECR Browser failed again to display any conserved sequence in the mouse/Xenopus and mouse/teleost alignments, not even in exonic regions, contrary to expectations as blastn comparisons of *Lama1* cDNA sequences (Ensembl) from mouse and zebrafish revealed 72% identity (E-value = 0.0). I suspected that the local alignment method applied by the ECR Browser is not optimal for the detection of conserved elements between highly divergent sequences that have undergone rearrangements and inversions (Brudno et al. 2003), and/or there were issues with the quality of the compared genomic sequences as the ECR Browser authomatically selects the aligned sequences. To circumvent this problem, I used the Shuffle-LAGAN (a hybrid global-local, or glocal alignment) algorithm at mVISTA, and performed customised comparisons of the genomic region containing the *Ptprm*, *Lrrc30*, *Lama1* and *Arhgap28* loci from mouse, human, opossum, chicken and zebrafish (see Chapter 2 for details of the analysis). This approach succeeded in the detection of conserved peaks in all of the alignments, including in the mouse/zebrafish comparison (Figure 4.2). In the latter alignment, most of the peaks correspond to conserved exons from the *Ptprm*, *Lrrc30*, *Lama1* and *Arhgap28* genes, while two of the identified peaks correspond to unique sequences in the non-coding regions, and were named here as CNEa and b (Figure 4.2; Table 4.1). Notably, CNEb is located within the 3rd intron of the *Lama1* gene.

The most informative alignment in the ECR Browser, in terms of the number of CNEs detected, was the mouse/opossum one. Therefore, I focused my studies on that comparison. Based on it, I identified 24 CNEs in the mouse sequence, numbered from 1 to 24 (Figure 4.3; Figure 4.4). CNE1 is positioned most proximally to the *Lama1* transcription start site and CNE24 is the most distal one. Interestingly, no CNE was identified within the intronic sequences of *Lama1* or in the intergenic space downstream of the *Lama1* 3' UTR. Instead, all CNEs are interspersed in a 361 kb interval upstream of the *Lama1* transcription start site. CNEs 1-3 are located in the intergenic space between *Lama1* and the *Lrrc30* gene, CNEs 4-22 are situated in the region between *Lrrc30* and the *Ptprm* gene while, notably, CNEs 23 and 24 are found in the first intron of *Ptprm* (Figure 4.3). The search for additional CNEs was not extended further downstream of the *Ptprm1* intron 1 for enough number of CNEs were already recovered for subsequent analyses.
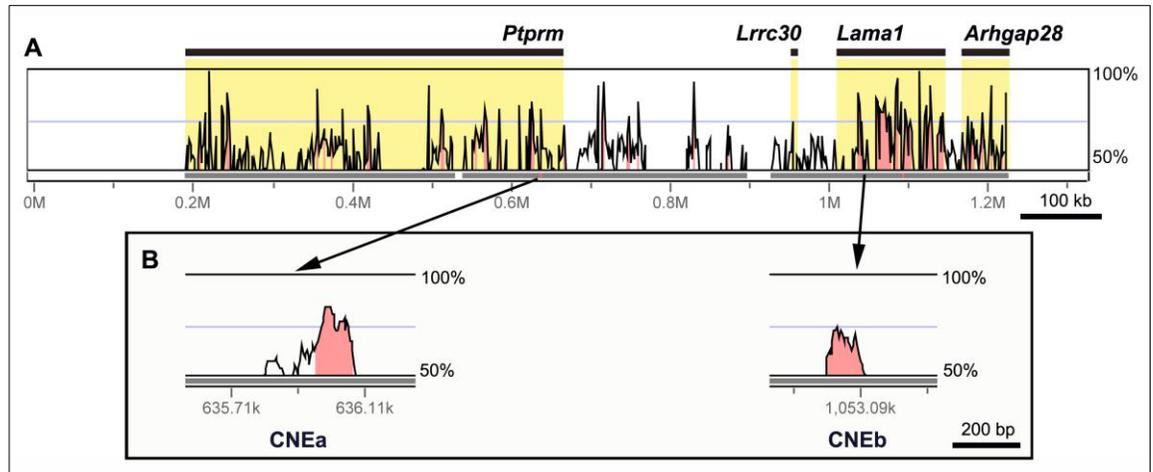
Figure 4.2. *Identification of CNEs in the vicinity of the Lama1 locus in a mouse/zebrafish alignment.* Comparison of the *Lama1*-containing genomic region, spanning from the 3'-end of *Ptprm* locus to the 5'-end of *Arhgap28*, from mouse and zebrafish, as generated in mVISTA. A) Full view of the alignment. Elements characterized with sequence identity ≥ 30% over a 30 bp window are displayed in *pink* colour. Most of the colored peaks correspond to exonic sequences. *Dark* bars above and *yellow* shadowing within the alignment plot indicate the extent of each gene locus. B) Detailed views of CNEa and b. CNEa corresponds to the 5'-end of CNE24, located in the 1st intron of *Ptprm*, while CNEb is located in the 3rd intron of the *Lama1* locus (see Table 4.1).

The pattern of distribution of some of the CNEs necessitates more detailed description. Accordingly, CNE1 is the nearest CNE to *Lama1* locus, situated 33.5 kb upstream of the *Lama1* transcription start site, or approximately 30 kb upstream of the known transcription-regulatory elements of *Lama1* (Niimi et al. 2003, 2004; Piccinni et al. 2004). CNE2, 3 and 4 are located close (within 10 kb) to the *Lrrc30* gene with CNE3 positioned only 12 bp 5' to the *Lrrc30* transcription start site, while CNE2 and 4 flank the *Lrrc30* gene (Figure 4.3). The close proximity of CNE3 to this poorly explored gene combined with results from the current study (see Chapter 6) raises interesting questions regarding CNE3's function.

Further upstream of the *Lama1* locus, CNEs 11-16 form a loose cluster where each CNE is separated from its nearest neighbour by less than 10 kb, while CNE22 is located about 6 kb 5' of the transcription start site of the *Ptprm* gene. Finally, as mentioned earlier, CNEs 23 and 24 are positioned within the first intron of *Ptprm* (Figure 4.3).

In terms of size (length of base pairs), the identified CNEs range from 137 bp to 953 bp which is consistent with the size reported in the literature (Elgar 2009; Pennacchio et al. 2006; Woolfe et al. 2005). It is important to indicate that in cases where two or more

94

promising peaks on the ECR Browser plot were separated by less than 150 bp of non-conserved sequence, the peaks together with the intervening regions were considered as part of one CNE, as is the case for CNEs 4, 5 and 13 (Figure 4.3; Figure 4.4). Based on the fact that transcription-regulatory elements may contain regions of high and low sequence conservation (Clarke et al. 2012), selecting only for the most highly-conserved regions could result in spurious activity during the functional assays, which would not reflect the genuine function of the entire regulatory element.

The ECR Browser-identified CNEs differ at the extent of sequence conservation, with CNE4 being the least conserved (62.1%) while CNEs 20 and 24 are the most conserved with 78.4% and 81.4% identity, respectively (Figure 4.4; Table 4.1). These percentages of sequence identity are the result of the mouse/opossum genome sequence alignment. Notably, however, CNEs 9, 14, 19, 23 and 24 do also appear conserved in the mouse/chicken comparison, indicating that the extent of sequence identity of these particular CNEs, shared between the mammalian and avian species, meets the established stringency criteria. In contrast, the rest of the CNEs are not identified in comparisons of the mouse with non-mammalian species sequences, and their description is based on the mouse/opossum alignment as stated earlier.

Figure 4.3. *Distribution of conserved non-coding elements (CNEs) in the vicinity of the mouse Lama1 gene based on the analysis from the ECR Browser.* (A) A diagram of the genomic region containing the CNEs identified in this study. To ease subsequent visualization in panel B, the whole region is divided in seven viewing windows shown as light-grey boxes and numbered from I to VII with Roman numerals. (B) A detailed view of the comparison between the mouse sequence and the genome of the opossum *M. domestica* and the chicken *G. gallus*, generated in the ECR Browser. On the left are shown the viewing windows (I – VII) with each window corresponding to a section of the mouse genomic region in (A) that is aligned with the opossum (bottom row) and chicken (upper row) (see Legend at the bottom). The percentages on the right reflect sequence conservation in the alignments. All CNEs identified in this study are highlighted in *cyan* (mouse/opossum) or *light-green* (mouse/chicken) in the alignment plots. Their position in the whole genomic region is indicated by *gray lines* that map the highlighted peaks from the alignments onto the diagrams of the genomic region, placed between windows I-II, III-IV and V-VI. The CNEs are numbered with Arabic numerals, from 1 to 24, depicted next to the connecting *grey lines*. At the bottom left is provided a scale bar of 10 kb to ease distance estimation in the ECR Browser plots. Below is shown a Legend that explains the meaning of the color code used in the Figure (see page 97).
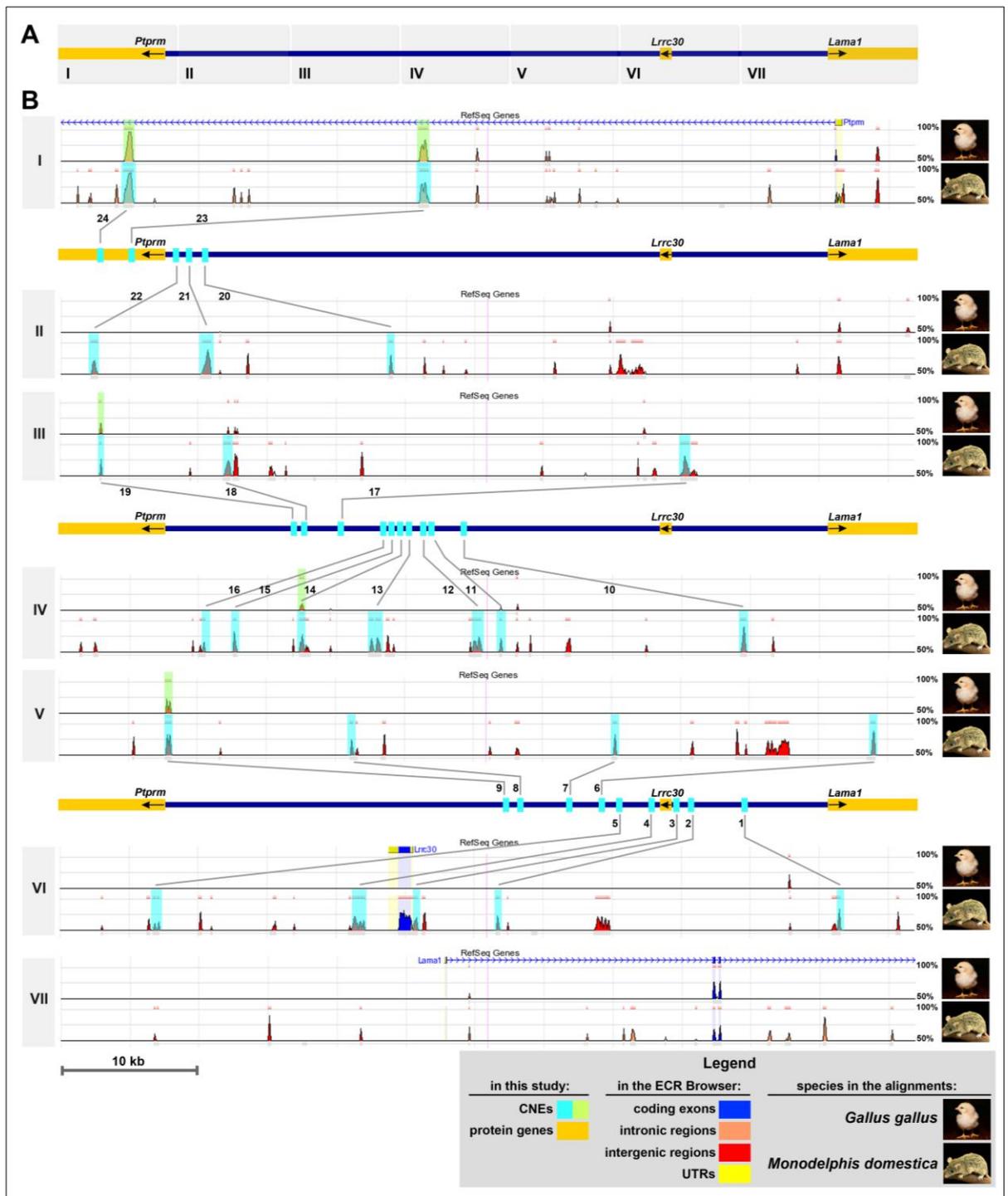
Figure 4.3. *Distribution of conserved non-coding elements (CNEs) in the vicinity of the mouse Lama1 gene based on the analysis from the ECR Browser*. Description is provided on the previous page (p. 95).

CNE	genomic coordinates on chromosome 17[1]	length (in bp)	% sequence identity[2]
a	67,302,811-67,302,924	114	80.0
b	67,719,841-67,719,952	112	68.8
1	67,663,410-67,663,687	278	71.9
2	67,638,755-67,638,982	228	67.1
3	67,632,731-67,633,050	320	68.8
4	67,628,306-67,629,259	953	62.1
5	67,613,886-67,614,330	444	64
6	67,603,803-67,604,119	317	77.6
7	67,585,118-67,585,462	345	68.5
8	67,566,070-67,566,263	194	65.5
9	67,552,672-67,553,105	434	72.4
10	67,532,119-67,532,491	373	73.2
11	67,514,616-67,514,817	202	65.8
12	67,512,661-67,513,331	671	64.7
13	67,505,209-67,506,030	822	64.8
14	67,500,119-67,500,500	382	67.3
15	67,495,323-67,495,591	269	70.6
16	67,493,075-67,493,309	235	63.4
17	67,466,906-67,467,545	640	66.6
18	67,433,815-67,434,304	490	67.6
19	67,424,809-67,424,945	137	75.9
20	67,371,939-67,372,109	171	78.4
21	67,368,753-67,369,379	627	72.1
22	67,360,683-67,361,162	480	65.6
23	67,323,788-67,324,481	694	70.2
24	67,302,361-67,303,096	736	81.4

Table 4.1. Features of the identified CNEs .[1]The coordinates are according to the GRCm38 assembly of the mouse genome (mm10 in the UCSC Browser). [2]The percentages reflect the conservation of the mouse elements with the corresponding opossum sequences (for CNEs 1-24) or zebrafish sequences (for CNEs a and b).

4.3.2. Some CNEs contain binding sequence motifs for GLI and ZIC factors

Each of the CNEs is characterised by a unique nucleotide sequence which contains potential motifs (sites) for interaction with DNA-binding proteins. The group of DNA-binding proteins of particular importance for this study is the GLI family of transcription factors. According to my hypothesis for a direct role of SHH signalling in the regulation of *Lama1* transcription, I predicted the existence of GLI-binding sites within the *Lama1* locus and/or its vicinity, and

specifically within CNEs, as CNEs have been reported to be a preferential hub for TFBS with regulatory functions (Hemberg et al. 2012; Sato et al. 2012).
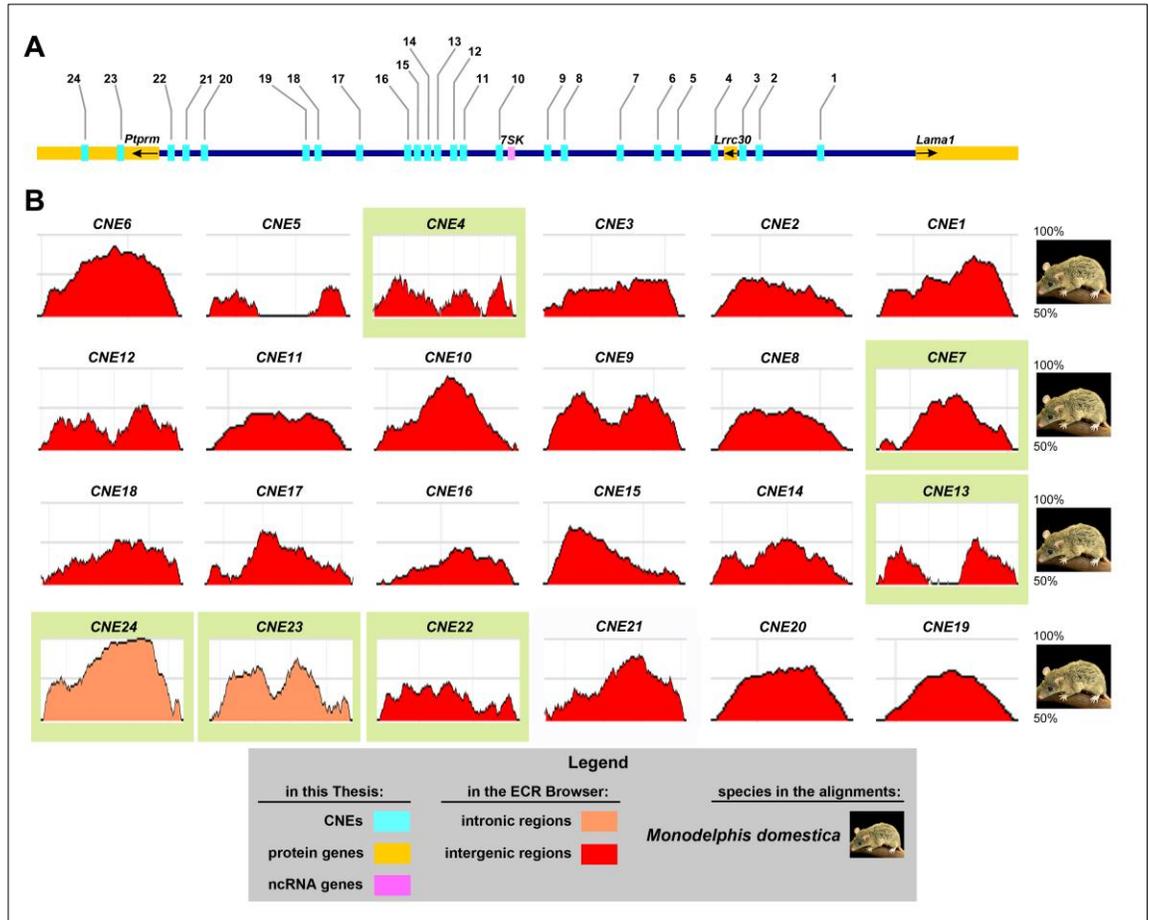
Figure 4.4. *Conservation profile of individual CNEs (1-24) based on the mouse/opossum genome comparison from the ECR Browser.* (A) A map of the location of the 24 CNEs in relation to the mouse *Lama1* locus and neighbouring loci (*black arrows* show direction of transcription). (B) Detailed conservation plots for each CNE as generated by the ECR Browser for alignments of the mouse and opossum genomes. The percentages on the right reflect sequence conservation in the alignment. CNEs highlighted in light-green (4, 7, 13, 22, 23 and 24) contain one or more GLI/ZIC binding motifs, as discussed in the text. Below is shown a Legend that explains the meaning of the color code used in the Figure. Note that the lengths of the CNEs are not to scale; for detailed information on CNEs' lengths see Table 4.1.

To test this possibility, I submitted the pre-computed mouse/zebrafish and mouse/opossum alignments of each CNE for processing in the rVISTA2.0 tool (Loots and Ovcharenko 2004; see section 2.6.4 for details). This approach predicted the presence of GLI binding motifs in CNEs 13, 23 and 24 (see Table 4.2 and Figure 4.4). Notably, some of the GLI binding motifs in CNEs 13 and 23 overlapped with motifs for the ZIC transcription factors. This is not unexpected as previous studies have revealed that the minimal consensus

binding sequence of the ZIC transcription factors (5'-GGGTGGTC-3') is nearly identical to the consensus binding sequence of the GLI factors (5'-TGGGTGGTC-3') (Kinzler and Vogelstein 1990; Mizugishi et al. 2001). Moreover, gel shift assays demonstrated the ability of ZIC factors to bind to the consensus Gli-binding sequence albeit with lower affinity than the GLI factors (Aruga et al. 1994; Mizugishi et al. 2001). These highly similar binding properties could be explained by the fact that ZIC transcription factors contain a Zn-finger DNA binding domain with five Cys_2His_2 finger motifs, which is highly-homologous to the Zn-finger DNA-binding domain of the GLI factors (Aruga et al. 1996). Interestingly, the rVISTA tool also predicted that CNEs 4, 7 and 22 contain a ZIC motif but not a GLI one (Table 4.2).

CNE	GLI/ZIC motif	position in CNE	Expected ± SD	Over representation	Z-score	Gli/Zic motif presence in opossum	Gli/Zic motif conservation
4	ZIC	890-904 (+)	0.37 ± 0.61	2.70	0.21	no	no
7	ZIC	153-167 (+)	0.09 ± 0.31	10.53	1.31	no	no
13	GLI GLI GLI	68-82 (-) 295-309 (-) 759-773 (-)	0.98 ± 0.99	3.07	1.54	GLI (127-138) (+)	no
22	ZIC	412-426 (+)	0.18 ± 0.42	5.64	0.77	ZIC (324-332) (-)	yes
23	GLI GLI	146-160 (-) 273-287 (-)	0.71 ± 0.84	2.84	0.95	ZIC (545-553) (-)	no
24	GLI GLI	71-85 (-) 345-359 (+)	0.75 ± 0.86	2.67	0.87	no	no

Table 4.2. *Presence and conservation of GLI/ZIC motifs in some of the CNEs.* The current table lists the prediction of GLI/ZIC binding motifs within some of the CNEs, together with relevant statistics information, as obtained using the "Over-represented TFBSs" tool from Genomatix and rVISTA. The "Position in CNE" column shows the location of the motif in the mouse CNE sequence and the DNA strand, "+" (sense) and "−" (antisense). The "Expected ± SD" column displays the expected number of a given motif matches in an equally sized random genomic region, together with the standard deviation (SD). The "Over-representation" column indicates the fold factor of match numbers in the analysed sequence compared to an equally sized genomic region from the background, or found versus expected matches number. The "Z-score" is a measure of the statistical significance of the over-representation ratio; a Z-score above 2 or below -2 is considered statistically significant, corresponding to a p-value of ~ 0.05 (Genomatix; Sue et al. 2005). The "GLI/ZIC motif presense in opossum" column shows the presence and location of GLI/ZIC motifs within the opossum CNE sequence. The last column indicates conservation of GLI/ZIC motifs between mouse and opossum. Note that the ZIC motif in CNE22 is conserved.

However, none of the predicted binding motifs for GLI and ZIC factors are significantly over-represented in the CNEs relative to the genomic background (Table 4.2). Also, most of these GLI/ZIC motifs are not conserved in the mouse/opossum alignment, i.e. they do not meet the stringency criteria of the rVISTA algorithm described earlier (both in terms of position within the CNE and motif's sequence), and are identified only in the base

genome – that of the mouse (Table 4.2). An important exception is the ZIC binding site in CNE22, which shows both conserved motif sequence and position. Interestingly, despite the lack of conservation, the orthologous sequences of CNE13 and CNE23 in the opossum genome do harbor a GLI or a ZIC motif, respectively, but in a shuffled position within the CNE (see Table 4.2).

4.4. Discussion

Here, using phylogenetic footprinting, I reported on the identification of evolutionary conserved non-coding elements (CNEs), which may have putative regulatory functions, distributed in the genomic region of the mouse *Lama1* locus. Notably, some of these elements harbour GLI and/or ZIC binding motifs, which makes them candidates for mediating the control role of *Lama1* transcription by SHH signalling (Anderson et al. 2009).

4.4.1. Few of the identified CNEs are conserved with zebrafish sequences

Analysis of the *Lama1*-containing genomic region using the ECR Browser failed to identify conserved elements (even exons) in the mouse/zebrafish and mouse/fugu comparisons, even at low stringency settings (minimum 30% sequence identity over a minimum of 30 bp window) (Figure 4.1). Perhaps, this is due to the local alignment algorithm used by the ECR Browser, which is not suitable for the detection of conservation between highly divergent sequences (Brudno et al. 2003). In contrast, an alternative mouse/zebrafish glocal alignment using mVISTA under relaxed stringency, succeeded in detection of the conserved exons of *Ptprm*, *Lrrc30*, *Arhgap28* and *Lama1* genes, as well as in the identification of two CNEs, a and b, located in the 1st intron of the *Ptprm* gene and in the 3rd intron of the *Lama1* gene, respectively (Figure 4.2; Table 4.1). Thus, CNE a and b are candidates for regulatory elements of *Lama1*.

The scarcity of CNEs between mouse and zebrafish in the *Lama1*-containing genomic region is intriguing, when one considers the highly similar pattern of *Lama1* mRNA expression in the embryos of these species (Anderson et al. 2009; Zinkevich et al. 2006; Joseph Pickering's Thesis 2012) (Figures 1.3 and 1.4; Table 1.2; Figure 3.1). However, conservation of function does not necessarily correlate with or require conservation of sequence, as discussed later.

4.4.2. Some CNEs may represent long-range enhancers of *Lama1*

All of the CNEs identified in this study, except for CNEb, are located 5' from the transcription start site of the *Lama1* gene. The most proximal CNE to *Lama1* – CNE1, is ~ 33.5 kb upstream from the previously described basal promoter of the gene (Niimi et al. 2003, 2004; Piccinni et al. 2004), while the most distal element – CNEa (overlapping with the 5'- end of CNE24), is nearly 395 kb upstream of *Lama1*. Such distances from the *Lama1* locus do not negate the possibility that the CNEs may function as transcription-regulatory elements of *Lama1* for numerous studies have provided examples of regulatory elements, particularly enhancers, positioned tens or hundreds of kilobases away from their target genes (Gottgens et al. 2000; Uchikawa et al. 2003; Kundu et al. 2013). In an extreme case, the enhancer responsible for ZPA-specific expression of the *Shh* gene is situated nearly 1 Mb away from the *Shh* locus in intron 5 of the *Lmbr1* gene (Lettice et al. 2002; Lettice et al. 2003). Notably, CNEs 23 and 24, and CNEa, are located in intron 1 of the *Ptprm* gene, which is not inconsistent with a role in *Lama1* transcriptional regulation considering the case of the PZA *Shh* enhancer and the *Pax9* enhancer driving expression in the medial nasal process of the mouse embryo, which was found 8kb downstream of the last exon of *Pax9* - in the 7th intron of the *Slc25a21* gene (Santagati et al. 2003). Alternatively, these and other CNEs may be involved in the control of *Ptprm* and/or *Lrrc30* expression, instead, as discussed in Chapters 5 and 6. The remoteness of many enhancers from their target promoters hints for long-distance mechanisms of interactions that involve the establishment of physical contacts (chromatin looping) between the enhancer and the promoter (Lower et al. 2009; Dean 2011; Gibcus 2013). Hence, it is plausible that some of the CNEs detected in my study may function as long-range enhancer elements to modulate *Lama1* transcription.

4.4.3. Some CNEs may represent other classes of regulatory elements and/or ncRNA genes

Not all conserved non-coding sequences have enhancer properties. For instance, a conserved element may perform insulator functions as shown for the CTCF-dependent insulators flanking the mouse and human *β-globin* clusters (Farrell 2002), or the pan-vertebrate insulator shielding the *Hoxd* complex in mouse, chicken and zebrafish from the influence of the nervous system-specific enhancers of the *Evx2* gene (Kmita et al. 2002). Similarly, Glazko et al. (2003) found that nearly 11% of strongly conserved regions in mouse/human

alignments correspond to predicted matrix-attachment regions (MARs), huge fraction of which preceded the 5' ends of genes, suggesting a role in transcriptional regulation (Glazko et al. 2003). Alternatively, a CNE may exert a rather large-scale chromatin organisation function in chromosome condensation or replication (Cremer and Cremer 2001).

There is also a possibility that some of the CNEs in the vicinity of *Lama1* may operate via a RNA intermediate, i.e. that some CNEs are actually transcribed as non-coding RNAs (ncRNAs), which then embark on transcriptional regulation of target genes, as shown for the 2.2 kb ncRNA *HOTAIR* from the human *HOXC* locus which represses *HOXD* transcription *in trans* (Rinn et al. 2007). Such scenario for some of the identified CNEs in this study is plausible and consistent with the finding of pervasive non-coding transcription through-out the human genome (Birney et al. 2007). However, applying a novel algorithm for analysis of RNA-seq, ChIP-seq and DNaseI hypersensitivity data from human and mouse cell lines, Hemberg et al. (2012) reported that conserved elements are four times more likely to correspond to clusters of TFBSs than to unannotated ncRNAs, although it is also true that many of the TFBSs clusters were transcribed as low-abundant unspliced and non-polyadenylated RNAs (Hemberg 2012). In fact, this is consistent with recent studies which found that some enhancers are transcribed as enhancer RNA (eRNA) that appear not to be a by-product of accessible chromatin but to perform active function in enhancer-promoter looping, at least in the case of the oestrogen receptor α (ER-α) transcription factor and its target genes (Li et al. 2013).

Therefore, it is possible that some of the conserved elements that I have identified may not be *bona fide* enhancers but instead function as insulators, silencers, MARs or even ncRNA.

4.4.4. CNEs and gene synteny

That some of the identified CNEs might function as transcription-regulatory elements is suggested by the preservation of gene content and order in the surroundings of the *Lama1* locus across vertebrate phylogeny, a phenomenon known as "shared synteny" (Moreno-Hagelsieb et al. 2001). The constraints that impede synteny breakage could be a result of the requirements for regulatory elements in the transcriptional control of one or more genes in the cluster, which would impose negative selection pressure on any translocation event that separates a dependent gene from its regulatory elements (Ahituv et al. 2005; Engstrom et al.

2007; Irimia et al. 2012). The practical implication of this phenomenon for locus-specific TRE identification is that there is a much higher probability for a TRE of a gene of interest to reside in the syntenic block than outside of it (Haeussler 2011). Such scenario is plausible, especially when the expression pattern of the genes in the block is evolutionary conserved, as is the case for *Lama1* mRNA expression among mouse, zebrafish and chicken (Chapter 3 of this study; Joseph Pickering's Thesis 2012; Zinkevich et al. 2006). Shared synteny, combined with conserved expression pattern, suggest that some or all of the TREs controlling *Lama1* expression may reside in the vicinity of *Lama1*. Moreover, as mentioned earlier, CNEs 11-16 form a cluster in the territory between the *Lrrc30* and *Ptprm* loci (Figure 4.3). This is in agreement with the non-uniform distribution of ultraconserved elements in the mouse and human genomes (Bejerano et al. 2004) and with the clustering of developmentally-active conserved enhancers on the 3rd chromosome in *Drosophila* (Kundu et al. 2013), raising the possibility of similar regulatory properties for CNEs 11-16.

4.4.5. GLI binding motif-containing CNEs are candidate SHH-responsive enhancers

Most important in this study is the identification of several CNEs (CNEs 13, 23 and 24) that contain GLI binding motifs (Figure 4.4; Table 4.2), as these may represent the transcription-regulatory elements mediating the response to SHH signalling in the somites and neural tube (Anderson et al. 2009). Notably, few of the CNEs harbour binding motifs for the ZIC transcription factors (CNEs 4, 7, and 22) which have a highly homologous Zn-finger DNA binding domain compared to the GLI factors (Aruga et al. 1996). Given the fact that the binding motif for ZIC factors (5'-GGGTGGTC-3') is almost identical to the consensus motif for the GLI proteins (5'-TGGGTGGTC-3') (Kinzler and Vogelstein 1990; Mizugishi et al. 2001), it is not unlikely that the GLI TFs may have the potency to occupy both motifs *in vivo*. Therefore, CNEs 4, 7 and 22 are also candidates for the regulatory elements mediating the SHH signal in the mouse somites and neural tube. However, it is important to note that none of the identified GLI/ZIC motifs is significantly over-represented in the CNEs (Table 4.2), relative to genomic back-ground. Therefore the presence of GLI/ZIC motifs in CNEs 4, 7, 13, 22, 23 and 24 is not considered a strong indicator of their putative role as mediators of the SHH effects on *Lama1* transcription.

It is noteworthy to mention that despite the lack of conserved GLI/ZIC binding motifs, the orthologous sequences of CNE13 and CNE23 in the opossum genome do harbor a

GLI or a ZIC motif, respectively (see Table 4.2), albeit at a different position compared to the mouse orthologs. It is likely that during evolution, the exact place of the GLI/ZIC motif has shifted (shuffled) in the mouse or opossum genomes without necessarily affecting the putative regulatory function these CNEs might have. Such reorganization of TFBSs in developmental enhancers is well documented in the even-skipped (*eve*) stripe 2 enhancer (es2E) among multiple *Drosophila* species. There, the distribution and number of sites for the Bicoid, Hunchback, Krüppel, Knirps and Giant transcription factors have changed in the different species without changes in the regulatory output of the enhancer (Ludwig et al. 1998; Ludwig et al. 2000).

The identification of several conserved non-coding elements around the murine *Lama1* locus suggests that they may be involved in the transcriptional regulation of *Lama1* expression, functioning as enhancers, insulators, or silencers, among other possibilities. Notably, some CNEs contain GLI and/or ZIC binding motifs, although not significantly over-represented. Nevertheless, compared to the other CNEs, the GLI/ZIC motif-containing CNEs are putative candidates for the SHH-responsive somitic and neural enhancers of *Lama1*. Testing these hypotheses requires experimental examination of CNE function, as described in Chapters 5 and 6.

Chapter 5

in vitro analysis of CNE activity

5.1. Hypothesis and Aims

Numerous studies in various species have demonstrated the involvement of predicted conserved non-coding elements (CNEs) in the transcriptional regulation of nearby genes. Therefore, I hypothesised that the CNEs identified in the vicinity of the murine *Lama1* locus have transcriptional activities. I tested this in a cell culture system by performing transient cell transfections of reporter gene constructs driven by the CNEs, followed by measurements of reporter gene activity.

5.2. Introduction: transient cell transfection for analysis of CNE function

Once a candidate regulatory element has been predicted either by indirect computational methods based on sequence conservation or by direct biochemical profiling of chromatin states across the genome, the next essential step is the demonstration of its ability to modulate gene transcription.

A frequently employed method for relatively rapid screening of the functional properties of conserved (or non-conserved) non-coding sequences is the transient cell transfection assay, where reporter plasmids containing the candidate elements are introduced in cells maintained in culture (Naylor 1999). The activity of the tested element is assayed by its effect on the rate of transcription of the reporter gene, which in most protocols is inferred indirectly by measuring the concentration or enzymatic activity of the reporter's gene protein product (Bronstein et al. 1994). This approach is considered transient since the introduced plasmids remain episomal and rarely integrate into the host genome, which necessitates that the measurements of reporter activity must be taken within 24 – 72 hours post transfection (Carey and Smale 2000).

It is important to note that the host cells should maximally represent the tissue type where the putative element is expected to act considering the expression pattern of the endogenous gene of interest. This is especially relevant to assays of CNE function for it increases the probability that the host cells would express the full complement of transcription factors necessary for the activity of the candidate element (Carey and Smale 2000). In transfection experiments, the preferred path for introducing reporter plasmids in cells is via lipofection, where DNA is complexed with cationinc lipid compounds (Felgner et al. 1987). The most frequently used reporter encodes the firefly luciferase enzyme from *Photinus pyralis* (deWet et al 1987). The enzyme is active in cell lysates and catalyzes the

oxidation of exogenously supplied D-luciferin, which is accompanied by the release of light that can be quantified (Allard et al. 2008). The luciferase assay is very sensitive and spreads over a linear range of more than 7 orders of magnitude of luciferase concentration, which allows convenient measurements of a broad range of enzyme activities (TM040, Promega).

Numerous studies have employed the luciferase reporter system to examine the functional potential of putative TREs, including the few studies on the basal promoter and proximal enhancers of the murine and human *Lama1* gene (Niimi et al. 2003, Niimi et al. 2004; Piccinni et al. 2004; Niimi et al. 2006). Luciferase reporter-based analyses in cell culture are not limited to the identification of enhancers but can also reveal the function of silencer elements as demonstrated for a conserved repressor motif in the chicken *cardiac troponin T* gene (Tidyman et al. 2003) and the conserved Polycomb-dependent silencer region D11.12 between the human *HOXD11* and *HOXD12* genes (Woo et al. 2010).

These results illustrate that the highly sensitive and relatively rapid luciferase reporter-based assay is well suited for initial screening of the activity of candidate TREs in transient cell transfections. Nevertheless, several factors can influence the outcome of such an experiment (Table 5.1), leading to limitations, such as the requirement for high number of experimental replicates, the need to use endogenous promoters in the appropriate cell lines, as well as the difficulty of studying candidate enhancers that cannot function in isolation and/or elements requiring chromosomal integration and a particular chromatin environment (Carey and Smale 2000). However, some of these limitations can be mitigated with the appropriate controls (Table 5.1).

5.3. Results

5.3.1. Analysis of CNE function via transient cell transfection of luciferase reporter constructs

I decided to employ the transient cell transfection assay combined with luciferase reporters as an approach to screen several of the identified CNEs for potential enhancer activity. I focused my analyses on all CNEs containing GLI/ZIC binding motifs (CNEs 4, 7, 13, 22, 23 and 24) but also included CNEs 3, 5, 10, 15 and 19 (which do not harbor GLI/ZIC motifs) for these exhibited high sequence conservation in the mouse/opossum phylogenetic footprinting comparisons (see Table 4.1). Each of the selected CNEs was amplified from mouse genomic DNA (from the C57BL/6J mouse strain) in a standard PCR protocol and cloned upstream of

the SV40 promoter in the pGL3- Promoter plasmid (Promega), which harbours a modified firefly luciferase reporter gene *luc+*. It is important to note that the orientation of the CNEs in the reporter plasmid was maintained according to their orientation relative to the *Lama1* transcription start site on mouse chromosome 17 (Figure 5.1)

Factors leading to limitations of the assay	Controlled by	Controlled in this study
1. inefficient cell transfection	normalisation of the test reporter signal to the signal from a 2nd reporter plasmid (internal control), e. g. taking the ratio of firefly luciferase activity to Renilla luciferase activity.	yes
2. variation in cell proliferation, death and lysis between samples	using the same total cell lysate protein concentration in each assay	yes
3. inappropriate cell type	using a cell line that represents one or more of the tissue types normally expressing the gene of interest *in vivo*; ideally, the cell line should express the gene of interest.	yes/no
4. inapproapriate promoter	using the endogenous promoter of the gene of interest.	no
5. non-specific promoter induction	Comparing the activity of the tested candidate enhancer to the activity of a random genomic region.	no
6. reprorter-plasmid independent, endogenous luciferase activity	examining the luciferase activity in a lysate from non-transfected cells.	yes
7. aberrant transcription initiation from cryptic sites in the vector's backbone.	insertion of a polyA signal immediately up-stream of the promoter.	yes
8. promoter/promoter or enhancer/promoter interference or crosstalk	using an internal control plasmid with a different promoter or, ideally, by testing the candidate enhancer in stably transfected cells.	no

Table 5.1. *Potential factors that can influence the outcome of a transient cell transfection-based assay of candidate TREs.* Here are listed some of the most pertinent factors that can affect the results from a transient cell-transfection-based analysis of candidate TREs, together with proposed controls (Carey and Smale 2000; Schagat et al. 2007). The third column shows whether the relevant factors have been controlled for in the current study. Factors 3, 4, 5 and 8 are discussed in more detail in Section 5.4 of this chapter.
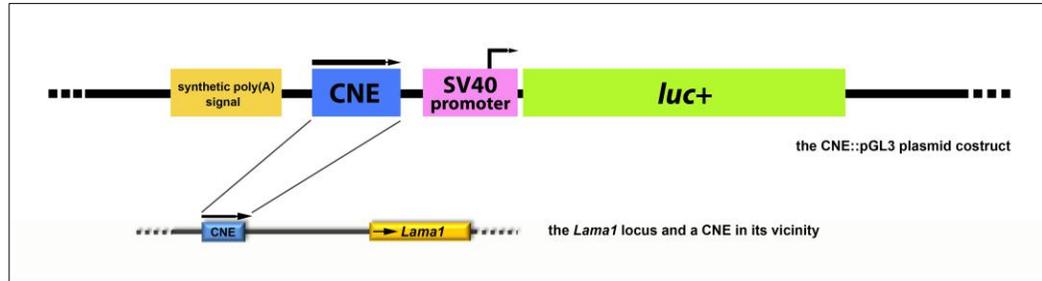
Figure 5.1. *A schematic diagram of CNE::pGL3 reporter plasmid constructs.* The top half of the diagram shows the arrangement of important functional elements in the firefly luciferase reporter plasmid pGL3-Promoter, including the location of the cloned CNE. Below is shown a diagram of the relationship of a CNE to the *Lama1* locus in terms of directionality (indicated by a *solid black arrow* above the blue CNE box). This feature was kept unchanged in the *CNE::pGL3* constructs.

The pGL3-Promoter plasmid is well suited for identification of mammalian elements with enhancer properties. The advantages of this reporter vector include optimized codon content of the *luc+* gene for efficient translation in mammalian cells, elimination of consensus transcription factor binding motifs from the coding sequence of *luc+* and, most importantly, the presence of a poly (A) site upstream of the tested element which serves as a transcription-pause signal to prevent background transcription initiation events from sequences in the vector's backbone (Promega 2008) (Figure 5.1).

I transfected each *CNE::pGL3* reporter plasmid individually into the mouse embryonic fibroblast cell line C3H10T1/2 (step 1 on Figure 5.2). This cell line has properties of mesenchymal multipotent stem cells, is SHH-responsive (Kinto et al. 1997), and can be directed into skeletal muscle, smooth muscle, osteogenic, chondrogenic and adipogenic cell lineages under different protocols (Haas et al. 2000; Bostrom et al. 2000; Shea et al. 2003; Tang et al. 2004). Notably, these cell types are also generated *in vivo* by progenitors descending from the somites (Christ and Scaal 2008). Therefore, the C3H10T1/2 line is a good *in vitro* model system to investigate mammalian somitic cell specification and differentiation. These advantages are especially relevant to the current study, which investigates the control of *Lama1* transcription by SHH in the murine somites and neural tube.
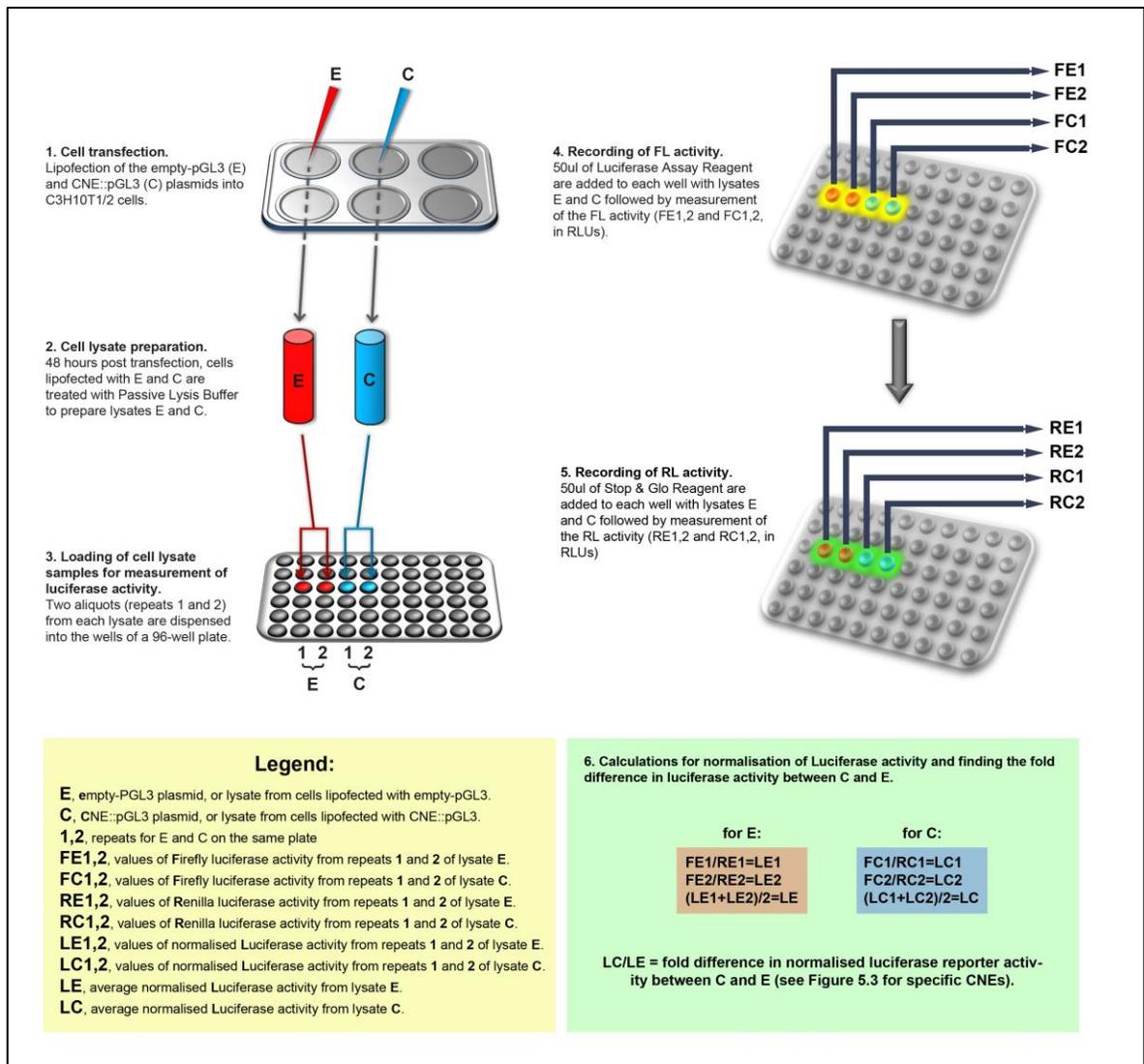
Figure 5.2. *A schematic diagram of the transient cell transfection procedure and the double luciferase assay used to screen for CNE activity.* (1) – (6) designate consecutive steps in the experimental and analytical approach accompanied with a short description of each step. The empty-pGL3 control plasmid and the lysate derived from cells transfected with this plasmid are highlighted in red; the CNE::pGL3 test plasmid and the lysate derived from cells transfected with this plasmid are highlighted in blue. The curved arrows emanating from the plates at steps (4) and (5) represent the recordings from the two luciferases, the firefly luciferase (FL) and Renilla luciferase (RL), respectively. A Legend panel is provided at the left bottom quadrant with explanation of used abbreviations. RLUs, **R**elative **L**ight **U**nits, are the units of measurement of luciferase activity.

In order to control for transfection efficiency, pipetting inconsistencies and toxicity, each *CNE::pGL3* test reporter plasmid was co-transfected with the *pRL-SV40* control reporter plasmid which encodes the *Renilla* sea pansy luciferase (Schagat et al. 2007). This enzyme uses a different substrate – coelenterazine (Matthews et al. 1977), than the firefly luciferase, allowing for the sequential recording of the activities of both enzymes in the same

sample (see Chapter 2 for details on the procedure). This permits the normalization of each sample by computing the ratio of test reporter activity (the firefly luciferase from *CNE::pGL3*) to internal control reporter activity (the *Renilla* luciferase from pRL). The latter approach minimizes or excludes sample-to-sample variability and enables comparisons of test reporter activity between cell samples transfected with *CNE::pGL3* and samples transfected with the empty-pGL3 (the parent pGL3 plasmid which does not contain any CNE) (Figure 5.2)

The activity of the luciferase reporters (measured in Relative Light Units, or RLUs) was recorded sequentially in the lysates from each CNE::pGL3 sample. At least three replicate transfections were performed for every *CNE::pGL3* construct including the empty-pGL3 control plasmid. In each replicate, I conducted two independent luciferase assays with separate aliquots from every cell lysate (step 3 on Figure 5.2). Then, I averaged the calculated FL/RL ratios for each CNE (LC in Figure 5.2) and compared these to the FL/RL values from the empty-pGL3 control sample (LE in Figure 5.2) to obtain the CNE::pGL3/empty-pGL3 ratios (LC/LE in Figure 5.2). This approach allowed me to determine the normalized fold change in activity between a *CNE::pGL3* construct and the empty-pGL3 control.

Next, I performed one-way analysis of variance (ANOVA) followed by a Dunnett's post-test using the $\log_{10}$ of the LC/LE ratios, which enabled the comparison of the mean of every LC/LE value to the LE/LE value (which equals 1). The fold change in luciferase activity and the results from the statistical analysis are shown in Figure 5.3.

Notably, several of the examined CNEs exhibited either higher or lower fold change in luciferase activity relative to the empty-pGL3 control (Figure 5.3). CNEs 10, 19, 21, 22 and 23 displayed increased luciferase activity. In contrast, CNEs 7 and 13 displayed decreased luciferase activity hinting for silencer-like properties (Figure 5.3). It is important to note that the ANOVA statistical analysis did not reveal significant differences between the mean RLUs of the tested *CNE::pGL3* constructs versus the empty-pGL3, except for *CNE7::pGL3*. The lack of statistical significance in the case of CNEs 10, 19, 21, 22 and 23 is likely the result of a small sample size (small number of replicates). To improve the data, higher number of replicates is needed.
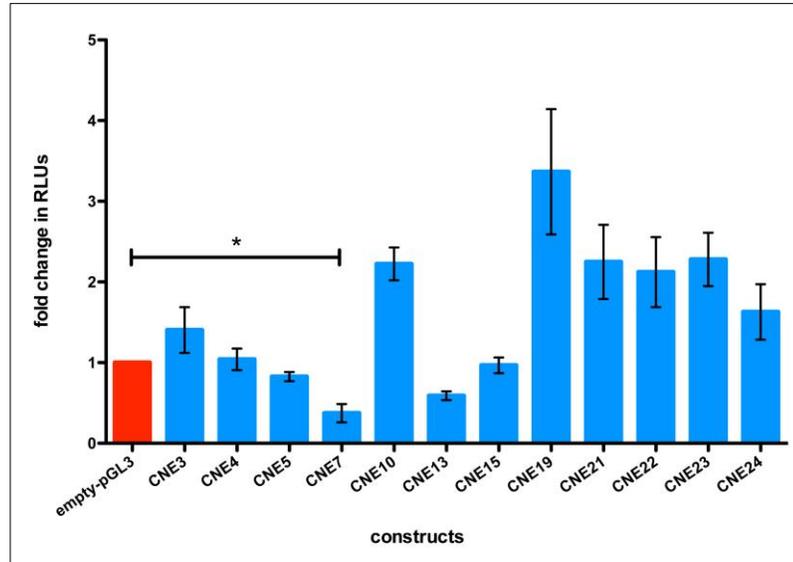
Figure 5.3. *Fold change in firefly luciferase activity driven by the tested CNEs.* A plot of the fold change in firefly luciferase activity (as measured in RLUs) driven by each CNE (shown by individual *blue* columns) compared to the activity of the empty-pGL3 control (*red* column), which equals 1. Black bars associated with each blue column represent the standard error of the mean (SEM) Statistical analyses were performed using One-Way ANOVA followed by Dunnett's post-test (P<0.0001).

In summary, the results from the transient cell transfections of *CNE::pGL3* luciferase reporter constructs in C3H10T1/2 cells do not suggest that some of the conserved elements possess transcriptional activity, except perhaps for CNE7 (Figure 5.3). The putative causes behind these results are discussed in more detail in Section 5.4.

5.3.2. *in silico* analysis of CNE correlation with molecular markers of chromatin state and TFBSs

Despite the weak or absent activity of the CNEs when tested in isolation from the rest of the genome, I decided to examine whether the CNEs associate with diagnostic characteristics of TREs, in particular with marks of occupied transcription factor binding sites, specific histone modifications and DNaseI hypersensitivity sites, while in their genomic context. First, I analysed the distribution of such marks in the genomic region encompassing all CNEs identified in this study, using the UCSC Genome Browser. Notably, several chromatin marks exhibit discrete, discontinuous enrichment along the entire locus; however, the majority of these marks do not co-localise with the CNEs identified in the current study (Figure 5.4). Intriguingly, *Lama1*-expressing organs (like brain and kidney) display similar chromatin mark profiles that are distinct from the mark profile of *Lama1*-non-expressing organs (such as spleen) (Figure 5.4).
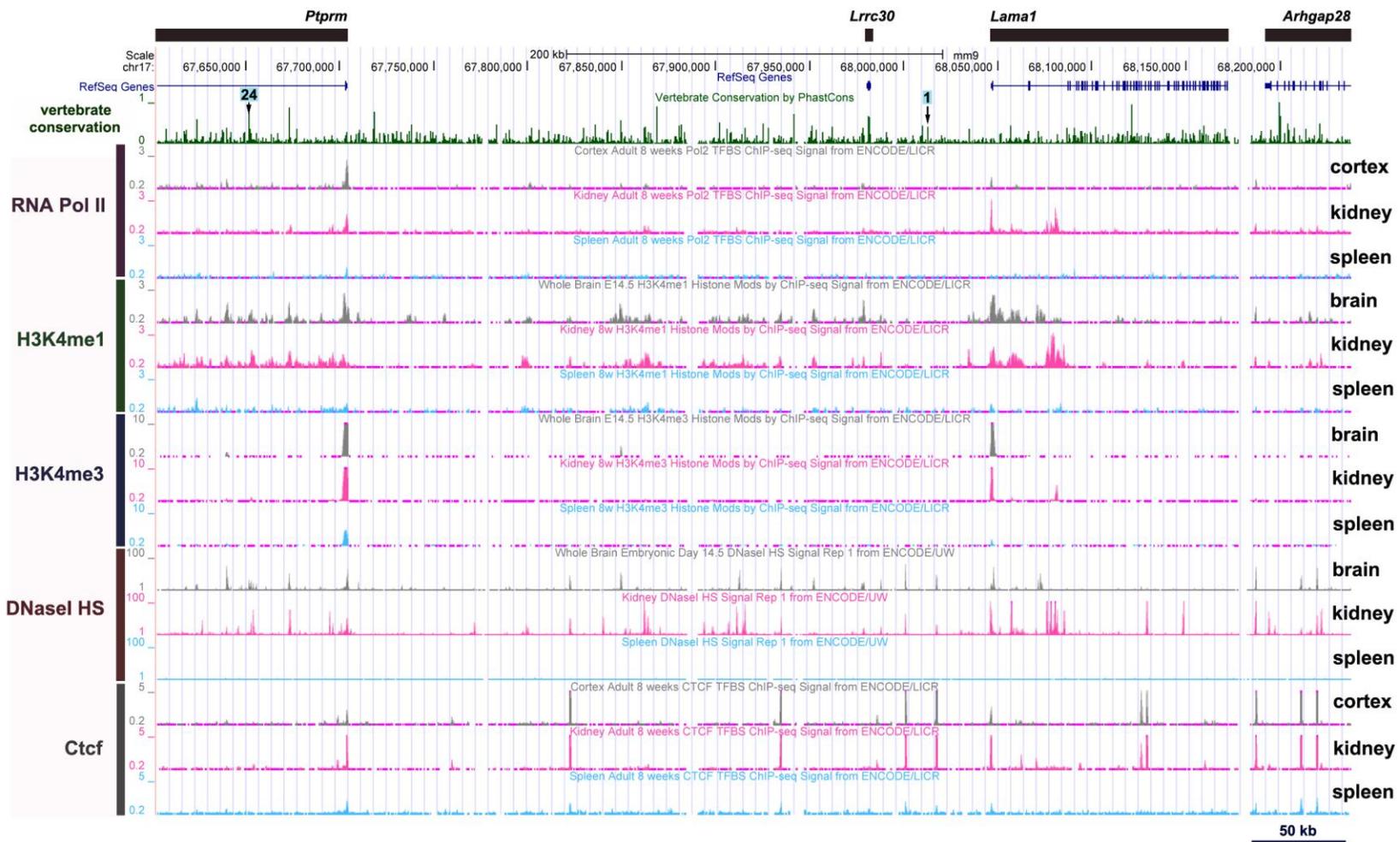
112

Figure 5.4. *Distribution of chromatin state marks in the genomic region including all 24 CNEs.* Description is provided on the next page (p. 114).

113

Figure 5.4. Distribution of chromatin state marks in the genomic region including all 24 CNEs. The distribution of RNA Pol II and CTCF binding, DNaseI hypersensitive sites, and H3K4me1 and H3K4me3 marks was analysed across the entire genomic region containing the identified CNEs, using the UCSC Genome Browser. The graph in the top row indicates sequence conservation generated by PhastCons. The plots on the subsequent rows display chromatin mark enrichment in brain, kidney and spleen. Notable peaks of most marks are observed in the promoter regions of *Lama1* and *Ptprm*. Interestingly, additional peaks of H3K4me1 and DNaseI hypersensitivity (DNaseI HS) are seen within the *Lama1* locus, whereas several peaks of CTCF enrichment and DNaseI HS are also found in the intergenic space between the *Lama1* and *Ptprm* loci. Importantly, of all CNEs identified in this study only CNE7 and CNE23 co-localise with enrichment peaks higher than 15% above background level (see Figures 5.5 and 5.6; Table 5.2). The rest of the CNEs show lower than 15% or no enrichment relative to background. The positions of CNE1 and CNE24 are indicated in *cyan boxes* on the sequence conservation row (see the previous page 113).

Next, I examined in more detail the association of each of the CNEs (1-24 and CNEa and b) with various chromatin marks. Here, I report on the findings made with CNEs 7 and 23 (Table 5.2) for the analyses of the other CNEs did not reveal significant enrichment for any transcription factor binding, histone modification or DNaseI hypersensitivity.

CNE7 is especially interesting (Figure 5.5 and Table 5.2) as its genomic position correlates with a high peak of CTCF (CCCTC-binding factor) occupancy in various organs such as the embryonic E14.5 brain and limb buds, adult cortex, kidney and heart, and with a somewhat lower CTCF peak in adult lung and cerebellum. This contrasts with the low or background levels of CTCF occupancy in the adult liver, small intestine, spleen, bone marrow, testes and mouse embryonic fibroblasts (Table 5.2 and Figure 5.5). Intriguingly, CNE7 also overlaps with DNaseI hypersensitive sites particularly in the adult heart, skeletal muscle, kidney and brain, as well as in the embryonic E11.5 mesoderm, forelimb and hindlimb buds, and the E14.5 brain. There are very low or no DNaseI peaks in the adult liver, colon, lung, cerebellum, adipose tissue, spleen and thymus (Table 5.2 and Figure 5.5). The putative implications of these features on CNE7's function are discussed later.

Analyses of CNE23 revealed the presence of the H3K4me1 histone modification associated with it in the adult olfactory bulb and kidney, and in the embryonic E14.5 brain and limbs (Table 5.2 and Figure 5.6). In addition, CNE23 correlates with DNaseI hypersensitive sites in the neonatal retina (highest peak), E14.5 brain, E11.5 mesoderm, E11.5 forelimb and hindlimb buds and the adult kidney.

Overall, the results from the *in silico* analyses of CNEs 7 and 23 hint for putative *in vivo* transcription-regulatory functions of these elements.

organ/tissue	CNE7		CNE23	
	CTCF[a]	DNaseI[b]	H3K4me1[c]	DNaseI[b]
E11.5 mesoderm		++		+
E11.5 FL and HL buds		++		+
E14.5 limb buds	++		+	
E14.5 brain	++	+	+	+
kidney 8w	++	+	+	+
skeletal muscle 8w		++		
heart 8w	++	++	+	
embryonic fibroblasts 8w	+/-			
bone marrow 8w	+/-			
thymus 8w		+/-		
spleen 8w	+/-	+/-		
adipose tissue 8w		+/-		
brown adipose tissue 24w			+	
testis 8w	+/-			
liver 8w	+/-	+/-		
colon 8w		+/-		
small intestine 8w	+/-		+	
lung 8w	+	+/-		
retina 1d				++
cortex 8w	++	+		
olfactory bulb 8w			++	
cerebellum 8w	+	+/-		

Table 5.2. *Correlation of TFBSs occupancy, histone modifications and nucleosome accessibility with CNEs 7 and 23 in embryonic and adult tissues and organs.* Legend: ++, high or moderate levels; +, low levels; +/-, very low levels or signal approaching background noise. Abbreviations: FL, forelimb; HL, hindlimb; 1d, 1 day postnatal; 8w, 8 weeks postnatal; 24w, 24 weeks postnatal; H3K4me1, monomethylation of Lysine 4 in histone 3; H3K27ac, acetylation of Lysine 4 in histone 3.

[a] data for TFBSs by ChIP-seq from the ENCODE/LICR track in the UCSC Browser.
[b] data for DNaseI hypersensitive sites by Digital DnaseI from the ENCODE/UW track in the UCSC Browser.
[c] data for histone modifications by ChIP-seq from the ENCODE/LICR track in the UCSC Browser.
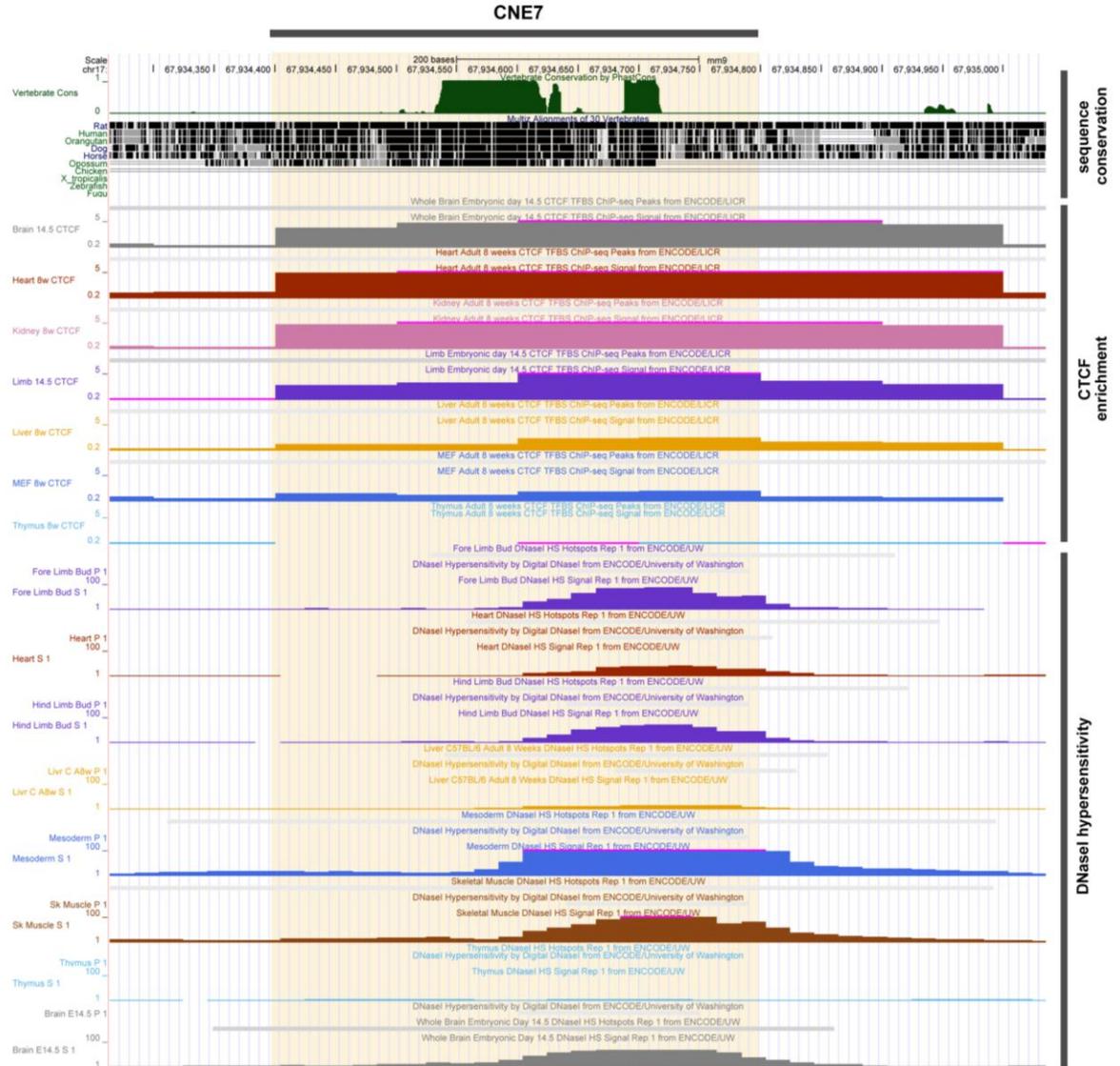
Figure 5.5. *Association of CNE7 with CTCF transcription factor occupancy and DNaseI hypersensitive sites as shown by the UCSC Genome Browser.* A graphical display of CNE7 sequence conservation, CTCF occupancy and DNaseI hypersensitivity generated by the UCSC Browser using the NCBI37/mm9 assembly of the mouse genome. Each row (or data track) from top to bottom presents features associated with the sequence at the corresponding genomic position (indicated by the coordinates at the top of the display). The subsets of tracks with thematically similar data (sequence conservation, CTCF enrichment and DNaseI hypersensitivity) are indicated on the right side of the diagram. The extent (in base pairs) of CNE7 is shown by the thick grey bar at the top of the display and highlighted in light orange along all data tracks.
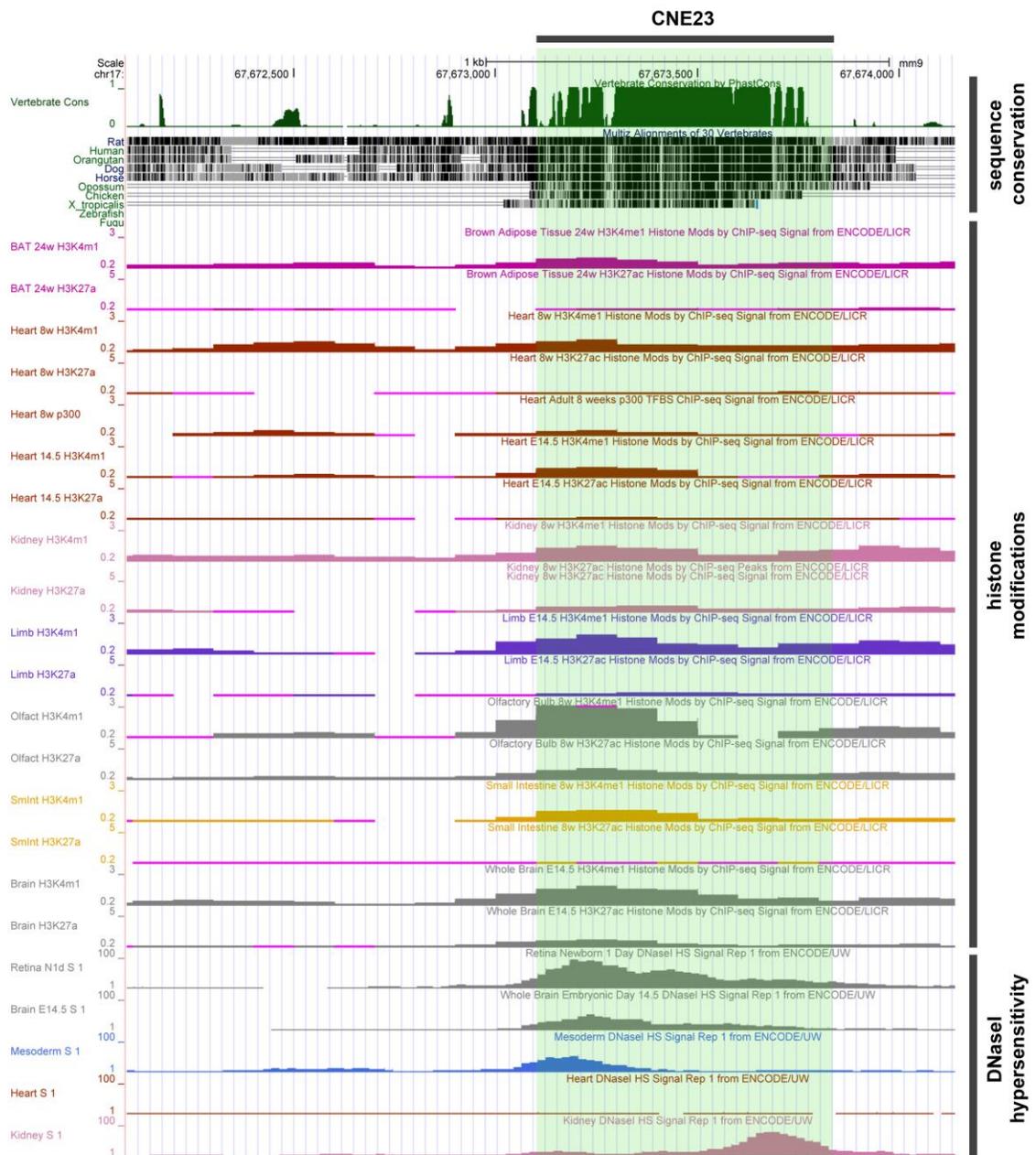
Figure 5.6. *Association of CNE23 with enhancer-enriched histone marks and DNaseI hypersensitive sites as shown by the UCSC Genome Browser.* A graphical display of CNE23 sequence conservation, histone modifications and DNaseI hypersensitivity generated by the UCSC Browser using the NCBI37/mm9 assembly of the mouse genome. Each row (or data track) from top to bottom presents features associated with the sequence at the corresponding genomic position (indicated by the coordinates at the top of the display). The subsets of tracks with thematically similar data (sequence conservation, histone modifications and DNaseI hypersensitivity) are indicated on the right side of the diagram. The extent (in base pairs) of CNE23 is shown by the thick grey bar at the top of the display and highlighted in light green along all data tracks.

5.4. Discussion

Here, I examined the transcriptional activity of CNEs 3, 4, 5, 7, 10, 13, 15, 19, 21, 22, 23 and 24 using a transient cell transfection approach combined with recordings of the activity of a firefly luciferase reporter driven by the CNEs (Figures 5.1, 5.2 and 5.3). Several of the CNEs, in particular CNE7, 10, 19, 21, 22 and 23, changed weakly the activity of the luciferase reporter compared to the negative control (Figure 5.3). However, it remains unclear whether this change is a result of genuine enhancer-like properties of the CNEs, or it is a stochastic effect of the CNEs on the SV40 promoter due to possible opportunistic interactions between transcription factors bound to the CNE and the RNA pol II machinery at the SV40 promoter. A control assay with a similarly-sized random genomic fragment is required in order to test the latter possibility, in which case a significantly higher activity is expected from the CNE, provided it is an enhancer, as compared to the random fragment. In addition, the ANOVA analyses of the mean fold change for each CNE did not show statistically significant differences from the empty pGL3 negative control (except for CNE7), which is a result of the weak and highly variable activity of the tested CNEs. Therefore, the current data do not provide a strong argument for the putative transcription regulatory activity of the tested CNEs, except for CNE7. Several factors may have contributed to this outcome, as discussed below.

5.4.1. CNEs and their transcription-regulatory activity in cell culture

The low luciferase activity driven by the tested CNEs (Figure 5.3), compared to results from other similar studies (Niimi et al. 2003; Hlawatsch et al. 2013), raises some questions about the experimental parametres and design.

One possibility for the low performance of the CNEs is that they might require the endogenous *Lama1* promoter for optimal activity. This is due to the phenomenon of enhancer-promoter specificity whereby the enhancers of one gene are incompatible with the promoters of other, neighbouring genes, and can direct transcription only from their target gene's promoter (Butler and Kadonaga 2001). Such enhancer/promoter dependency was previously demonstrated for the adjacent but divergently transcribed *gooseberry* (*gsb*) and *gooseberry neuro* (*gsbn*) genes, as well as for the *dpp* gene in *Drosophila* (Li and Noll, 1994; Merli et al. 1996). Using an enhancer-trap approach in *Drosophila*, Butler and Kadonaga

(2001) demonstrated that some transcriptional enhancers can direct reporter expression from a DPE-dependent core promoter, but not from a TATA-dependent core promoter, and vice versa, showing a preference for particular core promoter motifs. Specific enhancer-promoter preferences have also been described in yeast (Li et al. 2002) and sea urchins (Kobayashi et al. 2007). It is hypothesised that such specificity may enable some enhancers to activate transcription from the correct promoter in gene-dense genomic regions, or when the enhancer is located far (in tens or hundreds of kilobases) from the target gene (Butler and Kadonaga 2001).

Thus, it is possible that the weak or absent activity of the tested CNEs may be a result of their inability to effectively interact with the SV40 promoter in the pGL3 vector, as opposed to a putaive productive interaction with the endogenous *Lama1* promoter, because of promoter-specific sequence motif differences. Consistent with such an idea, the SV40 and *Lama1* promoters are composed of distinct sequence motifs - a TATA-box and a CpG island, respectively (Byrne et al. 1983; Piccinni et al. 2004). It can be hypothesised that some of the tested CNEs are transcriptional enhancers that have adapted to operate in conjunction with a CpG-rich promoters, but not with a TATA-box ones, analogously to the observations in *Drosophila* (Butler and Kadonaga 2001). Therefore, a better approach to testing the CNEs for enhancer-like properties is to design new *CNE::reporter* constructs where the reporter gene is driven by the endogenous *Lama1* promoter.

Second, the low activity of the tested CNEs might be due to the requirement for specific transcription factors and/or inducing signals that were not presented in the C3H10T1/2 cell culture system. As mentioned above, analyses of the existing literature suggested that C3H10T1/2 cells exhibit properties of mesoderm-derived cells (Haas and Tuan 2000; Bostrom et al. 2000; Shea et al. 2003; Tang et al. 2004), such that they are perhaps an optimal environment for preliminary identification of the mesodermally-active *Lama1* enhancers. Additional support for the choice of C3H10T1/2 cells could be provided by a demonstration of their ability to endogenously express *Lama1* mRNA. Furthermore, C3H10T1/2 cells may not be the ideal system to assess the transcription regulatory activity of CNEs with non-mesodermal tissue-specificity, such as candidate enchancers operating in the neural tube. The use of other cell lines, combined with transgenesis assays, would be the most efficient strategy to address this limitation.

119

Another possible explanation is that some of the tested CNEs might have other roles than that of enhancers or silencers, like matrix attachment regions (MARs) or locus control regions (LCRs), for instance, the properties of which are difficult to demonstrate in transient assays for such demonstration requires chromosomal integration of the tested construct (Dean 2011; Fraser and Grosveld, 1998). This scenario is consistent with the observed lack of enrichment for enhancer-associated chromatin marks on most CNEs (Section 5.3.2). In regard to LCRs, these genomic regions are defined based on their ability to confer high levels of integration-site independent expression of a linked transgene (Fraser and Grosveld 1998). As previously suggested (Carey and Smale, 2000), the most effective way to uncover the function of putative LCRs is to generate several independent stable cell lines or transgenic animal lines where it is expected that each one has integrated the transgene construct at random locations in the genome. If the candidate element acts as a locus control region, then more lines (in principle, all of them) should express the transgene than would have been expected for an enhancer element. This is so for LCRs confer protection from heterochromatic modifications that could silence the transgene (Kioussis and Festenstein, 1997). Therefore, a further examination of some CNEs in transgenic animals or stable cell lines in culture would address this possibility.

Also, the low activity of tested CNEs might be a result of interference or cross-talk between the enhancer and/or promoter on the test plasmid (the pGL3) and the promoter on the internal control plasmid (the pRL). The high copy number of reporter plasmids may have led to competition for transcription factors present at limited quantities, which in turn may have resulted in low level occupancy of their binding sites and inefficient reporter gene transcription (Carey and Smale 2000; Mercola et al. 1985; Promega 2008). However, examination of the FL and RL activities does not indicate that such interference has occurred in the current study. Perhaps, the best way to circumvent this limitation is to test each *CNE::reporter* construct in stably-transfected C3H10T1/2 cells (Carey and Smale 2000).

In summary, one or a combination of several uncontrolled factors may have contributed to the weak activity of the CNEs when tested in transient transfection conditions (Table 5.1). Therefore, the data could be improved by substituting the SV40 promoter for the endogenous *Lama1* promoter, by the use of other cell lines, by increasing the number of experimental replicates, and by the inclusion of a comparison with random genomic

elements. Alternatively, the CNEs could be tested in stable cell lines or in transgenic animals, as described in Chapter 6.

5.4.2. CNE7 and its silencer properties in cell culture

Interestingly, CNE7 significantly decreased the activity of the luciferase reporter compared to the empty-pGL3 control (Figure 5.3). This behavior hints to putative silencer properties for CNE7. If this were the case, CNE7 would be expected to operate via a long-range mechanism *in vivo* (Hampsey et al. 2011), as its nearest potential target, the *Lrrc30* gene, is situated nearly 46 kb away from CNE7. Such a scenario is consistent with previous luciferase reporter studies combined with a C3 (or **C**hromosome **C**onformation **C**apture) assay, which revealed the existence of silencer elements in the vicinity of the human *MECP2* gene that are able to directly interact with the *MECP2* promoter despite it being located ~130 kb away of the silencers (Liu and Francke 2006). There is evidence that long-distance interactions between silencers and target promoters are facilitated by DNA looping mediated by the Ume6 and TFIIB factors, which are responsible for the recruitment of the repressively-acting Isw2 chromatin remodeler to target promoters in *S. cerevisiae* (Yadon et al. 2013). It would be interesting to see whether CNE7 acts via a long-range chromosome looping mechanism and which sequence it targets. Are there any target sequences within or nearby the *Lama1* promoter? Both questions could be addressed by the circular chromosome conformation capture (4C) technique (Zhao et al. 2006).

It is also important to note that CNE7 harbours a ZIC binding motif, although not a significantly over-represented one (Chapter 4, Table 4.2), and it is possible that the silencer properties of CNE7 might be mediated by the binding of a ZIC- and/or GLI-repressors forms. This putative dependence on the ZIC motif can be addressed by generating mutant versions of CNE7 with substitutions in the ZIC motif combined with over-activation or down-regulation of the hedgehog signalling pathway.

5.4.3. CNE7 and CTCF occupancy

Another interesting finding about CNE7 was its association with CTCF occupancy in several organs and tissues (Figure 5.4 and Table 5.2). Consistent with these findings, further examination of CNE7 by the MatInspector and FrameWorker *in silico* tools (Cartharius et al. 2005) revealed the presence of a CTCF binding motif on CNE7 that is conserved across

therian mammals (placentals and marsupials), which is indicated on Figure 5.7. However, this motif does not appear to be significantly over-represented in CNE7 compared to the genomic background (Number of matches: 1; Expected: 0.28±0.53; Over-representation: 3.54; Z-score: 0.41; see Legend to Table 4.2 for details on Z-scores). Nonetheless, the peak of CTCF enrichment (as well as the peak of DNaseI hypersensitivity) is centered on the conserved block of CNE7 that harbours the CTCF motif (compare Figure 5.7 with Figure 5.5). These observations hint for a possible role of CTCF in the activity of CNE7 and such hypothesis could be tested in cell transfection experiments with a mutant CNE7 which lacks the CTCF binding motif.
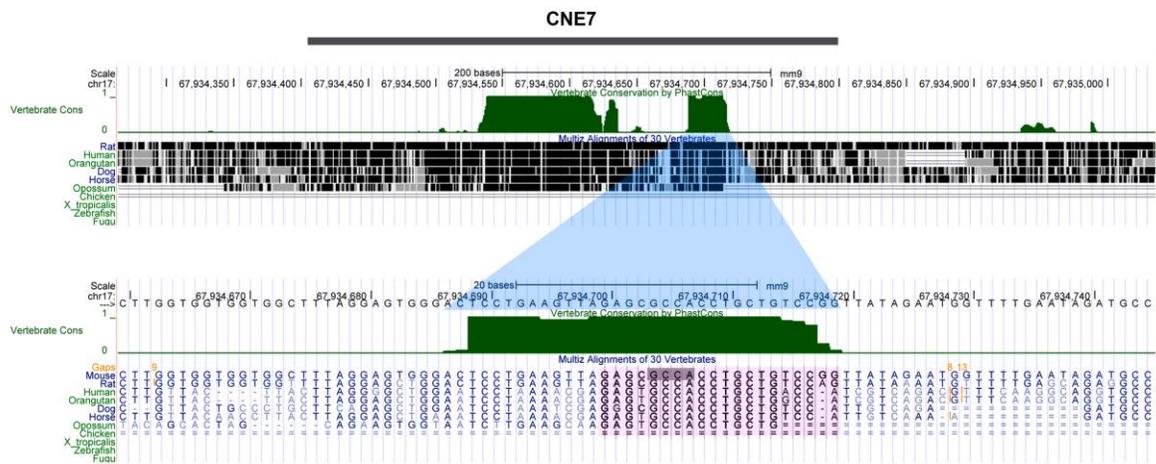
***Figure 5.7.** The CTCF binding motif in CNE7 is conserved across mammals.* A display of the sequence conservation along CNE7 as generated in the UCSC Browser using the NCBI37/mm9 assembly of the mouse genome. The extent of CNE7 is indicated by a thick grey bar at the top of the diagram. The conserved sub-region of CNE7 that contains the CTCF binding motif (highlighted in *pink*) is shown enlarged at the bottom half of the figure. The core of the CTCF motif is highlighted in *grey* and is derived from sequence analysis of CNE7 using the MatInspector tool from the Genomatix Software Suit (Cartharius et al. 2005).

CTCF is a Zn-finger domain containing protein (Burcin et al. 1997) that binds DNA in a sequence-specific manner to motifs with the core consensus CCCTC (Lobanenkov et al. 1990), and is the only known factor that mediates insulator function in vertebrates (Bell et al. 1999; Bell and Felsenfeld, 2000). Genome wide studies of CTCF binding often reveal its localization at the boundaries of transcriptionally active and repressed chromatin (Barski et al. 2007; Cuddapah et al. 2009), and some observations suggest that CTCF exerts its insulator functions by changing chromosome conformation, facilitated by interactions with the cohesin complex (Nativio et al. 2009; Wendt et al. 2008). In addition, there is also

evidence that CTCF -bound insulators can physically associate with enhancers (Handoko et al. 2011) presenting another possible mechanism of enhancer-blocking activity by insulators.

However, in my study CNE7 was cloned only 10 bp 5' of the SV40 promoter in the pGL3 plasmid. Therefore, it is unlikely that CNE7 acted as a classical insulator in this case. Rather, CNE7 would be expected to directly interfere with transcription initiation from the nearby SV40 promoter. Therefore, one can envision a hypothetical scenario, where a CTCF - bound CNE7 directly interacts with its target promoter to silence it *in vivo*, analogously to the repressive interaction of a CTCF -bound CpG-rich element with the mouse *Pax6* P0 promoter prior to glial differentiation of embryonic stem cells (Gao et al. 2011). Whatever the case, it would be interesting to examine whether a CTCF-dependent cohesin-mediated looping process is responsible for the silencing effect of CNE7 on *luc+* expression. This can be tested by a ChIP assay with antibodies against CTCF and the cohesin components SA2 or SMC1/3 (Xiao et al. 2011), in conjunction with the circular chromosome conformation capture method in a C3H10T1/2 cell line stably-transfected with either the wild type *CNE7::pGL3* construct or a version of it with loss-of-function mutations in the CTCF binding motif.

The levels of CTCF occupancy at CNE7 differ in different tissues

It is intriguing that CTCF occupancy of CNE7 varies in different organs and developmental stages. As can be seen in Figure 5.5 and Table 5.2, CTCF levels are higher in the E14.5 embryonic brain and limbs, and in the adult heart and kidneys as compared to the low levels in the adult liver and thymus. If CTCF directly binds to its motif in CNE7, then the binding affinity and/or the activity of bound CTCF may differ in different tissues, and this may affect the putative insulator/silencer activity of CNE7. This is consistent with studies showing that CTCF binding to insulators can be negatively influenced by methylation of cytosines in the binding sites (Hark et al. 2000; Engel et al. 2006), and that the activity of CTCF can be modulated through physical interactions with co-factors, such as CP190 and the thyroid hormone receptor (Lutz et al. 2003; Weth et al. 2010; Wood et al. 2011).

The high levels of CTCF enrichment in the embryonic E14.5 brain and adult kidney, where *Lama1* is expressed (Miner et al. 2004; Sorokin et al. 1997), are puzzling. It is possible that in these structures CNE7 might positively affect *Lama1* expression *in vivo*, as suggested by a recent demonstration of the ability of CTCF /cohesin complexes to bind both the

promoter and enhancers regions of the human *PCDHA* gene facilitating chromatin loop formation and subsequent transcriptional activation (Guo et al. 2012). Alternatively, and in line with the observed silencer properties of CNE7 in culture, this element might in fact function as a CTCF -dependent silencer *in vivo* to quantitatively modulate the rate of transcription from the *Lama1* promoter in tissues where *Lama1* is expressed (embryonic brain, kidney). This is not unlikely for Lum and Lee (2001) showed the presence of a silencer element immediately upstream of the strong TATA-less promoter of the human *HMGB1* gene, which was able to reduce by nearly 6-fold the activity of the highly potent promoter, thus keeping in check the ubiquitous basal levels of *HMGB1* expression in non-proliferating cells (Lum and Lee, 2001).

In summary, the emergent picture of CTCF is that of a multifunctional factor engaged in diverse nuclear activities not only in enhancer-blocking processes and chromatin boundary demarcation but also in transcriptional activation and repression (Phillips and Corces, 2009; Nikolaev 2009). Such diverse roles can be achieved by post-translational modifications of CTCF, like sumoylation (Kitchen and Schoenherr, 2010), by interactions with different co-factors (Lutz et al. 2000; Wood et al. 2011), and by modulating the access to its binding motifs on DNA (Hark et al. 2000), for instance. A comprehensive *in vivo* study that combines BAC-reporter transgenesis, transcription factor binding site mutagenesis and chromosome conformation capture analyses will provide detailed understanding of the precise mechanisms of CNE7 action in a chromosomal context and the role of CTCF.

5.4.4. CNE7 and DNaseI hypersensitivity

CNE7 is also characterised by a peak of DNaseI hypersensitivity. As mentioned earlier, such hypersensitivity marks not only enhancer and promoters but also insulators and silencers (Boyle 2008), which is consistent with the silencer-like behavior of CNE7 in culture. That silencer elements feature such hypersensitivity is demonstrated by a study, which combined phylogenetic footprinting, DNaseI hypersensitivity and luciferase reporter assays to describe the regulatory elements of the human *FSHR* gene, which is specifically expressed in the testicular Sertoli and ovarian granulosa cells (Hermann and Heckert, 2005). Four hypersensitive sites (DHS1-4) were identified, the third of which corresponding to an OCT-1-dependent silencer element in the first intron of *FSHR*, which represses *FSHR* expression in non-gonadal cells (Hermann and Heckert, 2005).

Another interesting aspect of CNE7 is the differential pattern of DNaseI hypersensitivity across different tissues and organs (Figure 5.5 and Table 5.2). For instance, CNE7 features a high peak of hypersensitivity in the embryonic E14.5 brain, E11.5 mesoderm, E11.5 fore- and hindlimb buds and adult skeletal muscles, but a low peak of hypersensitivity in the adult heart, liver and thymus (Figure 5.5 and Table 5.2). The variation in DNaseI hypersensitivity between these structures may reflect the extent of CNE7's activity *in vivo*. As described by Boyle et al. (2008), DNaseI hypersensitivity is not a binary property but represents a continuous range of chromatin accessibility, with the same region showing different "levels" of hypersensitivity in different tissues and conditions. In agreement to this, low or non-expressed genes in CD4+ T-cells showed weaker hypersensitivity at their transcription start sites compared to highly expressed genes (Boyle et al. 2008).

Remarkably, the tissue pattern of DNaseI hypersentivity at CNE7 corresponds to the pattern of CTCF occupancy in different tissue. Hence, tissues/organs with high levels of CTCF occupancy like the embryonic limbs and brain have also higher hypersensitivity (Figure 5.5). It is tempting to hypothesize that CTCF binding on CNE7 depends on chromatin accessibility which might require the binding of a pioneer transcription factor. For instance, binding of PU.1 to an intronic enhancer in *Pax5* leads to reduced nucleosomal occupancy at the locus and up-regulation of *Pax5* expression in B-cell development (Decker et al. 2009; Guertin et al. 2013).

5.4.5. CNE23 and its target gene

Although CNE23 did not significantly enhance reporter gene expression *in vitro* (Figure 5.3), an analysis of this CNE in the UCSC Browser revealed association with the enhancer-enriched histone mark H3K4me1 (Heintzman 2007), as well as overlap with DNaseI hypersensitive sites in diverse embryonic and adult organs and tissues (Figure 5.6 and Table 5.2). The location of CNE23 in the first intron of the *Ptprm* gene suggests the possibility that this element might be devoted to the transcriptional regulation of *Ptprm*, rather than to the regulation of *Lama1*. Such an assumption is consistent with numerous studies in the mouse and other metazoans that have mapped the location of tissue-specific enhancers in the first intron of their target genes. For instance, in the mouse, a 499 bp N-box-dependent enhancer activates *acetylcholinesterase* (*AChE*) gene expression in skeletal muscle cells (Chan et al. 1999); a SOX9-dependent element drives *Col11a2* activity in cartilage (Liu et al. 2000);

a 1.7 kb TBX1-, NKX2.5-, ISLET1-dpendent region directs *Foxa10* expression in the anterior heart field (Watanabe et al. 2012), while a GATA3-dependent enhancer ensures *N-myc* gene expression in the branchial arches (Potvin et al. 2010). Studies of transcriptional regulation in other species also provide examples for enhancers located in the first intron of their target genes: a cartilage-specific enhancer of the *col2a1* gene in *Xenopus* (Kerney et al. 2010); a photoreceptor-specific enhancer of the *CiPax6* gene driving expression in the sensory vesicle of *Ciona intestinalis* (Irvine et al. 2008), and a 334 bp enhancer for early embryonic expression of the *HpOtxL* gene (a member of the *orthodenticle-related* gene family) in the sea urchin *Hemicentrotus pulcherrimus* (Hayashibara et al. 2004). Thus, CNE23 could display tissue-specific activity, in particular in the heart, as suggested by the location of CNE23 in the first intron of the *Ptprm* gene and the endogenous expression of *Ptprm* in the adult myocardium (Koop et al. 2003). However, as evident from Figure 5.6, the peak height for H3K4me1 and p300 coactivator enrichment in both embryonic and adult hearts is not convincingly high to suggest cardiac activity. Moreover, the acetylation mark H3K27ac, which is associated with active enhancers (Rada-Iglesias et al. 2011; Cotney et al. 2012) is absent at CNE23 in the heart (Figure 5.6), which further argues against cardiac-specific function of this CNE.

Alternatively, CNE23 might be involved in the regulation of *Lama1*, as cases exist of genes being regulated by intronic enhancers positioned in another gene located up to a 1 Mb away (Lettice et al. 2003). BAC-reporter studies combined with deletion analyses and a chromosome conformation capture experiment in several relevant cell types will help to uncover the target, or targets, of CNE23.

5.4.6. CNE23 and chromatin features
The presence of H3K4me1 marks not restricted to a specific tissue/organ, but observed in various structures derived from all three germ layers like the embryonic limb, the olfactory bulbs and small intestine (Figure 5.6), makes it difficult to deduce the putative expression pattern driven by CNE23 *in vivo*. Perhaps, CNE23's role is to boost the activity of tissue-specific elements, as previously shown for some regulatory modules of the sea urchin genes *endo16* and *cyIIIa* (Kirchhamer et al. 1996; Coffman and Davidson 2001; Davidson 2006).

Nevertheless, the combined profile of relatively low peaks of H3K4me1, H3K27ac and DNaseI hypersensitivity suggests that CNE23 might be a genuine regulatory element *in*

vivo, with enhancer-like properties, but it also suggests that in most tissues CNE23 is not in an active state, except perhaps in the adult olfactory bulbs (Figure 5.6). The latter structures are characterised by high H3K4me1 occupancy combined with relatively higher levels of H3K27ac than in the rest of the tissues, and the joint presence of these two marks hints for active state of CNE23 in the olfactory bulbs. Interestingly, while there are no reports of *Ptprm* expression in the brain (Koop et al. 2003), *Lama1*expression is observed in the murine embryonic brain well into the E14.5 stage (Miner et al. 2004). Thus, a *Lama1*-devoted function of CNE23 cannot be ruled out.

It is plausible that in the majority of the tissues, CNE23 is kept in an inactive, or poised state (Calo and Wysocka 2013), which is characterised by an enrichment for H3K4me1, p300 and H3K27me3 occupancy, and the absence H3K27ac marks (Creyghton et al. 2010; Rada-Iglesias et al. 2011) and is observed in enhancers regulating forelimb- versus hindlimb-specific developmental genes, for instance (Cotney et al. 2012). Interestingly, poised enhancers become enriched in H3K27ac marks and depleted in H3K27me3 marks upon cell specification and differentiation (Creyghton et al. 2010; Rada-Iglesias et al. 2011). Thus, one possibility is that CNE23 is in a poised state in most tissues.

Finally, an appropriate *in vivo* transgenesis assay with a CNE23-driven gene reporter would potentially address the question of the transcription-regulatory potential of CNE23.

In this study, I performed transient cell transfections of *CNE::reporter* constructs in cell culture to screen for putative regulatory functions of the conserved elements. The assay's results however, do not convincingly support the hypothesis that the tested CNEs are enhancers or silencers *in vitro*, perhaps except for CNE7, and the speculated causes for these results were discussed. In addition, database analyses of CNEs' chromatin state do not indicate correlation of the CNEs with transcription-regulatory element marks, except for CNEs 7 and 23. However, the lack of such correlation might be due to the fact that the chromatin state data have been obtained mostly from adult tissues and cell cultures, conditions under which an embryonic TRE might be inactive and difficult to predict. Therefore, in order to test the putative regulatory activity of the CNEs in more native environment, I adopted a different strategy, namely the examination of CNE function *in vivo*, in transgenic animals, as described in the next Chapter. Such an approach provides the

additional advantage of informing about the spatio-temporal activity of the candidate elements.

Chapter 6

Functional screening of mouse CNEs in transiently-transgenic zebrafish embryos

6.1. Hypothesis and aims

The analyses of CNE function in cell culture suggested a weak putative silencer activity of CNE7, but failed to convincingly demonstrate enhancer/silencer properties of the rest of the tested CNEs. Also, this assay did not provide information about tissue-specificity of the tested candidate regions. Furthermore, as the optimal activity of many transcription regulatory elements requires the integration of inputs from several signalling pathways (Davidson 2006), which can only be recapitulated *in vivo*, it is possible that my *in vitro* assay may have failed to detect the putative activity of the tested CNEs. To address this, a GFP reporter-based screening of a subset of the mouse CNEs was performed in transient transgenic zebrafish embryos. Such strategy is promising for, despite the lack of mouse/zebrafish conserved elements at the *Lama* locus, this gene features a largely conserved expression pattern in both species (this study; Anderson et al. 2009; Joseph Pickering's Thesis 2012). This suggests that conserved *trans*-acting inputs regulate *Lama1* in the mouse and zebrafish, implying that the murine CNEs may be able to respond to the transcription factor milieu in the fish embryo (Chatterjee et al. 2011; Fisher et al. 2006).

6.2. Introduction

6.2.1. The *Tol2* transposon system in the analysis of TREs *in vivo*

There are several strategies for the generation of transgenic zebrafish expressing fluorescent protein reporter genes. The simplest one relies on the injection of plasmid DNA carrying the reporter construct into fertilized eggs, which leads to high mosaicism in F0 and low frequency of germ line transmission of the transgene - about 5% of injected fish produce transgenic offspring (Stuart et al. 1988; Long et al. 1997). A more efficient method relies on the injection of pseudotyped retroviral vectors at the blastula stage, which results in nearly 100% of the injected fish becoming founders (Lin et al. 1994). However, the handling and modifying of retroviral vectors is laborious making this approach unsuitable for the rapid screening of a large number of candidate regulatory elementsin transiently transgenic fish embryos (Kawakami, 2007). Another highly efficient and relatively simple strategy for zebrafish transgenesis uses the autonomous *Tol2* DNA transposon from the medaka fish (*Oryzias latipes*) (Kawakami et al. 1998).The *Tol2* element is nearly 4.7 kb in length and encodes a transposase consisting of 649 amino acid residues (Kawakami and Shima. 1999;

Kawakami et al. 2000). The enzyme is functional in all tested vertebrate models (zebrafish, *Xenopus*, chicken, mouse) and is able to mobilise non-autonomous *Tol2* elements lacking the transposase gene but retaining the two termini at the ends of the transposon (Kawakami et al. 2000; Kawakami et al. 2004; Sato et al. 2007; Kawakami and Noda 2004). *Tol2* is transposed via a "cut-and-paste" mechanism - the transposase catalyses excision of the element from its original site followed by random integration at a new site of the genome (Kawakami et al. 2000). The minimal 200 bp and 150 bp sequences at the left and right termini flanking the *Tol2* transposase gene, respectively, are critical in this process. These termini contain 12 bp terminal inverted repeats and subterminal regions which are necessary and sufficient for transposition (Urasaki et al. 2006). Thus, any foreign DNA that is cloned in between the minimal sequences can be mobilized and integrated into the host genome, via co-injection of the Tol2-based construct with *Tol2* transposase mRNA into one cell-stage embryos (Kawakami 2007).

This is especially advantageous in examining the activity of candidate regulatory elements driving *GFP* reporter expression in transient assays for it results in early integration of the foreign construct into the host genome, which minimizes transgene mosaicism eventually leading to more consistent pattern of tissue- and/or stage-specific activation of the reporter gene (Fisher et al. 2006). Several studies have demonstrated the efficiency of the Tol2 system for testing putative regulatory elements in zebrafish embryos (Navratilova et a. 2009; Royo et al. 2011; Zelenchuk and Bruses 2011; Ikle et al. 2012; Ritter et al. 2012). For instance, Ikle et al. (2012) used *in silico* approaches followed by Tol2-mediated transgenesis in zebrafish to identify mouse and fish enhancers of *Hand2* that modulate its expression in the ventral pharyngeal arches, while Ritter et al. (2012) employed the Tol2 strategy to investigate the function of candidate TREs in stable zebrafish lines based on phylogenetic comparisons between the human and zebrafish genomes. Interestingly, the combined analyses revealed the existence of tissue-specific enhancers located in the protein-coding regions of their target genes (Ritter et al. 2012). Remarkably, the exonic enhancers were characterized by H3K4me1 enrichment similarly to the typical intergenic non-coding enhancers and are three times more likely to harbor this modification as compared to the rest of the exonic non-enhancer sequence (Ritter et al. 2012).

131

In summary, the use of GFP reporter-based transient analysis aided by Tol2-mediated reporter construct integration in zebrafish embryos is an effective strategy for the rapid screening of mouse candidate transcription regulatory elements.

6.3. Results.

Based on the successful results from the studies described above, I decided to employ transient transgenesis with *mouse CNE::GFP* reporter constructs in the zebrafish embryo to test *in vivo* the performance of a subset of the identified CNEs. I chose to analyse the activity of CNEs 3, 6, 7, 9, 10, 14, 19, 21, 23 and 24 either because they were examined in the *in vitro* tests (CNEs 7, 10, 19, 21, 23 and 24) (Figure 5.3), or because they displayed conservation not only with the opossum sequence but with the chicken as well, and as such are candidates for transcription regulatory elements (CNEs 3, 6, 9 and 14).
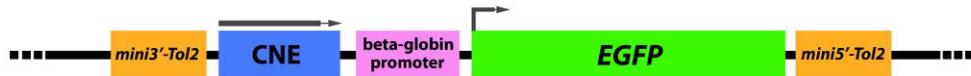
Figure 6.1. *A schematic diagram of a portion of the CNE::EGFP reporter constructs.* The individual functional elements in the reporter construct are depicted as colored rectangular boxes. The arrow above the "CNE" box indicates the orientation of the CNE according to its genomic orientation relative to the *Lama1* transcription start site. The curved arrow at the start of the "*EGFP*" box signifies the direction of transcription. Note that the whole "*CNE::β-globin promoter::EGFP*" unit is flanked by the minimal terminal elements of the Tol2 transposon (*orange boxes*).

Each CNE was individually cloned in its original genomic orientation (relative to the *Lama1* transcription start site) into an *EGFP*-reporter plasmid (McDonald et al. 2010; Yu et al. 2011), immediately upstream of the minimal human β-globin promoter (β-globinP), which is used to drive expression of the *EGFP* reporter gene (Figure 6.1). Importantly, the reporter vector contains the minimal *Tol2* terminal elements flanking the *CNE::EGFP* construct (Figure 6.1), which ensures that the whole *CNE::β-globinP::EGFP* unit can be efficiently integrated into the host embryo genome upon co-injection of the *CNE::EGFP* plasmids with *Tol2* transposase mRNA. Before injection of the *CNE::EGFP* plasmids, I confirmed that the empty *EGFP*-reporter plasmid exhibits no cryptic enhancer activity (Figure 6.2; n=157).
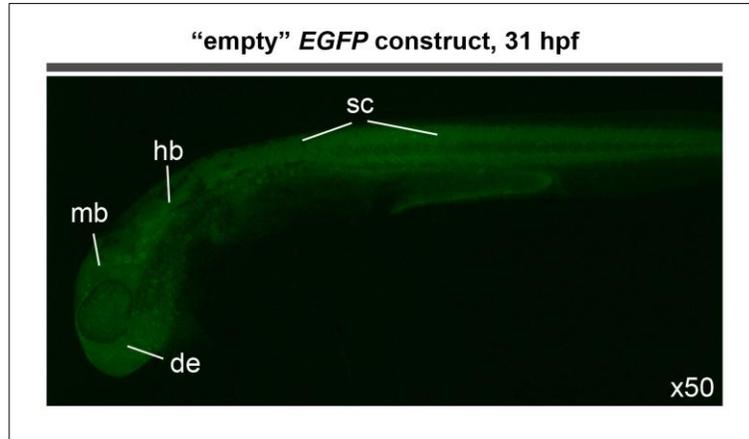
Figure 6.2. *The "empty" EGFP construct does not activate reporter gene expression.* Shown is an embryo at 31 hpf (n=157) that has been injected with "empty" EGFP plasmid, which does not contain any CNE. The embryo shows no EGFP expression, demonstrating the utility of the construct to reliably assess the activity of candidate enhancer elements. Abbreviations: de, diencephalon; hb, hindbrain; mb, midbrain; sc, spinal cord.

Two independent injections in zebrafish embryos at the one-cell stage were performed for each *CNE::EGFP* plasmid (two preps of each were tested), followed by the analysis of EGFP expression in a total of 300 to 500 injected embryos at 6, 24, 31, 48 and 72 hours post fertilization (hpf). The results of the experiments are summarized in Table 6.1 and Figure 6.3. Surprisingly, most of the screened CNEs, except for CNE3, directed weak and highly mosaic reporter expression: the few EGFP-marked cells were not restricted to a single tissue (with very few exceptions, as described in Table 6.1) but were dispersed through the whole body in the positive embryos, in a highly variable (inconsistent) pattern between the embryos. Based on these data, I concluded that CNEs 6, 7, 9, 10, 14, 19, 21, 23 and 24 were unable to function as tissue-specific transcriptional enhancers, at least in the current study (Table 6.1; Figure 6.3). The putative causes behind these results are discussed in Section 6.4.

	CNE3	CNE6	CNE7	CNE9	CNE10	CNE14	CNE19	CNE21	CNE23	CNE24
Total number of analysed embryos[#]	511	396	412	324	370	435	387	369	401	422
Number of EGFP positive embryos*	292	3	1	13	3	6	10	2	8	9
Number of embryos with tissue-specific EGFP expression[¥]	289	0	0	1	1	0	0	0	0	2
Number of EGFP negative embryos	219	393	411	311	367	429	377	367	393	413

Table 6.1. *Numbers of CNE-injected embryos*. # indicates the total number of analysed injected embryos, excluding the dead ones. * indicates that although the constructs with CNEs 6, 7, 9, 10, 14, 19, 21, 23 and 24 produced EGFP-positive embryos, the reporter's expression in these embryos was weak, restricted to few cells in different tissues and was highly variable between embryos. In contrast, CNE3 directed strong tissue-specific EGFP expression in the skeletal muscles in ~ 57% (289) of the analysed embryos (511); only 3 out of the 292 postive CNE3-injected embryos showed weak and highly mosaic EGFP expression (not indicated in the Table). ¥ indicates the number of embryos with tissue-specific EGFP expression. Most of the EGFP[+] CNE3-injected embryos featured tissue-specific reporter expression in the skeletal musculature; only one of the EGFP[+] CNE9-injected embryos showed restricted expression in the gut; one EGFP[+] CNE10-injected embryo had notochord-specific reporter expression; and only two of the EGFP[+] CNE24-injected embryos exhibited specific expression in the heart and skeletal muscles, respectively.
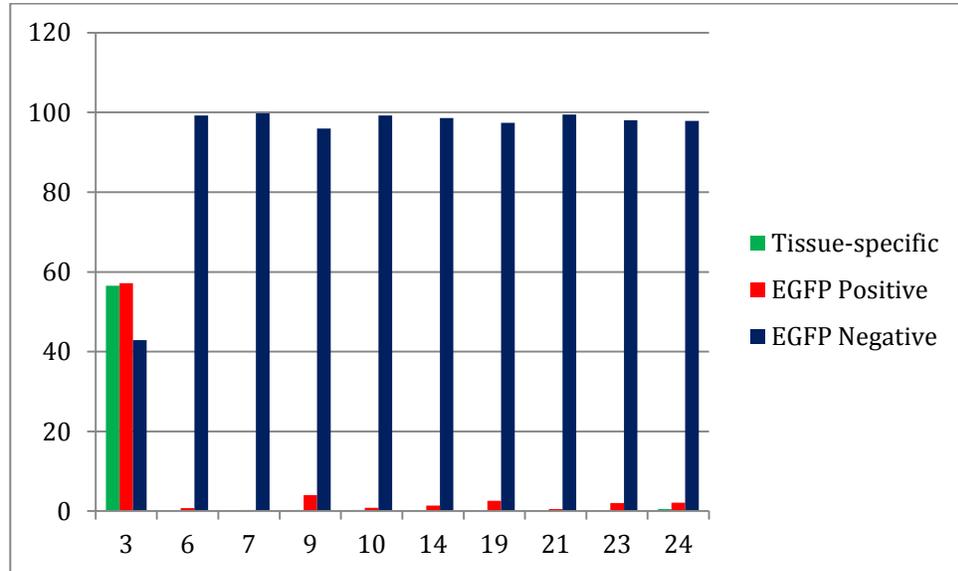
Figure 6.3. *Summary of the analyses of CNEs in transiently transgenic zebrafish*. A graphical representation (in percentages) of the results from Table 6.1. Note that most of the EGFP[+] CNE3-injected embryos feature tissue-specific EGFP expression, while the remaining CNEs directed weak, highly mosaic and variable expression in a few of the injected embryos.

Of all murine CNEs that were tested in the zebrafish embryo, only CNE3 directed strong and consistent tissue-specific expression of the *EGFP* reporter gene, particularly in the skeletal muscle cells (Table 6.1; Figure 6.3; Figure 6.4). At 24 hpf, reporter gene expression driven by CNE3 was weak in the myotome (data not shown). However, a moderate reporter signal was first observed at 24 hpf in the myotome alone but by 72 hpf, strong EGFP expression was observed in both myotomal and cranial skeletal muscles (Figure 6.4).
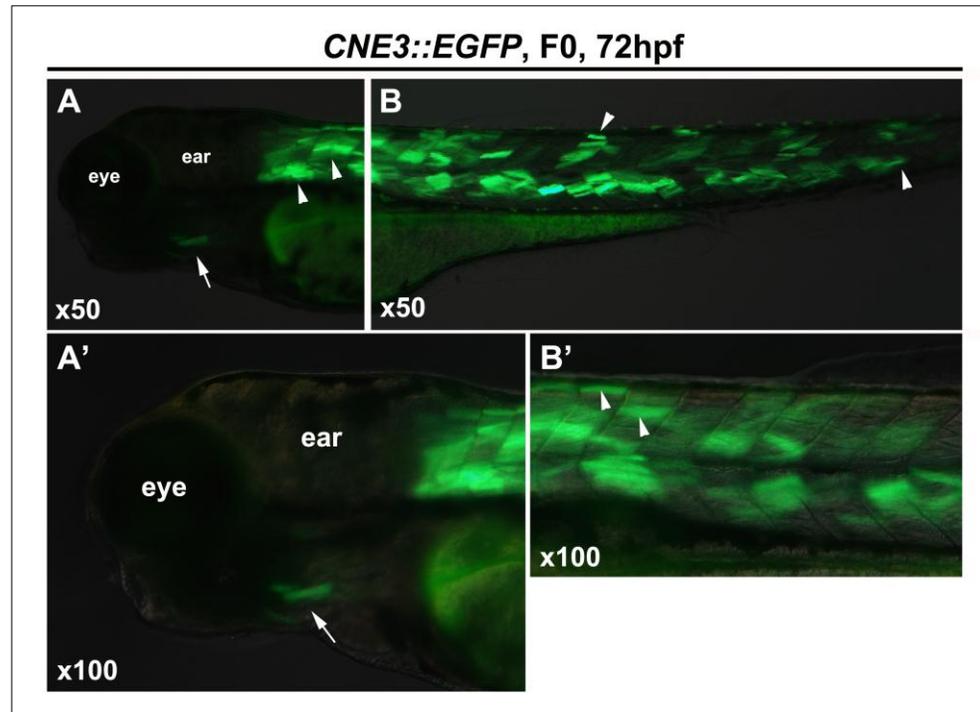
Figure 6.4. *CNE3 directs EGFP expression in the skeletal muscles of zebrafish embryos.* Panels A and B present the pattern of EGFP activity driven by CNE3 in the anterior and posterior halves of the same F0 embryo, respectively, at 60 hpf. A' and B' show magnified views of A and B. EGFP is specifically observed in the myotomal (*white* arrowheads) and cranial (*white* arrows) skeletal muscle cells.

Such muscle-specific activity prompted me to perform further *in silico* analyses of CNE3 for the presence of binding motifs for the myogenic regulatory factors (MRFs), which would explain the *in vivo* activity of CNE3. Remarkably, both rVISTA and MatInspector results showed the presence of overlapping binding motifs containing a conserved pan-amniote E-box with the sequence 5'-CAGCTG-3' for the basic helix-loop-helix myogenic regulatory transcription factors MyoD and Myog (Murre et al. 1989; Chaudhary and Skinner, 1999) (Figure 6.5). However, this MRF binding motif is not significantly over-represented in CNE3 relative to genomic background (Number of matches: 2; Expected: 0.49±0.70; Over-representation: 4.10; Z-score: 1.45; see Legend to Table 4.2 for details on Z-scores).

Intriguingly, detailed examination of CNE3 in the UCSC Genome Browser revealed that CNE3 overlaps with high peaks of MyoD and Myogenin occupancy in the C2C12 mouse myoblast cell line, and with a broad peak of DNaseI hypersensitivity in adult mouse skeletal muscles (Figure 6.6). In contrast, there is no DNaseI hypersensitivity over CNE3 in non-skeletal muscle tissues, like the adult heart, retina and cerebellum, which is in agreement with the absence of EGFP expression in these structures in the transiently-transgenic fish embryos (Figure 6.4; Figure 6.6).
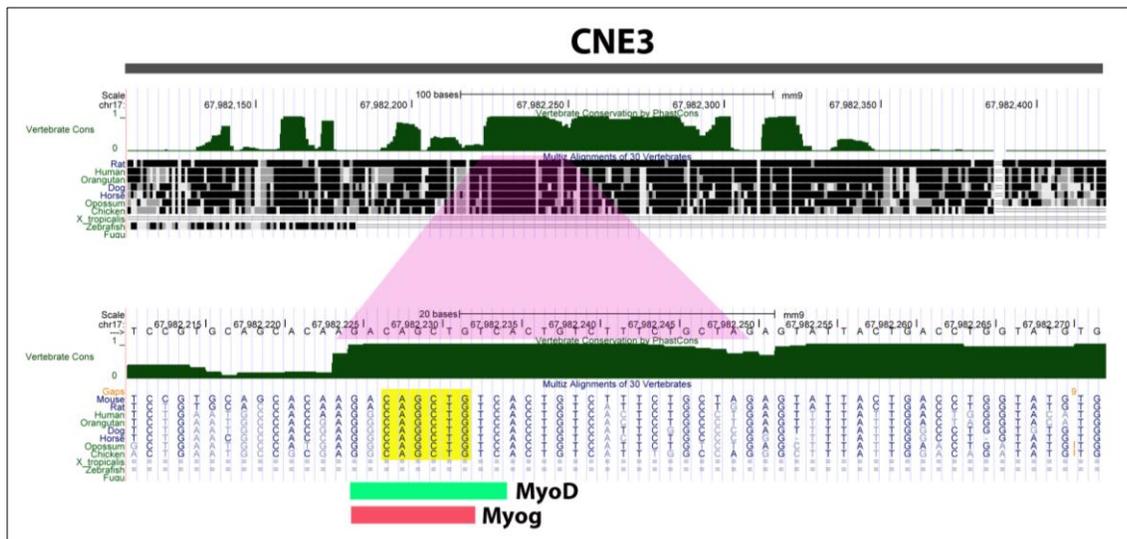
Figure 6.5. *CNE3 harbours a conserved E-box containing binding motif for myogenic regulatory factors (MRFs).* A display of the sequence conservation along CNE3 as generated in the UCSC Browser. The extent of CNE3 is indicated by a thick grey bar at the top of the diagram. The conserved sub-region of CNE3 that contains the E-box and the MRF binding motif is shown magnified at the bottom half of the figure. The conserved "CAGCTG" E-box is highlighted in *yellow* and is derived from sequence analysis of CNE3 using the rVISTA tool (Ovcharenko et al. 2004). Predicted binding sites for MYOD and myogenin are indicated in *green and pink* rectangles, respectively.

6.4. Discussion

The *in vivo* EGFP reporter-based screen in zebrafish of a subset of the CNEs near the murine *Lama1* locus revealed that CNE3 has skeletal muscle-specific enhancer properties (Figure 6.4), while the remaining of the analysed CNEs failed to display strong and consistent reporter gene expression (Table 6.1; Figure 6.3). Nevertheless, one or two of the zebrafish embryos injected with CNEs 9, 10 and 24 displayed tissue-specific *EGFP* expression. The latter observation is most likely unrelated to the CNE but is an enhancer-trapping effect where the β-globin promoter has fallen under the control of a nearby enhancer at the integration site. The lack of enhancer activity of mouse CNEs 6, 7, 9, 10, 14, 19, 21, 23 and

24 in zebrafish embryos raises questions about the putative function of these CNEs in their genomic context in the mouse and about the nature of the transcription factor environment in the host species, as discussed later.
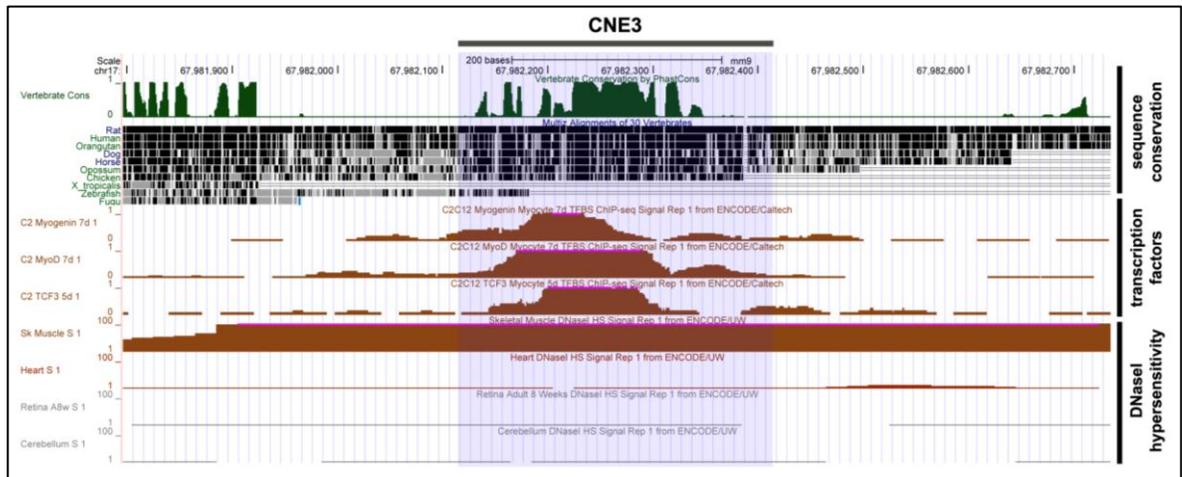
Figure 6.6. Association of CNE3 with myogenic regulatory factors (MRFs) and DNaseI hypersensitive sites. A graphical display of CNE3 sequence conservation, transcription factor occupancy and DNaseI hypersensitivity generated in the UCSC Browser using the NCBI37/mm9 assembly of the mouse genome. Each row (or data track) from top to bottom presents features associated with the sequence at the corresponding genomic position (indicated by the coordinates at the top of the display). The subsets of tracks with thematically similar data (sequence conservation, transcription factors and DNaseI hypersensitivity) are indicated on the right side of the diagram. The extent (in base pairs) of CNE3 is shown by the thick *grey bar* at the top of the display and highlighted in *light blue* along all data tracks.

In contrast to the other tested CNEs, CNE3 specifically up-regulated EGFP expression in both myotomal and cranial skeletal muscle cells in ~ 57% of the injected embryos (Figure 6.3; Figure 6.4). Notably, CNE3 harbours an evolutionary conserved E-box sequence and is occupied by the myogenic regulatory factors MYOD and myogenin in the murine C2C12 myoblast cell line, and also maintains an open chromatin configuration in adult skeletal muscles that is suggestive of transcription-regulatory activity in the CNE3 locus (Figure 6.5; Figure 6.6). Taken together, the results from the *in vivo* and *in silico* analyses suggest that CNE3 may act as a muscle-specific enhancer in the mouse.

6.4.1. Most of the tested CNEs failed to drive EGFP expression in transient transgenesis in the zebrafish

As described earlier, mouse CNEs 6, 7, 9, 10, 14, 19, 21, 23 and 24 failed to drive strong and consistent expression of the *EGFP* reporter in transgenic zebrafish embryos (Table 6.1; Figure 6.3). Several factors could explain these results.

1) One possibility, which was discussed in Chapter 5, is that some regulatory elements are selective in terms of promoter specificity and perform optimally only in conjunction with the endogenous promoter of their target gene, as shown in *Drosophila* (Li and Noll 1994; Merli et al. 1996). Consistent with this explanation, the human β-globin promoter in the EGFP (-) vector is dissimilar to the *Lama1* promoter: the former contains TATA- and CACCC-boxes, whereas the latter is comprised of a CpG-rich sequence (Piccinni et al. 2004). Thus, the tested CNEs might have a preference to the *Lama1*-promoter, instead of the β-globin promoter. To address this issue, the β-globin promoter could be substituted for the basal promoter of the murine *Lama1* gene (Niimi et al. 2003; Piccinni et al. 2004) and new transgenic analyses could be performed in the zebrafish.

2) Another possibility is that some elements cannot function in isolation, that is - when tested individually outside of their genomic context, and require synergistic interactions with additional elements to fine-tune enhancer output. Such mechanism has been demonstrated for the enhancers modulating hypothalamic expression of *shh* in zebrafish (Ertzer et al. 2007), the enhancers controlling the troponin I genes in mice (Guerrero et al. 2010), as well as for the booster-like proximal regulatory region (PRR) that cooperates with the distal regulatory region (DRR) of the murine *MyoD* gene to provide high levels of skeletal muscle-specific expression (Tapscott et al. 1992). Perhaps the most comprehensive way to address this issue is to employ transgenesis with BAC constructs harbouring deletions of individual CNEs.

3) Alternatively, as discussed in Chapter 5, some CNEs may not function as enhancers but as insulators or silencers. For instance, Royo et al. (2011) characterized three insulators from the human *IRXB* genomic cluster in transgenic zebrafish. However, the *EGFP* reporter construct used for the *in vivo* analyses in my study is not adequate to detect such elements as the empty vector generates basal (almost undetected) levels of transcription (Figure 6.2). Instead, an *EGFP* reporter driven by a strong enhancer (like the mid-brain-specific Z48 enhancer of the cardiac actin gene) could be employed for the identification of insulator elements, where the candidate insulator is cloned between the enhancer and the minimal promoter (Bessa et al. 2009).

4) It is also plausible that the tested CNEs, which are not conserved in the zebrafish genome, are clade-specific and can function only in mammalian tissues, implying that the

trans-environment in zebrafish embryos is unable to provide regulatory inputs to these CNEs. This phenomenon is thought to be a consequence of long evolutionary divergence times (as between teleosts and mammals) leading to lineage specific *cis-trans* coevolution, or developmental system drift (Gordon and Ruvinsky 2012; True and Haag 2001). This results in genes, with otherwise conserved expression pattern between two species, being regulated by divergent *cis*-regulatory elements and divergent transcription factor inputs. In effect, an enhancer-swap experiment may lead to failure of reporter expression because the host *trans* environment is unable to properly interpret the donor's *cis* information (Ariza-Cosano et al. 2012). Such *cis-trans* compensatory coevolution has been suggested from studies of the neurogenic ectoderm enhancers (NEEs) in drosophilid flies, where each species – *melanogaster*, *pseudoobscura* and *virilis*, has accumulated adaptive parallel species-specific changes in their neurogenic ectoderm enhancers, most likely in response to changes in the transcription factor milieu of each species (Crocker et al. 2008). For instance, the eve stripe 2 enhancer from *D. yacuba*, *D. erecta* and *D. pseudoobscura* directed *lacZ* reporter expression patterns in transgenic *D. melanogaster* embryos that are identical to the pattern driven by the *D. melanogaster eve* stripe 2 enhancer (Ludwig et al. 1998). However, the orthologs of the eve stripe enhancers from the phylogenetically distant sepsid flies did not precisely recapitulate the drosophilid pattern when tested in *D. melanogaster* (Hare et al. 2008). If this were the case with CNEs 6, 7, 9, 10, 14, 19, 21, 23 and 24, it would be more appropriate to test their activity in transgenic mouse embryos, instead.

5) Finally, some of the CNEs might be active only at later developmental stages that were not covered by the 3 days period of screening. It could be that during the assay period of 72 hours, some CNEs were in an inactive state due to repressive chromatin configuration for instance, which is remodeled at later stages thus allowing the regulatory element to interact with various transcription factors. Such a phenomenon was observed in the regulation of the immunoglobulin J (*IGJ*) gene during the antigen-driven stages of B-cell development, where the IL-2 cytokine induces opening of the enhancer chromatin, which enables STAT5 to bind to its motifs and activate the *IGJ* gene (Kang et al, 1998). Similarly, the ontogenetic progressive activation and repression of the genes in the human β-globin cluster depend on stage-specific regulatory elements in their promoters and the nearby locus control region (LCR) (Levings and Bungert 2002; Orkin 1995). If this is the case with the

tested CNEs, a longer monitoring in transgenic fish might be required to detect any potential enhancer activity.

6.4.2. CNE3 acts as a muscle-specific enhancer

As reported above, CNE3 contains a conserved E-box that is bound by the MYOD and myogenin transcription factors in myoblast culture (Figure 6.5; Figure 6.6). MYOD and myogenin, together with MYF5 and MYF6 (MRF4), are myogenic basic helix-loop-helix (bHLH) transcription factors, well known for their roles in vertebrate skeletal muscle cell commitment and differentiation (Buckingham et al. 2003; Pownall et al. 2002). MYOD is a powerful activator of the myogenic programme as its forced expression is sufficient to convert not only fibroblasts but also liver, pigment and neuronal cultured cells into skeletal muscle (Weintraub et al. 1989), while myogenin functions downstream of *MyoD* and is essential for the activation of skeletal muscle differentiation genes and muscle formation *in vivo*, as shown by the severe reduction of all skeletal muscles and neonatal lethality in *Myog*-deficient mice (Hasty et al. 1993; Nabeshima et al. 1993).

MYOD, as well as the other bHLH myogenic factors, binds to the consensus CANNTG E-box sequence (Chaudhary and Skinner 1999). MYOD strongly activates reporter constructs carrying a pair of E-boxes partly due to inter-protein interactions that stabilise the MYOD-DNA complex (Weintraub et al., 1990), and the nearly ubiquitous E-proteins (like TCF3) have been shown to participate in such oligomerisation with MYOD via the HLH domains (Lassar et al. 1991). It is particularly interesting and relevant to this study that binding sites for non-bHLH proteins like MEIS1, SP1 and MEF2 can substitute for the second E-box, and thus contribute to the stability of the bound MYOD-DNA complex (Tapscott 2005; Knoepfler et al. 1999; Biesiada et al. 1999). Remarkably, CNE3 contains a conserved MEIS1 binding site that partially overlaps the E-box, as analysed by rVISTA, suggesting the possibility that the transcription factor activity of the E-box-bound MYOD/myogenin is augmented by cooperative heterotypic interaction with MEIS1 (Knoepfler et al. 1999). It would be interesting to test whether CNE3's muscle-specific activity requires the conserved E-box and the MEIS1 binding motif by performing reporter analyses using mutant versions of CNE3 with loss-of-function substitutions at these sites.

6.4.3. CNE3 and its putative target gene

The pattern of reporter gene expression driven by CNE3 raises the puzzling question about its target gene in the murine genome. Indeed, mine and previous studies in our laboratory have shown that the major domains of *Lama1* expression in the mouse embryo are the neural tube, the meso- and metanephric kidneys, the presomitic mesoderm and the somitic sclerotome (Chapter 3 of this study; Anderson et al. 2009; Miner et al. 2004). *Lama1* mRNAs are observed throughout the somite in the newly formed 2-3 somites, but in mature somites *Lama1* expression is down-regulated in the dermomyotome and remains expressed at low levels in the sclerotome (Andersone et al. 2009; see also Chapter 3 of this study). Therefore, there is so far no indication of *Lama1*expression associated with skeletal muscles in the embryo, although laminin α1 has been observed associated with the myotube ends in developing intercostal muscles at E11.5 and E15.5 (Patton et al. 1997). However, this study did not provide information on the cell type expressing *Lama1* at the muscle/rib junction, leaving the possibility that laminin α1 may be produced by the costal or tendon cells (both derived largely from the sclerotome (Christ and Scaal 2008)), but accumulates at the surface of the myocytes, putatively interacting with integrin and dystroglycan receptors (Anderson et al. 2009; Bajanca et al. 2006). As for expression in the adult, Falk et al. (1999) and Patton et al. (1997) report that laminin α1 is undetectable in adult skeletal muscles. However, contrary to these reports, recent studies in our laboratory detected laminin α1 expression at the sites of activated satellite cells in injured adult skeletal muscles (Shantisree Rayagiri, unpublished data), raising the possibility that CNE3 may be involved in the transcriptional activation of *Lama1* in regenerating adult muscles.

Alternatively, CNE3's close proximity to the transcription start site of the poorly characterized *Lrrc30* gene (the two loci are just 12 bp away) (Figure 4.3), hints for a possible role of CNE3 in the control of skeletal muscle expression of *Lrrc30*. In relation to such scenario, it is interesting to note that adult human skeletal muscles express *LRRC30* mRNA, as inferred from RNA-seq data in the Illumina Body Map project (GeneCards®), and also the zebrafish *lrrc30* gene is clearly expressed in the myotome (Thisse et al. 2004). The latter observation is particularly intriguing, as it implies that the murine CNE3 may be responsible for the activation/maintenance of a skeletal-muscle specific pattern of *Lrrc30* expression that is conserved across vertebrates. In order to support this view, it would be necessary to

141

examine *Lrrc30* mRNA expression in the mouse and test whether CNE3 up-regulates reporter gene expression in the skeletal muscles of transgenic mouse embryos. The latter experiment is important for previous studies have demonstrated that although conserved with fish, some human enhancers generate different expression patterns when tested in zebrafish and mice, which might indicate the existence of evolutionary divergent *trans*-environments (Ariza-Cosano et al. 2012).

It cannot be excluded that CNE3 might also function as a shared enhancer, driving transcriptional activation from the promoters of both the *Lrrc30* and *Lama1* genes. Such mechanism has been previously reported for the salivary gland-specific expression of the *pig-1* and *sgs-4* genes in *Drosophila*, which share a SEBP1-binding enhancer (Hofmann and Lehmann, 1998), for the co-expression of achaete and scute genes in the proneural clusters of *Drosophila* (Gomez-Skarmeta et al. 1995), and suggested for the synchronic co-expression of the *Myf5* and *Mrf4* genes in the most ventral part of the murine thoracic somites (Carvajal et al. 2001). Performing circular chromosome conformation capture experiments (4C) will facilitate answering the question of CNE3's target genes. Alternatively, one may generate transgenic mouse embryos carrying a BAC clone which spans the *Lrrc30-Lama1* region, including CNE3, where two different reporter genes – *lacZ* and *PALP (human placental alkaline phosphatase* gene), are inserted downstream of the *Lrrc30* and *Lama1* promoters, respectively. Deleting CNE3 from the BAC construct and assaying reporters' expression would enable to determine whether *Lama1* or *Lrrc30*, or both, are regulated by CNE3. Such method has been employed to examine the complex *cis*-regulatory apparatus of the murine *Mrf4-Myf5* locus, for instance (Carvajal et al. 2008)

The results from the *in vivo* analysis of CNEs around the murine *Lama1* locus in transgenic zebrafish revealed that only one of the elements, CNE3, has tissue-specific enhancer properties, while the rest of the sequences failed to drive strong and consistent reporter expression. This poor activity is similar to the results from the *in vitro* luciferase assays and could be due to endogenous promoter requirements, the need for synergistic interactions with other elements, the divergent transcription environment of the host, activity at later developmental stages, or because the CNEs are not transcriptional enhancers but silencers or insulators. Importantly, the outcome of this assay did not succeed in informing about the transcription regulatory elements of *Lama1* that mediate its expression under the

142

influence of SHH. Therefore, another approach is necessary to unravel these elements as described in the next chapter.

Chapter 7

Identification of a neural-specific enhancer in intron 1 of the murine *Lama1* gene

7.1. Hypothesis and Aims

Transgenic analyses in zebrafish of conserved non-coding elements from the vicinity of the murine *Lama1* locus did not reveal any candidate enhancers except for CNE3. Importantly, none of the tested CNEs directed reporter expression in the neural tube or early somites, two sites where *Lama1* transcription requires SHH signals in the mouse embryo (Anderson et al. 2009; this study). Consequently, I hypothesized that the SHH-sensitive enhancers directing *Lama1* transcription in the neural tube and somites might not be conserved. To uncover these elements, I employed an alternative approach that does not depend on sequence conservation but involves the association of genomic sequence with particular chromatin features such as occupied transcription factor binding sites, specific histone modifications and DNaseI hypersensitivity. Here, I report on the analysis of available data from ChIP-based studies, which led me to identify an enhancer element in intron 1 of mouse *Lama1*.

7.2. Introduction

The correlation of DNaseI hypersensitivity and particular histone modifications with active developmental enhancers has been used effectively for identification of the latter, as described in Chapter 1. Therefore, in order to identify potential SHH-regulated enhancers of the murine *Lama1* gene, I carried out an investigation of the available literature focusing on studies of the genomic occupancy of GLI transcription factors.

In search of direct target genes of SHH involved in distal autopod development, Vokes et al. (2008) performed chromatin immunoprecipitation with E11.5 mouse embryo limb buds that expressed conditionally a Flag-tagged version of GLI3, GLI3[Flag]. 5274 GLI3[Flag]-bound genomic regions (GBRs) with mean length of 854 base pairs were identified, followed by demonstration that some of the regions mediated GLI-dependent transcriptional regulation of *Prdm1*, *Gli1*, *Gremlin* and *Hand2* in the limb buds of transgenic mouse embryos (Vokes et al. 2008). Surprisingly, 16 of the GBRs corresponded to a subset of 25 GLI1-occupied neural-specific enhancers that have been described earlier in neuralised embryoid bodies by the same team (Vokes et al. 2007). Interestingly, these enhancers were non-functional in the limb buds as they were enriched for the repressive chromatin mark H3K27me3, and none of their target genes - *Nkx2.1*, *Nkx2.2* or *Foxa2*, were expressed in the limb buds of wild type or *Gli3[-/-]* embryos. This led the authors to conclude that although GLI3 is not involved in silencing of these SHH-dependent neural enhancers in the limbs, it

was nonetheless able to gain access to the GLI motifs in their DNA sequence (Vokes et al. 2008), suggesting that GLI activator and GLI repressor forms, as well as different GLI factors, exhibit similar binding specificities *in vivo*.

7.3. Results

7.3.1. Identification of a GLI-bound region in the 1st intron of the murine *Lama1* gene

Encouraged by the observation of GLI3^{Flag}'s ability to bind otherwise inactive neural enhancers in the developing limbs, I screened the full list of 5274 GLI3^{Flag}-bound genomic regions that was provided in the Supplementary Data Set 1 document by Vokes et al. (2008). I identified a 907 base pair peak with Rank No. 1926 located within intron 1 of the murine *Lama1* locus, approximately 200 bp downstream of the 3'-end of exon 1, and containing not one but three GLI binding motifs (Figure 7.1) (SuppDataSet1, Vokes et al. 2008). Detailed sequence analyses with MatInspector (Cartharius et al. 2005) revealed the presence of another two putative GLI-binding sites within this 907 bp region (Figure 7.1). The latter motifs however were not reported in Vokes's study. Thus, this intronic region harbours five GLI binding motifs in total, where motifs 1, 2 and 5 were annotated by Vokes et al. (2008), while motifs 3 and 4 were found in this study.

Based on the fact that a ~1 kb region within the 1st intron of *Lama1* contains five GLI binding motifs, some or all which are occupied by GLI3 *in vivo*, I hypothesized that this region might be a functional transcription-regulatory element, and even more, that it could be a promising candidate for mediating the effects of SHH on *Lama1* transcription in the mouse somites and/or neural tube. Thus, I performed PCR with mouse genomic DNA to isolate a 1038 bp amplicon (which I named "a1-NSE", for reasons explained below) containing the 907 bp GLI-bound element, followed by cloning of a1-NSE into the same Tol2-based *EGFP*-reporter vector used for the functional screening of CNEs described in the previous chapter (Figure 7.2).

146
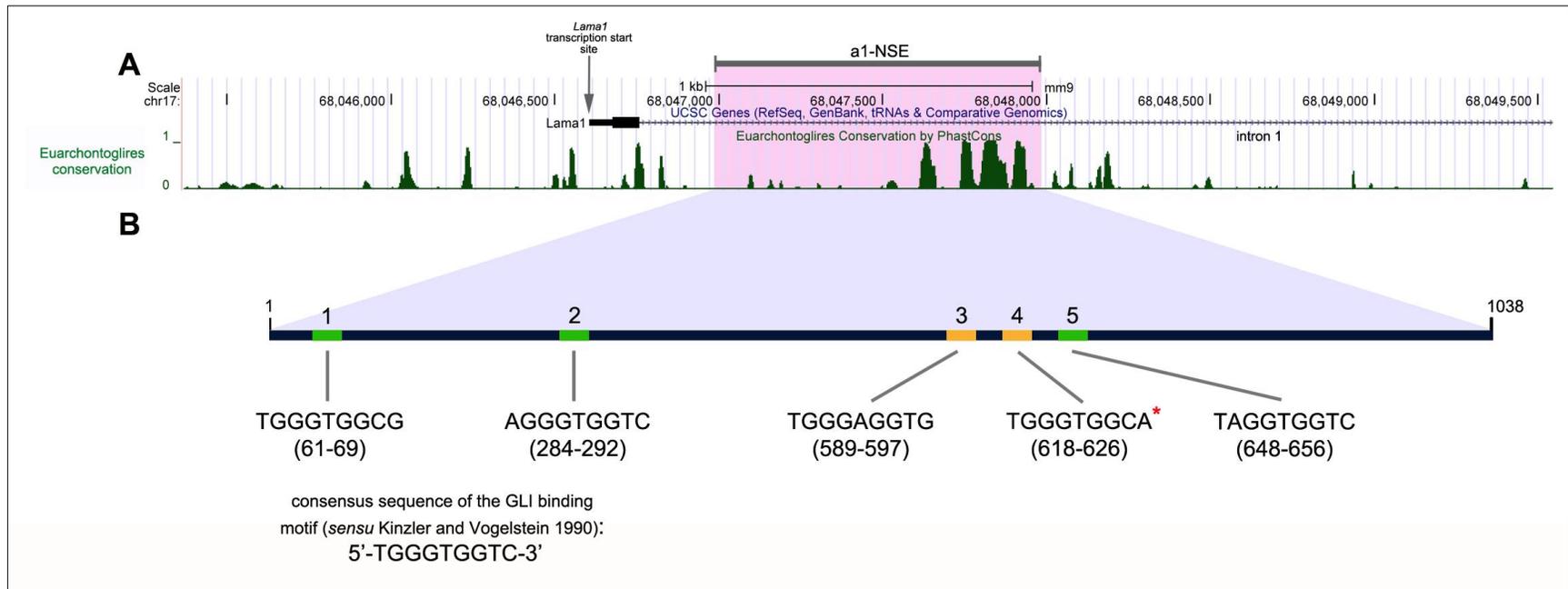
Figure 7.1. *Identification of a GLI-occupied region in intron 1 of the murine Lama1 gene.* Panel A displays the 5' end of the *Lama1* gene with its first exon and part of the 19.3 kb-long intron 1 harbouring the a1-NSE element, as analysed in the UCSC Genome Browser. Below the locus is presented a phastCons plot of sequence conservation over the whole genomic region. Panel B is a schematic representation of a1-NSE indicating the relative positions of individual GLI binding motifs (1-5) and their sequence. Motifs 1, 2 and 5 were annotated by Vokes et al. (2008) and are indicated in *dark green*, while motifs 3 and 4 uncovered in this study are highlighted in *orange*. * GLI motif 4 is on the opposite strand relative to the other four GLI motifs. The GLI binding motif consensus sequence is indicated below.

147

Figure 7.2. *A schematic representation of the a1-NSE::EGFP reporter construct.* a1-NSE (*blue box*) is cloned upstream of the human *β-globin promoter:EGFP* reporter cassette, in the same orientation as the direction of transcription in the *Lama1* locus (indicated by a *straight grey arrow* above the blue box). A *curved grey arrow* at the beginning of the *EGFP* gene indicates the direction of *EGFP* transcription. *mini3'-* and *mini5'-Tol2* elements (*orange boxes*) flank the *a1-NSE::EGFP* construct and enable its integration in the zebrafish genome by Tol2 transposase.

7.3.2. Activity of a1-NSE in transgenic zebrafish

I microinjected 30 ng/µL *a1-NSE::EGFP* construct, together with 20 ng/µL *Tol2* mRNA, in 1-cell stage wild type zebrafish embryos and examined its activity in transient transgenic zebrafish (F0) at different developmental time-points. In total, 413 injected embryos were analysed, 259 of which were EGFP-positive. The majority, 254, exhibited consistent expression restricted to the ventral neural tube only (Figure 7.3), whereas the remaining of the EGFP[+] embryos had weak and highly mosaic expression in varios organs, including the neural tube (Table 7.1). Those F0 fish that showed consistent tissue-specific expression of EGFP in the ventral neural tube (n= 254) were selected and left to develop until sexual maturity, when 40 of them were screened in paired out-crosses with wild type fish for germ-line transmission of the *a1-NSE::EGFP* construct. Only 4 of the screened adults (10%) generated EGFP[+] F1 clutches, in which the frequency of reporter-expressing embryos was 25-35%, depending on the clutch. Importantly however, all of the EGFP[+] F1 embryos displayed reporter expression restricted to the ventral neural tube; no expression was observed in any other organ.

Here, I report on the enhancer activity of a1-NSE as observed in F1 transgenic zebrafish for the reporter pattern in F1 embryos (Figures 7.4 and 7.5) was highly similar to that in F0 transient transgenic embryos (Figure 7.3), although less mosaic and less intensive, which is usually encountered in enhancer transgenic experiments in zebrafish (compare Figures 7.3 and 7.4).
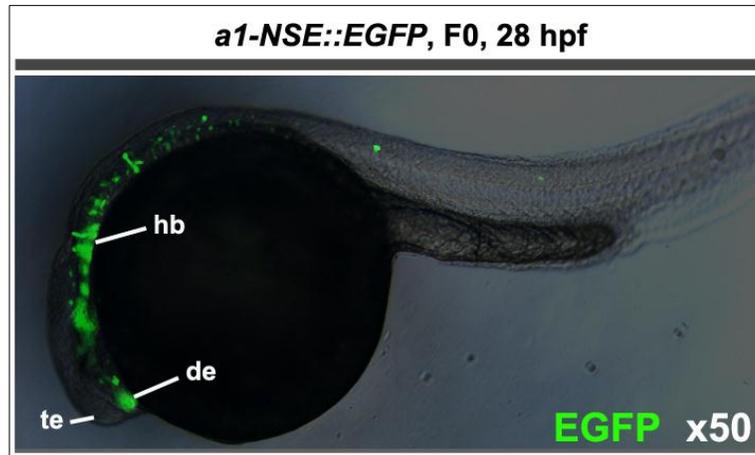
Figure 7.3. *a1-NSE directs EGFP expression in the neural tube of F0 transgenic zebrafish embryos.* Abbreviations: de, diencephalon; hb, hindbrain; te, telencephalon.

Total number of analysed injected embryos	413
Total number of EGFP+ embryos	259
Number of embryos with weak and variable EGFP expression	5
Number of embryos with strong EGFP signal in the NT	254
Number of embryos with strong EGFP signal in the ventral NT	254
Number of injected fish grown to adulthood	230
Number of F0 screened	40
Number of F0 founders	4
Total number of EGFP+ embryos in F1	112
Frequency of EGFP+ embryos in F1	25 – 35%

Table 7.1. *Quantitative data from the Tol2 transposase-mediated transgenesis of the a1-NSE::EGFP construct in zebrafish.* The "Total number of analysed injected embryos" excludes the dead ones. "Strong EGFP signal in NT" indicates high number of cells expressing high levels of EGFP in the neural tube (NT), as contrasted to the five EGFP+ embryos featuring only weak reporter expression in a small number of cells in multiple locations in the body.

In F1 embryos, faint EGFP signal was detected at ~ 24 hpf only in the anterior neural tube, but by 31 hpf reporter expression has intensified and was observed only in the neural tube, spanning from the rostral tip of the diencephalon anteriorly and continuing posteriorly into a narrow longitudinal stripe of EGFP+ cells in the spinal cord (Figure 7.4). Curiously, the EGFP signal was restricted to the ventral half of the neural tube and was absent from the telencephalon. Transverse cross sections of the brain revealed that EGFP was expressed by a

dorso-ventral patch of cells that was merely 2-3 cells thick (Figure 7.4). Later on, at 52 hpf, EGFP expression expanded dorsally consistent with the dorso-ventral growth of the brain and spinal cord but was, interestingly, restricted to the ventricular zone of the neural tube (Figure 7.5).
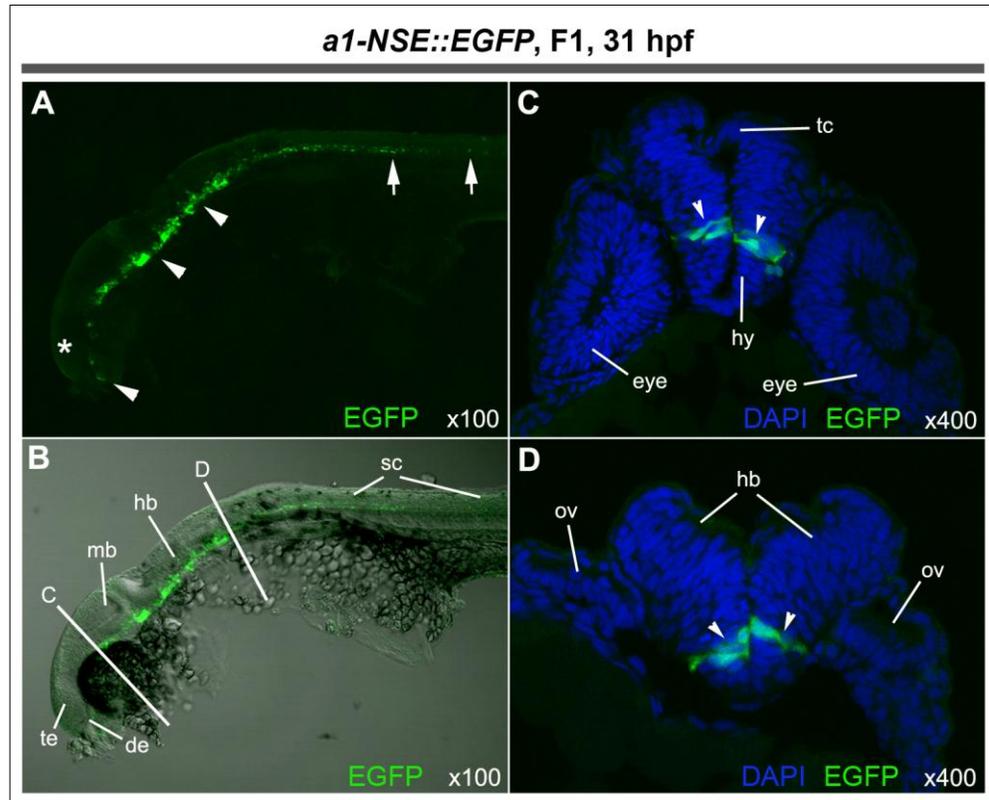
Figure 7.4. EGFP expression in the neural tube of 31 hpf F1 a1-NSE::EGFP-transgenic zebrafish embryos (n=63). (A, B) whole-mount images, (C, D) transverse sections through the optic (C) and otic (D) levels of the brain as shown in (B). *White arrowheads* in (A, C, D) indicate the expression of EGFP in the ventral diencephalon, midbrain and hindbrain; *white arrows* in (A) show the continuation of EGFP expression in the ventral spinal cord. *White asterisk* in (A) highlights the absence of EGFP signal in the telencephalon Abbreviations: de, diencephalon; hb, hindbrain; hy, hypothalamus; mb, midbrain; ov, otic vesicle; sc, spinal cord; te, telencephalon; tc, tectum.

Thus, the mouse a1-NSE non-coding element behaves as a neural-specific enhancer when tested in a heterologous system – the zebrafish embryo. Hence, a1-NSE received its current name for its position in intron 1 of *Lama1* (**a1-**) and the fact that it harbours **N**eural-**S**pecific **E**nhancer properties.

Figure 7.5. *EGFP expression in the neural tube of 52 hpf F1 a1-NSE::EGFP-transgenic zebrafish embryos (n=49).* (A, B) whole-mount images, (C, D) transverse sections through the optic (C) and posterior hindbrain (D) levels of the brain as shown in (B). *White arrowheads in (A, C, D) indicate the expression of EGFP in the ventricular zone of the brain; white arrows in (A) show the continuation of EGFP expression in the ventral spinal cord. White asterisk in (A) highlights the absence of EGFP signal in the telencephalon. Abbreviations: hb, hindbrain; hy, hypothalamus; sc, spinal cord; te, telencephalon; tg, tegmentum; tc, tectum.*
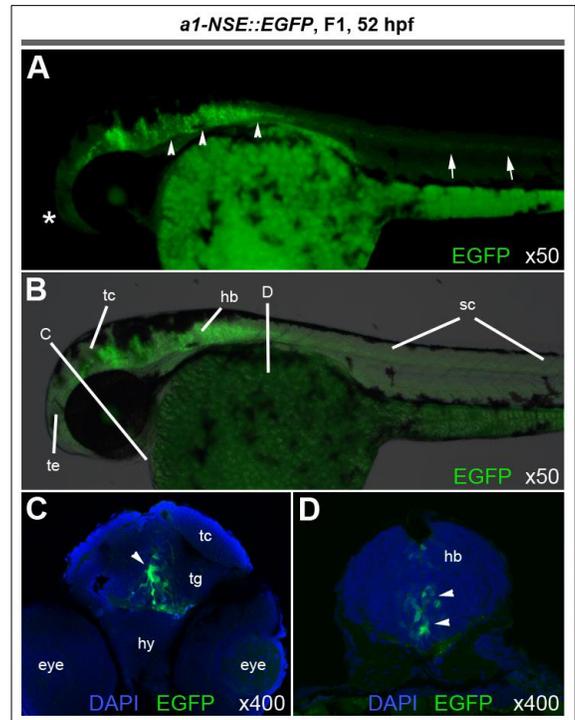
7.3.3. Activity of a1-NSE in transgenic mouse embryos

The activity of the murine a1-NSE element in zebrafish embryos is reminiscent of *Lama1* mRNA expression in the mouse neural tube (see Figures 3.1 and 3.2), leading me to examine a1-NSE function in its endogenous environment in the mouse embryo. For this, I cloned a1-NSE upstream of the *lacZ* reporter gene driven by the minimal human β-globin promoter (Figure 7.6). The *lacZ* cassette encodes a β-galactosidase (β-Gal) reporter with nuclear localization signal enabling more precise labeling of *lacZ*-expresing cell, while the human β-globin promoter has lower basal transcriptional activity compared to alternative promoters
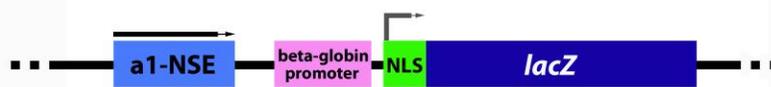
Figure 7.6. *A schematic representation of the a1-NSE::lacZ reporter construct.* a1-NSE (*blue* box) is cloned upstream of the human *β-globin promoter:lacZ* reporter cassette, in the same orientation as the direction of transcription in the *Lama1* locus (indicated by a straight *grey* arrow above the blue box). A curved *grey* arrow at the beginning of the *lacZ* gene indicates the direction of *lacZ* transcription. β-galactosidase is targeted to the nucleus as a result of a nuclear localisation signal sequence at the 5'-end of *lacZ* (*light green* box).

(like the thymidine kinase promoter), thus reducing the probability of stochastic non-specific reporter expression (Coy et al. 2011; Yee and Rigby 1993).

The linearized *a1-NSE::lacZ* construct was introduced into mouse oocytes via pro-nuclear microinjection, performed by our collaborators in IMCB Singapore, according to

Figure 7.7. *β-Gal expression in the neural tube of E9.0 (n=1) and E9.25 (n=3) a1-NSE::lacZ-transgenic mouse embryos.* (A) whole-mount lateral view of the E9.0 transgenic embryo. (B) magnified lateral view of the head region of the same embryo indicating the presence of β-Gal in the CNS but not the 1st pharyngeal arch and heart. (C) dorsal view of the same embryo as in (A) showing reporter expression in the neural tube but not in somites. (D, E, F) whole-mount lateral views of E9.25 embryos No. 1, 2 and 3, respectively, showing consistent β-Gal expression in the diencephalon, mesencephalon, hindbrain and neural tube, but not in the telencephalon. *Black* arrow in (B, D, E and F) indicates weaker β-Gal expression at the midbrain/hindbrain boundary. Abbreviations: de, diencephalon; hb, hindbrain; hrt, heart; mes, mesencephalon; nt, neural tube; pa1, 1st pharyngeal arch; psm, presomitic mesoderm; som, somites; spc, spinal cord; te, telencephalon.
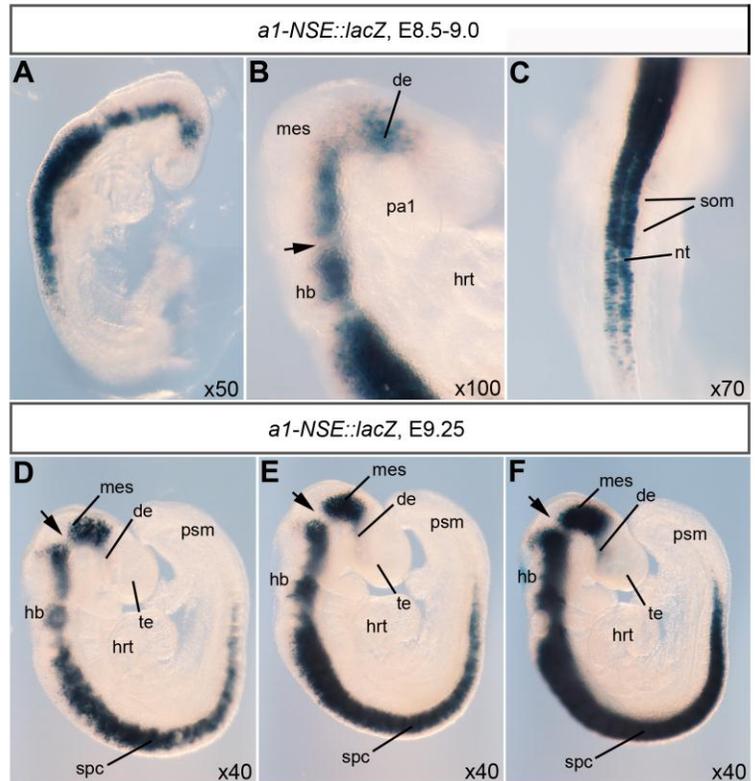
standard mouse transgenesis procedures (Brown and Corbin 2002), and the resultant transgenic embryos were analysed at several developmental stages. The earliest transgenic embryo obtained was at stage E8.5-E9.0 (n=1). Remarkably, a1-NSE drove β-Gal expression solely in the neural tube, as in transgenic zebrafish (Figure 7.7A-C). Similar expression was observed at E9.25 (n=3) (Figure 7.7D-F), and at E9.5 (n=1) (Figure 7.8). At E9.5, the transgenic embryo displayed β-Gal staining in the diencephalon, mesencephalon, hindbrain and along the neural tube, but absent from the telencephalon, the caudal-most end of the tube and curiously, from the isthmic region between mesencephalon and hindbrain. Moreover, somites, presomitic mesoderm, neprhric tissues and head mesenchyme, which all express *Lama1*, lacked β-Gal staining (Figure 7.8), further corroborating the neural specificity of a1-NSE.

152

Figure 7.8. *β-Gal expression in the neural tube of E9.5 (n=1) a1-NSE::lacZ-transgenic mouse embryo.*
(A-G) whole-mount images. (H) and (I) are near-horizontal sections of the head, while (J) is a transverse section at the forelimb level. (A) a lateral view of the whole embryo; (B) magnified view of the head region showing β-Gal expression in the diencephalon, mesencephalon and hindbrain but not in the telencephalon and the pharyngeal arches. (C) frontal view of the head indicating the characteristic ventro-lateral stripes (*orange* arrows) of β-Gal signal in the diencephalon. (D) dorsal view of the head showing absence of reporter expression in the floor-plate and at the midbrain/hindbrain boundary (*black* arrows). (E) dorsal view of the trunk region revealing that β-Gal is confined to the neural tube and is absent from somites. (F) dorsal view of the caudal end of the embryo, showing similar pattern as in (E). (G) magnified lateral view of the caudal region showing the posterior limit of β-Gal expression and absence of the latter from the paraxial mesoderm; the newly-formed somite (sI) is outlined.
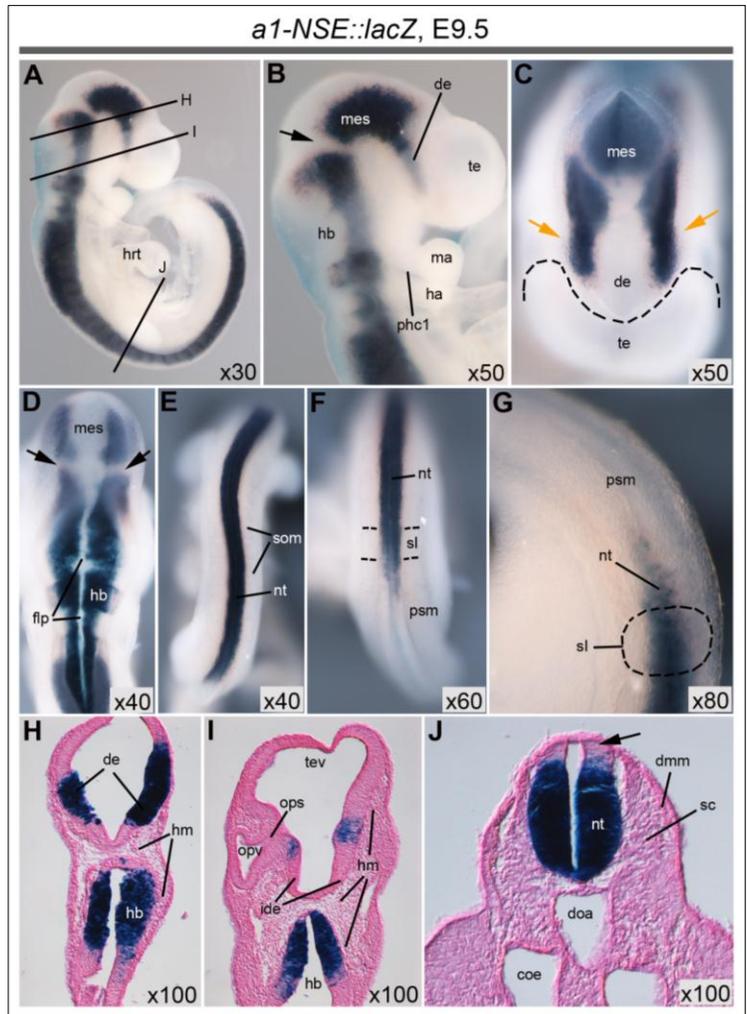
Sections (H) and (I) reveal that reporter expression is indeed restricted to CNS as head mesenchyme is devoid of signal. (J) β-Gal expression is absent from the dorsal neural tube (indicated by a *black* arrow), and the somites. Abbreviations: coe, coelom; doa, dorsal aorta; de, diencephalon; dmm, dermomyotome; flp, floor-plate; ha, hyoid arch; hb, hindbrain; hm, head mesenchyme; hrt, heart; ma, mandibular arch; mes, mesencephalon; nt, neural tube; ops, optic stalk; opv, optic vesicle; psm, presomitic mesoderm; sc, sclerotome; sI, newly formed somite; som, somites; te, telencephalon; tev, telencephalic vesicle.

At E10.5 (n=1), the activity of a1-NSE in the neural tube remained similar to that observed at earlier stages, with strong β-Gal expression in all regions of the CNS except the telencephalon, dorsal isthmus and caudal neural tube (Figure 7.9). However, I observed additional expression in several non-neural domains like the heart, sclerotomes, the 1st pharyngeal pouch, in the limb bud mesenchyme and cloacal area (Figure 7.9B, E, H), raising the possibility that reporter expression in these domains might also be under the control of a1-NSE. However, examination of later stage transgenic mouse embryos, revealed that this is

unlikely to be the case as these embryos featured β-Gal expression mainly in neural structures, but not in the heart or in any somitic derivatives (Figure 7.10). Thus, an E12.5 embryo (n=1) exhibited a weaker but nevertheless CNS-restricted β-Gal staining in all regions of the brain and spinal cord, except for the telencephalon and caudal-most end of the neural tube, respectively (Figure 7.10A-C).

Figure 7.9. *β-Gal expression in the neural tube of E10.5 (n=1) a1-NSE::lacZ-transgenic mouse embryo.* (A-H) whole-mount images. (I-K) sections as indicated in (A). (A) a side view of the whole embryo, indicating the consistent pattern of β-Gal expression in the central nervous system similar to that observed at earlier stages (compare with Figure 7.7 and 7.6). (B) frontal view of the whole embryo. (C) magnifiedlateral view of the head region. β-Gal is expressed in the diencephalon, mesencephalon and hindbrain, but not in the telencephalon. Intriguingly, the first pharyngeal cleft is also positive. (D) dorsal view of the head showing the absence of reporter expression in the floor-plate. (E) side view of the inter-limb region showing expression in the neural tube and in the sclerotome (black arrowheads). (F) dorsal view of the inter-limb region. (G) Dorsal view of the tailbud region; the location of the newly formed somite (sI) is outlined. (H) side view of the tailbud region, showing expression in the ectoderm of the cloacal area. In panels (E, F, I and J) anterior is to the top. (I) horizontal section through the head revealing β-Gal expression in the neural tube but not in the head mesenchyme or otic vesicles. (J) horizontal section through the pharyngeal region showing β-Gal expression in the

endoderm of the first pharyngeal pouch (*green* arrows). (K) transverse section at the forelimb level. β-Gal staining is present in the neural tube where it appears to be reduced in the mantle layer (*white* asterisk), and is absent from the floor-plate and the dorsal neural tube (*black* arrow); the vertebral mesenchyme (derived from the sclerotome) below the neural tube also expresses β-Gal, while the rest of the somitic derivatives lack reporter expression. Abbreviations: cla, cloacal area; doa, dorsal aorta; de, diencephalon; flb, forelimb bud; flp, floor-plate; ha, hyoid arch; hb, hindbrain; hm, head mesenchyme; hrt, heart; ma, mandibular arch; mes, mesencephalon; mtm, myotome; nt, neural tube;ov, otic vesicles;phc1, 1st pharyngeal cleft; php1, 1st pharyngeal pouch; psm, presomitic mesoderm; sI, newly formed somite; som, somites; te, telencephalon; vrtm, vertebral mesenchyme.
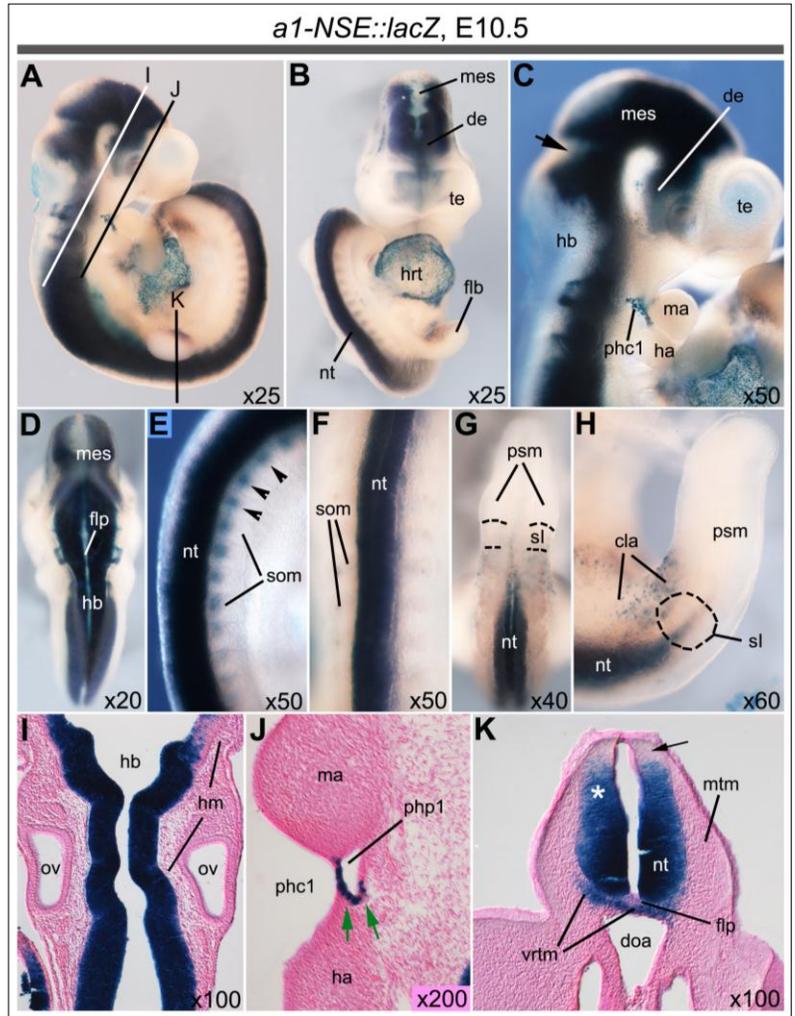
Figure 7.10. *β-Gal expression in the neural tube of E12.5 (n=1) and E13.5 a1-NSE::lacZ-transgenic mouse embryos.* (A-D) whole-mount images and transverse section of the E12.5 embryo.(A) lateral, (B) frontal and (C) dorsal views of the whole embryo, showing β-Gal expression restricted to the post-telencephalic central nervous system. (D) transverse section at the forelimb level revealing reporter expression restricted to the ventricular zone of the ventral half of the spinal cord (*black* arrow). (E-J) and (K-R) are image sets from E13.5 embryos No.1 and 2, respectively. (E, K) lateral, (F, L) frontal and (G, M) dorsal views of the embryos. (H) frontal section through the head of embryo No. 1 as indicated in (E); β-Gal is expressed at multiple sites in the telencephalon, eyes and jaws.

(N) frontal section through the head of embryo No. 2 revealing reporter expression in the telencephalon that is complementary to the pattern observed in embryo No. 1 (H). (I) and (J) are transverse sections of the trunk of embryo No. 1 at forelimb and hindlimb levels, respectively. β-Gal is expressed not only in the ventricular zone, but also in the mantle layer of the spinal cord; in addition, the dorsal root ganglia are also β-Gal-positive. (O) and (P) are transverse sections of the trunk of embryo No. 2 at forelimb and interlimb levels, respectively. Similarly to embryo No. 1 (I, J), β-Gal is present in both the ventricular zone and in the mantle layer of the spinal cord. However, in contrast to embryo No. 1, there is no reporter expression in dorsal root ganglia. (Q) and (R), proximal and distal, respectively, transverse sections of embryo No. 2 at the level of the tail. Abbreviations: amg, amygdala; ctx, neocortex, drg, dorsal root ganglia; epl, ependimal layer of spinal cord; hi, hippocampus; lge, lateral ganglionic eminence; lns, lens;mes, mesencephalon; meo, medulla oblongata;mntl, mantle layer of spinal cord; mrgl, marginal layer of spinal cord; poa, postoptic area of hypothalamus; rtn, retina; spc, spinal cord; te, telencephalon; th, thalamus;tng,tongue;vrtc,vertebral-condensation.

155
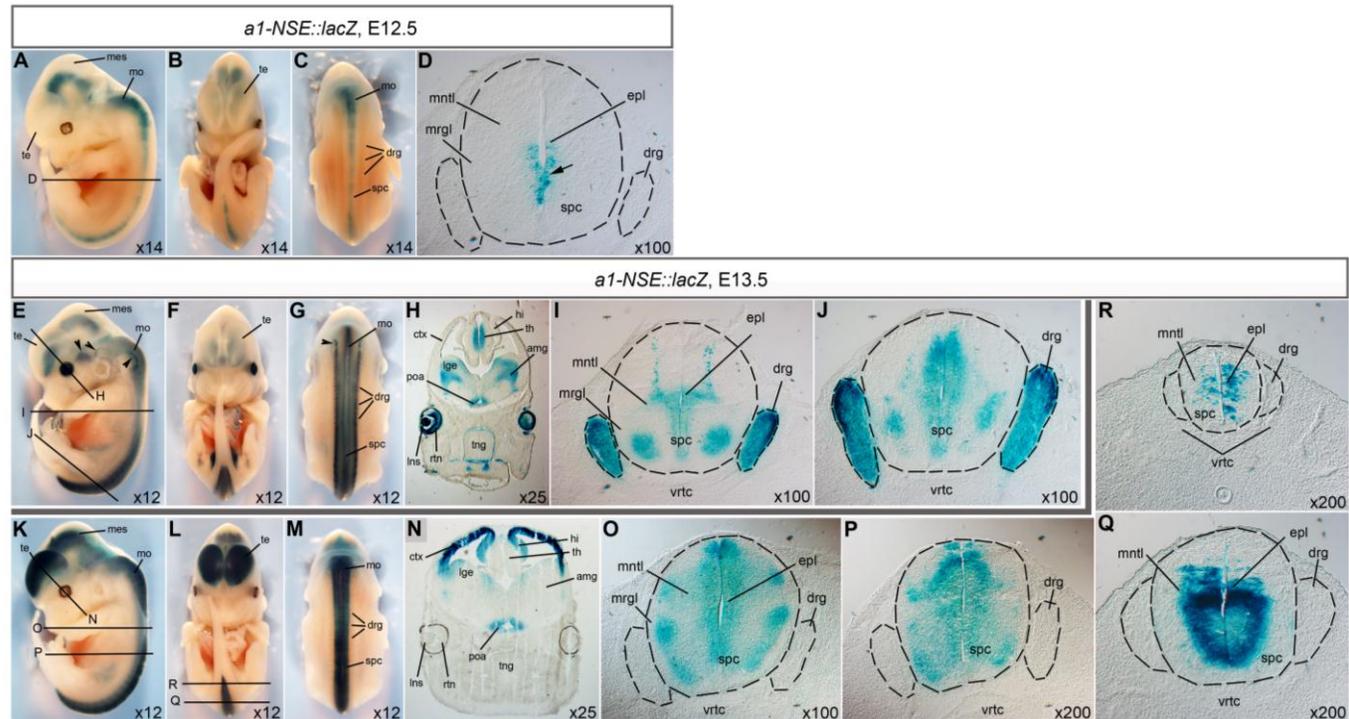
Interestingly, the β-Gal signal in the spinal cord of the E12.5 embryo was limited to the ventricular zone (Figure 7.10D), reminiscent to a1-NSE's activity in transgenic fish and to the endogenous *Lama1* mRNA expression pattern. Similarly, E13.5 embryos (the latest stage examined; n=2) displayed strong expression in all regions of the CNS including, surprisingly, parts of the telencephalon (Figure 7.10E-G and K-M). However, the telencephalic β-Gal staining was not identical but rather complementary in the two littermate embryos − one of the transgenics showed staining in the neocortex and hippocampus but lacked β-Gal in the amygdala (Figure 7.10N), while the opposite pattern was observed in the second embryo (Figure 7.10H), which also showed heavy β-Gal expression in the dorsal root ganglia, cranial nerves, retina and the lens (Figure 7.10G, H, I). Otherwise, β-Gal staining in the post-telencephalic CNS was largely consistent between the two E13.5 embryos (Figure 7.10I-J and O-R). This suggests that β-Gal expression observed in the telencephalon and non-CNS structures in the two E13.5 and the single E10.5 transgenic embryos is due to transgene-integration-site effects, as it is common in such types of assays in the mouse. However, one domain of β-Gal expression was consistent among all nine transgenic embryos − that of the post-telencephalic CNS, and it was highly similar to the pattern of *Lama1* mRNA expression in wild type mouse embryos.

In summary, a1-NSE behaves as a tissue-specific enhancer directing reporter gene expression specifically in the central nervous system of transgenic mouse and zebrafish embryos. The results of these functional assays combined with the presence of occupied GLI binding motifs, suggest that a1-NSE may operate as a CNS-specific enhancer of the murine *Lama1* gene *in vivo*, and that perhaps its activity is modulated by SHH.

7.3.4. *in silico* analyses of a1-NSE

To gain additional knowledge about the mechanisms underlying a1-NSE's tissue-specific transcriptional control, I performed detailed *in silico* sequence analyses using MatInspector, the VISTA Genome Browser and the Mouse Genome Informatics database. MatInspector revealed the presence of 240 transcription factor binding motifs within a1-NSE. Interestingly, in addition to five GLI motifs described above, a1- NSE harbours putative binding sites for several other transcription factors with prominent roles in neural development like the bHLH

proteins MASH1 and neurogenin, the HMG-box-containing SOX3, SOX6 and SOX9 factors, the POU-domain transcriptional activators, the homeodomain protein LHX3 and the Zn-finger transcriptional repressor REST (also known as NRSF), among many others (Figure 7.11; Table 7.2).

Motif	Number of matches	Expected ± SD	Over - representation	Z-score
GLI1/2/3	5	1.05 ± 1.03	4.74	3.36
FOXH1	4	1.95 ± 1.4	2.05	1.11
FOXJ1	4	8.38 ± 2.88	0.48	-1.69
HMX1/2	5	9.5 ± 3.07	0.53	-1.63
LEF1	2	3.46 ± 1.86	0.58	-1.05
LHX3	1	7.07 ± 2.65	0.14	-2.48
MASH1/NEUROG	2	1.38 ± 1.17	1.45	0.1
PAX3	2	0.94 ± 0.97	2.13	0.58
PAX6	2	2.09 ± 1.44	0.96	-0.41
POU2F1/3F3	4	7.59 ± 2.74	0.53	-1.49
POU3F2/4F1	4	7.89 ± 2.8	0.51	-1.57
POU6F1	3	4.73 ± 2.17	0.63	-1.03
REST	1	0.69 ± 0.83	1.45	-0.23
RXRA/B/G	5	3.16 ± 1.78	1.58	0.75
SOX3/6/9	15	8.65 ± 2.93	1.73	2

Table 7.2. *Binding motifs within a1-NSE of transcription factors involved in CNS development.* The current table lists the binding motifs of transcription factors with roles in CNS development, together with relevant statistics information, as obtained using the "Over-represented TFBSs" tool from Genomatix. Note that most motifs are presented more than once within the a1-NSE sequence (e.g. GLI), and that in some of these cases more than one paralog from a particular TF family is considered by Genomatix in the analysis (e.g. GLI, RXR, POU factors, etc). The "Expected ± SD" column displays the expected number of a given motif matches in an equally sized random genomic region, together with the standard deviation (SD). The "Over-representation" column indicates the fold factor of match numbers in the analysed sequence compared to an equally sized genomic region from the background, or found versus expected matches number. The "Z-score" is a measure of the statistical significance of the over-representation ratio; a Z-score above 2 or below -2 is considered statistically significant, corresponding to a p-value of ~ 0.05 (Genomatix; Sue et al. 2005).
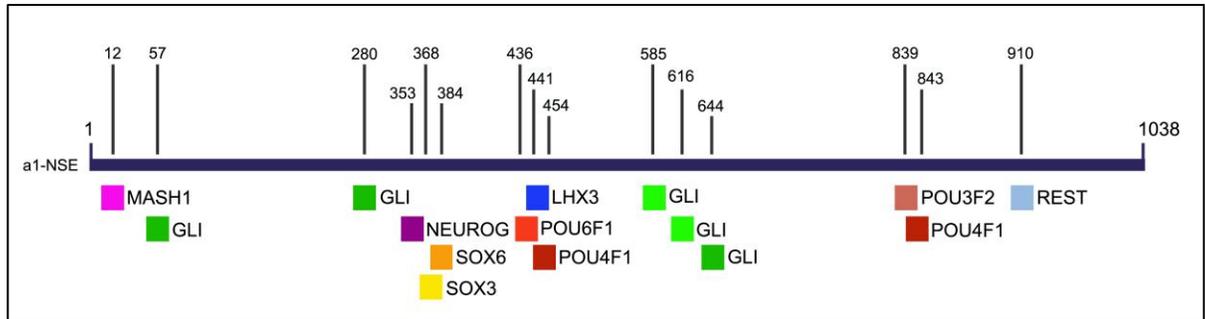
Figure 7.11. *Map of transcription factor binding motifs in a1-NSE.* A schematic representation of a1-NSE as a dark blue bar with relative locations of individual binding motifs for transcription factors with important roles in neural development shown bellow in different colours. Numbers above the bar indicate the position of the first base pair of each motif relative to the beginning (1) and end (1038) of a1-NSE.

Notably, a1-NSE, and as a matter of fact – the whole *Lama1*'s intron 1, does not display high sequence conservation across vertebrates, and even across Mammalia. Customized glocal alignment of the murine *Lama1*'s intron 1 sequence with the corresponding sequences from human, opossum, chicken and zebrafish revealed limited conservation in the mouse/human and mouse/opossum comparisons, whereas no conservation was detected with chicken and zebrafish, even at very low stringency conditions – 70% sequence identity over a 30 bp window (Figure 7.12A, B). Despite the limited sequence conservation or lack of it, the 1st introns of the mouse, human, chicken and zebrafish *Lama1* orthologs share some common TFBSs (Table 7.3).

Although the murine *Lama1*'s intron 1 exhibits low total conservation with the corresponding human and opossum sequences, some discrete regions appear to be conserved in the mouse/human alignment, including regions within a1-NSE (Figure 7.12A, B, C). DiAlign TF analyses (Genomatix) showed that none of these regions contain conserved GLI-binding motifs. The same analyses however, revealed that a1-NSE harbors several other conserved transcription factor binding motifs, some of which are putative binding sites for factors involved in vertebrate CNS development, such as FOXH1, HBP1, NMYC, PAX6/8, POU3F2, POU3F3, POU4F1, RXRA/B/G and SOX9 (Table 7.4).

158
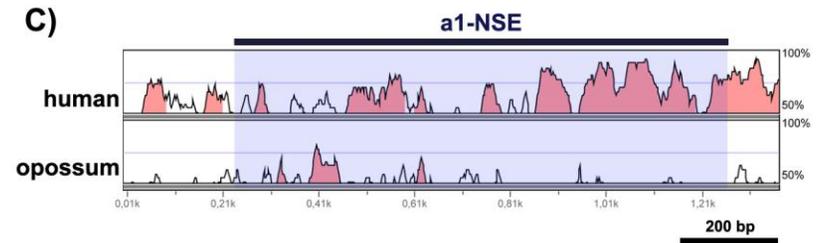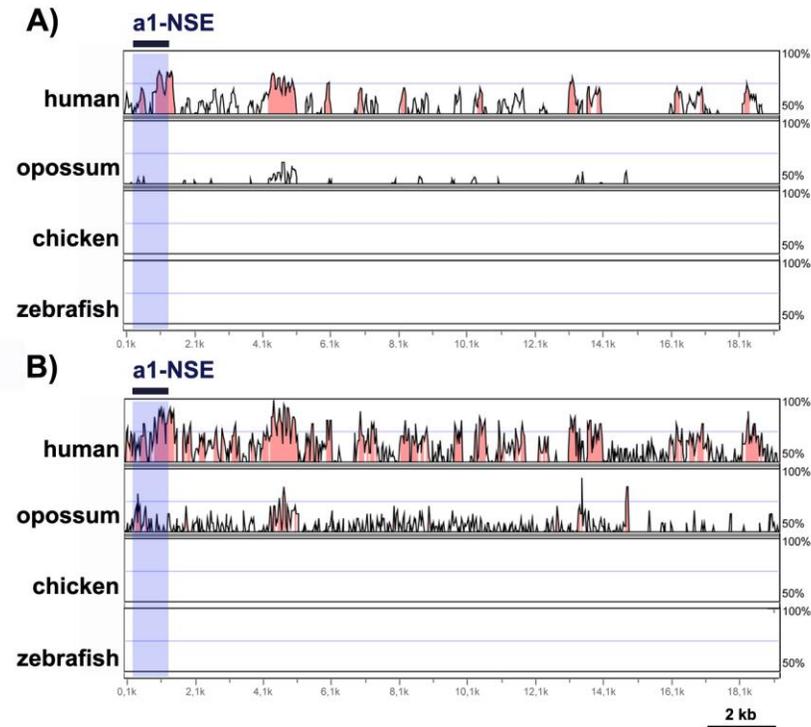
Figure 7.12. *Phylogenetic footprinting analysis of Lama1's intron 1.* The 1st intron of the murine *Lama1* gene was aligned with the corresponding 1st intron of the *Lama1* orthologs from human, opossum, chicken and zebrafish using the mVISTA tool. (A) An alignment performed under default stringency settings (minimum 70% sequence identity over a minimum of 100 bp window). (B) In order to allow for the detection of short conserved sequences in the size range of individual transcription factor binding motifs, the same alignment was repeated under decreased stringency settings (minimum 70% sequence identity over a minimum of 30 bp window). Note the absence of sequence conservation in the mouse/chicken and mouse/zebrafish alignments. (C) A close-up view of the first 1, 350 base pairs of the murine *Lama1*'s intron 1, containing the a1-NSE element, aligned with human and opossum sequences under low stringency settings (minimum 70% sequence identity over a minimum of 30 bp window). A *dark blue* bar with labeling on top indicates the extent of a1-NSE in the alignments in panels (A), (B) and (C). A *black* scale bar is provided below the alignments.

A.

Binding motif	p-value	Common	mouse	human	chicken	zebrafish
DDIT3 (DNA-damage inducible transcript 3)	6.50E-05	4	1	21	2	3
ZFP628 (zinc finger protein 628)	7.99E-05	4	4	9	4	1
NFE2L1 (nuclear factor, erythroid derived 2,-like 1)	9.36E-05	4	4	6	3	3
PAX1 (paired box 1)	0.000101	4	2	3	3	2
ATF4 (activating transcription factor 4)	0.000105	4	5	14	5	2
HINFP (histone H4 transcription factor)	0.000113	4	1	9	1	1
ABL1 (c-abl oncogene 1, non-receptor tyrosine kinase)	0.000117	4	10	14	3	4
HOMEZ (homeodomain leucine zipper-encoding gene)	0.000117	4	5	6	3	1
SALL1 (spalt-like transcription factor 1)	0.000155	4	2	20	3	4
GTF3C (general transcription factor III C)	0.000267	3	0	4	1	1
PAX9 (paired box 9)	0.000340	4	3	7	7	1
THAP1 (THAP domain containing, apoptosis associated protein 1)	0.000340	4	4	14	5	1
TAF (TATA box binding protein (TBP)-associated factor)	0.000409	4	3	5	5	7
ZFP110 (zinc finger protein 110)	0.000480	2	1	0	1	0
MLX (MAX-like protein X)	0.000488	4	6	5	2	3
ZBTB26 (zinc finger and BTB domain containing 26)	0.000538	4	3	5	5	1
ZBTB7A (zinc finger and BTB domain containing 7a)	0.000538	4	8	9	2	2
ZBTB33 (zinc finger and BTB domain containing 33)	0.000564	4	6	10	7	7
OSR (odd-skipped related)	0.000678	4	6	4	8	1
ZSCAN21 (zinc finger and SCAN domain containing 21)	0.002370	4	6	12	15	6

B.

Binding motif	p-value	Common	mouse	human	chicken	zebrafish
PAX 4, 6, 7 (paired box 4, 6 and 7)	0.003675	4	11	39	20	27
ZIC (zinc finger protein of the cerebellum family)	0.006014	4	8	22	6	5
PAX3 (paired box 3)	0.022713	4	19	32	18	9
REST (RE1-silencing transcription factor)	0.027427	4	21	25	11	4
NKX6.1 (NK6 homeobox 1)	0.036085	4	44	101	63	78
SMAD (SMAD family member)	0.036885	4	21	49	11	6
ETS (E26 avian leukemia oncogene family member)	0.040725	4	92	153	73	30
GLI (GLI-Kruppel family member)	0.043061	4	34	48	16	3
NEUROD (neurogenic differentiation family member)	0.043905	4	20	32	20	10

Table 7.3. *Transcription factors binding motifs common to the 1ˢᵗ introns of Lama1 from mouse, human, chicken and zebrafish*. The 1ˢᵗ introns of the *Lama1* genes from mouse, human, chicken and zebrafish were analysed for common TF binding motifs. A) The top 20 statistically significant common motifs matches are displayed only, out of 37 significant motif matches in total. B) Some of the statistically significant common motifs matches are for TFs with roles in vertebrate CNS development, although these motifs matches are not included in top 20. The p-value indicates "the probability to obtain an equal or greater number of sequences with a motif match from a randomly drawn sample of the same size as the set of input sequences. The lower the p-value, the higher is the statistical significance (importance) of the observed common motif match" (Genomatix). The "Common" column shows the number of input sequences (4 input sequences in this study) with a common motif. The columns named "mouse", "human", "chicken", and "zebrafish" show the number of individual motif matches within the input sequence from each species.

Binding motif	Position
FOXH1 (forkhead box H1)	934 – 950 (-)
GATA1 (GATA binding protein 1)	945 – 957 (-)
HBP1 (high mobility group box transcription factor 1)	178 - 202 (+) 674 – 698 (+) 825 – 849 (-)
HIC1 (hypermethylated in cancer 1)	787 – 799 (-)
HMGA (high mobility group AT-hook)	561 – 585 (+) 837 – 861 (-) 932 – 956 (+)
HNF1B (HNF1 homeobox b)	940 – 956 (+)
IKZF1 (IKAROS family zinc finger 1)	706 – 718 (-)
MECOM (MDS1 and EVI1 complex locus)	604 – 620 (+)
Nanog (Nanog homeobox)	179 – 197 (+) 831 -849 (-)
NFIL3 (nuclear factor, interleukin 3, regulated)	934 – 954 (-)
NMYC (v-myc myelocytomatosis viral related oncogene, neuroblastoma derived)	765 – 781 (+)
PAX6 (paired box 6)	839 – 857 (+)
PAX8 (paired box 8)	828 – 856 (-)
POU2F1 (POU domain, class 2, transcription factor 1)	828 – 842 (+)
POU3F2 (POU domain, class 3, transcription factor 2)	839 – 857 (+)
POU3F3 (POU domain, class 3, transcription factor 3)	838 – 852 (-)
POU4F1 (POU domain, class 4, transcription factor 1)	843 – 861 (+)
RREB1 (ras responsive element binding protein 1)	781 – 795 (-)
RXR (retinoid X receptor)	205 - 229 (-)
SOX9 (SRY (sex determining region Y)-box 9)	781 – 805 (-)
SPIC (Spi-C transcription factor (Spi-1/PU.1 related)	928 – 948 (+)
USF1 (upstream transcription factor 1)	686 – 702 (-)

Table 7.4. *Transcription factor binding motifs within a1-NSE that are conserved between mouse and human.* Here are listed the transcription factor binding motifs that are identified as sequence conserved within a1-NSE, as obtained from the mouse/human alignment (displayed at figure 7.11C) using the DiAlign TF tool (Genomatix). The "Position" column shows the location of each putative motif within a1-NSE as base-pair coordinates; the (+) and (-) signs refer to the "sense" and "antisense" strand, respectively. TFs highlighted in *cyan* feature roles in vertebrate CNS development.
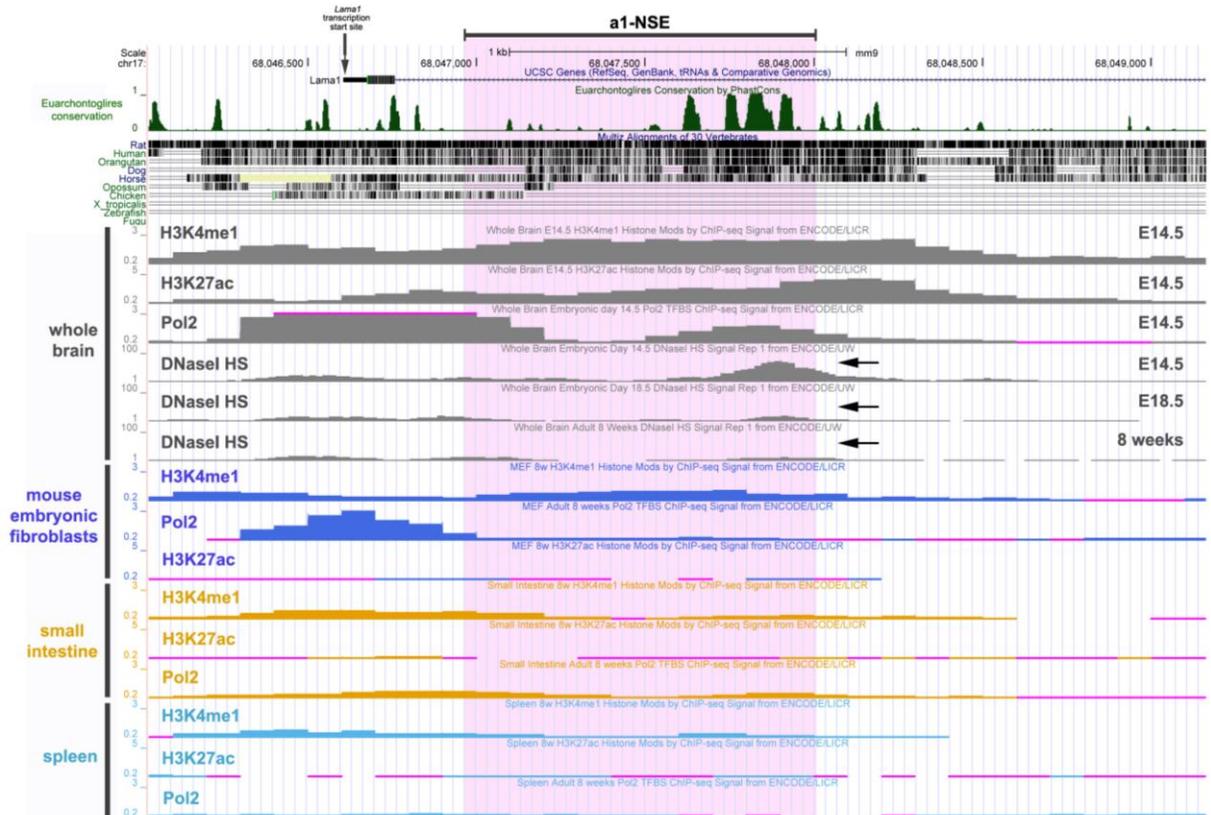
Figure 7.13. *Chromatin state profile of a1-NSE in different cell types/organs and developmental stages.* a1-NSE features high and broad peaks of H3K4me1 H3K27ac histone marks, as well as RNA pol II occupation and DNaseI hypersensitivity in brain tissue but not in other organs, as revealed by analysis in the UCSC Genome Browser. Horizontal black arrows point at the peak of DNaseI hypersensitivity that gradually attenuates with progression of development.

An important insight into a1-NSE's function is provided by its association with certain chromatin features, as revealed by the examination of a1-NSE in the UCSC Genome Browser. Remarkably, a1-NSE correlates with a broad peak of H3K4me1 and H3K27ac histone modifications, which mark active enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2011), in the brain of E14.5 mouse embryos but not in embryonic fibroblasts, small intestine or spleen (Figure 7.13). This further supports the idea that a1-NSE is a neural-specific enhancer. Moreover, there is a peak of DNaseI hypersensitivity over a1-NSE whose pattern undergoes progressive temporal changes suggesting that the a1-NSE locus is more accessible to nuclease digestion at E14.5 but becomes virtually inaccessible in the adult brain (black arrows in Figure 7.13). Another curious finding is the enrichment for RNA Pol II binding not only around the *Lama1* promoter but also in a1-NSA itself, and this is only

162

observed in the brain but not in other tissues/organs. Taken together, the data gained through *in silico* analyses strongly suggest that a1-NSE operates as a neural-specific enhancer of the murine *Lama1* gene that is active during central nervous systems development (CNS).

7.4. Discussion

Here I reported on the identification of a GLI-bound enhancer element, a1-NSE, located in the 1st intron of the mouse *Lama1* gene, which drives neural-specific reporter expression in the CNS of transgenic zebrafish and mouse embryos, suggesting conserved transcription factor environments in both species. Consistent with this, a1-NSE contains binding sites for transcription factors implicated in vertebrate neural development, and also features chromatin marks of active enhancers in the brain but not in non-neural tissues. Yet, a1-NSE displays no sequence conservation between mouse and zebrafish, although intron 1 of the mouse *Lama1* gene shares common TF binding motifs with that of zebrafish.

7.4.1. GLI-binding sites and a1-NSE's function

a1-NSE harbours a cluster of five GLI-binding motifs within ~ 1 kb of sequence, some or all of which are probably occupied by GLI factors in the developing neural tube. Each of the five motifs closely resembles the GLI consensus "TGGGTGGTC" sequence (Kinzler and Vogelstein 1990) but none is a perfect match to it (Figure 7.1), which is not unusual as most transcription factors bind to a spectrum of related sequences (Zhang et al. 2006). Motifs 3 and 4 were not annotated by Vokes et al. (2008) but were predicted by MatInspector. Perhaps, any of the three GLI factors in the mouse could bind to any of the 5 sites, as GLI1/2/3 show very similar binding affinities *in vitro* (Hallikas et al. 2006). Interestingly, Vokes et al. (2007) reported that 24% of GLI1-target genes in neural cell cultures contain their GLI1-ocuppied sites within the first or second intron, which is consistent with the location of the five GLI motifs of a1-NSE in the first intron of *Lama1* (Figure 7.1).

Moreover, the density of GLI binding sites within a1-NSE is significantly higher than what is expected by random distribution of GLI motifs in the genome (Z-score = 3.36, see Table 7.1; Hallikas et al. 2006; Vokes et al. 2008), which hints for their possible functional significance in a1-NSE activity. In fact, this phenomenon - the presence of multiple binding motifs for the same transcription factor within a regulatory element, is known as "homotypic clustering", and is proposed to increase the thermodynamic probability of transcription factor

163

binding in the process of one-dimensional diffusion of transcription factors along DNA (Gorman and Greene 2008). Interestingly, Gotea et al. (2010) reported that homotypic clusters are 25-times over-represented in developmental enhancers, and it is suggested that clustering may facilitate cooperative interactions between the DNA-bound factors, which serves to reduce gene expression noise by increasing the number of required transcriptional activation steps (Segal et al. 2008), resulting in sharper gene expression patterns. This model is supported by the inability of Bicoid (K57R) mutant to engage in homotypic protein-protein cooperative interactions which lead to failure in establishing the sharp posterior boundary of hunchback gene expression and gross morphological defects in the head and thorax of mutant *Drosophia* flies (Lebrecht et al. 2005).

The presence of occupied GLI motifs in a1-NSE suggests that GLI proteins participate in its neural-specific enhancer function. Such assumption is supported by the conserved expression pattern of *Gli* genes in the vertebrate neural tube of, where all *Gli* are expressed throughout the neuroectoderm in early stages of development, before *Gli1* mRNA becomes restricted to the ventral half of the neural tube (excluding the floor plate), and *Gli2* and *Gli3* to the dorsal half of the tube (Hui et al. 1994; Sasaki et al. 1997; Vanderlaan et al. 2005). To test whether some or all of the GLI motifs are required for the activity of a1-NSE, loss-of-function base pair substitutions were introduced and the effects examined in transient transgenic zebrafish embryos, as described in the next chapter.

7.4.2. ZIC transcription factors and a1-NSE

It is also plausible that the activity of GLI proteins bound to their motifs in a1-NSE could be modulated by ZIC (**zi**nc finger protein of the **c**erebellum) transcription factors. ZIC proteins (ZIC 1-5) bind the GLI consensus motif sequence *in vitro* (Mizugishi et al. 2001), due to a $(Cys_2His_2)_5$ Zn-finger DNA-binding domain of ZIC that is highly homologous to the DNA-binding domain of GLI (Aruga et al. 1996) as discussed in Chapter 4. ZIC and GLI proteins can also physically interact through Zn-fingers 3-5 of their DNA-binding domains and this interaction facilitates the nuclear translocation of GLI proteins (Koyabu et al. 2001). Interestingly, cell culture assays indicate that once in the nucleus, ZIC and GLI proteins synergistically enhance each other's ability to trans-activate reporter gene expression, which is again dependent on Zn-fingers 3-5 in both proteins (Koyabu et al. 2001).

Thus, a model has emerged where ZIC factors are modulators of GLI-mediated SHH signaling responses *in vivo*. For instance, in the central nervous system *Zic1, Zic2* and *Zic3* genes promote early neuroectoderm specification in *Xenopus* (Mizuseki et al. 1998; Nakata et al. 1998), *Zic2, Zic3* and *Zic5* are essential for neuropore closure in mice (Nagai et al. 2000; Inoue et al. 2004), and *Zic2* function is required for division of the prosencephalon (Brown et al. 1998; Nagai et al. 2000), for proper proliferation of granule cell precursors in the anterior cerebellum (Aruga et al. 1998; 2002), and for correct contralateral projection of retinal ganglion cells (Herrera et al. 2003). This plethora of functions is reflected in the expression of *Zic* genes during neural development where all *Zic1-5* have similar expression pattern restricted to the dorsal ½ or 1/3 of the midbrain, hindbrain and spinal cord, and to the allar plates of the telencephalon (Nagai et al. 1997).

Studies in zebrafish, *Xenopus*, chick and mouse embryos reveal that both antagonistic and cooperative interactions exist between GLI and ZIC factors, and the particular mode of interaction is probably determined by cell lineage, extracellular signals and/or by intrinsic propensities of different ZIC factors to activate some GLI proteins but to inhibit others upon physical contact when bound to DNA (Aruga et al. 1993; Aruga et al. 1999; Brewster et al. 1998). Based on this, it is possible that ZIC could bind to one or more GLI motifs in a1-NSE and contribute to its neural enhancer function through direct interaction with GLI factors. The close proximity between GLI motifs 3-5 in a1-NSE favours such a model, which can be tested by immunoprecipitating ZIC bound to a1-NSE in a ChIP experiment *in vivo*, or by an electrophoretic mobility shift assay (EMSA) *in vitro*.

7.4.3. a1-NSE exhibits chromatin features of a tissue-specific enhancer

Inspection of a1-NSE in the UCSC Genome Browser revealed that in brain tissue a1-NSE is enriched for the histone modifications H3K4me1 and H3K27ac and displays opened chromatin configuration as shown by the peak of DNaseI hypersensitivity at E14.5 embryonic brain but not in other cell types/organs (Figure 7.13). As discussed in Section 5.4.6., H3K4me1 is over-represented in both poised and active enhancers (May et al. 2011), while H3K27ac is considered as a reliable mark of active versus poised state of developmental enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2011). A peak of DNaseI hypersensitivity within a1-NSE spans conserved sub-regions 4-5, indicating that this particular part of a1-NSE is in a more accessible chromatin state (black arrows in Figure

7.13). Remarkably, however, the peak gradually disappears as development proceeds. This is consistent with the idea that the accessibility of regulatory regions varies in time and in different tissues, depending on the presence/absence of pioneer transcription factors which can elicit chromatin decompaction and facilitate the binding of other transcription factors (Guertin and Lis 2013). The temporal pattern of a1-NSE accessibility is also consistent with the expression pattern of *Lama1* mRNA, which is transcribed in the embryonic and foetal central nervous system but gradually disappears at later stages (Miner et al. 2004).

The low levels or absence of chromatin features in other cells/organs – embryonic fibroblasts, intestine and spleen (Figure 7.13), suggests strongly that a1-NSE is a developmentally-regulated neural-specific enhancer, and this is consistent with results from a genome-wide study of the chromatin states of promoters, insulators and enhancers in various human cell types, revealing that chromatin signatures of promoters and insulators are largely invariant across different tissues, while enhancers exhibit high cell-type-specific histone modification patterns (Heintzman et al. 2009).

Another intriguing feature of a1-NSE is its enrichment for bound RNA Pol II (Figure 7.13). Several recent studies have demonstrated that some enhancers associate with Pol II and are transcribed as non-coding RNA called "enhancer" RNA (eRNAs) (Koch et al. 2011). However, it is unclear whether all eRNAs are functionally relevant or are by-products of stochastic initiation events due to random collision of Pol II with accessible chromatin (Natoli and Andrau 2012; Struhl 2007). Nevertheless, there are studies, which hint or demonstrate important biological roles for some eRNAs. For instance, *Evf-2* is a spliced, multi-exonic ncRNA that is transcribed from the ultraconserved enhancer element *ei* located between the *Dlx5* and *Dlx6* loci in the mouse genome, and is induced by SHH signaling (Feng et al. 2006). *Evf-2*eRNA appears to act *in trans* by forming a complex with DLX2 leading to enhancement of its transactivation potential. In turn, the *Evf-2*-bound DLX2 promotes the activation of the *Dlx5/6* genes by binding to the same ultraconserved element, *ei*, that encodes *Evf-2*. Thus, it appears that *Evf-2* acts in feed-forward regulatory loop (Feng et al. 2006). Another key study on a genome-wide scale uncovered signal-induced transcription of thousands of CBP-bound enhancers in murine cortical neurons (Kim et al. 2010). These neuronal enhancers associate with RNA pol II in an activity-dependent manner leading to bi-directional transcription, the levels of which correlate with the levels of mRNA

166

synthesis from nearby genes, suggesting that transcription at enhancers promotes target gene expression (Kim et al. 2010). Thus, it is plausible that the neural-specific activity of a1-NSE might be mediated by an eRNA intermediate. It would be highly interesting to investigate this possibility further, by injecting *a1-NSE::EGFP*-transgenic zebrafish embryos with siRNA and morpholinos complementary to the sense- or antisense strands of a1-NSE.

7.4.4. a1-NSE and the entire intron 1 of *Lama1* display limited sequence conservation

Intron 1 of the murine *Lama1* gene, together with a1-NSE, are not conserved in the mouse/chicken and mouse/zebrafish comparisons, whereas the mouse/human and mouse/opossum alignments show some conserved regions (Figure 7.12). The lack of conservation with zebrafish *lama1*'s intron 1 is intriguing, especially in light of the similar function of a1-NSE when tested in zebrafish and mouse embryos, and because the 1st intron of zebrafish *lama1* contains CNS-active enhancer sequences (Joseph Pickering's Thesis 2012). However, evolutionary conservation of enhancer function despite interspecific sequence divergence appears to be a relatively common phenomenon, as introduced in Section 1.10.2.1 and elaborated further on in Section 9.2, and could be explained by the co-occurrence of common TFBSs establishing "regulatory grammar" (Senger et al. 2004). Consistent with this, the 1st introns of mouse and zebrafish *Lama1* orthologs share common binding motifs of many TFs, some of which are involved in neural development, including GLI, ZIC, NEUROD and PAX proteins (Table 7.3B). Therefore, one can hypothesise that when tested in transgenic zebrafish embryos, a1-NSE displays a neural-specific activity that is similar to the activity in transgenic mouse embryos, for the TF environment in the developing zebrafish CNS is similar to the TF environment in the murine CNS, despite the lack of a recognisable "a1-NSE orthologous element" in the 1st intron of fish *lama1*. This hypothesis predicts that a reporter gene would exhibit neural expression in transgenic mouse embryos when driven by intron 1 of the zebrafish *lama1*, and this could be tested experimentally.

7.4.5. Neural specificity of a1-NSE function

The presence of GLI binding motifs cannot alone explain the neural specificity of a1-NSE, as a *GFP* reporter driven only by a cluster of eight tandem consensus GLI binding sites efficiently activates reporter expression upon HH pathway activation not only in neural

167

lineages but also in other SHH-responsive cells, such as the somitic cells (Stamataki et al. 2005; Kahane et al. 2013). Therefore, additional transcription factor inputs must be present to restrict a1-NSE's activity to the neural tube, and not to the somites, limb buds or cranial mesenchyme, which are also dependent on HH signaling (Borycki et al. 1998; Hu and Helms 1999; McGlinn and Tabin 2006). Consistent with this idea, a1-NSE harbours putative binding sites for transcription factors with established functions in CNS development (Table 7.2; Guillemot 2007; Stolt and Wegner 2010). Some of the putative sites appear to be conserved between mouse and human (Table 7.4).

However, most of the predicted TFBSs do not exhibit statistically significant over- (or under-) representation in a1-NSE, except for the GLI, and SOX3/6/9 motifs (Table 7.2), hinting for potential *in vivo* function of these motifs. Interestingly, the LHX3 motif features significant under-representation – instead of 7 expected LHX3 motifs, a1-NSE contains only a single such motif (Table 7.2), suggesting for some functional importance of the exclusion of LHX3 motifs from a1-NSE. Despite the lack of statistical significance for the majority of CNS-development-related motifs in a1-NSE, the motifs could still be genuine TFBSs *in vivo*, for many transcription factors engage in physical interactions between each other and the bound regulatory DNA element, modulating target gene transcription in a combinatory manner (Davidson 2006). Therefore, although individual TF binding motifs are not significantly over-represented in a1-NSE, the combination of several different motifs could be.

a1-NSE contains binding motifs for the proneural basic helix-loop-helix transcription factors neurogenin and MASH1, as well as for several SOX and POU-domain proteins (Figure.7.11). Neurogenins 1, 2 and 3 and MASH1 are predominantly expressed in the ventricular zone of the prospective spinal cord and brain starting from E8.5 in mouse development (Guillemot and Joyner 1993; Sommer et al. 1996), and the SOXB1 class of proteins (SOX1, 2, 3) are specifically expressed in the ventricular zone of the mouse neural tube together with many POU factors (He et al. 1989; Uchikawa et al. 1999). Notably, these transcriptions factors are known to synergise in the regulation of target genes. For instance, SOX1/2/3 and POU3F2 are essential for the synergistic activation of the 30 bp neural enhancer of the nestin gene in the developing murine spinal cord via binding to adjacent sites in the regulatory element (Tanaka et al. 2004). Analogously, synergistic interaction between

MASH1 and the Class III POU proteins POU3F2 and POU3F2 is requied for the expression of *Delta1* in the murine dorsal spinal cord and telencephalon (Castro et al. 2006). In addition, cross-regulatory interactions exist between these transcription factors. For instance, neurogenins and MASH1 trigger neuronal differentiation by first repressing the transcription of the SOXB1 class of genes, which are required for maintenance of pluripotency and self-renewal in neural stem cells (Bylund et al. 2003).

Thus, one can envision a model where combinatorial inputs from GLI and/or ZIC, proneural bHLH, POU domain, SOX and homeodomain transcription factors contribute synergistically to the neural specificity of a1-NSE. This model can be tested by deletion of the motifs for SOX, POU or bHLH transcription factors from a1-NSE and examining the effects in transient transgenic zebrafish injected with the mutant a1-NSE-reporter construct.

However, in addition to cooperative interactions between transcriptional activators, the specificity of a1-NSE could be also determined by the binding of a global repressor, as it is often the case in embryonic patterning (Davidson 2006). Consistent with this, a1-NSE contains a binding motif for the **RE**1-**s**ilencing **t**ranscription factor, or REST (also known as NRSF) (Table 7.2). REST is a repressor of neuronal genes in non-neural tissues and in undifferentiated neural precursor cells (Lunyak and Rosenfeld 2005). Inactivation of *Rest* in mouse embryos or over-expression of a dominant negative form of REST (dnREST) in chicken embryos leads to ectopic activation of neuronal marker genes in the myotome, epicardium and in the limb ectoderm, or to premature activation of neural differentiation genes in the ventricular zone of the spinal cord (Chen et al. 1998).

REST acts via binding to the relatively long and rare 21 bp repressor element-1 (RE1) motif, also known as NRSE (Neuron-Restrictive Silencer Element) (Mori et al. 1992), in the promoters or enhancers of genes encoding proteins of differentiated neurons (Brivanlou 1998; Sun et al. 2005). REST recruits a complex that adds repressive marks to the associated chromatin (Andres et al. 1999; Lakowski et al. 2006). In effect, this results in silencing of the REST-bound promoter or enhancer and repression of target gene transcription. Expression of *Rest* is also consistent with its role as major repressor of neuronal genes in non-neural tissues, as it is ubiquitously expressed at E8.5 and E9.5 in mouse embryos (Chen et al. 1998), but in neural tissues its levels are strongly down-regulated in the transition between stem- and intermediate progenitors with complete dissociation from RE1 motifs at the onset of neuronal

differentiation (Ballas et al. 2005). Surprisingly, a role of Rest in modulation of Hh signaling in zebrafish neural tube patterning was described by Gates et al. (2010). Zebrafish embryos treated with morpholinos against *rest* exhibit moderate ventralisation of the neural tube, with dorsal expansion of *ptch1*, *foxa2*, and *nkx2.2a* expression, reminiscent of up-regulated Hh signaling. Analysis of *gli* gene expression in *rest* morphants uncovered that *gli2a* was overexpressed, indicating that rest acts on the Hh pathway by repressing *gli2a* expression (Gates et al. 2010).

Thus, it is plausible that the neural-specific activity of a1-NSE is, at least in part, conferred by REST-mediated silencing in non-neural tissues, while a1-NSE is active in the CNS due to down-regulation of REST expression or inhibition of its binding to the RE1 motif on a1-NSE. This assumption can be tested by deletion of the RE1 motif from a1-NSE and observing the effects on EGFP reporter expression in transient transgenic zebrafish.

In summary, the neural-specific pattern of a1-NSE activity is most likely determined by the combined effects of neural-specific transcriptional activators and global, ubiquitous repressors whose activity is however downregulated in the central nervous system.

7.4.6. a1-NSE's activity in transgenic zebrafish

The Tol2 transposase-mediated transient transgenesis of *a1-NSE::EGFP* into F0 zebrafish embryos appears to have been successful for nearly 62% of the analysed injected embryos were EGFP-positive, an yield that is consistent with previous reports (Navratilova et al. 2009; Royo et al. 2011; Ikle et al. 2012). However, only 4 founders were found among the screened EGFP$^+$ fish (n=40), resulting in germ-line transmission rate of 10%, which is lower than the reported average rate of 30 – 70% for Tol2-based transgenesis (Kawakami 2007), suggesting for reduced activity of the a1-NSE enhancer in the F1 generation. A possible cause behind this reduction in activity could be a transgene silencing phenomenon previously described in zebrafish (Akitake et al. 2011; Goll et al. 2009), where CpG-rich sequences in the transgene construct are particularly susceptible to DNA methylation leading to epigenetic silencing of transgene expression (Goll et al. 2009). Interestingly, the first 200 base pairs of a1-NSE are part of an extensive 717 bp long CpG island (Figure 7.14), suggesting the possibility that the *a1-NSE::EGFP* transgene might have been subjected to DNA methylation in some of the F0 zebrafish, hence the low rate of detected founders (10%).

Nevertheless, the 4 founders gave rise to F1 clutches where the frequency of EGFP[+] embryos ranged from 25% to 35%, which is in the range of the 0.3-100% frequency reported in other studies (Kawakami et al. 2004; Kawakami 2007; Urasaki et al. 2006). Notably, all EGFP[+] F1 embryos featured reporter expression restricted only to the ventral neural tube, with no expression observed in other structures, suggesting that a1-NSE's activity is resistant to the potential influence of by-stander enhancers at the genome integration sites in different F1 embryos.
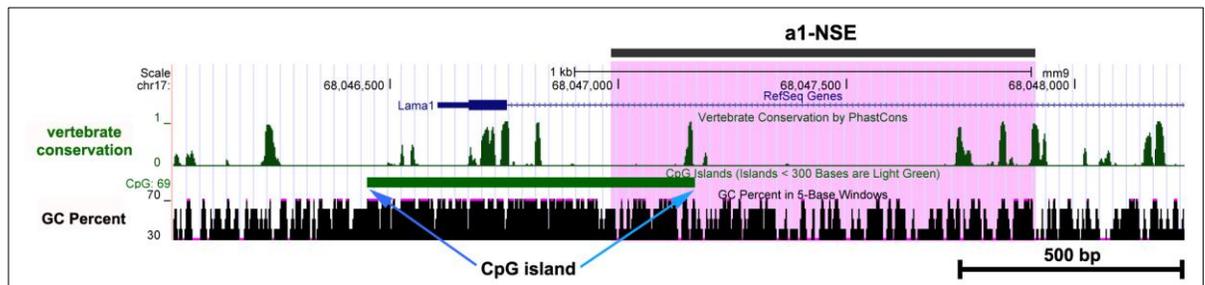
Figure 7.14. *a1-NSE partially overlaps with a CpG-rich sequence at the start of the murine Lama1 gene.* The *pink* shading indicates the extent of a1-NSE.

7.4.7. a1-NSE activity and *Lama1* expression

The reporter expression pattern driven by the murine a1-NSE in transgenic mouse and zebrafish embryos is very similar to the neural expression pattern of endogenous *Lama1* in mouse and zebrafish embryos (this study; Joseph Pickering Thesis 2012) (Figure 7.15). *Lama1* is transcribed at higher levels by cells in the ventricular/subventricular zone (V/SVZ) of the neural tube in zebrafish and mouse, and in the mouse embryo this expression becomes progressively ventrally restricted, such that *Lama1* mRNA is eventually excluded from the dorsal neural tube, as well as from the floor plate (Figure 3.1; Figure 3.2). Remarkably, the pattern of β-Gal signal in the *a1-NSE::lacZ* transgenic mouse embryos is very similar to that of endogenous *Lama1*, although there was a somewhat broader reporter signal that spanned the whole medio-lateral width of the neural tube (Figure 7.15A-D; Figure 7.8J; Figure 7.9K). This could be due to the high stability of β-galactosidase protein (β-Gal) (or alternatively – to the stability of the color product), which may have marked cells at the ventricular zone initially, but was subsequently transmitted to daughter cells that migrated radially to form the mantlelayer of the neural tube (Kriegstein and Alvarez-Buylla 2009). Consistent with such explanation, the E10.5 transgenic mouse embryo featured weaker expression of β-Gal in the mantle layer as compared to the E9.5 embryo (white asterisk in Figure 7.9K), and in the

E12.5 transgenic embryo reporter expression was confined solely to the ventricular zone (black arrow in Figure 7.10D), reminiscent to the endogenous expression of *Lama1* mRNA at that stage (Miner et al. 2004), and this is most likely consequence of the eventual degradation of the β-Gal protein or a dilution effect.
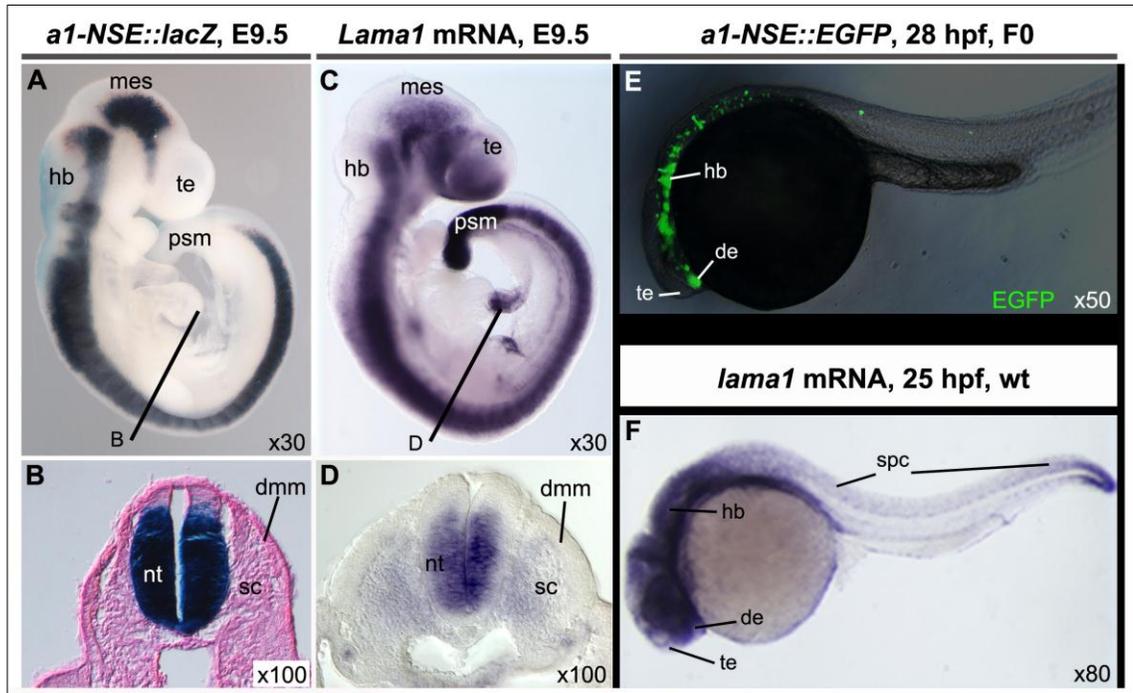
Figure 7.15. *The a1-NSE-driven reporter expression pattern is very similar to the neural expression pattern of endogenous Lama1.* Comparison between the *a1-NSE::lacZ* E9.5 transgenic (A, B) and a wild type E9.5 mouse embryo stained for *Lama1* mRNA (C, D). (B) and (D) are sections at forelimb level. (E) The a1-NSE-driven EGFP expression pattern in a 28 hpf transiently transgenic zebrafish embryo is similar to the endogenous *lama1* mRNA expression pattern in the anterior neural tube of a 25 hpf wild type zebrafish embryo. There is significant overlap between the pattern of β-Gal staining (A) and *Lama1* mRNA expression (C) along the murine neuraxis, except for the telencephalon, where a1-NSE is inactive, while endogenous *Lama1* is expressed there. Note the absence of reporter expression in non-neural tissues (like the sclerotome and presomitic mesoderm) in both transgenic mouse and zebrafish embryos. Abbreviations: de, diencephalon; dmm, dermomyotome; hb, hindbrain; mes, mesencephalon; nt, neural tube; psm, presomitic mesoderm; sc, sclerotome; spc, spinal cord; psm, presomitic mesoderm; te, telencephalon; wt, wild type. (Image in F is kindly provided by Dr. Joseph B. Pickering, University of Sheffield)

Alternatively, the expanded β-Gal signal could be a result of the absence of binding sites in a1-NSE for putative repressors, which might function to silence endogenous *Lama1* expression in the mantle layer, resulting in persistent *lacZ* reporter expression beyond the V/SVZ. In relation to this, I observed a complex expression pattern of β-Gal in multiple sites in the mantle layer of the spinal cord in the two E13.5 transgenic embryos (Figure 7.10I-J, O-P), where *Lama1* mRNA does not appear to be expressed (Miner et al. 2004). This could again be an outcome of the lack of essential repressor inputs on a1-NSE. Given the presence

of potential binding sites for numerous proneural transcriptional activators (Figure 7.11), it is possible that the ectopic *lacZ* expression results from permissive conditions.

Similar reasons may explain the expression of β-Gal in the telencephalon of the two E13.5 embryos (Figure 7.10E-H, K-N), as a1-NSE does not appear to possess telencephalic activity since none of the transgenic embryos at the earlier stages showed expression of β-Gal in the telencephalon (Figures 7.7, 7.8 and 7.9). Also, this staining is inconsistent between the two E13.5 embryos (Figure 7.10H and N), and may result from the usurpation of the β-globin promoter in the *a1-NSE::lacZ* construct by the enhancers of a bystander gene at the integration site. This second possibility is hinted by the suspicious β-Gal signal in the dorsal root ganglia and in the trigeminal ganglion, two sites where endogenous *Lama1* mRNA is not detected (Figure 3.2; Anderson et al. 2009; Miner et al. 2004), in one of the E13.5 transgenics. In this case, the transgene has probably integrated in the vicinity of a neural crest-specific developmental enhancer, as ganglia in the peripheral nervous system are derived from the neural crest (D'amico-Martel and Noden 1983). Alternatively, a1-NSE may harbour genuine telencephalic activity at E13.5, as endogenous *Lama1* is expressed in the cortex at that stage (Miner et al. 2004), and examination of additional E13.5 or later transgenic embryos may clarify this issue.

Similarly, the E10.5 transgenic embryo also displayed non-CNS β-Gal expression in the heart, sclerotomes, first pharyngeal pouch, and limb mesenchyme. Although *Lama1* is expressed in the sclerotome and first pharyngeal pouch at E10.25 (Figure 3.2), it is unlikely that β-Gal expression in these structures corresponds to a real transcriptional activity of a1-NSE as none of the other transgenics showed *lacZ* activity in these domains. Most likely, expression at these sites is due to integration effects, or to the absence of repressor motifs on a1-NSE that would normally prevent a1-NSE activity there.

The pattern of β-Gal expression along the neural tube of *a1-NSE*-transgenic mouse and zebrafish embryos (Figures 7.7, 7.8A-C, 7.9A) coincides strikingly to the antero-posterior expression pattern of *Shh* along the neuraxis (Echelard et al. 1993; Krauss et al. 1993). In the presumptive spinal cord and hindbrain of mouse embryos, *Shh* transcription is limited to the ventral midline, while in the midbrain and forebrain it extends ventro-laterally. This is particularly evident in the forebrain at E9.5, as *Shh* is expressed as two ventro-lateral stripes in the rostral diencephalon (Echelard et al. 1993). This very specific expression

173

domain overlaps with the pattern of β-Gal expression in the E9.5 transgenic mouse embryo (orange arrows in Figure 7.8C), suggesting that a1-NSE, like endogenous *Lama1*, may be responsive to Shh signals secreted from the ventral neural tube (Figures 3.1, 3.2 and 3.3; Anderson et al. 2009). Thus, a1-NSE may be the regulatory element mediating the Shh effects on *Lama1* expression in the neural tube.

The pattern of EGFP driven by a1-NSE in the zebrafish ventral neural tube is reminiscent to the expression pattern of genes that are positive targets (or Class II genes) of Hh signaling, such as *nkx2.2a*, *olig2* and *nkx6.1* (Gates et al. 2010; Guner et al. 2007; Park et al. 2002) (Figure 7.4). Moreover, the activity of a1-NSE in transgenic zebrafish is similar to the pattern of endogenous *Lama1* expression in both mouse and fish embryos, and this is particularly evident at stage 52 hpf transgenic zebrafish where EGFP signal is confined largely to the ventricular zone of the brain (Figure 7.5). Interestingly, as mentioned earlier, in both species, reporter expression is excluded from the dorsal neural tube (Figures 7.4 and 7.5; black arrows in Figures 7.8J; 7.9K; 7.10D) and this is most likely due to repressive inputs by GLI3 repressor form in the mouse and by both Gli2a and Gli3 repressor forms in zebrafish, which although expressed throughout the dorsoventral extent of the neural tube at early stages, later become confined to the dorsal half of the tube where they repress Shh-target genes (Persson et al. 2002; Tyurina et al. 2005; Vanderlaan et al. 2005). It is also possible that BMP signaling in the dorsal neural tube contributes to silencing of a1-NSE. Indeed SMAD proteins have been shown to physically interact with the GLI3 repressor form in cultured mouse cells (Liu et al. 1998) and with Gli2a repressor during zebrafish somite patterning (Maurya et al. 2011).

In summary, the neural-specific activity of the murine a1-NSE in transgenic mouse and zebrafish embryos resembles the endogenous expression of *Lama1* in both species, suggesting that a1-NSE may be the *cis*-regulatory element responsible for directing *Lama1* transcription in the murine central nervous system. Moreover, the fact that a1-NSE can operate in a homologous manner when tested in fish and mammals hints for conservation of the signaling pathways and transcription factor inputs impinging on a1-NSE. The presence of GLI binding motifs combined with the characteristic activity of a1-NSE along the neuraxis that parallels the expression of Shh in the notochord and floor plate, strongly suggest that

SHH signaling might be a modulator of a1-NSE's function, and this hypothesis is addressed in the next chapter.

Chapter 8

Regulation of a1-NSE's transcriptional activity
by the Hh signalling pathway

8.1. Hypothesis and aims

The presence of five GLI binding motifs, and the GLI3 binding at a1-NSE *in vivo* (Vokes et al. 2008), combined with the specific reporter expression pattern in the neural tube of transgenic mouse and zebrafish embryos, raises the hypothesis that the activity of a1-NSE might be regulated by the HH signalling pathway. Here, I examined this assumption by performing Hh signalling loss- and gain-of-function experiments in transient transgenic zebrafish embryos that express EGFP under the control of the murine a1-NSE enhancer. In addition, I tested whether the activity of a1-NSE depends on intact GLI binding motifs by performing transient transgenesis in zebrafish embryos using a mutated version of a1-NSE (mut a1-NSE).

8.2. Cyclopamine treatment abolishes the activity of a1-NSE

Treatment of whole embryos, tissue explants or cells in culture with small molecule inhibitors of the Hh signalling pathway is a straightforward initial approach to address whether the function of a particular transcription regulatory element depends on active Hh signalling. Cyclopamine is the oldest and most widely used antagonist of HH signalling, which at certain concentrations can effectively block Hh signal transduction when added to zebrafish embryo medium water (Wolff et al., 2003). Therefore, I decided to treat zebrafish embryos injected with the *a1-NSE::EGFP* construct with cyclopamine in order to examine whether the function of a1-NSE requires active HH signalling.

8.2.1. Introduction

Cyclopamine is a steroidal alkaloid first isolated from the lilly *Veratrum californicum* following an investigation on the etiology of cyclopia epidemic in sheep during the 1950's in the United States (Binns et al., 1963; Keeler and Binns, 1968). Subsequently, it was revealed that administration of cyclopamine to gastrulation-stage amniote embryos triggered a complex set of midline facial and neurological defects ranging from microcephaly, ocular hypotelorism (closer than normal midline proximity of the eyes) to fusion of the olfactory placodes, synopthalmia and alobar holoprosencephaly (undivided prosencephalon), which characterise the cyclopic condition (Keeler and Binns, 1968; Incardona et al., 1998). Notably, these disruptions were highly similar to the midfacial and brain malformations observed in *Shh*-deficient mouse embryos (Chiang et al. 1996) and human patients with mutations in the

177

SHH gene (Roessler et al. 1996). Eventually, it was demonstrated that the teratogenic effects of cyclopamine were caused by attenuation of Hh signalling, which was evidenced by perturbations of the dorsal-ventral patterning of the neural tube and somites in chicken embryos (Cooper et al. 1998; Incardona et al. 1998), and cyclopia, loss of the horizontal myoseptum and pectoral fin patterning disruptions in zebrafish (Neumann et al. 1999).

Cyclopamine exerts its antagonistic effects on Hh signalling by direct binding to the transmembrane heptahelical bundle of smoothened (SMO), which is thought to trigger inhibitory conformational changes in SMO structure (Chen et al., 2002). Although, it remains elusive how cyclopamine mechanistically affects downstream functions of smoothened, it appears this is not through inhibition of SMO trafficking to the primary cilium (Wilson et al., 2009).

Multiple studies in zebrafish have effectively used cyclopamine to down-regulate the Hh signalling pathway (Neumann et al. 1999; Neumann et al. 2000; Wolff et al. 2003; Gering et al. 2005). For instance, Wolff et al. (2003) showed that increasing concentrations of administered cyclopamine lead to perturbations in myotomal cell fate acquisition of progressively more superficial fibres. For instance, the differentiation of the medially-located muscle pioneers, which are most sensitive to Hh signalling, can be blocked by 5 uM cyclopamine, whereas the differentiation of the superficial slow fibers (SSFs) is only affected at 20-30 uM cyclopamine (Wolff et al. 2003), indicating that muscle pioneer progenitors require higher levels of Hh activity than the SSF progenitors (Jackson and Ingham, 2013).

8.2.2. Results

Prior to cyclopamine treatment of *a1-NSE::EGFP*-injected zebrafish, I first tested the potency of the cyclopamine stock solution I used. To do this, I divided a population of wild type (AB strain) zebrafish embryos at the 30%-epiboly stage (equivalent to ~ 5 hpf stage) into two groups. One group (n=193) was incubated in 50 μM cyclopamine solution in embryo medium, while the second group (n=227) served as a negative control where, instead of cyclopamine an equivalent volume of 98% ethanol was added to the embryo medium. The experiment was terminated after 24 hours, or at the prim-16 stage (31 hpf) followed by analysis of the embryos' morphology.

Notably, the cyclopamine-treated embryos displayed typical morphological abnormalities, indicating attenuated Hh signalling, such as microcephaly, ocular

hypotelorism (medial ocular proximity), aberrant shape of the prosencephalon, deficient retinal pigmentation and U-shaped myotomes (Brand et al. 1996; Chen et al. 2001) (Figure 8.1E, F, G and H). In contrast, all control embryos appeared normal (Figure 8.1A, B, C and D). These preliminary observations indicate that 50 uM cyclopamine disrupts effectively Hh signalling in zebrafish embryos. Thus, similar concentration can be used to treat *a1-NSE::EGFP*-injected embryos.

Figure 8.1. *Treatment of zebrafish embryos with 50 μM cyclopamine is sufficient to disrupt Hh signalling.* (A-D) control embryos (n=193) and (E-H) cyclopamine-treated embryos (n=227) were analysed at 31hpf for the presence of Hh-mediated effects. Treatment of wild type fish with 50uM cyclopamine effectively disrupts Hh signalling as indicated by microcephaly (*double-headed arrow* in E), lack of retina pigmentation (*arrows* in F), aberrant prosencephalon morphology (*arrowheads* in G) and U-shaped somites (an asterisk and outline of an individual somite in H). In contrast, the control treated embryos showed normal phenotype (A-D).
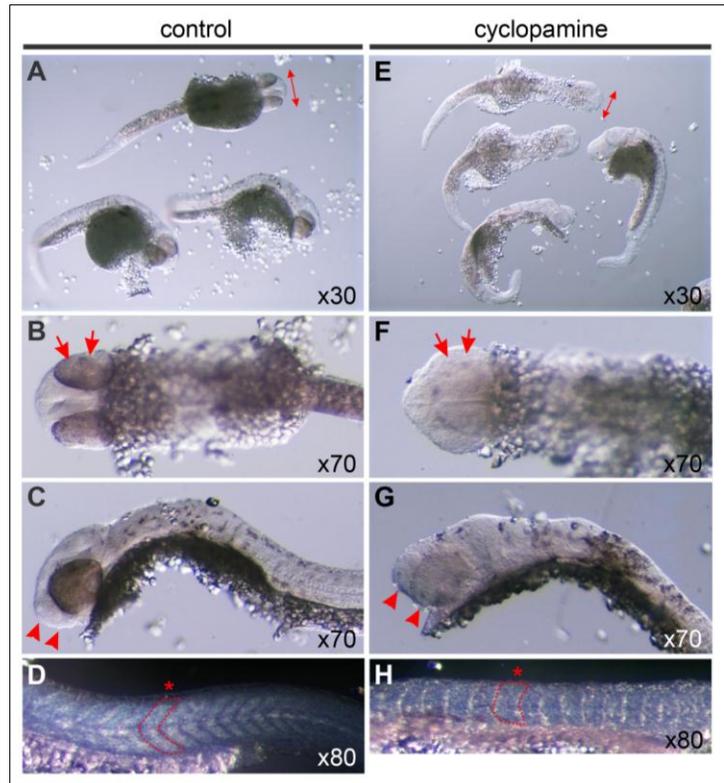
In order to test whether cyclopamine treatment affects the activity of a1-NSE, I microinjected the *a1-NSE::EGFP* plasmid together with *Tol2* mRNA in 1-cell stage wild type zebrafish embryos (see Chapter 7). Approximately five hours after injection, at the 30%-epiboly stage, the total population of injected embryos was divided in two groups. One group was incubated in 50 μM cyclopamine solution in E3 embryo medium, while the second group served as a negative control and incubated with an equivalent volume of 98% ethanol added to the embryo medium. The treatment was carried out for 24 hours until the prim-16 stage, or 31 hpf, when the experiment was terminated, and the fish embryos were fixed and prepared for analysis. The results of the experiment are shown in Table 8.1 and Figures 8.2 and 8.3.

Remarkably, most of the cyclopamine treated embryos exhibited complete absence of EGFP signal; instead, 6.5% showed only fragmentary (patchy) EGFP staining with reduced intensity remaining only in the anterior CNS (Figure 8.2B and B' and Figure 8.3). Importantly, none of the embryos displayed the specific a1-NSE-driven EGFP expression pattern restricted to the ventral half of the neural tube, and extending from the tip of the rostral diencephalon to the caudal spinal cord (see also Figure 7.4 and 7.5), to which I refer here as the "normal" or "wild type, wt" expression pattern (compare Figure 8.2A and A' with Figure 8.2B and B'; Figure 8.3). In contrast, 63.1% of the control embryos exhibited normal expression of EGFP in the ventral half of the brain, while only 4.6% of control embryos showing patchy/reduced reporter expression (Figure 8.2A, A'; Figure 8.3).

These results indicate that cyclopamine interferes with the activity of the murine a1-NSE in transient transgenic zebrafish embryos, and strongly suggest that the interference is due to failure in the activation of the Hh signalling pathway.

	repeat 1		repeat 2	
	control	cyclopamine	control	cyclopamine
Total No. of embryos	244	187	213	205
No. of embryos with normal expression	159	0	130	0
No. of embryos with patchy/reduced expression	11	16	10	9
% embryos with normal expression	65.16	0	61.03	0
% embryos with reduced expression	4.51	8.56	4.69	4.39
% embryos with no expression	30.33	91.44	34.28	95.61

Table 8.1. *Numbers and percentages of control and cyclopamine-treated zebrafish embryos with normal or patchy/reduced a1-NSE-driven EGFP expression from two independent replicate experiments.*
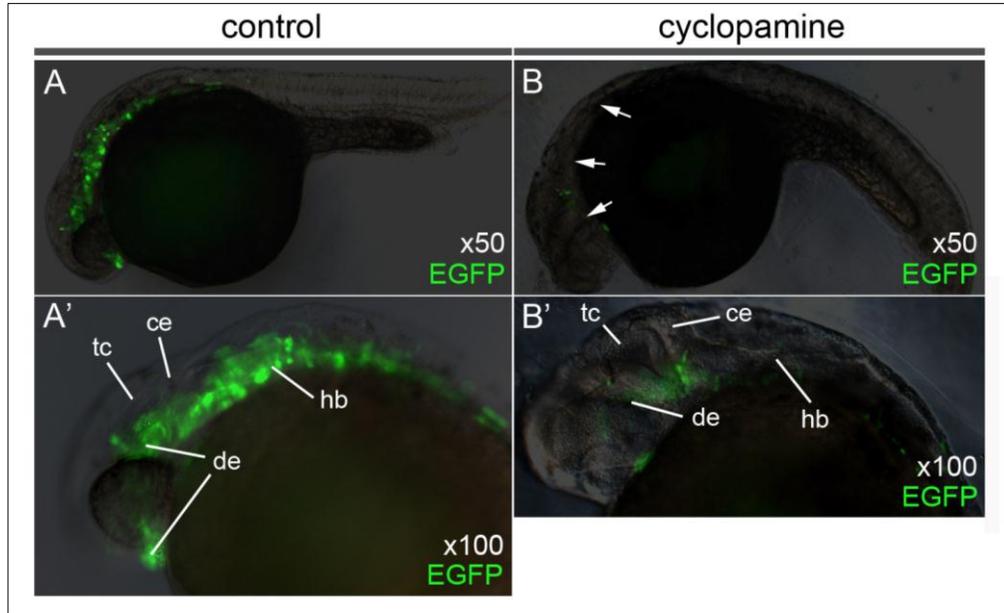
Figure 8.2. *Treatment of zebrafish embryos with 50 μM cyclopamine disrupts a1-NSE's activity.* EGFP fluorescence from control (A, A') and cyclopamine treated zebrafish embryos (B,B') at 31 hpf. Note the severe reduction in EGFP expression in the brain of cyclopamine-treated fish (white arrows B). Higher magnification reveals a small patch of residual EGFP signal in the ventral mesencephalon and anterior hindbrain (B'). Abbreviations: ce, cerebellum; de, diencephalon; hb, hindbrain; tc, tectum.
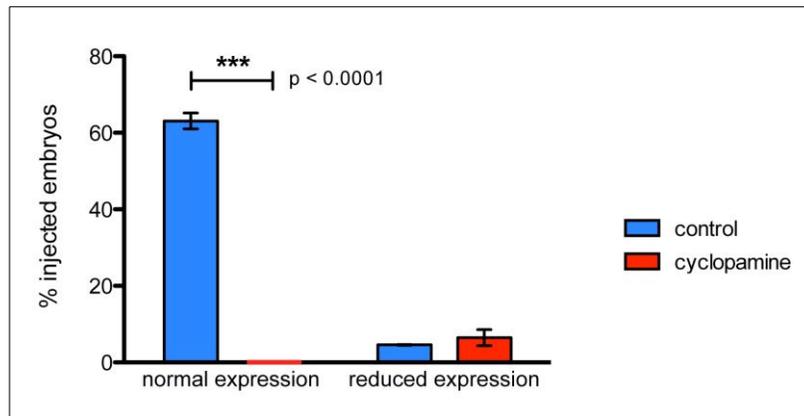
Figure 8.3. *Treatment of zebrafish embryos with 50 μM cyclopamine disrupts a1-NSE's activity (Quantitative Analysis).* Graphical representation of the results from two replicate experiments showing the percentages of injected embryos with normal and reduced/patchy pattern of a1-NSE-directed EGFP expression from both control (blue-colored columns) and cyclopamine-treated groups (red-colored columns). "Normal expression pattern" of EGFP is defined in the text. Note that the percentage of embryos with complete absence of EGFP signal in the neural tube in both the control and cyclopamine-treated groups is not indicated on the graph. Importantly, the mean number of control fish with normal EGFP expression is significantly higher than the mean number of cyclopamine-treated fish with normal EGFP pattern (which equal zero), as shown statistically by unpaired two-tailed t-test in GraphPad Prism. The black bars above the columns represent the standard error of the mean (SEM) for each column.

8.3. Smoothened is required for the activity of a1-NSE

Smoothened is an essential component of HH signal pathway transduction that is negatively modulated by patched in the absence of HH (Beachy et al. 2010; Yavari et al. 2010). Therefore, an alternative and genetic approach to test whether a1-NSE function depends on activated Hh signalling is to inject the *a1-NSE::EGFP* construct into smoothened-deficient zebrafish embryos, which are characterised by abrogated Hh signalling (Barresi et al. 2000; Chen et al. 2001).

8.3.1. Introduction.

Smoothened (SMO) is a seven-transmembrane domain containing protein that is conserved across eumetazoans and is required for the response of cells to hedgehog signals (Alcedo et al. 1996; van den Heuvel and Ingham 1996). Studies in *Drosophila* have shown that the stability and subcellular localisation of smoothened are regulated by patched in a non-stoichiometric way (Denef et al. 2000; Ingham et al. 2000), perhaps by modulating the transport of small lipid molecule activators or inhibitors of Smo (Taipale et al. 2002). Binding of Hedgehog to Patched causes phosphorylated Smo to move from intracellular membranes to the cell surface where it activates the Hh signalling pathway, and forced localisation of Smo to the plasma membrane leads to constitutive activation of Hh target genes transcription (Nakano et al. 2004; Zhu et al. 2003). Studies in mice and zebrafish demonstrated that vertebrate smoothened is trafficked to the primary cilium upon engagement of patched by HH and this translocation to the cilium is required for activation of HH signalling by SMO (Corbit et al. 2005; Kim et al. 2010; Glazer et al. 2010).

The zebrafish smoothened (*smo*) gene was first described in analyses of the *smu*[-/-] (**s**low-**mu**scle-omitted) mutant fish embryos, which displayed complete loss of *eng*[+] muscle pioneers, severe reduction in the number of slow muscle fibers, ventrally curved tails, U-shaped somites, partial cyclopia, loss of cranial cartilages, reduced pectoral fin size, lack of secondary motor neurons, loss of the lateral floor plate and parts of the ventral forebrain (Barresi et al. 2000; Varga et al. 2001). All *smu* mutant defects were remarkably similar to those observed in Hh signalling deficient fish mutants, like the *syu* (*sonic you*) and *yot* (*you-too*), encoding *shha* and *gli2a*, respectively (Karlstrom et al. 1999; Schauerte et al. 1998), and it was eventually demonstrated that *smu* encodes the zebrafish smoothened ortholog (Varga et al. 2001).

Another study using retroviral mutagenesis in zebrafish obtained two loss-of-function alleles of *smo*, *smo^hi229* and *smo^hi1640*, both of which have a 6 kb pro-viral insertion in the first exon of the *smo* gene at positions -110 and +220 relative to the start codon, respectively (Amsterdam et al., 1999; Chen et al., 2001). The two alleles are considered null as both RT-PCR and *in situ* hybridisation in 26 hpf embryos of either mutant failed to detect the presence of *smo* mRNA (Chen et al. 2001). Importantly, both *smo^hi229* and *smo^hi1640* mutants displayed morphological and gene expression abnormalities that are indicative of perturbed Hh signalling, like U-shaped somites, reduced floor plate and *foxa2* expression, partial cyclopia, microcephaly, agenesis of cranial cartilages, almost complete absence of *ptch1* expression, lack of *myoD* expression in the paraxial mesoderm, absence of muscle pioneers, reduced number of slow muscle fibers at 24 hpf, etc (Chen et al. 2001). Subsequently, the *smo^hi1640* mutant has been widely used in multiple studies (Aanstad et al. 2009; Cunliffe 2004; Elworthy et al. 2008). For instance, Bergeron et al. (2008) employed a combination of DNA microarrays and *in situ* hybridisation using *smo^hi1640* and other mutants to uncover genes previously unknown to be Hh-responsive, such as *neuroD*, *follistatin*, *wt1a*, *irx1b* and *claudin b*, and genes negatively controlled by Hh signalling such as *smarca2* (Bergeron et al. 2008).

Based on the results from previous studies, I decided to take advantage of the *smo^hi1640* mutant zebrafish line (Chen et al. 2001) as an *in vivo* system to address a1-NSE's requirement for active Hh signalling.

8.3.2. Results.

In order to test whether the activity of a1-NSE requires the presence of intact *smoothened* (*smo*), I co-injected the *a1-NSE::EGFP* plasmid with *Tol2* mRNA and 400 ng/μL *H2B-mCherry* mRNA (which served as a marker to control for injection efficiency) into 1-cell stage zebrafish embryos that were obtained from incrosses of *smo^hi1640/+* parents, and all subsequent analyses were performed at the 31 hpf stage.

As expected for typical Mendelian inheritance, the injected embryos segregated into two phenotypic classes: approximately ¾ of the fish displayed a wild type phenotype and these were assumed to represent embryos with *smo^+/+* or *smo^hi1640Tg/+* genotype (hereafter named "siblings") (Table 8.2 and Figure 8.4B and C), while the remaining ¼ of the embryos exhibited U-shaped somites, reduced head size and mild cyclopia - morphological defects

that are observed in the *smo*[hi1640Tg/hi1640Tg] mutant embryos and are also a general characteristic of disrupted Hh signalling pathway (Brand et al. 1996; Chen et al. 2001) (Figure 8.4E and F). Therefore, the latter embryos are considered to represent animals with the *smo*[hi1640Tg/hi1640T g]genotype, which are loss-of-function *smo* mutants (hereafter named *smo*[-/-]; Table 8.2) (Chen et al. 2001).

Interestingly, none of the *smo*[-/-] embryos displayed a normal pattern of EGFP expression (as defined earlier in the text). Instead, 96.2% of these embryos lacked any EGFP expression, whereas the remaining 3.8% *smo*[-/-] embryos showed only patchy and reduced signal in the brain (Figure 8.4D and Figure 8.5). In contrast, nearly 67% of the siblings featured a normal expression of EGFP in the brain, with only 2% of sibling embryos displaying patchy/reduced expression in that domain (Figure 8.4A and Figure 8.5). The remaining 31% were not EGFP-positive (data not shown). The consistent lack of EGFP expression in *smo*[-/-]embryos is not due to failure of injection as evidenced by the presence of H2B-mCherry staining in the cell nuclei in both wild type and mutant embryos (Figure 8.4B and E). H2B-mCherry was chosen as an injection control as it incorporates efficiently into nucleosome core particles, due to the human histone 2B moiety, while the mCherry fluorescent protein enables *in vivo* cell labelling without perturbations of the cell cycle and interference with normal development (Kanda et al. 1996; Shaner et al. 2004).

	repeat 1		repeat 2	
	smo[+/+], *smo*[+/-]	*smo*[-/-]	*smo*[+/+], *smo*[+/-]	*smo*[-/-]
Total No. of embryos	112	34	156	41
No. of embryos with normal expression	77	0	102	0
No. of embryos with patchy/reduced expression	2	1	3	2
% embryos with normal expression	68.75	0	65.38	0
% embryos with reduced expression	1.79	2.94	1.92	4.88
% embryos with no expression	29.46	97.06	32.7	95.12

Table 8.2. *Numbers and percentages of smo*[+/+]*, smo*[+/-]*and smo-/- zebrafish embryos with normal or patchy/reduced a1-NSE-driven EGFP expression from two independent replicate experiments.*
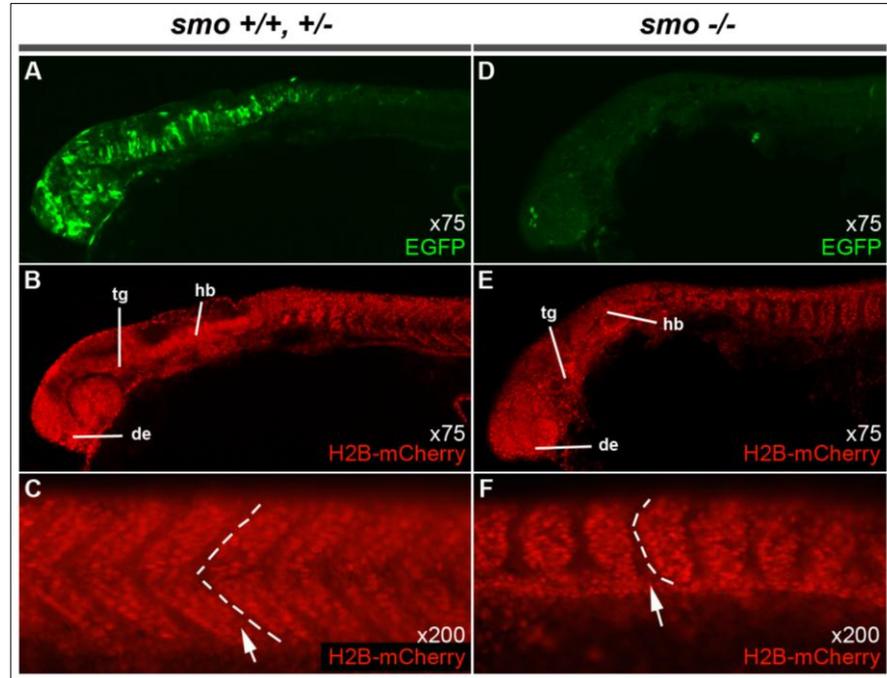
Figure 8.4. *smoothened (smo) is required for the activity of a1-NSE.* a1-NSE-driven EGFP expression is absent in *smo*[-/-] embryos (D) compared to *smo*[+/+], *smo*[+/-] embryos (A) at 31 hpf. (B, C) and (E, F) – antibody staining of the positive control H2B-mCherry indicates similar injection efficiency in both the *smo*[+/+], *smo*[+/-] and *smo*[-/-] fish. Note the characteristic U-shaped somites in the *smo*[-/-] embryo (white outline and arrow in F) compared to the chevron-, or V-shaped somites in the fish with wild type phenotype (white outline and arrow in C). Note also the abnormal morphology of the anterior diencephalon in (E), which is another feature of *smo*-defficient embryos. Abbreviations: de, diencephalon; hb, hindbrain; tg, tegmentum (ventral half of the midbrain).
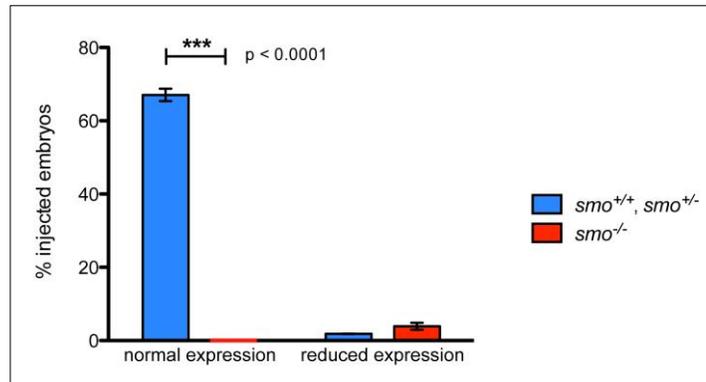
Figure 8.5. *smoothened (smo) is required for the activity of a1-NSE (Quantitative Analysis).* Graphical representation of the results from two replicate experiments showing the percentages of injected *smo* embryos with normal and reduced/patchy pattern of a1-NSE-directed EGFP expression from both *smo*[+/+], *smo*[+/-] (blue-colored columns) and *smo*[-/-] (red-colored columns) embryos. "Normal expression pattern" of EGFP is defined in the text. Note that the percentage of embryos with complete absence of EGFP signal in the neural tube in both "*smo*[+/+], *smo*[+/-]" and "*smo*[-/-]" groups is not indicated on the graph. Importantly, the mean number of *smo*[+/+], *smo*[+/-] fish with normal EGFP expression is significantly higher than the mean number of *smo*[-/-] fish with normal EGFP pattern (which equal zero), as shown statistically by unpaired two-tailed t-test in GraphPad Prism. The black bars above the columns represent the standard error of the mean (SEM) for each column.

185

The results from the injections of *a1-NSE::EGFP* in *smo* mutant zebrafish indicate that intact smoothened is necessary for the activity of a1-NSE. Combined with the effect of treating zebrafish embryos with cyclopamine, this strongly supports the conclusion that active Hh signalling is required for the function of a1-NSE.

8.4. A dominant negative form of PKA causes ectopic activation of a1-NSE.

Protein kinase A (PKA) is a negative modulator of the HH signalling pathway that lies downstream of smoothened, and forms part of the complex that targets GLI proteins for processing through its kinase activity (Tempe et al. 2006). Consequently, loss-of-function of PKA leads to increased and ectopic transcription of Hh target genes (Jiang and Struhl 1995; Concordet et al. 1996). Therefore, to address whether constitutively active Hh signalling is sufficient to activate a1-NSE at ectopic sites in the embryo, I co-injected zebrafish embryos with the *a1-NSE::EGFP* construct and mRNA encoding a dominant negative form of PKA.

8.4.1. Introduction.

PKA is a cAMP-dependent serine-threonine kinase that is expressed in many tissues and is involved in the regulation of multiple cellular processes during growth and embryonic development, response to stress, memory and apoptosis (Song et al. 2003; Taylor et al. 2004). At low levels of cAMP, PKA is a tetrameric holoenzyme (R_2C_2) consisting of two catalytic (C) subunits bound to a dimer of inhibitory regulatory (R) subunits, which renders the complex inactive. When the intracellular concentration of cAMP increases, the regulatory subunits bind cAMP. This destabilises their interaction with the catalytic subunits, and the holoenzyme dissociates into one $R_2(cAMP)_4$ dimer and two monomeric catalytic subunits (Su et al. 1995). The activity of the catalytic monomers is further increased by autophosphorylation, which makes them capabale to phosphorylate their target proteins in the cytosol and nucleus (Voet 2004).

Some of the many targets of PKA are the GLI proteins. It has been shown that *Drososphila* Cubitus interruptus (Ci) and vertebrate GLI proteins contain a cluster of sites downstream of the Zn-finger domain that are sequentially phosphorylated by PKA, glycogen synthase kinase 3 (GSK3) and casein kinase 1 (CK1) (Jiang and Struhl 1998; Tempe et al. 2006). This recruits Slimb/beta-TrCP, a substrate-specific component of the SCF-type E3 ubiquitin ligase, that eventually results in the partial proteasome-mediated removal of the activation-domain-containing C-terminus of Ci/GLI, leaving the N-terminal half, which

contains the repressor and DNA-binding domains of Ci/GLI (Pan et al. 2009; Price and Kalderon 2002; Smelkinson et al. 2007). Thus, in the absence of HH ligand, PKA phosphorylates Ci/GLI, resulting in the generation of Ci^R/GLI^R (repressor forms of Ci/GLI) and repression of HH target gene transcription (Aza-Blanc et al. 1997).

That basal levels of PKA negatively modulate the HH signalling pathway in the absence of HH ligand was first demonstrated in the patterning of *Drosophila* wing and leg imaginal discs, where the hh-responsive genes *dpp* and *wg* are ectopically induced in PKA-deficient clones in the anterior wing/leg disc compartment (Jiang and Struhl 1995; Li et al. 1995; Pan and Rubin 1995). Subsequently, knocking down PKA activity in zebrafish embryos by injection of mRNA encoding a dominant negative form of the regulatory subunit of PKA (dnPKA) resulted in up-regulation and ectopic expression of Hh target genes, like *ptch1, myoD* and *nkx2.2* in the lateral mesoderm and dorsal neural tube (Concordet et al. 1996). This demonstrated the conserved role of PKA in both *Drosophila* and vertebrate HH signalling. In a similar study, it was shown that dnPKA leads to ventralisation of the zebrafish neural tube indicated by the dorsal expansion of the floor plate-restricted *foxa2* mRNA transcription and the elevated and ectopic expression of the motor-neuron specific *islet1* in the hindbrain and anterior spinal cord (Hammerschmidt et al. 1996), which was reminiscent to the ventralising effects of shh overexpression (Hammerschmidt et al. 1996; Krauss et al. 1993; Ungar et al. 1996).

The dominant negative form of PKA used in zebrafish by Concordet et al. (1996) contains a point mutation in the cAMP-binding site of the regulatory subunit (R), which locks the inhibitory R_2 dimer to the catalytic subunits irrespective of elevated cAMP, resulting in severe reduction or loss of catalytic PKA subunit activity (Clegg et al. 1987).

Based on these observations, I decided to use *dnPKA* mRNA injections to up-regulate the Hh signalling pathway in zebrafish embryos and examine whether constitutive activation of the pathway is sufficient to induce *a1-NSE::EGFP* at ectopic locations in the embryo.

8.4.2. Results

In order to examine whether ectopic activation of the Hh signalling pathway was sufficient to direct a1-NSE-driven expression of EGFP, I co-injected the *a1-NSE::EGFP* plasmid with *Tol2* mRNA and with 400 ng/µL *dnPKA* mRNA in 1-cell stage wild type zebrafish embryos, and analysed the injected embryos at the 31 hpf stage. Another group of wild type embryos

served as a negative control where the *dnPKA* mRNA was substituted for an equivalent volume of water.

However, prior to the planned experiment, I tested the efficiency of the *dnPKA* mRNA preparation to up-regulate the Hh signalling pathway by injecting *dnPKA* mRNA into embryos of the *Tg(olig2:EGFP)* transgenic zebrafish line (Shin et al. 2003). This line recapitulates the endogenous pattern of *olig2* expression in the progenitors of motor neurons and oligodendrocytes in the embryonic spinal cord (Shin et al. 2003). *Olig2* is a class II (positive target) gene of SHH in the embryonic neural tube, and it has been demonstrated in zebrafish and mice that HH signalling is both required for and sufficient to up-regulate *Olig2* expression (Lu et al. 2000; Nery et al. 2002; Park et al. 2002). In control *Tg(olig2:EGFP)* fish embryos, the EGFP signal at 31 hpf appears as a narrow longitudinal stripe in the ventral spinal cord (Figure 8.6D), located in close proximity to the sources of Shh (Krauss et al. 1993). Notably, in the embryos that received *dnPKA* mRNA, this domain is expanded dorsally in the neural tube (Figure 8.6H), which is consistent with *Olig2* expression observed in the neural tube of *Wnt1::Shh* transgenic or $Ptch1^{-/-}$; $Hhip1^{-/-}$ knockout mouse embryos with ectopic or up-regulated SHH signalling, respectively (Rowitch et al. 1999; Lu et al. 2000; Jeong and McMahon, 2005). This observation demonstrates the efficiency of the *dnPKA* mRNA preparation to up-regulate the Hh signalling pathway.

Having confirmed the potency of my *dnPKA* mRNA preparation, I proceeded with co-injecting wild type zebrafish embryos with the *a1-NSE::EGFP* plasmid and *Tol2* mRNA in the presence or absence (negative control) of *dnPKA* mRNA, the results of which are presented in Table 8.3 and Figures 8.6 and 8.7. As expected, a large fraction (56%) of the embryos which received *dnPKA* mRNA displayed ectopic expansion of EGFP expression in the dorsal half of the neural tube (Figure 8.6E and G), while only a minor fraction (2.2%) of the *dnPKA* mRNA-injected embryos exhibited normal EGFP pattern restricted to the ventral half of the neural tube (Table 8.3; Figure 8.7). In contrast, none of the control embryos showed ectopic expression of the EGFP reporter in the dorsal half of the neural tube; instead, 65.8% of these embryos displayed normal pattern of EGFP signal, which was specifically restricted to the ventral half of the neural tube (Figure 8.6A and C; Figure 8.7).

	repeat 1		repeat 2	
	control	*dnPKA* mRNA	control	*dnPKA* mRNA
Total No. of embryos	208	111	189	151
No. of embryos with normal expression	125	2	135	4
No. of embryos with ectopic expression	0	55	0	94
% embryos with normal expression	60.1	1.80	71.43	2.65
% embryos with ectopic expression	0	49.55	0	62.25

Table 8.3. *Numbers and percentages of control and dnPKA mRNA-injected zebrafish embryos with normal or ectopic a1-NSE-driven EGFP expression from two independent replicate experiments.*

Figure 8.6. *Blocking PKA function causes ectopic activation of a1-NSE.* a1-NSE-driven EGFP expression pattern in control (A-C) and 400 ng/µL *dnPKA* mRNA-injected (E-G) wild type embryos at 31 hpf. (A) and (E) show anti-EGFP antibody staining of a control and a *dnPKA* mRNA-injected embryo, respectively. (B) and (F) show bright field images of a control and *dnPKA* mRNA-injected embryos, respectively, while (C) and (G) are merged images of A+B and E+F, respectively. Note the ectopic expansion of EGFP expression in the dorsal half of the neural tube of the *dnPKA* mRNA-injected embryo (E; indicated by a white bar in G) as contrasted to the ventrally-restricted normal expression of EGFP in the control embryo (A; indicated by a white bar in C). (H) shows the potency of dnPKA to up-regulate Hh signalling evidenced by the expansion of the EGFP$^+$ domain (white arrows in H) in *olig2:EGFP* zebrafish embryos that received *dnPKA* mRNA as contrasted to the narrow stripe of EGFP expression in control embryos (white arrows in D). Abbreviations: de, diencephalon; hb, hindbrain; te, telencephalon.
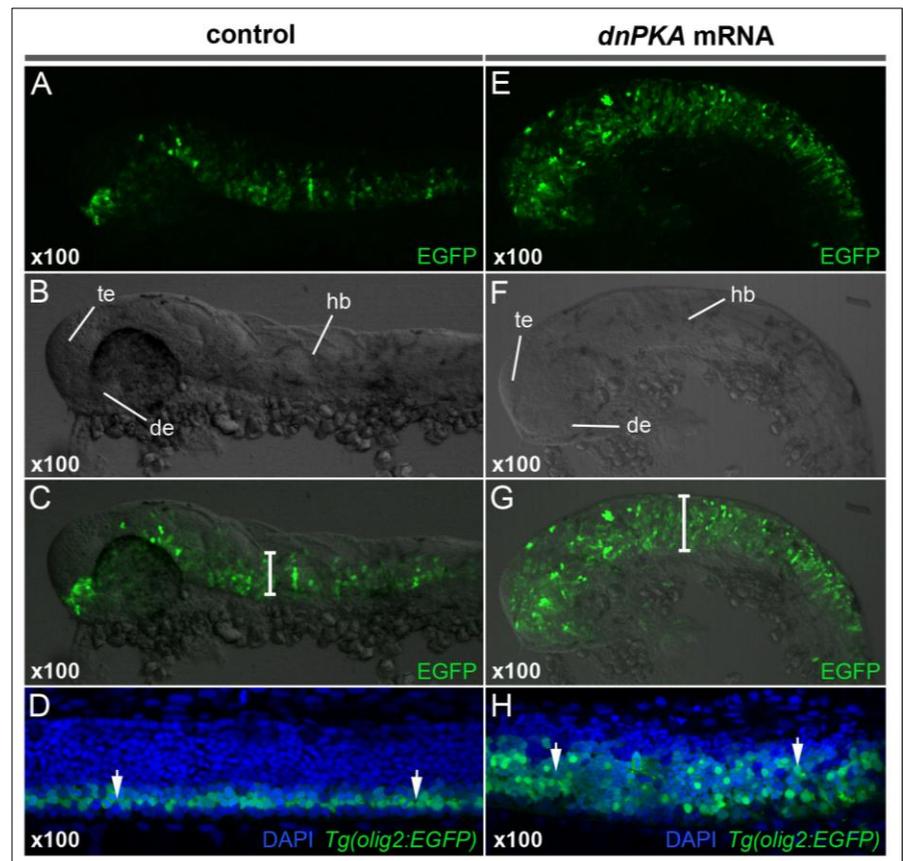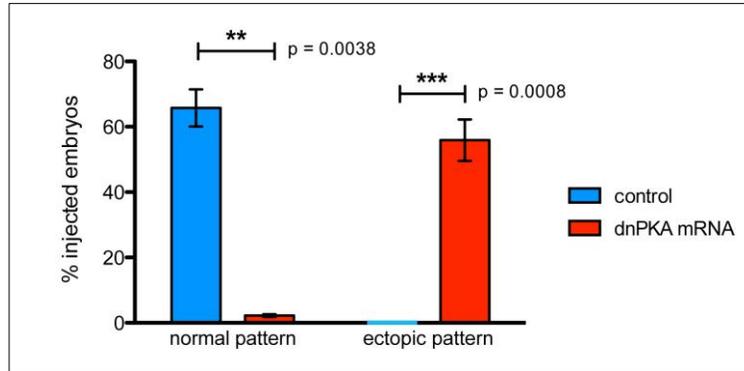
Figure 8.7. *Blocking PKA function causes ectopic activation of a1-NSE (Quantitative Analysis).* Graphical representation of the results from two replicate experiments showing the percentages of embryos with normal and ectopic pattern of a1-NSE-directed EGFP expression from both control (blue-colored columns) and *dnPKA* mRNA-injected (red-colored columns) wild type embryos at 31 hpf. "Normal expression pattern" of EGFP is defined in the text. Importantly, the mean number of *dnPKA* mRNA-injected fish with ectopic EGFP expression is significantly higher than the mean number of controlfish with ectopic EGFP pattern (which equal zero). The opposite is true for fish with normal EGFP expression - the mean number of *dnPKA* mRNA-injected fish with normal EGFP expression is significantly lower than the mean number of control fish with normal EGFP pattern. The unpaired two-tailed t-test in GraphPad Prism was performed to statistically analyse the data. The black bars above the columns represent the standard error of the mean (SEM) for each column.

These findings indicate that dnPKA is sufficient to cause activation of the *a1-NSE::EGFP* construct at ectopic sites in the neural tube of zebrafish embryos, and strongly suggest that the observed effect is due to constitutive up-regulation of the Hh signalling pathway in the dorsal neural tube.

8.5. Overexpression of *shh* mRNA causes ectopic activation of a1-NSE

Although inactivation of PKA function effectively up-regulates the HH signalling pathway leading to ectopic activation of HH-responsive genes, the possibility that disruption in other signalling pathways contributes to the ectopic pattern cannot be excluded. In fact, multiple studies have shown that PKA is involved in some aspects of FGF (Baron et al. 2000), WNT (Chen et al. 2005; Gallegos et al. 2012) and BMP signalling (Ghayor et al. 2009; Liu et al. 2005; Sakai et al. 2006). Therefore, in order to confirm that the ectopic activation of a1-NSE observed in the dorsal neural tube of *dnPKA* mRNA-injected embryos is specifically caused by the constitutive up-regulation of Hh signalling, I co-injected the *a1-NSE::EGFP* plasmid together with zebrafish *shha* mRNA.

8.5.1. Introduction

The zebrafish ortholog of the mouse SHH gene, *shha*, exhibits a dynamic expression pattern with the first transcripts detected in the converging axial mesoderm at 60%-epiboly, and later during somitogenesis (from ~10.5 hpf to 22-24 hpf) in the neural tube from the tip of the diencephalon rostrally to the floor plate caudally (Krauss et al. 1993). By 36 hpf, *shha* mRNAs are reduced in the ventral spinal cord remaining high in the floor plate at the tail bud level, and in the fore- and hindbrain (Krauss et al. 1993). The zebrafish genome encodes another *shh* gene – *shhb* (previously known as *tiggy winkle hedgehog*) that displays similar expression in the neural tube but becomes down-regulated in the axial mesoderm earlier than *shha* (Ekker et al. 1995). The other two members of the zebrafish Hh family – *dhh*, *ihha* and *ihhb* (synonym of *echidna hedgehog*) are predominantly expressed by mesoderm-derived skeletal tissues (Currie et al. 1996; Avaron et al. 2006).

Zebrafish embryos with *shha* deficiency (the *syu* (*sonic you*) group of mutants) feature various defects in somite and neural tube patterning (Schauerte et al. 1998; van Eeden et al. 1996): they lack *eng*[+] muscle pioneers, which leads to the absence of horizontal myoseptum; *myoD* expression in the adaxial cells is strongly reduced; the *fkd4*[+] lateral floor plate cells are absent, but medial floor plate development is unaffected (as indicated by unperturbed expression of *F-spondin2* and *col2a1*); primary and secondary motor neurons are present in *sya* mutants but the extension, branching and patterning of their axons is abnormal (Schauerte et al. 1998). It is interesting that the abnormalities in neural tube patterning in *shha* mutant zebrafish are milder than those of *Shh*-null mouse embryos, which have complete absence of *Foxa2*[+] floor plate cells and motor neurons (Chiang et al. 1996). This might be due to partial compensation by *shhb* and *ihhb* in the zebrafish (Schauerte et al. 1998).

Conversely, constitutive expression of *shha* mRNA in wild type zebrafish embryos leads to alterations in neural tube patterning, such as: ectopic dorsal expansion of *foxa2* expression in the diencephalon, midbrain, hindbrain and anterior spinal cord; dorsal expansion of other Class II genes whose expression in the neural tube is positively-regulated by Hh signaling, like *olig2*, *nkx2.6* and *nkx2.9*; severe dorsal depletion in the expression of Class I target genes, like *pax3*, *pax7*, *pax6a*, *dbx1a* and *dbx2a*, which are normally repressed by Hh only ventrally; failure of brain ventricles formation; absence of the lens and reduced

ocular pigmentation (Guner et al. 2007; Krauss et al. 1993; Ekker et al. 1995). Somitic development is also affected in *shha-* and *ihha*-overexpressing fish embryos: *myoD* mRNA, which is normally restricted to the adaxial cells, expands laterally through the whole width of the paraxial mesoderm, the number of *eng*[+] pioneer cells is strongly increased, while the horizontal myosepta exhibit irregular morphology (Hammerschmidt et al. 1996).

These studies demonstrate that the administration of excessive amounts of *shha* mRNA in early zebrafish embryos results in the constitutive up-regulation of Hh signaling in responsive tissues. Therefore, I decided to use *shha* mRNA to up-regulate Hh signaling and to assess whether this approach affects a1-NSE activity in a similar manner as dnPKA does.

8.4.2. Results

The *a1-NSE::EGFP* plasmid was co-injected with *Tol2* mRNA and 400 ng/µL *shha* mRNA into 1-cell stage wild type zebrafish embryos, and injected embryos were analysed at the 31 hpf stage. Interestingly, and consistent with the result from the dnPKA mRNA-injections, nearly 57% of the fish embryos that received *shha* mRNA exhibited an expansion of EGFP signal in the dorsal half of the neural tube, while 3% displayed a normal pattern of EGFP (Table 8.4; Figure 8.8E, F; Figure 8.9). In contrast, none of the control embryos showed ectopic expression of EGFP in the dorsal neural tube; instead, 61% of those embryos had normal pattern of EGFP signal restricted to the ventral half of the neural tube (Table 8.4; Figure 8.8A, B: Figure 8.9). These results support my previous conclusion and indicate that constitutively active Hh signalling is sufficient to activate the *a1-NSE::EGFP* construct at ectopic locations in the neural tube.

	repeat 1		repeat 2	
	control	*shh* mRNA	control	*shh* mRNA
Total No. of embryos	200	172	157	192
No. of embryos with normal expression	124	7	94	4
No. of embryos with ectopic expression	0	98	0	109
% embryos with normal expression	62	4.07	59.87	2.08
% embryos with ectopic expression	0	56.98	0	56.77

***Table 8.4.** Numbers and percentages of control and shha mRNA-injected zebrafish embryos with normal or ectopic a1-NSE-driven EGFP expression from two independent replicate experiments.*
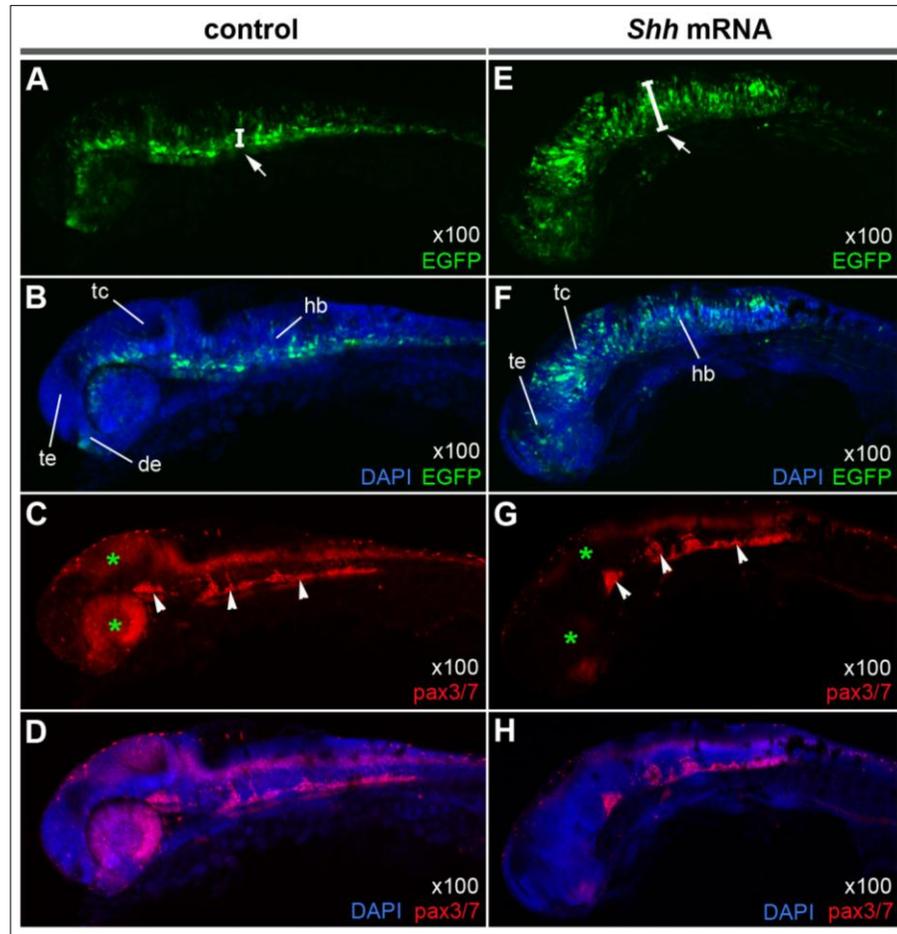
Figure 8.8. *Over-expression of shha causes ectopic activation of a1-NSE.* a1-NSE-driven EGFP expression pattern in wid type control embryos (A-D), and in wild type embryos injected with 400 ng/µL *shha* mRNA (E-H), at 31 hpf. (A) and (E) show anti-EGFP antibody staining of a control and a *shha* mRNA-injected embryo, respectively. (B) and (F) show merged DAPI and anti-EGFP staining of a control and a *shha* mRNA-injected embryo, respectively. (C) and (G) indicate anti-Pax3/Pax7staining in control and shha mRNA injected embryos, respectively, while (D) and (H) are the corresponding merged images of DAPI and anti-Pax3/Pax7 staining. Note the ectopic expansion of EGFP expression in the dorsal half of the neural tube of the *shha* mRNA-injected embryo (indicated by a white bar in E) as contrasted to the ventrally-restricted normal expression of EGFP in the control embryo (indicated by a white bar in A). (G) and (H) show the potency of excess shha to up-regulate Hh signalling evidenced by the down-regulation of Pax3/Pax7 expression in the optic tectum and eye (green asterisks) as contrasted to normal expression of Pax3/Pax7 in control embryos (C and D). White arrowheads in (C) and (G) indicate Pax3/Pax7 expressing neural crest- and/or placode-derived glial and neuronal precursors of the cranial ganglia (Minchin and Hughes 2008; Schlosser and Ahrens 2004). Abbreviations: de, diencephalon; hb, hindbrain; tc, optic tectum; te, telencephalon.
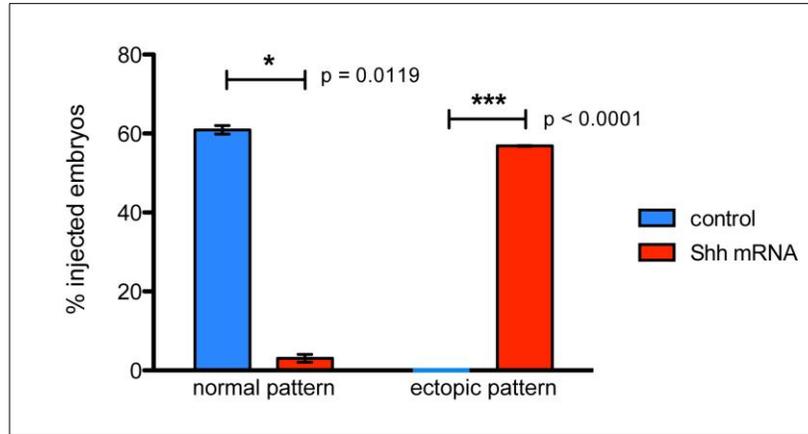
Figure 8.9. Over-expression of shha causes ectopic activation of a1-NSE (Quantitative Analysis). Graphical representation of the results from two replicate experiments showing the percentage of embryos with normal and ectopic pattern of a1-NSE-directed EGFP expression from both control (blue-colored columns) and *shha* mRNA-injected (red-colored columns) wild type embryos at 31 hpf. "Normal expression pattern" of EGFP is defined in the text. Importantly, the mean number of *shha* mRNA-injected fish with ectopic EGFP expression is significantly higher than the mean number of control fish with ectopic EGFP pattern (which equal zero). The opposite is true for fish with normal EGFP expression - the mean number of *shha* mRNA-injected fish with normal EGFP expression is significantly lower than the mean number of control fish with normal EGFP pattern. The unpaired two-tailed t-test in GraphPad Prism was performed to statistically analyse the data. The black bars above the columns represent the standard error of the mean (SEM) for each column.

Changes in the expression of Pax3/Pax7 in *shha* mRNA-injected fish were used to monitor the effect of *shha* mRNA. Indeed, *pax3* and *pax7* are Class I Hh responsive genes, which are restricted to the dorsal half of the neural tube by repressive Shh signals, and participate in the specification of dorsal interneuron progenitors (Bergeron et al. 2008; Guner et al. 2007; Liem et al. 1995; Luo et al. 2006; Ericson et al. 1997). Notably, zebrafish embryos that received *shha* mRNA displayed strong reduction in Pax3Ppax7 expression in the optic tectum, eyes and hindbrain (Figure 8.8G and H), such that expression was limited to the dorsal-most compartments of the neural tube. This result is consistent with previous studies and clearly demonstrates the efficiency of my *shha* RNA preparation to up-regulate the Hh signalling pathway.

8.6. The activity of a1-NSE requires intact GLI binding motifs

As described in Chapter 1, GLI transcription factors mediate the response to Hh signals in all vertebrates (Hui and Angers 2011). GLI proteins bind DNA at the consensus motif sequence 5'-TGGGTGGTC-3' and operate as bi-functional transcription factors – either as GLIR (repressor) or GLIA (activator), depending on the absence or presence of HH signals, respectively (Kinzler and Vogelstein 1990; Sasaki et al. 1999).

194

The zebrafish genome encodes four *gli* genes: *gli1*, *gli2a*, *gli2b* and *gli3*. A major site of Gli function is the developing CNS, where all four *gli* paralogs show complex expression patterns. At early stages, *gli2a* and *gli3* mRNA transcription initiates throughout the neuroectoderm but by 15 hpf they are restricted to the dorsal neural tube, whereas *gli1* displays a complementary expression pattern with high mRNA levels in the ventral neural tube and absence of expression in the floor plate and dorsal neural tube (Karlstrom et al. 1999; Karlstrom et al. 2003; Tyurina et al. 2005; Vanderlaan et al. 2005). Mutations and morpholino-mediated knockdowns of zebrafish *gli* result in multiple abnormalities of CNS development, as detailed in Chapter 1.

Taking together the facts that all four zebrafish *gli* genes are transcribed in the neural tube and that deficiency in their expression leads to aberrations in the transcription of Hh-responsive genes and neuronal development, I decided to test whether the CNS-specific activity of a1-NSE depends on the five GLI binding motifs within its sequence. If the GLI binding motifs are essential for a1-NSE function, then their inactivation should impair a1-NSE-mediated EGFP expression.

8.6.1. Results

I used the SequenceShaper *in silico* tool from the Genomatix Software Suite (Cartharius et al. 2005) to design base pair substitutions in each of the five GLI binding motifs within a1-NSE. The parametres of SequenceShaper were adjusted to avoid the elimination or creation of binding motifs for other transcription factors. The GLI motifs received from 2 to 3 point mutations each such that the core and matrix similarity scores of every GLI motif were severely reduced, which is a predictor of decreased affinity of the sequence to its binding protein (as indicated in Table 8.5). The mutated a1-NSE, with all five GLI binding motifs harbouring substitutions is referred to as *mut a1-NSE* in this study, while the un-modified a1-NSE is called *wt a1-NSE*. The so designed mut a1-NSE was obtained by customised gene synthesis from Genscript, and cloned into the EGFP-reporter vector.

In order to test whether the activity of a1-NSE depends on the GLI binding motifs, I co-injected the *mut a1-NSE::EGFP* plasmid with *Tol2* mRNA and 400 ng/μL *H2B-mCherry* mRNA into 1-cell stage wild type zebrafish embryos, and analysed the embryos at the 31 hpf stage. Concomitantly, a second group of wild type embryos received the unmodified *wt-a1-NSE::EGFP* construct and these served as control.

motif	sequence	core / matrix similarity	position in a1-NSE
consensus	TGGGTGGTC	-	-
motif 1	ACAGTGGGTGGCGAG	1 / 0.884	57 - 71
motif 1 *	ACACTGAGTGTCGAG	0.225 / 0.157	57 - 71
motif 2	ATCCAGGGTGGTCTA	1 / 0.937	280 - 294
motif 2 *	ATGCATGTTGGTCTA	0.016 / 0.155	280 - 294
motif 3	GAGATGGGAGGTGGA	1 / 0.879	585 - 599
motif 3 *	GAGAAGGTAGGTGGA	0.016 / 0.258	585 - 599
motif 4	ACTCTGGGTGGCATG	1 / 0.890	616 - 630
motif 4 *	ACTCTAGGTGACATC	0.257 / 0.161	616 - 630
motif 5	CTGCTAGGTGGTCAC	1 / 0.895	644 - 658
motif 5 *	ATGCTAGGTAGTGAC	0.012 / 0.140	644 - 658

Table 8.5. Sequence features of the five GLI binding motifs within a1-NSE, indicating the changes caused by the introduced mutations in mut a1-NSE. Here are displayed the sequences of the five GLI binding motifs according to MatInspector, together with the consensus GLI motif (Kinzler and Vogelstein 1990). For comparative purposes, the regions of the predicted motifs that overlap the consensus motif are shaded in *light blue*; *yellow* shading indicates the positions in the predicted motifs that differ from the consensus sequence. Motif names with *asterisks* label the mutated motifs. Positions in *red* indicate the base substitutions in the mutated motifs, as designed by SequenceShaper. A *green* bar underlines the core of each predicted motif (according to MatInspector). Since a particular transcription factor usually binds to a range of related sequences, a weight matrix is generated for each set of related sequences and represents "the complete nucleotide distribution for each single position" in the set (Cartharius et al., 2005). The "core" of a motif, in the context of SequenceShaper and MatInspector (Cartharius et al., 2005), is a sequence of 4 consecutive base-pairs that is conserved in at least 90% of all sequences used to define a weight matrix. The highest core and matrix similarity score that a predicted motif can obtain is a 1, and for the five wild type GLI motifs in a1-NSE these scores range from 1 to 0.884, which is sufficiently high. In contrast, each of the mutated GLI motifs features core and matrix similarity scores in the range of 0.012 to 0.258, which indicates that the mutated sequences are highly dissimilar from the core and weight matrix of GLI factors and this predicts decreased affinity (or lack of thereof) of the mutated Gli motifs for their cognate GLI proteins. Note that when bound to DNA, GLI factors protect a region of 23-24 base pairs, as shown by DNase footprinting (Kinzler and Vogelstein 1990). However, most of the GLI protein-DNA contatcts are established within the 9-mer regions (shaded in *light blue*), with some additional contacts as far as 5 base pairs upstream of the 9-mer sequence (Pavletich and Pabo 1993), which is perhaps the reason why MatInspector and SequenceShaper operate with extended GLI motif sequences (15 bp), instead of the 9-bp region only.

Remarkably, none of the embryos that received the mut-a1-NSE presented a normal EGFP expression pattern. Instead, 1% embryos showed only mosaic (patchy) and reduced EGFP expression (Table 8.6 and Figures 8.10C and 8.11). In contrast, most control embryos exhibited normal pattern of EGFP expression in the ventral half of the neural tube, with only 3% showing reduced/patchy EGFP expression (Table 8.6 and Figures 8.10A and 8.11). The presence of H2B-mCherry in all/most cell nuclei of both mut-a1-NSE- and control-injected embryos indicates that the absence of EGFP expression in mut-a1-NSE-injected fish is not due to injection inefficiency (Figure 8.10B and D).

These results demonstrate that the activity of a1-NSE requires intact GLI binding motifs.

	repeat 1		repeat 2	
	wt-a1-NSE (control)	mut-a1-NSE	wt-a1-NSE (control)	mut-a1-NSE
Total No. of embryos	273	160	199	264
No. of embryos with normal expression	126	0	119	0
No. of embryos with reduced expression	4	1	9	4
% embryos with normal expression	46.15	0	59.80	0
% embryos with reduced expression	1.47	0.63	4.52	1.52
% embryos with no expression	52.38	99.37	35.68	98.48

Table 8.6. *Numbers and percentages of control and mut a1-NSE::EGFP-injected zebrafish embryos with normal, reduced and absent a1-NSE-driven EGFP expression from two independent replicate experiments.*
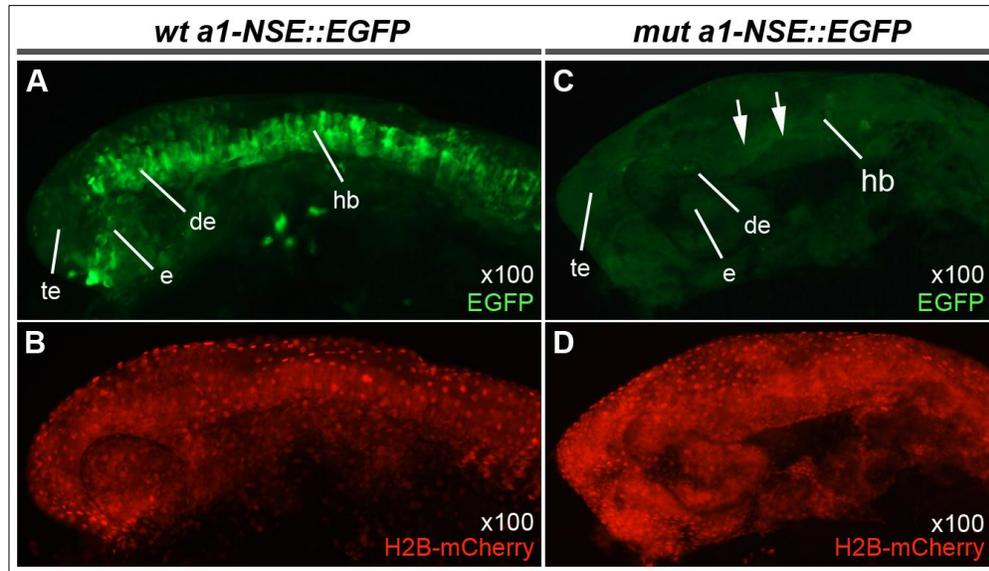
Figure 8.10. *Intact GLI-binding motifs are essential for a1-NSE's activity.* Immunofluorescent images of wild type embryos injected with 30 ng/µL wt a1-NSE::EGFP plasmid (A, B), and wild type embryos injected with 30 ng/µL *mut a1-NSE::EGFP* plasmid (C, D), at 31 hpf. In both cases, the reporter plasmids were co-injected with 400 ng/µl *H2B-mCherry* mRNA. Note the inability of mut a1-NSE to activate EGFP expression in the neural tube (white arrows in C), in contrast to the situation in control embryos (A). Importantly, this difference is not caused by injection inefficiency as both groups express H2B-mCherry (B and C). Abbreviations: de, diencephalon; e, eye; hb, hindbrain; te, telencephalon.
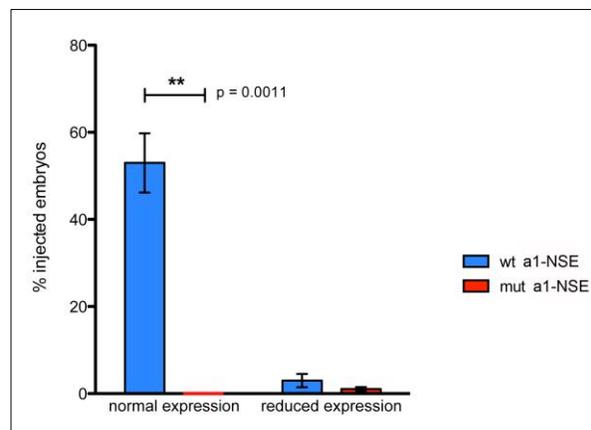
Figure 8.11. *Intact GLI-binding motifs are essential for a1-NSE's activity (Quantitative Analysis).* Graphical representation of the results from two replicate experiments showing the percentage of embryos injected with wt a1-NSE::EGFP plasmid (control) and of those embryos injected with mut a1-NSE. "Normal expression pattern" of EGFP is defined in the text. Note that the percentage of embryos with complete absence of EGFP signal in the neural tube in both the "wt a1-NSE" and "mut a1-NSE" groups is not indicated on the graph. Importantly, the mean number of control zebrafish with normal EGFP expression is significantly higher than the mean number of "mut a1-NSE" zebrafish with normal EGFP pattern (which equal zero), as shown statistically by unpaired two-tailed t-test in GraphPad Prism. The black bars above the columns represent the standard error of the mean (SEM) for each column.

8.7. Intact GLI binding motifs are required for ectopic activation of a1-NSE by dnPKA

There is evidence that in some systems HH signalling is not mediated by the GLI transcription factors. This phenomenon is commonly known as "non-canonical" Hh signalling of which two types have been recognised: type I is mediated by patched but does not involve smoothened, whereas type II non-canonical HH signaling operates through smoothened but independently of GLI (Brennan et al. 2012). For instance, Ptch1 exhibits smoothened-independent pro-apoptotic effects in 293T cells and anti-mitotic effects in urinary and epidermal epithelia in the absence of HH ligands (Adolphe et al. 2006; Chinchilla et al. 2010; Jenkins et al. 2007). An example of type II non-canonical HH signalling is the promotion of vascular tubulogenesis by HH ligands in human umbilical vein and microvascular endothelial cells, which appears to be independent from GLI-induced transcription, but relies on smoothened and RAC1-mediated regulation on RHOA activity (Chinchilla et al. 2010). Conversely, some GLI-dependent cellular responses may be independent of HH signalling. Dennler et al. (2007) showed that transcription of GLI1 (a positive target gene of HH signalling) in cyclopamine-treated human dermal fibroblasts is uncoupled from SHH-smoothened but is triggered by SMAD3-dependent activation of GLI2, instead. Similarly, it was demonstrated that the cytokine osteopontin (OPN) regulates the activation and subcellular distribution of GLI1 via Akt-GSK3β signalling in human neoplastic cells and this process is insensitive to cyclopamine inhibition of SMO (Das et al. 2013).

Therefore, I tested whether the ectopic effects of dnPKA on a1-NSE activity require intact GLI motifs. I predicted that if there was such requirement, then *dnPKA* mRNA would fail to trigger *EGFP* expression from the mut-a1-NSE.

8.7.1. Results

In order to test the dependence of dnPKA-induced ectopic activation of a1-NSE on intact GLI binding motifs, I performed three different parallel injections at the 1-cell stage of wild type zebrafish embryos. One group of embryos received the *wt-a1-NSE::EGFP* construct alone; a second group received the *wt-a1-NSE::EGFP* construct together with 400 ng/μL *dnPKA* mRNA; and a third group of embryos were injected with the *mut-a1-NSE::EGFP* construct together with 400 ng/μL *dnPKA* mRNA. All embryos received 20 ng/μL *Tol2*

mRNA, as well. Then, the embryos were analysed at the 31 hpf stage (Table 8.7; Figures 8.12 and 8.13).

Interestingly, dnPKA was unable to ectopically activate *EGFP* driven by the mut a1-NSE, either in the dorsal or ventral halves of the neural tube, in contrast to the embryos that received wt a1-NSE plus *dnPKA* mRNA, which displayed ectopic EGFP expression in the dorsal neural tube (compare Figure 8.12C and E). It is unlikely that the failure of EGFP expression was caused by issues during microinjection for expression of Pax7 appeared reduced in the optic tectum and dorsal hindbrain in the embryos that received *mut a1-NSE::EGFP* with *dnPKA* mRNA (Figure 8.12B, D and F), suggesting that the dominant negative form of PKA was able to up-regulate the Hh signalling pathway. These observations suggest that the effects of Hh signalling on a1-NSE function are directly mediated by the GLI binding motifs, as mutations in these sites render a1-NSE refractory to even constitutively-active Hh signalling.

	wt-a1-NSE	wt-a1-NSE + *dnPKA* mRNA	mut-a1-NSE + *dnPKA* mRNA
Total No. of embryos	194	137	175
No. of embryos with normal expression	141	2	0
No. of embryos with patchy expression	3	0	2
No. of embryos with ectopic expression	0	68	0
% embryos with normal expression	72.68	1.46	0
% embryos with reduced expression	1.55	0	1.14
% embryos with ectopic expression	0	49.64	0

Table 8.7. *Numbers and percentages of zebrafish embryos injected with wt a1-NSE::EGFP alone or with dnPKA mRNA, and of mut a1-NSE::EGFP + dnPKA mRNA, showing the effects on EGFP reporter expression as derived from a single experiment.*
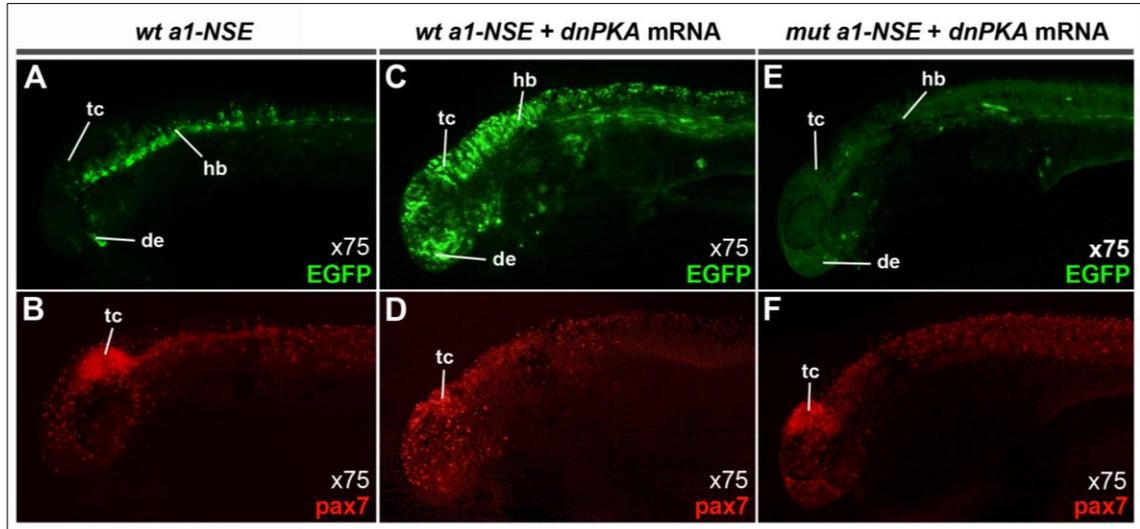
Figure 8.12. *Intact GLI binding motifs are required for activation of EGFP expression by dnPKA.* Immunofluorescent images of: (A, B) a "control 1" embryo injected with 30n g/µL wt a1-NSE; (C, D) a "control 2" embryo co-injected with 30 ng/µL wt a1-NSE::EGFP plasmid and 400 ng/µL *dnPKA* mRNA; (E, F) an embryo co-injected with 30 ng/µL mut a1-NSE::EGFP plasmid and 400 ng/µL *dnPKA* mRNA; all embryos are at stage 31 hpf. (E) Up-regulated Hh signalling by dnPKA is insufficient to activate EGFP expression in the absence of intact GLI-binding sites, in contrast to "control 2" embryos (C), suggesting that the effects of Hh signalling on a1-NSE's activity are mediated by the GLI-binding sites. (B, D, F) anti-Pax7 antibody staining showing the efficiency of injection for *dnPKA* mRNA-injected embryos (D and F) exhibit reduced Pax7 expression in the dorsal brain, in contrast to "control 1" embryos (B) which show normal expression. Abbreviations: de, diencephalon; hb, hindbrain; tc, tectum.
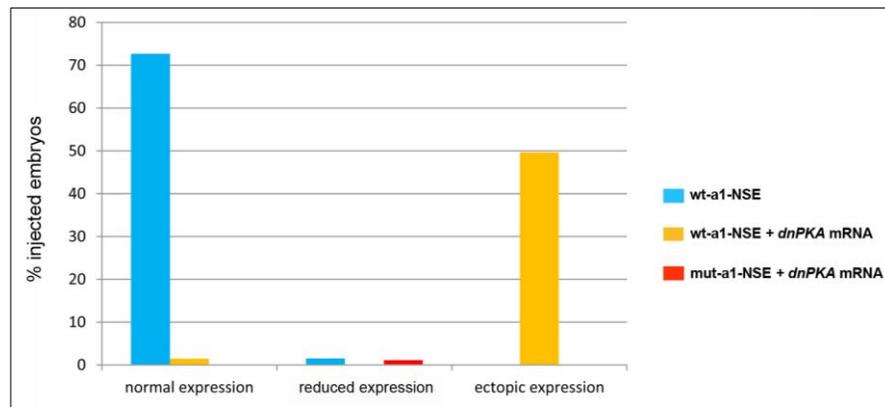
Figure 8.13. *Intact GLI binding motifs are required for activation of EGFP expression by dnPKA (Quantitative Data).* Graphical representation of the results from a single experiment showing the percentage of embryos injected with wt a1-NSE::EGFP plasmid, with wt a1-NSE::EGFP plasmid + *dnPKA* mRNA, and with mut a1-NSE::EGFP plasmid + *dnPKA* mRNA. "Normal", "reduced" and "ectopic" patterns of EGFP expression are defined in the text. Note that the percentage of embryos with complete absence of EGFP signal in the neural tube in all three groups is not indicated on the graph. Importantly, embryos injected with *dnPKA* mRNA are unable to generate either ectopic or normal pattern of EGFP expression driven from mut a1-NSE.

201

8.8. Discussion

Here, I demonstrated that the function of a1-NSE requires active Hh-signalling pathway. In *smo* zebrafish mutants and cyclopamine-treated wild type fish a1-NSE fails to drive *EGFP* expression. Moreover, constitutive activation of the pathway either by interfering with PKA function or via *shha* over-expression is sufficient to ectopically direct *a1-NSE::EGFP* expression in the dorsal neural tube. Finally, a1-NSE' response to Shh signalling maps to the GLI binding motifs that are essential for reporter gene expression.

8.8.1. Active Hh signalling is required for a1-NSE's activity

Cyclopamine treatment led to complete abrogation of the normal pattern of a1-NSE-driven EGFP expression in the ventral neural tube (Figure 8.2), analogously to the down-regulation of the Hh target genes *olig2* and *nkx6.1* in cyclopamine-treated wild-type zebrafish (Guner et al. 2007; Park et al. 2002). However, few cyclopamine-treated *a1-NSE::EGFP*-injected embryos retained patches of reduced EGFP signal in the brain which could be due to incomplete tissue penetration of cyclopamine and/or fortuitous activation of the β-globin promoter in the *a1-NSE::EGFP* construct by Hh-independent enhancers at the site of chromosomal integration (Wilson et al. 1990).

Likewise, a1-NSE was inactive in *smo$^{-/-}$* embryos (Figure 8.4D). This is reminiscent of the defects in the expression of class II Hh targets like *gli1*, *ptch1*, and *nkx2.2*, reported previously (Karlstrom et al. 2003; Varga et al. 2001). However, 3 out of 75 *smo*-null fish had small clusters of EGFP-expressing cells in the brain. It is possible that in these cases residual smoothened activity from maternally deposited *smo* mRNA was sufficient to activate a1-NSE in few cells of the mutant embryos, and Chen et al. (2001) show that there is abundant supply of maternal *smo* mRNA. Maternally deposited smoothened is sufficient to transduce Hh signals at early stages (but not at the 18-somites stage), as demonstrated by the ability of *smo* mutants to increase the expression of *nkx2.2* in the neuroectoderm upon injection of *shha* mRNA (Varga et al., 2001).

8.8.2. Intact GLI-binding motifs are required for a1-NSE's activity

The GLI-binding motifs appear to be essential for a1-NSE's activity as none of the *mut-a1-NSE::EGFP*-injected zebrafish embryos succeeded in generating a normal pattern of EGFP expression (Figure 8.10C; Figure 8.11), probably because Gli proteins failed to bind to the mutated GLI motifs. However, 1% of the *mut-a1-NSE::EGFP*-injected embryos (n=5) exhibited few patches of EGFP signal in the neural tube ("reduced expression" in Figure 8.11), which is most likely the result of chromosome integration-site effects, as discussed above. Although the a1-NSE element has been shown to be occupied by GLI factors *in vivo* (Vokes et al. 2008), it remains to be demonstrated that GLI proteins directly bind to the GLI motifs, and that GLI factors are essential for a1-NSE's activity. Direct binding could be tested by DNase I footprinting where radioactively labelled wild type or mutant a1-NSE is incubated with purified GLI proteins, or by EMSA where short radioactively labelled oligonucleotide sequences carrying either the wild type or mutant GLI motifs are incubated with purified GLI proteins. If GLI proteins directly bind to the GLI motifs in a1-NSE, it could be expected that these motifs will be protected from the action of the DNase, or will form detectable gel-shifted complexes in EMSA. A comparison between the latter EMSA experiment with purified GLI proteins and an EMSA with nuclear extracts from embryos or GLI-expressing cell lines, will inform about whether the GLI factors are the only proteins able to cccupy the GLI motifs in a1-NSE, or there are other proteins capable of binding these motifs. In addition, a requirement for the GLI proteins could be tested via the analysis of a1-NSE-driven reporter expression in *Gli*-deficient embryos or cell lines (*Gli* mutants, or RNAi-mediated knockdowns).

Alternatively, it could be that binding of factors other than GLI is affected in mut a1-NSE. The GLI binding motifs identified by MatInspector overlap with motifs of other transcription factors, some of which are expressed in the developing CNS, such as PAX3 and HMX2 (Table 8.8). Thus, the absence of reporter activity from the mut a1-NSE-driven construct might be caused by the abrogation of DNA binding of some of the factors listed in Table 8.8. This possibility can be addressed by 1) testing whether any of these factors is required for a1-NSE's activity via an analysis of a1-NSE-driven reporter expression in embryos deficient for the factor (either mutants, morpholino- or siRNA-treated embryos),

and 2) by testing whether any of the factors binds to a1-NSE in transgenic embryos via performing a ChIP analysis.

GLI motif	Overlapping motifs
1	**NR5A2** (nuclear receptor subfamily 5, group A, member 2) **AR** (androgen receptor) **NR3C1** (nuclear receptor subfamily 3, group C, member 1; glucocorticoid receptor) **PAX3** (paired box 3)
2	**KLF1** (Kruppel-like factor 1 (erythroid)) **SALL2** (sal-like 2 (Drosophila)) **ZBTB7A** (zinc finger and BTB domain containing 7a) **E2F** (E2F transcription factor)
3	**HMGA1** (high mobility group AT-hook 1) **ZFP263** (zinc finger protein 263)
4	**MECOM** (MDS1 and EVI1 complex locus) **HIC1** (hypermethylated in cancer 1) **SALL2** (sal-like 2 (Drosophila)) **GATA1** (GATA binding protein 1)
5	**ZBTB7A** (zinc finger and BTB domain containing 7a) **HMX2** (H6 homeobox 2) **MESP1** (mesoderm posterior 1) **PAX3** (paired box3)

Table 8.8. *Transcription factor binding motifs that overlap with the GLI motifs in a1-NSE.* Each GLI motif within a1-NSE partially overlaps with predicted motifs of other transcription factors.

In conclusion, based on the current data one cannot rule out the possibility that the introduced substitutions into one or more of the five GLI motifs did actually create new binding sites for some transcriptional repressors, such that it was recruitment of these repressors that led to the failure of mut-a1-NSE activity, and not the abrogation of GLI factor binding per se, although SequenceShaper deliberately avoided the creation of other motifs. This scenario can be examined by generating a variant mut-a1-NSE with complete deletions of the GLI motifs to test whether absence of the sites also perturbs a1-NSE's function.

8.8.3. a1-NSE activity depends on the canonical Shh signalling pathway

The findings in this chapter suggest that it is the canonical SHH-signalling (the signalling cascade involving smoothened and GLI functions) that is responsible for a1-NSE's activity, instead of the non-canonical pathways described in Section 8.6 (Brennan et al. 2012; Jenkins 2009). Such non-canonical Hh signalling is hypothesised in the GLI-independent modulation of cell motility in C3H10T1/2 fibroblasts treated with SHH, where the latter acts as a chemoattractant and induces reorganisation of the actin cytoskeleton followed by lamellipodia formation within 10 minutes of SHH administration, a response more than 36-fold faster than GLI-mediated responses (Bijlsma et al. 2007). In another study, Bourikas et al. (2005) showed that SHH expressed in the chicken embryo floor plate acts as a

chemorepellent for commissural axons (CAs) in a smoothened-independent manner as cyclopamine has no effect on CA migration and these axons do not express either Patched or Smoothened (Bourikas et al. 2005). However, such non-canonical mechanisms are less likely to account for a1-NSE's function as I showed the requirement for smoothened (Figure 8.4), and for the presence of intact GLI-binding motifs (Figure 8.10) for a1-NSE's activity.

However, the current data do not exclude the possibility of parallel involvement of inputs from other signalling systems, such as the BMP, Wnt and retinoic acid pathways. For instance, the precise dorso-ventral boundaries of *Nkx2.2* expression in the murine spinal cord depend on the combinatorial binding of GLI activator and TCF4 repressor inputs to an enhancer located 2 kb upstream of the promoter (Lei et al. 2006; Ulloa and Marti 2010). Interestingly, a1-NSE harbours LEF1/TCF motifs suggesting that WNT signalling may be involved in a1-NSE's function by inhibiting its activity in the dorsal neural tube. Moreover, the effect of WNT signals may be indirect via regulation of *Gli* gene expression, as previously shown for *Gli3* in the neural tube (Alvarez-Medina et al. 2008; Yu et al. 2008), and for *Gli2* and *Gli3* in the somites (Borycki et al. 2000). The role of WNT and/or BMP signalling pathways in a1-NSE' function can be tested in *a1-NSE::EGFP* transgenic zebrafish embryos treated with morpholinos against components of these pathways, or in zebrafish mutants.

8.8.4. Shh signalling may directly regulate a1-NSE's activity, via the GLI factors

Up-regulation of Hh signalling by both dnPKA and shha overexpression resulted in ectopic expansion of a1-NSE-driven EGFP expression in the dorsal half of the neural tube (Figure 8.6E, G and 8.8E). This is highly reminiscent to the effects on *ptch1*, *gli1* and *foxa* mRNA expression, which in untreated embryos is confined to the ventral half of the neural tube, but their expression expands dorsally upon injection of *shha* or *dnPKA* mRNA (Karlstrom et al. 2003; Krauss et al. 1993). It is suggested that *Ptch1*, *Gli1* and *Foxa2* are direct targets of HH signalling for all three of them harbour essential GLI binding motifs within their promoters (Vokes et al. 2007), as does a1-NSE as well (Figure 7.1). This, combined with the inability of mut a1-NSE to drive ectopic reporter expression in conditions of up-regulated Hh signalling (Figure 8.12), suggests that the activity of a1-NSE may be directly dependent on Shh signalling mediated by the GLI transcription factors.

Chapter 9

Final Discussion

9.1. A summary of results

In this study, I reported the tissue-specific expression pattern of the murine *Lama1* gene in the neural tube, somites, presomitic mesoderm, head mesenchyme and nephric structures, and also confirmed the essential role of SHH in the control of *Lama1* transcription in the somites and neural tube. Furthermore, I revealed that SHH is differentially required for the maintenance of *Lama1* expression in the neural tube and the activation of *Lama1* in the somites. Finally, expression of *Lamb1* and *Lamc1* does not depend on SHH. Interestingly, I showed that the expression pattern of chicken *Lama1* (*cLama1*) is highly similar to the pattern of its mouse ortholog, except for the striking absence of *cLama1* mRNA in the presomitic mesoderm. Thus, *Lama1* displays conserved expression pattern in evolution hinting for the conservation of signalling and transcriptional regulatory mechanisms controlling its activity.

Using bioinformatics, I identified 25 conserved non-coding elements (CNE1-24, and CNEb) distributed within a 361 kb region located upstream of the murine *Lama1* locus (except for CNEb located in the 3rd intron of the murine *Lama1* gene), six of which contained putative GLI/ZIC binding sites, raising the possibility that some of them may mediate the effects of SHH on *Lama1* expression. Luciferase reporter assays in cultured fibroblasts suggested that one of the GLI-motif-containing elements – CNE7, may possess silencing activity, whereas the remaining of the tested CNEs failed to significantly change reporter's activity. Similarly, when tested *in vivo* in transient transgenic zebrafish embryos the CNEs failed to display enhancer activity, except for CNE3, which directed reporter gene expression in skeletal muscles.

Data mining for genome-wide occupancy of GLI transcription factors led to the identification of a GLI3-bound non-conserved 1 kb region, a1-NSE, located in the first intron of the murine *Lama1* gene (Vokes et al. 2008). Remarkably, when tested in transgenic mouse and zebrafish embryos, a1-NSE directs tissue-specific reporter expression in the neural tube in a pattern highly reminiscent to the expression of the endogenous *Lama1* mRNA. Functional analyses of a1-NSE in transient transgenic zebrafish embryos revealed that Shh signalling is both required and sufficient for a1-NSE's transcriptional activity. Moreover, mutations in the GLI binding motifs within a1-NSE demonstrated that they are essential for a1-NSE's activity and its response to Hh signalling. Taken together, these results suggest that

a1-NSE might be a SHH-dependent enhancer responsible for *Lama1* gene expression in the neural tube of mouse embryos.

9.2. Sequence conservation versus functional conservation of *Lama1* enhancers

Although the 25 CNEs identified initially by bioinformatics, and a1-NSE identified by data mining, are conserved within mammals, none share sequence similarities with the zebrafish genome, except for CNEa and CNEb, whose putative activities have not been examined yet. Nevertheless, both CNE3 and a1-NSE behave as tissue-specific enhancers *in vivo*, generating consistent reporter expression patterns, at least for a1-NSE, in both mouse and zebrafish transgenics. Therefore, I hypothesise that these two elements are examples of the so called type of "billboard" enhancer (Arnosti and Kulkarni 2005).

Depending on the mode of input computation within enhancers, two models of enhancer structure have been proposed – the "enhanceosome" and "billboard" models (Arnosti and Kulkarni 2005), which are the two contrasting extremes of a continuum, where most enhancers in nature fall in between those two models (Meireles-Filho and Stark 2009). Enhanceosome-like enhancers (EEs) consist of spacing-, orientation-, and arrangement-sensitive transcription factor binding sites (TFBSs) (Merika and Thanos 2001; Valentine et al. 1998). Most TFBSs within EEs are essential for their cognate transcription factors bind cooperatively and produce a supra-molecular structure with specific geometry, which is required for transcriptional activation (Thanos and Maniatis 1995). Because of this cooperative combinatorial requirement, EEs regulate their target genes in an ON/OFF mode (or bimodal)(Sutherland et al. 1997). Also, because of the constraints on TFBSs organization, enhanceosome-like enhancers are expected to be highly sequence-conserved across species (Carroll et al. 2004). The best known example of such elements is the virus-induced enhancer of the human interferon-beta gene (IFN-β) which receives cooperative inputs from NF-kappaB, IRF-3/IRF-7 and ATF-2/c-JUN (Panne et al. 2007)

In contrast, "billboard" enhancers (BBE) consist of independently- (or quasi-independently) acting TFBSs or short clusters of TFBSs, each of which is able to bind its transcription factor and interact with the basal transcription machinery independently from the others (Arnosti and Kulkarni 2005). The key is that the contribution of individual TFBSs within a "billboard" enhancer is summed up to give a total output (Gao and Finkelstein et al. 1998). Thus, the transcription rate of a target gene driven by such enhancers exhibits

continuous (or rheostatic) profile depending on the transcription factor occupancy in the enhancer - there are not just two simple on/off responses (Biggar and Crabtree 2001; Rossi et al. 2000). Importantly, "billboard" enhancers can tolerate considerable turnover in sequence (like changes in spacing, orientation, order and even composition of TFBSs), without affecting their function. What matters at the end is that the sum of all transcription factor inputs should be unchanged (in terms of final output) (Arnosti and Kulkarni 2005). Thus, new TFBSs for functionally equivalent transcription factors may evolve without modifying the output (Ludwig et al. 1998). Therefore, because of their inherent plasticity, "billboard" enhancers are expected to evolve faster and display considerable sequence variation between species (Arnosti et al. 1996; Ludwig et al. 1998), sometimes even hindering their identification by sequence comparison, and several examples of such enhancers were described in Section 1.10.2.1 (Chapter 1).

Based on this, a1-NSE appears to exhibit characteristics of a "billboard" enhancer – its sequence is highly divergent, even within mammals, and yet it generates similar expression outputs in mouse and in a heterologous species – the zebrafish (Figure 7.14). The characteristic operational organization of "billboard" enhancers, and possibly of a1-NSE as well, allows them to quantitatively modulate the transcription rate of their target genes, and hence the amount of synthesized protein. This may be important in the case of laminin α1 expression, where the amount of laminin α1 production in the neural tube may have implications for the rate of basement membrane assembly (Anderson et al. 2009), sequestration of signaling molecules and cell migration (Condic and Letourneau 1997; Huttenlocher et al. 1995).

Remarkably, although I could not identify a "zebrafish a1-NSE" element, earlier research from our laboratory demonstrated that the 9.8 kb intron 1 of the zebrafish *lama1* locus contains almost all of the *cis*-regulatory information required for proper tissue-specific expression of *lama1*, including the enhancer(s) for expression in the neural tube (Joseph Pickering's Thesis 2012). However, this neural tube element lies in the second half of the intron – from position +4416 to +9779, unlike a1-NSE, which is located within the first 1.5 kilobases of the 19.3 kb intron 1 of the murine *Lama1* locus. Nevertheless, zebrafish intron 1 of *lama1* contains three putative binding sites for Gli factors and two conserved regions (conserved with the Fugu genome but not with mammals) (Joseph Pickering's Thesis 2012).

Thus, it appears that despite the great extent of sequence divergence between mammalian and teleost genomes, the enhancer(s) for neural expression of laminin α1 are still present in intron 1 of the locus in both clades, albeit in a shuffled configuration (Sanges et al. 2006). Moreover, the observations from zebrafish suggest that some of the other regulatory sequences of the mouse *Lama1* gene may also be located in intron 1. This requires further studies using bacterial artificial chromosome-based approaches coupled with serial deletions in intron 1 of *Lama1*.

Taken together, the case of a1-NSE highlights the evolutionary plasticity of enhancer sequences and demonstrates that methods alternative to sequence comparisons are instrumental in identifying the full complement of transcription-regulatory regions of a gene.

9.3. A potential for interaction between a1-NSE and some CNEs

An attractive possibility is that the neural-specific activity of a1-NSE in its native context in the mouse genome might be modulated through interactions with other *Lama1* regulatory elements, including some CNEs. Such relationships between transcription regulatory elements are well established in the literature and a prominent example include several enhancer elements located within the 600 kb "gene desert" on one side of the mammalian *HoxD* cluster which interact between each other and with the *HoxD* loci by chromatin looping to ensure proper patterning of the autopod (Montavon et al. 2011). Similarly, the MAF- and BACH1-bound MARE elements in the locus control region of the human β-globin genes physically interact and this appears to be crucial for β-globin gene expression (Yoshida et al. 1999). Such physical interactions between regulatory elements can boost or attenuate the expression of target genes, or alter their spatial and temporal pattern of expression (Frankel 2012).

In this regard, I would like to consider the possibility of a putative communication between a1-NSE and CNE7. As described and discussed in Chapter 5, CNE7 exhibits weak but statistically significant transcriptional-silencer activity and contains a conserved CTCF motif. As discussed, there is evidence for the involvement of CTCF in developmental gene repression, among other transcription-related processes (Gao et al. 2011). Thus, it is possible that CTCF-bound CNE7 interacts with a1-NSE to shut down its enhancer activity in cells undergoing neural differentiation, but not in neural stem cells in the ventricular zone and neural progenitors in the sub-ventricular zone, where some inhibitory mechanisms (perhaps

DNA methylation of the CTCF site) would prevent CNE7 from communicating with a1-NSE. This could explain the lack of *Lama1* mRNA expression in the mantle layers of the central nervous system, which contains differentiated neurons and glia, and the wide-spread β-Gal staining in the spinal cord of the two E13.5 *a1-NSE::lacZ* transgenic mouse embryos. In the latter case, the absence of CNE7 from the *a1-NSE::lacZ* construct might have allowed persistent activity of a1-NSE in differentiated neural cells.

9.4. SHH might directly regulate the expression of an ECM gene

To my knowledge, this study is the first to suggest a direct GLI-mediated role of SHH signalling in the transcriptional regulation of an extracellular matrix component during mouse embryogenesis. However, Dr. Xin Gang Wang from Prof. Philip Ingham's laboratory (IMCB, Singapore) showed that Gli2a binds to regions close to the *lamc1* gene promoter in zebrafish, and that abrogation of Hh signalling results in down-regulation of *lamc1* mRNA expression in both the neural tube and somites (Wang et al. 2013). This is in contrast to the mouse, where *Lamc1* expression does not require SHH activity (see Chapter 3 of this study), suggesting differences in the mechanisms regulating laminin expression in mammals and teleosts.

There are studies, although not in neural cells, showing direct activation of *Lamc1* transcription by interleukin-1-beta via NF-kappaB (O'Neill et al. 1997), or by TGF-β signals mediated by cooperative interaction between TEF3 and SMAD transcription factors (Kawata et al. 2002), whereas *Lamb1* transcription in F9 teratocarcinoma cells depends on retinoic acid-induced binding of RAR-alpha/beta receptors to the promoter of *Lamb1* (Vasios et al. 1991). Perhaps, the activation of different laminin genes by distinct signalling pathways ensures that a particular laminin trimer is synthesised only when the producing cells have received several combinatorial inputs, thus decreasing the chance of stochastic/aberrant energy-demanding synthesis of these crucial ECM components, caused by noise in the developmental system.

Through its induction of *Lama1*expression in the neural tube via a1-NSE, SHH may execute two tasks. First, it enables the synthesis of an important adhesive and signalling component of the pial basement membrane, as described in Chapter 1, and second, Shh may indirectly enhance/inhibit the availability of diffusible signalling molecules, including its own, to neuroepithelial cells (Borycki 2013). The latter phenomenon could operate through

the association of laminins with heparan sulphate proteoglycans (Hantaz-Ambroise et al. 1987), which are known to modulate the diffusion and morphogen gradient properties of SHH, WNT and BMP ligands (Belenkaya et al. 2004; Bornemann et al. 2004; Han et al. 2004), and such indirect role of laminin γ1-mediated restriction of BMP signalling has already been demonstrated in the zebrafish somites (Dolez et al. 2011), as described in Chapter 1.

9.5. Is a1-NSE essential for expression of *Lama1* in the murine central nervous system?

My analyses of a1-NSE activity in both mouse and zebrafish embryos indicate that this enhancer is active in the whole neuraxis – from the rostral-most part of the diencephalon to the caudal most regions of the spinal cord, whereas data argue against activity in the telencephalon. The latter finding suggests the existence of another, distinct enhancer directing *Lama1* expression in the telencephalon, as endogenous *Lama1* mRNA is detected in this domain (Chapter 3 of this study; Miner et al. 2004). The genomic location of this enhancer is currently unknown but as discussed above, extrapolating from studies in zebrafish (Section 9.2), it could be located in intron 1 of the murine *Lama1* locus, and BAC-reporter deletion studies might help to uncover it. Thus, when summed together a1-NSE, the putative telencephalic enhancer and also the hypothetical SHH-independent enhancer for initiation of *Lama1* transcription in the neural tube, should be able to recapitulate the full pattern of endogenous *Lama1* expression in the murine CNS.

However, a critical question remains unanswered – is a1-NSE required for *Lama1* expression in its native genomic context *in vivo*? Most enhancer elements are absolutely required for tissue-specific activation of their target genes, as clearly evidenced by pathological conditions in human patients carrying inactivating mutations in critical enhancers of genes like *ATOH7* (Ghiasvand et al. 2011), *EGR2* (Funalot et al. 2012), *SHH* (Lettice et al. 2003), *SOX9* (Benko et al. 2009), and *IRF6* (Rahimov et al. 2008,) to name a few. Thus, would deletion of a1-NSE (and its human counterpart) result in developmental abnormalities of the central nervous system as those described by Heng et al. (2011) and Ichikawa-Tomikawa et al. (2012) in *Lama1*[-/-] mutant mice (see Chapter 1)? I hypothesise that a1-NSE is essential for normal neural development, as explained below.

Many genes are regulated by seemingly redundant elements termed "shadow enhancers" which when tested individually in transgenic animals generate highly similar or

identical patterns of reporter expression (Hong et al. 2008; Jeong et al. 2006; Frankel et al. 2010). A revealing example comes from a study in *Drosophila* by Frankel et al. (2010), where expression of the cuticle patterning gene shavenbaby (*svb*) is regulated by multiple enhancers. The combined pattern of the distant enhancers D2 and Z overlaps with the pattern driven by the proximal enhancers A and E, and all are synchronously active in the cuticle, suggesting that they may be redundant (Frankel et al. 2010). Interestingly, mutant flies with deletion of D2 and Z have wild type phenotype under normal temperature, which strikingly contrasts with the severe reduction of trichome formation when the same mutants are grown at extreme temperatures. Analogous perturbations of cuticle morphogenesis in D2-Z deletion mutant flies was observed on a *wg/+* genetic background but not on a *+/+* background (Frankel et al. 2010) and similar dependencies were documented in the regulation of the snail gene in *Drosophila* (Perry et al. 2010). What has emerged from these studies is that "shadow enhancers" are essential for ensuring robust expression of their target genes when the developmental system is faced with extreme environmental conditions (like high temperature) or is under genetic background stress due to the presence of suboptimal variation elsewhere in the genome (Hong et al. 2008; Frankel 2012).

This is important for interpreting future experiments on a1-NSE function, where one might delete the element from its genomic context to study its requirement for development. Two outcomes are possible: 1) the knockout animal displays abnormalities of CNS development associated with lack of *Lama1* mRNA in neural tissues, or 2) there are no discernible effects on CNS development and *Lama1* expression. The first scenario would directly demonstrate the importance of a1-NSE for *Lama1* expression in the neural tube, while the second scenario could be accounted for by the presence of a "shadow enhancer" somewhere else in the *Lama1* locus, which is functionally equivalent to a1-NSE. However, as discussed above, in such case a1-NSE and its "shadow" counterpart would seem dispensable in laboratory and normal conditions, but would be indispensable to buffer, or canalise, development in case of external or internal perturbations, which is frequently the case in the wild.

9.6. Concluding remarks
This study provides interesting novel insights regarding the transcriptional regulation of a laminin gene − *Lama1*, which is essential for embryonic and postnatal development in the

vertebrate embryo. My data suggest that some of the conserved non-coding elements around the murine *Lama1* locus are probably mammal-specific, that they may represent potential *cis*-regulatory elements devoted to the control of *Lama1* expression and/or to the regulation of other genome processes that are unrelated to transcription. Among these could be the enhancer mediating the SHH-dependent control on *Lama1* expression in the somites. Therefore, further understanding of the putative *in vivo* functions of the CNEs could be gained by reporter assays in transgenic mouse embryos.

Importantly, the existence of a GLI-occupied neural-specific SHH-responsive enhancer, a1-NSE, in the first intron of the murine *Lama1* gene suggests, but does not unequivocally proves, for a direct role of SHH in the expression of *Lama1* in the developing central nervous system, and thus provides some insight in the molecular mechanism behind the lack of *Lama1* mRNA expression in the neural tube of *Shh*-deficient mouse embryos (Anderson et al. 2009). However, my investigation hints for additional complexity in the control of *Lama1* expression in the CNS, manifested in the hypothetical existence of "initiator" and "telencephalic" enhancers. Future BAC-reporter analyses and genome-wide assays of transcription factor occupancy will help to uncover these elements. Additionally, the regulatory elements identified in this study could find potential applications in the exploration of developmental processes in normal and pathological conditions including cell lineage tracing, design of conditionally expressed transgenes and knockout animals, and even for therapeutic purposes to target diseased tissues in the future.

Finally, the weakly conserved a1-NSE element could be used as a model enhancer to gain further insight into the complex relationship between sequence and function, and into the constraints, as well as opportunities, such relationship imposes on organismal evolution.

References

Aanstad, P., N. Santos, et al. (2009). "The extracellular domain of Smoothened regulates ciliary localization and is required for high-level Hh signaling." <u>Curr Biol</u> **19**(12): 1034-1039.

Abbasi, A. A., Z. Paparidis, et al. (2007). "Human GLI3 intragenic conserved non-coding sequences are tissue-specific enhancers." <u>PLoS One</u> **2**(4): e366.

Adamska, M., D. Q. Matus, et al. (2007). "The evolutionary origin of hedgehog proteins." <u>Curr Biol</u> **17**(19): R836-837.

Adolphe, C., R. Hetherington, et al. (2006). "Patched1 functions as a gatekeeper by promoting cell cycle progression." <u>Cancer Res</u> **66**(4): 2081-2088.

Ahituv, N., S. Prabhakar, et al. (2005). "Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny." <u>Hum Mol Genet</u> **14**(20): 3057-3063.

Akitake C.M., Macurak M., Halpern M.E., Goll M.G. (2011). "Transgenerational analysis of transcriptional silencing in zebrafish." Dev Biol. Apr 15;352(2):191-201.

Alcedo, J., M. Ayzenzon, et al. (1996). "The Drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal." <u>Cell</u> **86**(2): 221-232.

Alexandre, C., A. Jacinto, et al. (1996). "Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins." <u>Genes Dev</u> **10**(16): 2003-2013.

Amigo, J. D., M. Yu, et al. (2011). "Identification of distal cis-regulatory elements at mouse mitoferrin loci using zebrafish transgenesis." <u>Mol Cell Biol</u> **31**(7): 1344-1356.

Amsterdam, A., S. Lin, et al. (1995). "The Aequorea victoria green fluorescent protein can be used as a reporter in live zebrafish embryos." <u>Dev Biol</u> **171**(1): 123-129.

Anderson, C., S. Thorsteinsdottir, et al. (2009). "Sonic hedgehog-dependent synthesis of laminin alpha1 controls basement membrane assembly in the myotome." <u>Development</u> **136**(20): 3495-3504.

Andres, M. E., C. Burger, et al. (1999). "CoREST: a functional corepressor required for regulation of neural-specific gene expression." <u>Proc Natl Acad Sci U S A</u> **96**(17): 9873-9878.

Arikawa-Hirasawa, E., H. Watanabe, et al. (1999). "Perlecan is essential for cartilage and cephalic development." <u>Nature genetics</u> **23**(3): 354-358.

Ariza-Cosano, A., A. Visel, et al. (2012). "Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression." <u>BMC Genomics</u> **13**: 713.

Arnosti, D. N., S. Barolo, et al. (1996). "The eve stripe 2 enhancer employs multiple modes of transcriptional synergy." <u>Development</u> **122**(1): 205-214.

Arnosti, D. N. and M. M. Kulkarni (2005). "Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards?" <u>J Cell Biochem</u> **94**(5): 890-898.

Aruga, J., T. Inoue, et al. (2002). "Zic2 controls cerebellar development in cooperation with Zic1." <u>J Neurosci</u> **22**(1): 218-225.

Aruga, J., O. Minowa, et al. (1998). "Mouse Zic1 is involved in cerebellar development." <u>J Neurosci</u> **18**(1): 284-293.

Aruga, J., K. Mizugishi, et al. (1999). "Zic1 regulates the patterning of vertebral arches in cooperation with Gli3." <u>Mech Dev</u> **89**(1-2): 141-150.

Aruga, J., T. Nagai, et al. (1996). "The mouse zic gene family. Homologues of the Drosophila pair-rule gene odd-paired." <u>J Biol Chem</u> **271**(2): 1043-1047.

Aruga, J., T. Tohmonda, et al. (2002). "Zic1 promotes the expansion of dorsal neural progenitors in spinal cord by inhibiting neuronal differentiation." <u>Dev Biol</u> **244**(2): 329-341.

Aruga, J., N. Yokota, et al. (1994). "A novel zinc finger protein, zic, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells." <u>J Neurochem</u> **63**(5): 1880-1890.

Aruga, J., A. Yozu, et al. (1996). "Identification and characterization of Zic4, a new member of the mouse Zic gene family." <u>Gene</u> **172**(2): 291-294.

Aumailley, M., L. Bruckner-Tuderman, et al. (2005). "A simplified laminin nomenclature." <u>Matrix biology : journal of the International Society for Matrix Biology</u> **24**(5): 326-332.

Avaron, F., L. Hoffman, et al. (2006). "Characterization of two new zebrafish members of the hedgehog family: atypical expression of a zebrafish indian hedgehog gene in skeletal elements of both endochondral and dermal origins." <u>Dev Dyn</u> **235**(2): 478-489.

Aza-Blanc, P., F. A. Ramirez-Weber, et al. (1997). "Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor." <u>Cell</u> **89**(7): 1043-1053.

Bai, C. B., W. Auerbach, et al. (2002). "Gli2, but not Gli1, is required for initial Shh signaling and ectopic activation of the Shh pathway." <u>Development</u> **129**(20): 4753-4761.

Bai, C. B. and A. L. Joyner (2001). "Gli1 can rescue the in vivo function of Gli2." <u>Development</u> **128**(24): 5161-5172.

Bai, C. B., D. Stephen, et al. (2004). "All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3." <u>Developmental cell</u> **6**(1): 103-115.

Bajanca, F., M. Luz, et al. (2006). "Integrin alpha6beta1-laminin interactions regulate early myotome formation in the mouse embryo." <u>Development</u> **133**(9): 1635-1644.

Bale, A. E. (2002). "Hedgehog signaling and human disease." <u>Annu Rev Genomics Hum Genet</u> **3**: 47-65.

Ballare, C., R. Zaurin, et al. "More help than hindrance: nucleosomes aid transcriptional regulation." <u>Nucleus</u> **4**(3): 189-194.

Ballas, N., C. Grunseich, et al. (2005). "REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis." <u>Cell</u> **121**(4): 645-657.

Bantignies F, Cavalli G. (2011). "Polycomb group proteins: repression in 3D." Trends Genet. 2011 Nov;27(11):454-64.

Baron, W., B. Metz, et al. (2000). "PDGF and FGF-2 signaling in oligodendrocyte progenitor cells: regulation of proliferation and differentiation by multiple intracellular signaling pathways." <u>Mol Cell Neurosci</u> **15**(3): 314-329.

Barresi, M. J., H. L. Stickney, et al. (2000). "The zebrafish slow-muscle-omitted gene product is required for Hedgehog signal transduction and the development of slow muscle identity." <u>Development</u> **127**(10): 2189-2199.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." <u>Cell</u> **129**(4): 823-837.

Baxter, E. W., F. Mirabella, et al. "The inducible tissue-specific expression of the human IL-3/GM-CSF locus is controlled by a complex array of developmentally regulated enhancers." <u>J Immunol</u> **189**(9): 4459-4469.

Beachy, P. A., M. K. Cooper, et al. (1997). "Multiple roles of cholesterol in hedgehog protein biogenesis and signaling." <u>Cold Spring Harb Symp Quant Biol</u> **62**: 191-204.

Beachy, P. A., S. G. Hymowitz, et al. "Interactions between Hedgehog proteins and their binding partners come into view." <u>Genes Dev</u> **24**(18): 2001-2012.

Beck, K., T. W. Dixon, et al. (1993). "Ionic interactions in the coiled-coil domain of laminin determine the specificity of chain assembly." Journal of molecular biology **231**(2): 311-323.

Beck, K., I. Hunter, et al. (1990). "Structure and function of laminin: anatomy of a multidomain glycoprotein." FASEB journal : official publication of the Federation of American Societies for Experimental Biology **4**(2): 148-160.

Bejerano, G., M. Pheasant, et al. (2004). "Ultraconserved elements in the human genome." Science **304**(5675): 1321-1325.

Belenkaya, T. Y., C. Han, et al. (2004). "Drosophila Dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans." Cell **119**(2): 231-244.

Bell, A. C. and G. Felsenfeld (2000). "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene." Nature **405**(6785): 482-485.

Bell, A. C., A. G. West, et al. (1999). "The protein CTCF is required for the enhancer blocking activity of vertebrate insulators." Cell **98**(3): 387-396.

Benko, S., J. A. Fantes, et al. (2009). "Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence." Nat Genet **41**(3): 359-364.

Benton M. J. (2005) "Vertebrate Paleontology", 3rd ed. Oxford: Blackwell Science Ltd

Berezikov, E., V. Guryev, et al. (2005). "Phylogenetic shadowing and computational identification of human microRNA genes." Cell **120**(1): 21-24.

Bergeron, S. A., L. A. Milla, et al. (2008). "Expression profiling identifies novel Hh/Gli-regulated genes in developing zebrafish embryos." Genomics **91**(2): 165-177.

Bery, A., B. Martynoga, et al. (2013). "Characterization of Enhancers Active in the Mouse Embryonic Cerebral Cortex Suggests Sox/Pou cis-Regulatory Logics and Heterogeneity of Cortical Progenitors." Cereb Cortex.

Bessa, J., J. J. Tena, et al. (2009). "Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish." Dev Dyn **238**(9): 2409-2417.

Bezakova, G. and M. A. Ruegg (2003). "New insights into the roles of agrin." Nature reviews. Molecular cell biology **4**(4): 295-308.

Biesiada, E., Y. Hamamori, et al. (1999). "Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter." Mol Cell Biol **19**(4): 2577-2584.

Biggar, S. R. and G. R. Crabtree (2001). "Cell signaling can direct either binary or graded transcriptional responses." Embo J **20**(12): 3167-3176.

Bilozur, M. E. and E. D. Hay (1988). "Neural crest migration in 3D extracellular matrix utilizes laminin, fibronectin, or collagen." Dev Biol **125**(1): 19-33.

Binns, W., L. F. James, et al. (1963). "A Congenital Cyclopian-Type Malformation in Lambs Induced by Maternal Ingestion of a Range Plant, Veratrum Californicum." Am J Vet Res **24**: 1164-1175.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Bishop, B., A. R. Aricescu, et al. (2009). "Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP." Nat Struct Mol Biol **16**(7): 698-703.

Bitgood, M. J., L. Shen, et al. (1996). "Sertoli cell signaling by Desert hedgehog regulates the male germline." Curr Biol **6**(3): 298-304.

Blanchette, M. and M. Tompa (2002). "Discovery of regulatory elements by a computational method for phylogenetic footprinting." Genome Res **12**(5): 739-748.

Blanco, J., M. Seimiya, et al. (2009). "Wingless and Hedgehog signaling pathways regulate orthodenticle and eyes absent during ocelli development in Drosophila." Dev Biol **329**(1): 104-115.

Blow, M. J., D. J. McCulley, et al. "ChIP-Seq identification of weakly conserved heart enhancers." Nat Genet **42**(9): 806-810.

Blow, M. J., D. J. McCulley, et al. (2010). "ChIP-Seq identification of weakly conserved heart enhancers." Nat Genet **42**(9): 806-810.

Bornemann, D. J., J. E. Duncan, et al. (2004). "Abrogation of heparan sulfate synthesis in Drosophila disrupts the Wingless, Hedgehog and Decapentaplegic signaling pathways." Development **131**(9): 1927-1938.

Borycki, A. G. "The myotomal basement membrane: insight into laminin-111 function and its control by Sonic hedgehog signaling." Cell Adh Migr **7**(1): 72-81.

Borycki, A. G., B. Brunk, et al. (1999). "Sonic hedgehog controls epaxial muscle determination through Myf5 activation." Development **126**(18): 4053-4063.

Borycki, A. G., L. Mendham, et al. (1998). "Control of somite patterning by Sonic hedgehog and its downstream signal response genes." Development **125**(4): 777-790.

Bostrom, K., Y. Tintut, et al. (2000). "HOXB7 overexpression promotes differentiation of C3H10T1/2 cells to smooth muscle cells." J Cell Biochem **78**(2): 210-221.

Boudreau, N., C. J. Sympson, et al. (1995). "Suppression of ICE and apoptosis in mammary epithelial cells by extracellular matrix." Science **267**(5199): 891-893.

Boyle, A. P., S. Davis, et al. (2008). "High-resolution mapping and characterization of open chromatin across the genome." Cell **132**(2): 311-322.

Brand, M., C. P. Heisenberg, et al. (1996). "Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain." Development **123**: 179-190.

Brand, M., C. P. Heisenberg, et al. (1996). "Mutations affecting development of the midline and general body shape during zebrafish embryogenesis." Development **123**: 129-142.

Brewster, R., J. Lee, et al. (1998). "Gli/Zic factors pattern the neural plate by defining domains of cell differentiation." Nature **393**(6685): 579-583.

Briscoe, J. and J. Ericson (1999). "The specification of neuronal identity by graded Sonic Hedgehog signalling." Semin Cell Dev Biol **10**(3): 353-362.

Briscoe, J., A. Pierani, et al. (2000). "A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube." Cell **101**(4): 435-445.

Briscoe, J., L. Sussel, et al. (1999). "Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling." Nature **398**(6728): 622-627.

Brivanlou, A. H. (1998). "Should the master regulator Rest in peace?" Nat Genet **20**(2): 109-110.

Bronstein, I., J. Fortin, et al. (1994). "Chemiluminescent and bioluminescent reporter gene assays." Anal Biochem **219**(2): 169-181.

Brown, S. A., D. Warburton, et al. (1998). "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired." Nat Genet **20**(2): 180-183.

Brudno M., Malde S., Poliakov A., Do C.B., Couronne O., Dubchak I., Batzoglou S. (2003). "Glocal alignment: finding rearrangements during alignment". Bioinformatics. 19. Suppl 1 (90001): i54–62.

Bryson-Richardson, R. J. and Currie, P. D. (2008). The genetics of vertebrate myogenesis. Nat Rev Genet 9, 632-46.

Buckingham, M., L. Bajard, et al. (2003). "The formation of skeletal muscle: from somite to limb." J Anat **202**(1): 59-68.

Buckingham, M. (2007). Skeletal muscle progenitor cells and the role of Pax genes. C R Biol 330, 530-3.

Bulger M, Groudine M. (2011) "Functional and mechanistic diversity of distal transcription enhancers." Cell. Feb 4;144(3):327-39.

Burcin, M., R. Arnold, et al. (1997). "Negative protein 1, which is required for function of the chicken lysozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF." Mol Cell Biol 17(3): 1281-1288.

Butler J.E., Kadonaga J.T. (2001). "Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs." Genes Dev. 2001 Oct 1; 15(19); 2515-9.

Bylund, M., E. Andersson, et al. (2003). "Vertebrate neurogenesis is counteracted by Sox1-3 activity." Nat Neurosci 6(11): 1162-1168.

Byrne B.J., Davis M.S., Yamaguchi J., Bergsma D.J., Subramanian K.N. (1983). "Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat." Proc Natl Acad Sci U S A; Feb; 80(3); 721-5.

Bystrom, B., I. Virtanen, et al. (2006). "Distribution of laminins in the developing human eye." Investigative ophthalmology & visual science 47(3): 777-785.

Calo, E. and J. Wysocka "Modification of enhancer chromatin: what, how, and why?" Mol Cell 49(5): 825-837.

Campos, L. S., D. P. Leone, et al. (2004). "Beta1 integrins activate a MAPK signalling pathway in neural stem cells that contributes to their maintenance." Development 131(14): 3433-3444.

Canettieri, G., L. Di Marcotullio, et al. (2010). "Histone deacetylase and Cullin3-REN(KCTD11) ubiquitin ligase interplay regulates Hedgehog signalling through Gli acetylation." Nature cell biology 12(2): 132-142.

Carey M. and Smale S. T. (2000) "Transcriptional Regulation in Eukaryotes: Concepts, Strategies and Techniques". Cold Spring Harbor Laboratory Press

Carpenter, D., D. M. Stone, et al. (1998). "Characterization of two patched receptors for the vertebrate hedgehog protein family." Proc Natl Acad Sci U S A 95(23): 13630-13634.

Carroll S. B., Grenier J., and Weatherbee S. (2004) "From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design", 2nd ed. Wiley-Blackwell

Cartharius, K., K. Frech, et al. (2005). "MatInspector and beyond: promoter analysis based on transcription factor binding sites." Bioinformatics 21(13): 2933-2942.

Carvajal, J. J., D. Cox, et al. (2001). "A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development." Development 128(10): 1857-1868.

Carvajal, J. J., A. Keith, et al. (2008). "Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5." Genes Dev 22(2): 265-276.

Caspary, T., M. J. Garcia-Garcia, et al. (2002). "Mouse Dispatched homolog1 is required for long-range, but not juxtacrine, Hh signaling." Curr Biol 12(18): 1628-1632.

Castro, D. S., D. Skowronska-Krawczyk, et al. (2006). "Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif." Dev Cell 11(6): 831-844.

Chalfie, M., Y. Tu, et al. (1994). "Green fluorescent protein as a marker for gene expression." Science 263(5148): 802-805.

Chan, R. Y., C. Boudreau-Lariviere, et al. (1999). "An intronic enhancer containing an N-box motif is required for synapse- and tissue-specific expression of the acetylcholinesterase gene in skeletal muscle fibers." Proc Natl Acad Sci U S A 96(8): 4627-4632.

Chandrasekhar, A., H. E. Schauerte, et al. (1999). "The zebrafish detour gene is essential for cranial but not spinal motor neuron induction." Development 126(12): 2727-2737.

Chang, D. T., A. Lopez, et al. (1994). "Products, genetic linkage and limb patterning activity of a murine hedgehog gene." Development **120**(11): 3339-3353.

Chatterjee, S., G. Bourque, et al. "Conserved and non-conserved enhancers direct tissue specific transcription in ancient germ layer specific developmental control genes." BMC Dev Biol **11**: 63.

Chaudhary, J. and M. K. Skinner (1999). "Basic helix-loop-helix proteins can act at the E-box within the serum response element of the c-fos promoter to influence hormone-induced promoter activation in Sertoli cells." Mol Endocrinol **13**(5): 774-786.

Chen, A. E., D. D. Ginty, et al. (2005). "Protein kinase A signalling via CREB controls myogenesis induced by Wnt proteins." Nature **433**(7023): 317-322.

Chen, J. K., J. Taipale, et al. (2002). "Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened." Genes Dev **16**(21): 2743-2748.

Chen, J. K., J. Taipale, et al. (2002). "Small molecule modulation of Smoothened activity." Proc Natl Acad Sci U S A **99**(22): 14071-14076.

Chen, M. H., Y. J. Li, et al. (2004). "Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates." Genes Dev **18**(6): 641-659.

Chen, S. and V. G. Corces (2001). "The gypsy insulator of Drosophila affects chromatin structure in a directional manner." Genetics **159**(4): 1649-1658.

Chen, W., S. Burgess, et al. (2001). "Analysis of the zebrafish smoothened mutant reveals conserved and divergent functions of hedgehog activity." Development **128**(12): 2385-2396.

Chen, Z. F., A. J. Paquette, et al. (1998). "NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis." Nat Genet **20**(2): 136-142.

Chen, Z. L., V. Haegeli, et al. (2009). "Cortical deficiency of laminin gamma1 impairs the AKT/GSK-3beta signaling pathway and leads to defects in neurite outgrowth and neuronal migration." Developmental biology **327**(1): 158-168.

Cheng, Y. S., M. F. Champliaud, et al. (1997). "Self-assembly of laminin isoforms." J Biol Chem **272**(50): 31525-31532.

Chiang, C., Y. Litingtung, et al. (1996). "Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function." Nature **383**(6599): 407-413.

Christ, B., R. Huang, et al. (2007). "Amniote somite derivatives." Dev Dyn **236**(9): 2382-2396.

Christ, B. and M. Scaal (2008). "Formation and differentiation of avian somite derivatives." Adv Exp Med Biol **638**: 1-41.

Chuang, P. T. and A. P. McMahon (1999). "Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein." Nature **397**(6720): 617-621.

Clarke, S. L., J. E. VanderMeer, et al. "Human developmental enhancers conserved between deuterostomes and protostomes." PLoS Genet **8**(8): e1002852.

Clegg, C. H., L. A. Correll, et al. (1987). "Inhibition of intracellular cAMP-dependent protein kinase using mutant genes of the regulatory type I subunit." J Biol Chem **262**(27): 13111-13119.

Coffman, J. A. and E. H. Davidson (2001). "Oral-aboral axis specification in the sea urchin embryo. I. Axis entrainment by respiratory asymmetry." Dev Biol **230**(1): 18-28.

Cohen, J. and A. R. Johnson (1991). "Differential effects of laminin and merosin on neurite outgrowth by developing retinal ganglion cells." Journal of cell science. Supplement **15**: 1-7.

Coles, E. G., L. S. Gammill, et al. (2006). "Abnormalities in neural crest cell migration in laminin alpha5 mutant mice." Dev Biol **289**(1): 218-228.

Colognato, H. and P. D. Yurchenco (2000). "Form and function: the laminin family of heterotrimers." Developmental dynamics : an official publication of the American Association of Anatomists **218**(2): 213-234.

Concordet, J. P., K. E. Lewis, et al. (1996). "Spatial regulation of a zebrafish patched homologue reflects the roles of sonic hedgehog and protein kinase A in neural tube and somite patterning." Development **122**(9): 2835-2846.

Condic, M. L. and P. C. Letourneau (1997). "Ligand-induced changes in integrin expression regulate neuronal adhesion and neurite outgrowth." Nature **389**(6653): 852-856.

Cooper, A. F., K. P. Yu, et al. (2005). "Cardiac and CNS defects in a mouse with targeted disruption of suppressor of fused." Development **132**(19): 4407-4417.

Cooper, A. R., A. Taylor, et al. (1983). "Changes in the rate of laminin and entactin synthesis in F9 embryonal carcinoma cells treated with retinoic acid and cyclic amp." Developmental biology **99**(2): 510-516.

Cooper, M. K., J. A. Porter, et al. (1998). "Teratogen-mediated inhibition of target tissue response to Shh signaling." Science **280**(5369): 1603-1607.

Corbit, K. C., P. Aanstad, et al. (2005). "Vertebrate Smoothened functions at the primary cilium." Nature **437**(7061): 1018-1021.

Cotney, J., J. Leng, et al. "Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb." Genome Res **22**(6): 1069-1080.

Couly, G. F., P. M. Coltey, et al. (1993). "The triple origin of skull in higher vertebrates: a study in quail-chick chimeras." Development **117**(2): 409-429.

Cox, B., J. Briscoe, et al. (2010). "SUMOylation by Pias1 regulates the activity of the Hedgehog dependent Gli transcription factors." PloS one **5**(8): e11996.

Cremer, M., J. von Hase, et al. (2001). "Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells." Chromosome Res **9**(7): 541-567.

Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." Nat Rev Genet **2**(4): 292-301.

Creyghton, M. P., A. W. Cheng, et al. "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Crocker, J., Y. Tamori, et al. (2008). "Evolution acts on enhancer organization to fine-tune gradient threshold readouts." PLoS Biol **6**(11): e263.

Cuddapah, S., R. Jothi, et al. (2009). "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains." Genome Res **19**(1): 24-32.

Cunliffe, V. T. (2004). "Histone deacetylase 1 is required to repress Notch target gene expression during zebrafish neurogenesis and to maintain the production of motoneurones in response to hedgehog signalling." Development **131**(12): 2983-2995.

Currie, P. D. and P. W. Ingham (1996). "Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish." Nature **382**(6590): 452-455.

D'Amico-Martel, A. and D. M. Noden (1983). "Contributions of placodal and neural crest cells to avian cranial peripheral ganglia." Am J Anat **166**(4): 445-468.

Danen, E. H. and A. Sonnenberg (2003). "Integrins in regulation of tissue development and function." The Journal of pathology **201**(4): 632-641.

Das, S., R. S. Samant, et al. "Nonclassical activation of Hedgehog signaling enhances multidrug resistance and makes cancer cells refractory to Smoothened-targeting Hedgehog inhibition." J Biol Chem **288**(17): 11824-11833.

Dassule, H. R., P. Lewis, et al. (2000). "Sonic hedgehog regulates growth and morphogenesis of the tooth." Development **127**(22): 4775-4785.

Davidson, E. H. (2006). "The regulatory genome: Gene Regulatory Networks in Development and Evolution". Academic Press

Davidson, E. H. and D. H. Erwin (2006). "Gene regulatory networks and the evolution of animal body plans." Science **311**(5762): 796-800.

Dean, A. (2011). "In the loop: long range chromatin interactions and gene regulation." Brief Funct Genomics **10**(1): 3-10.

Decker, T., M. Pasca di Magliano, et al. (2009). "Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis." Immunity **30**(4): 508-520.

Denef, N., D. Neubuser, et al. (2000). "Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened." Cell **102**(4): 521-531.

Dennler, S., J. Andre, et al. (2007). "Induction of sonic hedgehog mediators by transforming growth factor-beta: Smad3-dependent activation of Gli2 and Gli1 expression in vitro and in vivo." Cancer Res **67**(14): 6981-6986.

Dermitzakis, E. T., A. Reymond, et al. (2002). "Numerous potentially functional but non-genic conserved sequences on human chromosome 21." Nature **420**(6915): 578-582.

Dessaud, E., A. P. McMahon, et al. (2008). "Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network." Development **135**(15): 2489-2503.

Di Rocco, G., A. Gavalas, et al. (2001). "The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function." J Biol Chem **276**(23): 20506-20515.

Dickmeis, T., C. Plessy, et al. (2004). "Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo." Genome Res **14**(2): 228-238.

Ding, Q., J. Motoyama, et al. (1998). "Diminished Sonic hedgehog signaling and lack of floor plate differentiation in Gli2 mutant mice." Development **125**(14): 2533-2543.

Dolez, M., J. F. Nicolas, et al. (2011). "Laminins, via heparan sulfate proteoglycans, participate in zebrafish myotome morphogenesis by modulating the pattern of Bmp responsiveness." Development **138**(1): 97-106.

Domogatskaya, A., S. Rodin, et al. (2012). "Functional diversity of laminins." Annual review of cell and developmental biology **28**: 523-553.

Drewell R. A. (2011) "Transcription Factor Binding Site Redundancy in Embryonic Enhancers of the Drosophila Bithorax Complex". G3, Genes Genomes Genetics. December; 1(7): 603–606

Duband, J. L., S. Dufour, et al. (1988). "The migratory behavior of avian embryonic cells does not require phosphorylation of the fibronectin-receptor complex." FEBS Lett **230**(1-2): 181-185.

Durbeej, M. (2010). "Laminins." Cell and tissue research **339**(1): 259-268.

Durbeej, M., L. Fecker, et al. (1996). "Expression of laminin alpha 1, alpha 5 and beta 2 chains during embryogenesis of the kidney and vasculature." Matrix biology : journal of the International Society for Matrix Biology **15**(6): 397-413.

Dziadek, M. and R. Timpl (1985). "Expression of nidogen and laminin in basement membranes during mouse embryogenesis and in teratocarcinoma cells." Developmental biology **111**(2): 372-382.

Echelard, Y., D. J. Epstein, et al. (1993). "Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity." Cell **75**(7): 1417-1430.

Edwards, M. M. and O. Lefebvre (2013). "Laminins and retinal vascular development." Cell adhesion & migration **7**(1): 82-89.

Edwards, M. M., E. Mammadova-Bach, et al. (2010). "Mutations in Lama1 disrupt retinal vascular development and inner limiting membrane formation." The Journal of biological chemistry **285**(10): 7697-7711.

Edwards, M. M., D. S. McLeod, et al. (2011). "Lama1 mutations lead to vitreoretinal blood vessel formation, persistence of fetal vasculature, and epiretinal membrane formation in mice." BMC developmental biology **11**: 60.

Ekblom, M., G. Klein, et al. (1990). "Transient and locally restricted expression of laminin A chain mRNA by developing epithelial cells during kidney organogenesis." Cell **60**(2): 337-346.

Ekblom, P., P. Lonai, et al. (2003). "Expression and biological role of laminin-1." Matrix biology : journal of the International Society for Matrix Biology **22**(1): 35-47.

Ekker, S. C., L. L. McGrew, et al. (1995). "Distinct expression and shared activities of members of the hedgehog gene family of Xenopus laevis." Development **121**(8): 2337-2347.

Ekker, S. C., A. R. Ungar, et al. (1995). "Patterning activities of vertebrate hedgehog proteins in the developing eye and brain." Curr Biol **5**(8): 944-955.

Elgar, G. (2009). "Pan-vertebrate conserved non-coding sequences associated with developmental regulation." Brief Funct Genomic Proteomic **8**(4): 256-265.

Elworthy, S., M. Hargrave, et al. (2008). "Expression of multiple slow myosin heavy chain genes reveals a diversity of zebrafish slow twitch muscle fibres with differing requirements for Hedgehog and Prdm1 activity." Development **135**(12): 2115-2126.

Endoh-Yamagami, S., M. Evangelista, et al. (2009). "The mammalian Cos2 homolog Kif7 plays an essential role in modulating Hh signal transduction during development." Curr Biol **19**(15): 1320-1326.

Engel, N., J. L. Thorvaldsen, et al. (2006). "CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus." Hum Mol Genet **15**(19): 2945-2954.

Engstrom, P. G., S. J. Ho Sui, et al. (2007). "Genomic regulatory blocks underlie extensive microsynteny conservation in insects." Genome Res **17**(12): 1898-1908.

Ericson, J., J. Muhr, et al. (1995). "Sonic hedgehog: a common signal for ventral patterning along the rostrocaudal axis of the neural tube." Int J Dev Biol **39**(5): 809-816.

Ericson, J., J. Muhr, et al. (1995). "Sonic hedgehog induces the differentiation of ventral forebrain neurons: a common signal for ventral patterning within the neural tube." Cell **81**(5): 747-756.

Ericson, J., P. Rashbass, et al. (1997). "Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling." Cell **90**(1): 169-180.

Ertzer, R., F. Muller, et al. (2007). "Cooperation of sonic hedgehog enhancers in midline expression." Dev Biol **301**(2): 578-589.

Essafi, A., A. Webb, et al. "A wt1-controlled chromatin switching mechanism underpins tissue-specific wnt4 activation and repression." Dev Cell **21**(3): 559-574.

Eugster, C., D. Panakova, et al. (2007). "Lipoprotein-heparan sulfate interactions in the Hh pathway." Developmental cell **13**(1): 57-71.

Evans, N. C., C. I. Swanson, et al. "Sparkling insights into enhancer structure, function, and evolution." Curr Top Dev Biol **98**: 97-120.

Falk, M., M. Ferletta, et al. (1999). "Restricted distribution of laminin alpha1 chain in normal adult mouse tissues." Matrix Biol **18**(6): 557-568.

Faro-Trindade I, Cook PR. (2006). "Transcription factories: structures conserved during differentiation and evolution." Biochem Soc Trans. Dec;34(Pt 6):1133-7.

Farrell, C. M., A. G. West, et al. (2002). "Conserved CTCF insulator elements flank the mouse and human beta-globin loci." Mol Cell Biol **22**(11): 3820-3831.

Felgner, P. L., T. R. Gadek, et al. (1987). "Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure." Proc Natl Acad Sci U S A **84**(21): 7413-7417.

Feng, J., C. Bi, et al. (2006). "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator." Genes Dev **20**(11): 1470-1484.

Feng, Y. Q., R. Warin, et al. (2005). "The human beta-globin locus control region can silence as well as activate gene expression." Mol Cell Biol **25**(10): 3864-3874.

Fisher, S., E. A. Grice, et al. (2006). "Conservation of RET regulatory function from human to zebrafish without sequence similarity." Science **312**(5771): 276-279.

Flanagan, L. A., L. M. Rebaza, et al. (2006). "Regulation of human neural precursor cells by laminin and integrins." J Neurosci Res **83**(5): 845-856.

Follows, G. A., R. Ferreira, et al. "Mapping and functional characterisation of a CTCF-dependent insulator element at the 3' border of the murine Scl transcriptional domain." PLoS One **7**(3): e31484.

Fox, J. W., U. Mayer, et al. (1991). "Recombinant nidogen consists of three globular domains and mediates binding of laminin to collagen type IV." The EMBO journal **10**(11): 3137-3146.

Frade, J. M., J. R. Martinez-Morales, et al. (1996). "Laminin-1 selectively stimulates neuron generation from cultured retinal neuroepithelial cells." Experimental cell research **222**(1): 140-149.

Frankel, N. "Multiple layers of complexity in cis-regulatory regions of developmental genes." Dev Dyn **241**(12): 1857-1866.

Frankel, N., G. K. Davis, et al. "Phenotypic robustness conferred by apparently redundant transcriptional enhancers." Nature **466**(7305): 490-493.

Fraser, P. and F. Grosveld (1998). "Locus control regions, chromatin activation and transcription." Curr Opin Cell Biol **10**(3): 361-365.

Frazer, K. A., L. Pachter, et al. (2004). "VISTA: computational tools for comparative genomics." Nucleic Acids Res **32**(Web Server issue): W273-279.

Fu, X. H., D. P. Liu, et al. (2002). "Chromatin structure and transcriptional regulation of the beta-globin locus." Exp Cell Res **278**(1): 1-11.

Fuda NJ, Ardehali MB, Lis JT. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature. Sep 10;461(7261):186-92.

Fujiwara, H., Y. Hayashi, et al. (2007). "Regulation of mesodermal differentiation of mouse embryonic stem cells by basement membranes." J Biol Chem **282**(40): 29701-29711.

Funalot, B., P. Topilko, et al. "Homozygous deletion of an EGR2 enhancer in congenital amyelinating neuropathy." Ann Neurol **71**(5): 719-723.

Futaki, S., Y. Hayashi, et al. (2004). "Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression." Molecular and cellular biology **24**(23): 10492-10503.

Galceran, J., S. C. Hsu, et al. (2001). "Rescue of a Wnt mutation by an activated form of LEF-1: regulation of maintenance but not initiation of Brachyury expression." Proc Natl Acad Sci U S A **98**(15): 8668-8673.

Gallegos, T. F., V. Kouznetsova, et al. "A protein kinase A and Wnt-dependent network regulating an intermediate stage in epithelial tubulogenesis during kidney development." Dev Biol **364**(1): 11-21.

Gao, J., J. Wang, et al. "Regulation of Pax6 by CTCF during induction of mouse ES cell differentiation." PLoS One **6**(6): e20954.

Gao, Q. and R. Finkelstein (1998). "Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen." Development **125**(21): 4185-4193.

Gaston K, Jayaraman PS. (2003). "Transcriptional repression in eukaryotes: repressors and repression mechanisms." Cell Mol Life Sci. Apr;60(4):721-41.

Gaszner, M. and G. Felsenfeld (2006). "Insulators: exploiting transcriptional and epigenetic mechanisms." Nat Rev Genet **7**(9): 703-713.

Gates, K. P., L. Mentzer, et al. "The transcriptional repressor REST/NRSF modulates hedgehog signaling." Dev Biol 340(2): 293-305.

Gering, M. and R. Patient (2005). "Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos." Dev Cell 8(3): 389-400.

Ghayor, C., M. Ehrbar, et al. (2009). "cAMP enhances BMP2-signaling through PKA and MKP1-dependent mechanisms." Biochem Biophys Res Commun 381(2): 247-252.

Ghiasvand, N. M., D. D. Rudolph, et al. (2011). "Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease." Nat Neurosci 14(5): 578-586.

Gibcus, J. H. and J. Dekker "The hierarchy of the 3D genome." Mol Cell 49(5): 773-782.

Gibcus, J. H. and J. Dekker (2013). "The hierarchy of the 3D genome." Mol Cell 49(5): 773-782.

Glazko, G. V., E. V. Koonin, et al. (2003). "A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions." Trends Genet 19(3): 119-124.

Goll M.G., Anderson R., Stainier D.Y., Spradling A.C., Halpern M.E. (2009). "Transcriptional silencing and reactivation in transgenic zebrafish." Genetics. Jul;182(3):747-55.

Gomez-Skarmeta, J. L., I. Rodriguez, et al. (1995). "Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs." Genes Dev 9(15): 1869-1882.

Gordon, K. L. and I. Ruvinsky (2012). "Tempo and mode in evolution of transcriptional regulation." PLoS Genet 8(1): e1002432.

Gorkin, D. U., D. Lee, et al. "Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes." Genome Res 22(11): 2290-2301.

Gorman, J. and E. C. Greene (2008). "Visualizing one-dimensional diffusion of proteins along DNA." Nat Struct Mol Biol 15(8): 768-774.

Gotea, V., A. Visel, et al. "Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers." Genome Res 20(5): 565-577.

Gottgens, B., L. M. Barton, et al. (2000). "Analysis of vertebrate SCL loci identifies conserved enhancers." Nat Biotechnol 18(2): 181-186.

Gros, J., M. Manceau, et al. (2005). "A common somitic origin for embryonic muscle progenitors and satellite cells." Nature 435(7044): 954-958.

Gros, J., M. Scaal, et al. (2004). "A two-step mechanism for myotome formation in chick." Dev Cell 6(6): 875-882.

Guerrero, L., R. Marco-Ferreres, et al. (2010). "Secondary enhancers synergise with primary enhancers to guarantee fine-tuned muscle gene expression." Dev Biol 337(1): 16-28.

Guillemot, F. (2007). "Spatial and temporal specification of neural fates by transcription factor codes." Development 134(21): 3771-3780.

Guillemot, F. and A. L. Joyner (1993). "Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system." Mech Dev 42(3): 171-185.

Guillemot, F., L. C. Lo, et al. (1993). "Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons." Cell 75(3): 463-476.

Gullberg, D., G. Sjoberg, et al. (1995). "Analysis of fibronectin and vitronectin receptors on human fetal skeletal muscle cells upon differentiation." Exp Cell Res 220(1): 112-123.

Guner, B. and R. O. Karlstrom (2007). "Cloning of zebrafish nkx6.2 and a comprehensive analysis of the conserved transcriptional response to Hedgehog/Gli signaling in the zebrafish neural tube." Gene Expr Patterns 7(5): 596-605.

Guo, Y., K. Monahan, et al. "CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice." Proc Natl Acad Sci U S A **109**(51): 21081-21086.

Haas, A. R. and R. S. Tuan (2000). "Murine C3H10T1/2 multipotential cells as an in vitro model of mesenchymal chondrogenesis." Methods Mol Biol **137**: 383-389.

Haeussler, M. and J. S. Joly "When needles look like hay: how to find tissue-specific enhancers in model organism genomes." Dev Biol **350**(2): 239-254.

Halfter, W., S. Dong, et al. (2002). "A critical function of the pial basement membrane in cortical histogenesis." The Journal of neuroscience : the official journal of the Society for Neuroscience **22**(14): 6029-6040.

Hall, P. E., J. D. Lathia, et al. (2008). "Laminin enhances the growth of human neural stem cells in defined culture media." BMC Neurosci **9**: 71.

Hallikas, O., K. Palin, et al. (2006). "Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity." Cell **124**(1): 47-59.

Hammerschmidt, M., F. Pelegri, et al. (1996). "Mutations affecting morphogenesis during gastrulation and tail formation in the zebrafish, Danio rerio." Development **123**: 143-151.

Hampsey, M., B. N. Singh, et al. "Control of eukaryotic gene expression: gene loops and transcriptional memory." Adv Enzyme Regul **51**(1): 118-125.

Han, C., T. Y. Belenkaya, et al. (2004). "Drosophila glypicans control the cell-to-cell movement of Hedgehog by a dynamin-independent process." Development **131**(3): 601-611.

Handoko, L., H. Xu, et al. "CTCF-mediated functional chromatin interactome in pluripotent cells." Nat Genet **43**(7): 630-638.

Hantaz-Ambroise, D., M. Vigny, et al. (1987). "Heparan sulfate proteoglycan and laminin mediate two different types of neurite outgrowth." J Neurosci **7**(8): 2293-2304.

Hare, E. E., B. K. Peterson, et al. (2008). "A careful look at binding site reorganization in the even-skipped enhancers of Drosophila and sepsids." PLoS Genet **4**(11): e1000268.

Hare, E. E., B. K. Peterson, et al. (2008). "Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation." PLoS Genet **4**(6): e1000106.

Hark, A. T., C. J. Schoenherr, et al. (2000). "CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus." Nature **405**(6785): 486-489.

Hasty, P., A. Bradley, et al. (1993). "Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene." Nature **364**(6437): 501-506.

Hayashibara, Y., K. Mitsunaga-Nakatsubo, et al. (2004). "The Otx binding site is required for the activation of HpOtxL mRNA expression in the sea urchin, Hemicentrotus pulcherrimus." Dev Growth Differ **46**(1): 61-67.

He, X., M. N. Treacy, et al. (1989). "Expression of a large family of POU-domain regulatory genes in mammalian brain development." Nature **340**(6228): 35-41.

Hebbes, T. R. and S. C. Allen (2000). "Multiple histone acetyltransferases are associated with a chicken erythrocyte chromatin fraction enriched in active genes." J Biol Chem **275**(40): 31347-31352.

Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.

Hemberg, M., J. M. Gray, et al. (2012). "Integrated genome analysis suggests that most conserved non-coding sequences are regulatory factor binding sites." Nucleic Acids Res **40**(16): 7858-7869.

Heng, C., O. Lefebvre, et al. (2011). "Functional role of laminin alpha1 chain during cerebellum development." Cell adhesion & migration **5**(6): 480-489.

Hermann, B. P. and L. L. Heckert (2005). "Silencing of Fshr occurs through a conserved, hypersensitive site in the first intron." Mol Endocrinol **19**(8): 2112-2131.

Hernandez-Vega, A. and C. Minguillon (2011). "The Prx1 limb enhancers: targeted gene expression in developing zebrafish pectoral fins." Dev Dyn **240**(8): 1977-1988.

Herold, M., M. Bartkuhn, et al. "CTCF: insights into insulator function during development." Development **139**(6): 1045-1057.

Herrera, E., L. Brown, et al. (2003). "Zic2 patterns binocular vision by specifying the uncrossed retinal projection." Cell **114**(5): 545-557.

Hidalgo, A. and P. Ingham (1990). "Cell patterning in the Drosophila segment: spatial regulation of the segment polarity gene patched." Development **110**(1): 291-301.

Higashijima, S., H. Okamoto, et al. (1997). "High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin." Dev Biol **192**(2): 289-299.

Hlawatsch, J., M. Karlstetter, et al. "Sterile alpha motif containing 7 (samd7) is a novel crx-regulated transcriptional repressor in the retina." PLoS One **8**(4): e60633.

Hofmann, A. and M. Lehmann (1998). "The transcriptional switch between the Drosophila genes Pig-1 and Sgs-4 depends on a SEBP1 binding site within a shared enhancer region." Mol Gen Genet **259**(6): 656-663.

Hogues, H., H. Lavoie, et al. (2008). "Transcription factor substitution during the evolution of fungal ribosome regulation." Mol Cell **29**(5): 552-562.

Hohenester, E., D. Tisi, et al. (1999). "The crystal structure of a laminin G-like module reveals the molecular basis of alpha-dystroglycan binding to laminins, perlecan, and agrin." Molecular cell **4**(5): 783-792.

Hohenester, E. and P. D. Yurchenco (2013). "Laminins in basement membrane assembly." Cell adhesion & migration **7**(1): 56-63.

Holmqvist, P. H. and M. Mannervik "Genomic occupancy of the transcriptional co-activators p300 and CBP." Transcription **4**(1): 18-23.

Hong, J. W., D. A. Hendrix, et al. (2008). "Shadow enhancers as a source of evolutionary novelty." Science **321**(5894): 1314.

Hozumi A, Yoshida R, Horie T, Sakuma T, Yamamoto T, Sasakura Y. (2013). "Enhancer activity sensitive to the orientation of the gene it regulates in the chordate genome." Dev Biol. Mar 1;375(1):79-91.

Hu, D. and J. A. Helms (1999). "The role of sonic hedgehog in normal and abnormal craniofacial morphogenesis." Development **126**(21): 4873-4884.

Hu, D. and R. S. Marcucio (2009). "A SHH-responsive signaling center in the forebrain regulates craniofacial morphogenesis via the facial ectoderm." Development **136**(1): 107-116.

Huang, C. C., D. H. Hall, et al. (2003). "Laminin alpha subunits and their role in C. elegans development." Development **130**(14): 3343-3358.

Huangfu, D. and K. V. Anderson (2005). "Cilia and Hedgehog responsiveness in the mouse." Proc Natl Acad Sci U S A **102**(32): 11325-11330.

Hui, C. C. and S. Angers "Gli proteins in development and disease." Annu Rev Cell Dev Biol **27**: 513-537.

Hui, C. C., D. Slusarski, et al. (1994). "Expression of three mouse homologs of the Drosophila segment polarity gene cubitus interruptus, Gli, Gli-2, and Gli-3, in ectoderm- and mesoderm-derived tissues suggests multiple roles during postimplantation development." Dev Biol **162**(2): 402-413.

Humke, E. W., K. V. Dorn, et al. (2010). "The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins." Genes Dev **24**(7): 670-682.

Huttenlocher, A., R. R. Sandborg, et al. (1995). "Adhesion in cell migration." <u>Curr Opin Cell Biol</u> **7**(5): 697-706.

Ichikawa-Tomikawa, N., J. Ogawa, et al. (2012). "Laminin alpha1 is essential for mouse cerebellar development." <u>Matrix biology : journal of the International Society for Matrix Biology</u> **31**(1): 17-28.

Ido, H., K. Harada, et al. (2004). "Molecular dissection of the alpha-dystroglycan- and integrin-binding sites within the globular domain of human laminin-10." <u>The Journal of biological chemistry</u> **279**(12): 10946-10954.

Ido, H., A. Nakamura, et al. (2007). "The requirement of the glutamic acid residue at the third position from the carboxyl termini of the laminin gamma chains in integrin binding by laminins." <u>The Journal of biological chemistry</u> **282**(15): 11144-11154.

Ikle, J. M., K. B. Artinger, et al. (2012). "Identification and characterization of the zebrafish pharyngeal arch-specific enhancer for the basic helix-loop-helix transcription factor Hand2." <u>Dev Biol</u> **368**(1): 118-126.

Imhof, A., X. J. Yang, et al. (1997). "Acetylation of general transcription factors by histone acetyltransferases." <u>Curr Biol</u> **7**(9): 689-692.

Incardona, J. P., W. Gaffield, et al. (1998). "The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction." <u>Development</u> **125**(18): 3553-3562.

Incardona, J. P., J. Gruenberg, et al. (2002). "Sonic hedgehog induces the segregation of patched and smoothened in endosomes." <u>Curr Biol</u> **12**(12): 983-995.

Ingham, P. W. and A. P. McMahon (2001). "Hedgehog signaling in animal development: paradigms and principles." <u>Genes Dev</u> **15**(23): 3059-3087.

Ingham, P. W., Y. Nakano, et al. (2011). "Mechanisms and functions of Hedgehog signalling across the metazoa." <u>Nat Rev Genet</u> **12**(6): 393-406.

Ingham, P. W., S. Nystedt, et al. (2000). "Patched represses the Hedgehog signalling pathway by promoting modification of the Smoothened protein." <u>Curr Biol</u> **10**(20): 1315-1318.

Inoue, T., M. Hatayama, et al. (2004). "Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives." <u>Dev Biol</u> **270**(1): 146-162.

Irimia, M., J. L. Royo, et al. "Comparative genomics of the Hedgehog loci in chordates and the origins of Shh regulatory novelties." <u>Sci Rep</u> **2**: 433.

Irimia, M., J. J. Tena, et al. "Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints." <u>Genome Res</u> **22**(12): 2356-2367.

Irvine, S. Q., V. C. Fonseca, et al. (2008). "Cis-regulatory organization of the Pax6 gene in the ascidian Ciona intestinalis." <u>Dev Biol</u> **317**(2): 649-659.

Jackson, H. E. and P. W. Ingham "Control of muscle fibre-type diversity during embryonic development: The zebrafish paradigm." <u>Mech Dev</u> **130**(9-10): 447-457.

Jacob, J. and J. Briscoe (2003). "Gli proteins and the control of spinal-cord patterning." <u>EMBO Rep</u> **4**(8): 761-765.

Jenkins, D., P. J. Winyard, et al. (2007). "Immunohistochemical analysis of Sonic hedgehog signalling in normal human urinary tract development." <u>J Anat</u> **211**(5): 620-629.

Jeong, J. and A. P. McMahon (2005). "Growth and pattern of the mammalian neural tube are governed by partially overlapping feedback activities of the hedgehog antagonists patched 1 and Hhip1." <u>Development</u> **132**(1): 143-154.

Jeong, Y., K. El-Jaick, et al. (2006). "A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers." <u>Development</u> **133**(4): 761-772.

Jessell, T. M. (2000). "Neuronal specification in the spinal cord: inductive signals and transcriptional codes." <u>Nat Rev Genet</u> **1**(1): 20-29.

Jia, J., L. Zhang, et al. (2005). "Phosphorylation by double-time/CKIepsilon and CKIalpha targets cubitus interruptus for Slimb/beta-TRCP-mediated proteolytic processing." <u>Dev Cell</u> **9**(6): 819-830.

Jiang, J. and G. Struhl (1995). "Protein kinase A and hedgehog signaling in Drosophila limb development." <u>Cell</u> **80**(4): 563-572.

Johnson, R. L., R. D. Riddle, et al. (1994). "Sonic hedgehog: a key mediator of anterior-posterior patterning of the limb and dorso-ventral patterning of axial embryonic structures." <u>Biochem Soc Trans</u> **22**(3): 569-574.

Johnson, R. L., R. D. Riddle, et al. (1994). "Mechanisms of limb patterning." <u>Curr Opin Genet Dev</u> **4**(4): 535-542.

Junion, G., M. Spivakov, et al. (2012). "A transcription factor collective defines cardiac cell fate and reflects lineage history." <u>Cell</u> **148**(3): 473-486.

Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. (2008) "The RNA polymerase II core promoter - the gateway to transcription." Curr Opin Cell Biol. Jun;20(3):253-9.

Kadoya, Y., K. Salmivirta, et al. (1997). "Importance of nidogen binding to laminin gamma1 for branching epithelial morphogenesis of the submandibular gland." <u>Development</u> **124**(3): 683-691.

Kahane, N., V. Ribes, et al. "The transition from differentiation to growth during dermomyotome-derived myogenesis depends on temporally restricted hedgehog signaling." <u>Development</u> **140**(8): 1740-1750.

Kanda, T., K. F. Sullivan, et al. (1998). "Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells." <u>Curr Biol</u> **8**(7): 377-385.

Kanduri, C., C. Holmgren, et al. (2000). "The 5' flank of mouse H19 in an unusual chromatin conformation unidirectionally blocks enhancer-promoter communication." <u>Curr Biol</u> **10**(8): 449-457.

Kanduri, C., V. Pant, et al. (2000). "Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive." <u>Curr Biol</u> **10**(14): 853-856.

Kang, C. J., C. Sheridan, et al. (1998). "A stage-specific enhancer of immunoglobulin J chain gene is induced by interleukin-2 in a presecretor B cell stage." <u>Immunity</u> **8**(3): 285-295.

Karlstrom, R. O., W. S. Talbot, et al. (1999). "Comparative synteny cloning of zebrafish you-too: mutations in the Hedgehog target gli2 affect ventral forebrain patterning." <u>Genes Dev</u> **13**(4): 388-393.

Karlstrom, R. O., T. Trowe, et al. (1996). "Zebrafish mutations affecting retinotectal axon pathfinding." <u>Development</u> **123**: 427-438.

Karlstrom, R. O., O. V. Tyurina, et al. (2003). "Genetic analysis of zebrafish gli1 and gli2 reveals divergent requirements for gli genes in vertebrate development." <u>Development</u> **130**(8): 1549-1564.

Kawakami, K. (2004). "Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element." <u>Methods Cell Biol</u> **77**: 201-222.

Kawakami, K. (2007). "Tol2: a versatile gene transfer vector in vertebrates." <u>Genome Biol</u> **8 Suppl 1**: S7.

Kawakami, K., K. Imanaka, et al. (2004). "Excision of the Tol2 transposable element of the medaka fish Oryzias latipes in Xenopus laevis and Xenopus tropicalis." <u>Gene</u> **338**(1): 93-98.

Kawakami, K., A. Koga, et al. (1998). "Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio." <u>Gene</u> **225**(1-2): 17-22.

Kawakami, K. and T. Noda (2004). "Transposition of the Tol2 element, an Ac-like element from the Japanese medaka fish Oryzias latipes, in mouse embryonic stem cells." <u>Genetics</u> **166**(2): 895-899.

Kawakami, K. and A. Shima (1999). "Identification of the Tol2 transposase of the medaka fish Oryzias latipes that catalyzes excision of a nonautonomous Tol2 element in zebrafish Danio rerio." <u>Gene</u> **240**(1): 239-244.

Kawakami, K., A. Shima, et al. (2000). "Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage." Proc Natl Acad Sci U S A **97**(21): 11403-11408.

Kawakami, K., H. Takeda, et al. (2004). "A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish." Dev Cell **7**(1): 133-144.

Kawakami, T., T. Kawcak, et al. (2002). "Mouse dispatched mutants fail to distribute hedgehog proteins and are defective in hedgehog signaling." Development **129**(24): 5753-5765.

Kawata, Y., H. Suzuki, et al. (2002). "bcn-1 Element-dependent activation of the laminin gamma 1 chain gene by the cooperative action of transcription factor E3 (TFE3) and Smad proteins." J Biol Chem **277**(13): 11375-11384.

Keeler, R. F. and W. Binns (1968). "Teratogenic compounds of Veratrum californicum (Durand). V. Comparison of cyclopian effects of steroidal alkaloids from the plant and structurally related compounds from other sources." Teratology **1**(1): 5-10.

Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kerney, R., B. K. Hall, et al. "Regulatory elements of Xenopus col2a1 drive cartilaginous gene expression in transgenic frogs." Int J Dev Biol **54**(1): 141-150.

Kim, H. R., J. Richardson, et al. "Gli2a protein localization reveals a role for Iguana/DZIP1 in primary ciliogenesis and a dependence of Hedgehog signal transduction on primary cilia in the zebrafish." BMC Biol **8**: 65.

Kim, H. R., J. Richardson, et al. (2010). "Gli2a protein localization reveals a role for Iguana/DZIP1 in primary ciliogenesis and a dependence of Hedgehog signal transduction on primary cilia in the zebrafish." BMC Biol **8**: 65.

Kim, J. and H. Kim "Recruitment and biological consequences of histone modification of H3K27me3 and H3K9me3." Ilar J **53**(3-4): 232-239.

Kim, J., J. E. Lee, et al. (2010). "Functional genomic screen for modulators of ciliogenesis and cilium length." Nature **464**(7291): 1048-1051.

Kim, K. and A. Kim "Sequential changes in chromatin structure during transcriptional activation in the beta globin LCR and its target gene." Int J Biochem Cell Biol **42**(9): 1517-1524.

Kim, T. K., M. Hemberg, et al. "Widespread transcription at neuronal activity-regulated enhancers." Nature **465**(7295): 182-187.

Kim, Y. W. and A. Kim "Histone acetylation contributes to chromatin looping between the locus control region and globin gene by influencing hypersensitive site formation." Biochim Biophys Acta **1829**(9): 963-969.

Kimura, H. "Histone modifications for human epigenome analysis." J Hum Genet **58**(7): 439-445.

Kinto, N., M. Iwamoto, et al. (1997). "Fibroblasts expressing Sonic hedgehog induce osteoblast differentiation and ectopic bone formation." FEBS Lett **404**(2-3): 319-323.

Kinzler, K. W. and B. Vogelstein (1990). "The GLI gene encodes a nuclear protein which binds specific sequences in the human genome." Mol Cell Biol **10**(2): 634-642.

Kioussis, D. and R. Festenstein (1997). "Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals." Curr Opin Genet Dev **7**(5): 614-619.

Kirchhamer, C. V., L. D. Bogarad, et al. (1996). "Developmental expression of synthetic cis-regulatory systems composed of spatial control elements from two different genes." Proc Natl Acad Sci U S A **93**(24): 13849-13854.

Kirchhamer, C. V., C. H. Yuh, et al. (1996). "Modular cis-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples." Proc Natl Acad Sci U S A **93**(18): 9322-9328.

Kise, Y., A. Morinaka, et al. (2009). "Sufu recruits GSK3beta for efficient processing of Gli3." Biochem Biophys Res Commun **387**(3): 569-574.

Kitchen, N. S. and C. J. Schoenherr "Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin." J Cell Biochem **111**(3): 665-675.

Klein, G., M. Langegger, et al. (1988). "Role of laminin A chain in the development of epithelial cell polarity." Cell **55**(2): 331-341.

Kleinjan, D. A., A. Seawright, et al. (2006). "Long-range downstream enhancers are essential for Pax6 expression." Dev Biol **299**(2): 563-581.

Kleinjan, D. A., A. Seawright, et al. (2001). "Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6." Hum Mol Genet **10**(19): 2049-2059.

Kleinman, H. K., I. Ebihara, et al. (1987). "Genes for basement membrane proteins are coordinately expressed in differentiating F9 cells but not in normal adult murine tissues." Developmental biology **122**(2): 373-378.

Kmita, M., N. Fraudeau, et al. (2002). "Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs." Nature **420**(6912): 145-150.

Knoepfler, P. S., D. A. Bergstrom, et al. (1999). "A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1." Nucleic Acids Res **27**(18): 3752-3761.

Kobayashi, A., Y. Watanabe, et al. (2007). "Real-time monitoring of functional interactions between upstream and core promoter sequences in living cells of sea urchin embryos." Nucleic Acids Res **35**(14): 4882-4894.

Koch, F. and J. C. Andrau "Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity." Transcription **2**(6): 263-268.

Koop, E. A., S. M. Lopes, et al. (2003). "Receptor protein tyrosine phosphatase mu expression as a marker for endothelial cell heterogeneity; analysis of RPTPmu gene expression using LacZ knock-in mice." Int J Dev Biol **47**(5): 345-354.

Koyabu, Y., K. Nakata, et al. (2001). "Physical and functional interactions between Zic and Gli proteins." J Biol Chem **276**(10): 6889-6892.

Krauss, S., J. P. Concordet, et al. (1993). "A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos." Cell **75**(7): 1431-1444.

Kriegstein, A. and A. Alvarez-Buylla (2009). "The glial nature of embryonic and adult neural stem cells." Annu Rev Neurosci **32**: 149-184.

Krotoski, D. M., S. E. Fraser, et al. (1988). "Mapping of neural crest pathways in Xenopus laevis using inter- and intra-specific cell markers." Developmental biology **127**(1): 119-132.

Kulkarni, M. M. and D. N. Arnosti (2005). "cis-regulatory logic of short-range transcriptional repression in Drosophila melanogaster." Mol Cell Biol **25**(9): 3411-3420.

Kumagai, C., T. Kadowaki, et al. (1997). "Disulfide-bonding between Drosophila laminin beta and gamma chains is essential for alpha chain to form alpha betagamma trimer." FEBS letters **412**(1): 211-216.

Kundu, M., A. Kuzin, et al. (2013). "cis-regulatory complexity within a large non-coding region in the Drosophila genome." PLoS One **8**(4): e60137.

Kurdistani, S. K. and M. Grunstein (2003). "Histone acetylation and deacetylation in yeast." Nat Rev Mol Cell Biol **4**(4): 276-284.

Kurukuti, S., V. K. Tiwari, et al. (2006). "CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2." Proc Natl Acad Sci U S A **103**(28): 10684-10689.

Kyrchanova, O., T. Ivlieva, et al. "Selective interactions of boundaries with upstream region of Abd-B promoter in Drosophila bithorax complex and role of dCTCF in this process." Nucleic Acids Res **39**(8): 3042-3052.

Lagha M., Bothma J.P., Levine M. (2012) "Mechanisms of transcriptional precision in animal development". Trends in Genetics, Volume 28, Issue 8, Pages 409-416

Lakowski, B., I. Roelens, et al. (2006). "CoREST-like complexes regulate chromatin modification and neuronal gene expression." J Mol Neurosci **29**(3): 227-239.

Lassar, A. B., R. L. Davis, et al. (1991). "Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo." Cell **66**(2): 305-315.

Lathia, J. D., B. Patton, et al. (2007). "Patterns of laminins and integrins in the embryonic ventricular zone of the CNS." J Comp Neurol **505**(6): 630-643.

Lebrecht, D., M. Foehr, et al. (2005). "Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila." Proc Natl Acad Sci U S A **102**(37): 13176-13181.

Lee, E. Y., H. Ji, et al. "Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis." Proc Natl Acad Sci U S A **107**(21): 9736-9741.

Lee, H., G. A. Shamy, et al. (2007). "Directed differentiation and transplantation of human embryonic stem cell-derived motoneurons." Stem cells **25**(8): 1931-1939.

Lee, J., K. A. Platt, et al. (1997). "Gli1 is a target of Sonic hedgehog that induces ventral neural tube development." Development **124**(13): 2537-2552.

Lee, J. J., S. C. Ekker, et al. (1994). "Autoproteolysis in hedgehog protein biogenesis." Science **266**(5190): 1528-1537.

Leek, J. P., T. P. Moynihan, et al. (1997). "Assignment of Indian hedgehog (IHH) to human chromosome bands 2q33-->q35 by in situ hybridization." Cytogenetics and cell genetics **76**(3-4): 187-188.

Lentz, S. I., J. H. Miner, et al. (1997). "Distribution of the ten known laminin chains in the pathways and targets of developing sensory axons." The Journal of comparative neurology **378**(4): 547-561.

Lettice, L. A., S. J. Heaney, et al. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Hum Mol Genet **12**(14): 1725-1735.

Lettice, L. A., T. Horikoshi, et al. (2002). "Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly." Proc Natl Acad Sci U S A **99**(11): 7548-7553.

Levine M. (2010). "Transcriptional enhancers in animal development and evolution." Curr Biol. Sep 14;20(17):R754-63.

Levings, P. P. and J. Bungert (2002). "The human beta-globin locus control region." Eur J Biochem **269**(6): 1589-1599.

Lewis, K. E., J. P. Concordet, et al. (1999). "Characterisation of a second patched gene in the zebrafish Danio rerio and the differential response of patched genes to Hedgehog signalling." Dev Biol **208**(1): 14-29.

Lewis, P. M., M. P. Dunn, et al. (2001). "Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1." Cell **105**(5): 599-612.

Li, J., J. Tzu, et al. (2003). "Laminin-10 is crucial for hair morphogenesis." The EMBO journal **22**(10): 2400-2410.

Li, S., D. Harrison, et al. (2002). "Matrix assembly, regulation, and survival functions of laminin and its receptors in embryonic stem cell differentiation." The Journal of cell biology **157**(7): 1279-1290.

Li, W., D. Notani, et al. (2013). "Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation." Nature **498**(7455): 516-520.

Li, W., J. T. Ohlmeyer, et al. (1995). "Function of protein kinase A in hedgehog signal transduction and Drosophila imaginal disc development." Cell **80**(4): 553-562.

Li, X. and M. Noll (1994). "Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo." Embo J **13**(2): 400-406.

Li, X., U. Talts, et al. (2001). "Akt/PKB regulates laminin and collagen IV isotypes of the basement membrane." Proc Natl Acad Sci U S A **98**(25): 14416-14421.

Li, X. Y., S. R. Bhaumik, et al. (2002). "Selective recruitment of TAFs by yeast upstream activating sequences. Implications for eukaryotic promoter structure." Curr Biol **12**(14): 1240-1244.

Libby, R. T., W. J. Brunken, et al. (2000). "Roles of the extracellular matrix in retinal development and maintenance." Results and problems in cell differentiation **31**: 115-140.

Liem, K. F., Jr., M. He, et al. (2009). "Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling." Proc Natl Acad Sci U S A **106**(32): 13377-13382.

Liem, K. F., Jr., G. Tremml, et al. (1995). "Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm." Cell **82**(6): 969-979.

Lin, S., N. Gaiano, et al. (1994). "Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish." Science **265**(5172): 666-669.

Litingtung, Y. and C. Chiang (2000). "Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3." Nat Neurosci **3**(10): 979-985.

Liu, G., A. Moro, et al. (2007). "The role of Shh transcription activator Gli2 in chick cloacal development." Dev Biol **303**(2): 448-460.

Liu, H., J. F. Margiotta, et al. (2005). "BMP4 supports noradrenergic differentiation by a PKA-dependent mechanism." Dev Biol **286**(2): 521-536.

Liu, J. and U. Francke (2006). "Identification of cis-regulatory elements for MECP2 expression." Hum Mol Genet **15**(11): 1769-1782.

Liu, Y., X. Cao, et al. (2007). "Fused-Costal2 protein complex regulates Hedgehog-induced Smo phosphorylation and cell-surface accumulation." Genes Dev **21**(15): 1949-1963.

Liu, Y., H. Li, et al. (2000). "Identification of an enhancer sequence within the first intron required for cartilage-specific transcription of the alpha2(XI) collagen gene." J Biol Chem **275**(17): 12712-12718.

Lobanenkov, V. V., R. H. Nicolas, et al. (1990). "A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene." Oncogene **5**(12): 1743-1753.

Long, Q., A. Meng, et al. (1997). "GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene." Development **124**(20): 4105-4111.

Loots, G. G. and I. Ovcharenko (2004). "rVISTA 2.0: evolutionary analysis of transcription factor binding sites." Nucleic Acids Res **32**(Web Server issue): W217-221.

Lopez-Martinez, A., D. T. Chang, et al. (1995). "Limb-patterning activity and restricted posterior localization of the amino-terminal product of Sonic hedgehog cleavage." Curr Biol **5**(7): 791-796.

Lower, K. M., J. R. Hughes, et al. (2009). "Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition." Proc Natl Acad Sci U S A **106**(51): 21771-21776.

Lu, Q. R., D. Yuk, et al. (2000). "Sonic hedgehog--regulated oligodendrocyte lineage genes encoding bHLH proteins in the mammalian central nervous system." Neuron **25**(2): 317-329.

Ludwig, M. Z., C. Bergman, et al. (2000). "Evidence for stabilizing selection in a eukaryotic enhancer element." Nature **403**(6769): 564-567.

Ludwig, M. Z., N. H. Patel, et al. (1998). "Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change." Development **125**(5): 949-958.

Lum, H. K. and K. L. Lee (2001). "The human HMGB1 promoter is modulated by a silencer and an enhancer-containing intron." Biochim Biophys Acta **1520**(1): 79-84.

Lunyak, V. V. and M. G. Rosenfeld (2005). "No rest for REST: REST/NRSF regulation of neurogenesis." Cell **121**(4): 499-501.

Luo, J., M. J. Ju, et al. (2006). "Regionalized cadherin-7 expression by radial glia is regulated by Shh and Pax7 during chicken spinal cord development." Neuroscience **142**(4): 1133-1143.

Lutz, M., L. J. Burke, et al. (2003). "Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor." Embo J **22**(7): 1579-1587.

MacDonald, R. B., M. Debiais-Thibaud, et al. "Functional conservation of a forebrain enhancer from the elephant shark (Callorhinchus milii) in zebrafish and mice." BMC Evol Biol **10**: 157.

Macgregor, G. R., G. P. Nolan, et al. (1991). "Use of Escherichiu coli (E. coli) lacZ (beta-Galactosidase) as a Reporter Gene." Methods Mol Biol **7**: 217-235.

Maeda, R. K. and F. Karch (2006). "The ABC of the BX-C: the bithorax complex explained." Development **133**(8): 1413-1422.

Mahadevan, L. C., A. C. Willis, et al. (1991). "Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors." Cell **65**(5): 775-783.

Malin, J., M. R. Aniba, et al. "Enhancer networks revealed by correlated DNAse hypersensitivity states of enhancers." Nucleic Acids Res **41**(14): 6828-6838.

Mansouri, A. (1998). "The role of Pax3 and Pax7 in development and cancer." Crit Rev Oncog **9**(2): 141-149.

Marigo, V., D. J. Roberts, et al. (1995). "Cloning, expression, and chromosomal location of SHH and IHH: two human homologues of the Drosophila segment polarity gene hedgehog." Genomics **28**(1): 44-51.

Marinic, M., T. Aktas, et al. "An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape." Dev Cell **24**(5): 530-542.

Martin, C. and Y. Zhang (2005). "The diverse functions of histone lysine methylation." Nat Rev Mol Cell Biol **6**(11): 838-849.

Martin, G. R. and R. Timpl (1987). "Laminin and other basement membrane components." Annual review of cell biology **3**: 57-85.

Martinelli, D. C. and C. M. Fan (2007). "Gas1 extends the range of Hedgehog action by facilitating its signaling." Genes Dev **21**(10): 1231-1243.

Matise, M. P. and A. L. Joyner (1999). "Gli genes in development and cancer." Oncogene **18**(55): 7852-7859.

Matthews, J. C., K. Hori, et al. (1977). "Substrate and substrate analogue binding properties of Renilla luciferase." Biochemistry **16**(24): 5217-5220.

Maurya, A. K., H. Tan, et al. "Integration of Hedgehog and BMP signalling by the engrailed2a gene in the zebrafish myotome." Development **138**(4): 755-765.

May, D., M. J. Blow, et al. "Large-scale discovery of enhancers from human heart tissue." Nat Genet **44**(1): 89-93.

McBride, D. J., A. Buckle, et al. "DNaseI hypersensitivity and ultraconservation reveal novel, interdependent long-range enhancers at the complex Pax6 cis-regulatory region." PLoS One **6**(12): e28616.

McDermott, A., M. Gustafsson, et al. (2005). "Gli2 and Gli3 have redundant and context-dependent function in skeletal muscle formation." Development **132**(2): 345-357.

McGlinn, E. and C. J. Tabin (2006). "Mechanistic insight into how Shh patterns the vertebrate limb." Curr Opin Genet Dev **16**(4): 426-432.

McKee, K. K., D. Harrison, et al. (2007). "Role of laminin terminal globular domains in basement membrane assembly." The Journal of biological chemistry **282**(29): 21437-21447.

Mercola, M., J. Goverman, et al. (1985). "Immunoglobulin heavy-chain enhancer requires one or more tissue-specific factors." Science **227**(4684): 266-270.

Merika, M. and D. Thanos (2001). "Enhanceosomes." Curr Opin Genet Dev **11**(2): 205-208.

Merli, C., D. E. Bergstrom, et al. (1996). "Promoter specificity mediates the independent regulation of neighboring genes." Genes Dev **10**(10): 1260-1270.

Methot, N. and K. Basler (2000). "Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus." Development **127**(18): 4001-4010.

Miner, J. H. (2008). "Laminins and their roles in mammals." Microscopy research and technique **71**(5): 349-356.

Miner, J. H., J. Cunningham, et al. (1998). "Roles for laminin in embryogenesis: exencephaly, syndactyly, and placentopathy in mice lacking the laminin alpha5 chain." The Journal of cell biology **143**(6): 1713-1723.

Miner, J. H., R. M. Lewis, et al. (1995). "Molecular cloning of a novel laminin chain, alpha 5, and widespread expression in adult mouse tissues." The Journal of biological chemistry **270**(48): 28523-28526.

Miner, J. H. and C. Li (2000). "Defective glomerulogenesis in the absence of laminin alpha5 demonstrates a developmental role for the kidney glomerular basement membrane." Dev Biol **217**(2): 278-289.

Miner, J. H., C. Li, et al. (2004). "Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation." Development **131**(10): 2247-2256.

Miner, J. H. and P. D. Yurchenco (2004). "Laminin functions in tissue morphogenesis." Annual review of cell and developmental biology **20**: 255-284.

Miner, Z. and M. Kulesz-Martin (1997). "DNA binding specificity of proteins derived from alternatively spliced mouse p53 mRNAs." Nucleic Acids Res **25**(7): 1319-1326.

Mitsuhashi, Y. and I. Hashimoto (2003). "Genetic abnormalities and clinical classification of epidermolysis bullosa." Archives of dermatological research **295 Suppl 1**: S29-33.

Mizugishi, K., J. Aruga, et al. (2001). "Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins." J Biol Chem **276**(3): 2180-2188.

Mizuseki, K., M. Kishi, et al. (1998). "Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction." Development **125**(4): 579-587.

Montavon, T., N. Soshnikova, et al. "A regulatory archipelago controls Hox genes transcription in digits." Cell **147**(5): 1132-1145.

Moreno-Hagelsieb, G., V. Trevino, et al. (2001). "Transcription unit conservation in the three domains of life: a perspective from Escherichia coli." Trends Genet **17**(4): 175-177.

Mori, N., C. Schoenherr, et al. (1992). "A common silencer element in the SCG10 and type II Na+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells." Neuron **9**(1): 45-54.

Motoyama, J., H. Heng, et al. (1998). "Overlapping and non-overlapping Ptch2 expression with Shh during mouse embryogenesis." Mech Dev **78**(1-2): 81-84.

Motoyama, J., J. Liu, et al. (1998). "Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus." Nat Genet **20**(1): 54-57.

Motoyama, J., T. Takabatake, et al. (1998). "Ptch2, a second mouse Patched gene is co-expressed with Sonic hedgehog." Nat Genet **18**(2): 104-106.

Muller, F., P. Blader, et al. (2002). "Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements." Bioessays **24**(6): 564-572.

Müller F., Tora L. (2013) "Chromatin and DNA sequences in defining promoters for transcription initiation." Biochim Biophys Acta. Nov 22.

Munshi, A., G. Shafi, et al. (2009). "Histone modifications dictate specific biological readouts." <u>J Genet Genomics</u> **36**(2): 75-88.

Murre, C., P. S. McCaw, et al. (1989). "Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence." <u>Cell</u> **58**(3): 537-544.

Murshed, M., N. Smyth, et al. (2000). "The absence of nidogen 1 does not affect murine basement membrane formation." <u>Molecular and cellular biology</u> **20**(18): 7007-7012.

Nagai, T., J. Aruga, et al. (2000). "Zic2 regulates the kinetics of neurulation." <u>Proc Natl Acad Sci U S A</u> **97**(4): 1618-1623.

Nagai, T., J. Aruga, et al. (1997). "The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation." <u>Dev Biol</u> **182**(2): 299-313.

Nagy A., Gertsenstein M., Vintersten K., Behringer R. (2003) "Manipulating the Mouse Embryo: A Laboratory Manual", 3rd ed. Cold Spring Harbor Laboratory Press

Nakano, Y., S. Nystedt, et al. (2004). "Functional domains and sub-cellular distribution of the Hedgehog transducing protein Smoothened in Drosophila." <u>Mech Dev</u> **121**(6): 507-518.

Nakata, K., T. Nagai, et al. (1998). "Xenopus Zic family and its role in neural and neural crest development." <u>Mech Dev</u> **75**(1-2): 43-51.

Nakaya, Y. and G. Sheng (2008). "Epithelial to mesenchymal transition during gastrulation: an embryological view." <u>Dev Growth Differ</u> **50**(9): 755-766.

Nativio, R., K. S. Wendt, et al. (2009). "Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus." <u>PLoS Genet</u> **5**(11): e1000739.

Natoli, G. and J. C. Andrau "Noncoding transcription at enhancers: general principles and functional models." <u>Annu Rev Genet</u> **46**: 1-19.

Navratilova, P., D. Fredman, et al. (2009). "Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes." <u>Dev Biol</u> **327**(2): 526-540.

Naylor, L. H. (1999). "Reporter gene technology: the future looks bright." <u>Biochem Pharmacol</u> **58**(5): 749-757.

Nelson, A. C. and F. C. Wardle "Conserved non-coding elements and cis regulation: actions speak louder than words." <u>Development</u> **140**(7): 1385-1395.

Neumann, C. J., H. Grandel, et al. (1999). "Transient establishment of anteroposterior polarity in the zebrafish pectoral fin bud in the absence of sonic hedgehog activity." <u>Development</u> **126**(21): 4817-4826.

Neumann, C. J. and C. Nuesslein-Volhard (2000). "Patterning of the zebrafish retina by a wave of sonic hedgehog activity." <u>Science</u> **289**(5487): 2137-2139.

Niimi, T., Y. Hayashi, et al. (2004). "SOX7 and SOX17 regulate the parietal endoderm-specific enhancer activity of mouse laminin alpha1 gene." <u>J Biol Chem</u> **279**(36): 38055-38061.

Niimi, T., Y. Hayashi, et al. (2003). "Identification of an upstream enhancer in the mouse laminin alpha 1 gene defining its high level of expression in parietal endoderm cells." <u>J Biol Chem</u> **278**(11): 9332-9338.

Niimi, T., Y. Hayashi, et al. (2006). "The Sp family of transcription factors regulates the human laminin alpha1 gene in JAR choriocarcinoma cells." <u>Biochimica et biophysica acta</u> **1759**(11-12): 573-579.

Nikolaev, L. G., S. B. Akopov, et al. (2009). "Vertebrate Protein CTCF and its Multiple Roles in a Large-Scale Regulation of Genome Activity." <u>Curr Genomics</u> **10**(5): 294-302.

Nikolova, G., N. Jabs, et al. (2006). "The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation." <u>Developmental cell</u> **10**(3): 397-405.

Nowak, S. J. and V. G. Corces (2004). "Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation." Trends Genet **20**(4): 214-220.

Nüsslein-Volhard C., and Dahm R., (2002) "Zebrafish: a practical approach". New York: Oxford University Press

O'Neill, B. C., H. Suzuki, et al. (1997). "Cloning of rat laminin gamma 1-chain gene promoter reveals motifs for recognition of multiple transcription factors." Am J Physiol **273**(3 Pt 2): F411-420.

Ochi, H. and Westerfield, M. (2007). Signaling networks that regulate muscle development: lessons from zebrafish. Dev Growth Differ 49, 1-11.

Ogbourne S, Antalis TM. (1998). "Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes." Biochem J. Apr 1;331 (Pt 1):1-14.

Oksenberg, N., L. Stevison, et al. "Function and regulation of AUTS2, a gene implicated in autism and human evolution." PLoS Genet **9**(1): e1003221.

Ordahl, C. P., E. Berdougo, et al. (2001). "The dermomyotome dorsomedial lip drives growth and morphogenesis of both the primary myotome and dermomyotome epithelium." Development **128**(10): 1731-1744.

Orkin, S. H. (1995). "Transcription factors and hematopoietic development." J Biol Chem **270**(10): 4955-4958.

Ovcharenko, I., M. A. Nobrega, et al. (2004). "ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes." Nucleic Acids Res **32**(Web Server issue): W280-286.

Pallerla, S. R., Y. Pan, et al. (2007). "Heparan sulfate Ndst1 gene function variably regulates multiple signaling pathways during mouse development." Developmental dynamics : an official publication of the American Association of Anatomists **236**(2): 556-563.

Pan, D. and G. M. Rubin (1995). "cAMP-dependent protein kinase and hedgehog act antagonistically in regulating decapentaplegic transcription in Drosophila imaginal discs." Cell **80**(4): 543-552.

Pan, Y. and B. Wang (2007). "A novel protein-processing domain in Gli2 and Gli3 differentially blocks complete protein degradation by the proteasome." J Biol Chem **282**(15): 10846-10852.

Pan, Y., C. Wang, et al. (2009). "Phosphorylation of Gli2 by protein kinase A is required for Gli2 processing and degradation and the Sonic Hedgehog-regulated mouse development." Dev Biol **326**(1): 177-189.

Panne, D., T. Maniatis, et al. (2007). "An atomic model of the interferon-beta enhanceosome." Cell **129**(6): 1111-1123.

Park, H. C., A. Mehta, et al. (2002). "olig2 is required for zebrafish primary motor neuron and oligodendrocyte development." Dev Biol **248**(2): 356-368.

Park, H. L., C. Bai, et al. (2000). "Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation." Development **127**(8): 1593-1605.

Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." Nat Rev Genet **10**(10): 669-680.

Parsons, M. J., S. M. Pollard, et al. (2002). "Zebrafish mutants identify an essential role for laminins in notochord formation." Development **129**(13): 3137-3146.

Patton, B. L., J. H. Miner, et al. (1997). "Distribution and function of laminins in the neuromuscular system of developing, adult, and mutant mice." J Cell Biol **139**(6): 1507-1521.

Paulsson, M., R. Deutzmann, et al. (1985). "Evidence for coiled-coil alpha-helical regions in the long arm of laminin." The EMBO journal **4**(2): 309-316.

Paulsson, M., K. Saladin, et al. (1988). "Binding of Ca2+ influences susceptibility of laminin to proteolytic digestion and interactions between domain-specific laminin fragments." European journal of biochemistry / FEBS **177**(3): 477-481.

Paulus, J. D. and M. C. Halloran (2006). "Zebrafish bashful/laminin-alpha 1 mutants exhibit multiple axon guidance defects." Dev Dyn **235**(1): 213-224.

Pavletich, N. P. and C. O. Pabo (1993). "Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers." Science **261**(5129): 1701-1707.

Pennacchio, L. A., N. Ahituv, et al. (2006). "In vivo enhancer analysis of human conserved non-coding sequences." Nature **444**(7118): 499-502.

Perris, R. and D. Perissinotto (2000). "Role of the extracellular matrix during neural crest cell migration." Mechanisms of development **95**(1-2): 3-21.

Perry, M. W., A. N. Boettiger, et al. "Shadow enhancers foster robustness of Drosophila gastrulation." Curr Biol **20**(17): 1562-1567.

Persson, M., D. Stamataki, et al. (2002). "Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity." Genes Dev **16**(22): 2865-2878.

Pfeffer, P. L., B. Payer, et al. (2002). "The activation and maintenance of Pax2 expression at the mid-hindbrain boundary is controlled by separate enhancers." Development **129**(2): 307-318.

Phillips, J. E. and V. G. Corces (2009). "CTCF: master weaver of the genome." Cell **137**(7): 1194-1211.

Piccinni, S. A., A. L. Bolcato-Bellemin, et al. (2004). "Kruppel-like factors regulate the Lama1 gene encoding the laminin alpha1 chain." J Biol Chem **279**(10): 9103-9114.

Plantman, S., M. Patarroyo, et al. (2008). "Integrin-laminin interactions controlling neurite outgrowth from adult DRG neurons in vitro." Mol Cell Neurosci **39**(1): 50-62.

Poschl, E., U. Schlotzer-Schrehardt, et al. (2004). "Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development." Development **131**(7): 1619-1628.

Potvin, E., L. Beuret, et al. "Cooperative action of multiple cis-acting elements is required for N-myc expression in branchial arches: specific contribution of GATA3." Mol Cell Biol **30**(22): 5348-5363.

Poulin, F., M. A. Nobrega, et al. (2005). "In vivo characterization of a vertebrate ultraconserved enhancer." Genomics **85**(6): 774-781.

Pownall, M. E., M. K. Gustafsson, et al. (2002). "Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos." Annu Rev Cell Dev Biol **18**: 747-783.

Price, M. A. and D. Kalderon (2002). "Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by Glycogen Synthase Kinase 3 and Casein Kinase 1." Cell **108**(6): 823-835.

Promega. (2008) "pGL3 Luciferase Reporter Vectors: Instructions For Use Of Products E1741, E1751, E1761 and E1771"

Rada-Iglesias, A., R. Bajpai, et al. "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Radakovits, R., C. S. Barros, et al. (2009). "Regulation of radial glial survival by signals from the meninges." The Journal of neuroscience : the official journal of the Society for Neuroscience **29**(24): 7694-7705.

Rahimov, F., M. L. Marazita, et al. (2008). "Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip." Nat Genet **40**(11): 1341-1347.

Ramalho-Santos, M., D. A. Melton, et al. (2000). "Hedgehog signals regulate multiple aspects of gastrointestinal development." Development **127**(12): 2763-2772.

Rasmussen, J. P., S. S. Reddy, et al. "Laminin is required to orient epithelial polarity in the C. elegans pharynx." Development **139**(11): 2050-2060.

Ravanpay, A. C., S. J. Hansen, et al. "Transcriptional inhibition of REST by NeuroD2 during neuronal differentiation." Mol Cell Neurosci **44**(2): 178-189.

Rea, S., F. Eisenhaber, et al. (2000). "Regulation of chromatin structure by site-specific histone H3 methyltransferases." Nature **406**(6796): 593-599.

Relaix, F., D. Rocancourt, et al. (2005). "A Pax3/Pax7-dependent population of skeletal muscle progenitor cells." Nature **435**(7044): 948-953.

Ribes, V. and J. Briscoe (2009). "Establishing and interpreting graded Sonic Hedgehog signaling during vertebrate neural tube patterning: the role of negative feedback." Cold Spring Harb Perspect Biol **1**(2): a002014.

Rifes, P. and S. Thorsteinsdottir "Extracellular matrix assembly and 3D organization during paraxial mesoderm development in the chick embryo." Dev Biol **368**(2): 370-381.

Rinn, J. L., M. Kertesz, et al. (2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." Cell **129**(7): 1311-1323.

Ritter, D. I., Z. Dong, et al. (2012). "Transcriptional enhancers in protein-coding exons of vertebrate developmental genes." PLoS One **7**(5): e35202.

Roessler, E., E. Belloni, et al. (1996). "Mutations in the human Sonic Hedgehog gene cause holoprosencephaly." Nat Genet **14**(3): 357-360.

Rohatgi, R., L. Milenkovic, et al. (2007). "Patched1 regulates hedgehog signaling at the primary cilium." Science **317**(5836): 372-376.

Rossi, F. M., A. M. Kringstein, et al. (2000). "Transcriptional control: rheostat converted to on/off switch." Mol Cell **6**(3): 723-728.

Rowitch, D. H., S. J. B, et al. (1999). "Sonic hedgehog regulates proliferation and inhibits differentiation of CNS precursor cells." J Neurosci **19**(20): 8954-8965.

Royo, J. L., C. Hidalgo, et al. (2011). "Dissecting the transcriptional regulatory properties of human chromosome 16 highly conserved non-coding regions." PLoS One **6**(9): e24824.

Rubin, J. B., Y. Choi, et al. (2002). "Cerebellar proteoglycans regulate sonic hedgehog responses during development." Development **129**(9): 2223-2232.

Ruel, L., A. Gallet, et al. (2007). "Phosphorylation of the atypical kinesin Costal2 by the kinase Fused induces the partial disassembly of the Smoothened-Fused-Costal2-Cubitus interruptus complex in Hedgehog signalling." Development **134**(20): 3677-3689.

Ruiz i Altaba, A. (1999). "Gli proteins encode context-dependent positive and negative functions: implications for development and disease." Development **126**(14): 3205-3216.

Sacilotto, N., R. Monteiro, et al. (2013). "Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development." Proc Natl Acad Sci U S A **110**(29): 11893-11898.

Sakai, D., T. Suzuki, et al. (2006). "Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development." Development **133**(7): 1323-1333.

Sander, M., S. Paydar, et al. (2000). "Ventral neural patterning by Nkx homeobox genes: Nkx6.1 controls somatic motor neuron and ventral interneuron fates." Genes Dev **14**(17): 2134-2139.

Sanges, R., E. Kalmar, et al. (2006). "Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage." Genome Biol **7**(7): R56.

Santagati, F., K. Abe, et al. (2003). "Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny." Genetics **165**(1): 235-242.

Sartorelli, P., C. Aprea, et al. (1997). "In vitro percutaneous penetration of methyl-parathion from a commercial formulation through the human skin." Occup Environ Med **54**(7): 524-525.

Sasaki, H., C. Hui, et al. (1997). "A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro." Development **124**(7): 1313-1322.

Sasaki, H., Y. Nishizaki, et al. (1999). "Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling." Development **126**(17): 3915-3924.

239

Sasaki, T., J. Takagi, et al. (2010). "Laminin-121--recombinant expression and interactions with integrins." Matrix biology : journal of the International Society for Matrix Biology **29**(6): 484-493.

Sato, S., K. Ikeda, et al. (2012). "Regulation of Six1 expression by evolutionarily conserved enhancers in tetrapods." Dev Biol **368**(1): 95-108.

Sato, Y., T. Kasai, et al. (2007). "Stable integration and conditional expression of electroporated transgenes in chicken embryos." Dev Biol **305**(2): 616-624.

Schagat T., Paguio A., and Kopish K. (2007) "Normalizing Genetic Reporter Assays Approaches and Considerations for Increasing Consistency and Statistical Significance". Promega Corporation

Schauerte, H. E., F. J. van Eeden, et al. (1998). "Sonic hedgehog is not required for the induction of medial floor plate cells in the zebrafish." Development **125**(15): 2983-2993.

Scheele, S., A. Nystrom, et al. (2007). "Laminin isoforms in development and disease." Journal of molecular medicine **85**(8): 825-836.

Schittny, J. C. and P. D. Yurchenco (1990). "Terminal short arm domains of basement membrane laminin are critical for its self-assembly." J Cell Biol **110**(3): 825-832.

Seifert, A. W., Z. Zheng, et al. "Sonic hedgehog controls growth of external genitalia by regulating cell cycle kinetics." Nat Commun **1**: 23.

Semina, E. V., D. V. Bosenko, et al. (2006). "Mutations in laminin alpha 1 result in complex, lens-independent ocular phenotypes in zebrafish." Dev Biol **299**(1): 63-77.

Senger, K., G. W. Armstrong, et al. (2004). "Immunity regulatory DNAs share common organizational features in Drosophila." Mol Cell **13**(1): 19-32.

Seo, K. W., Y. Wang, et al. (2006). "Targeted disruption of the DM domain containing transcription factor Dmrt2 reveals an essential role in somite patterning." Developmental biology **290**(1): 200-210.

Shaner, N. C., R. E. Campbell, et al. (2004). "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein." Nat Biotechnol **22**(12): 1567-1572.

Shea, C. M., C. M. Edgar, et al. (2003). "BMP treatment of C3H10T1/2 mesenchymal stem cells induces both chondrogenesis and osteogenesis." J Cell Biochem **90**(6): 1112-1127.

Shei GJ, Broach JR. (1995). "Yeast silencers can act as orientation-dependent gene inactivation centers that respond to environmental signals." Mol Cell Biol. Jul;15(7):3496-506.

Shi, Y., F. Lan, et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." Cell **119**(7): 941-953.

Shin, J., H. C. Park, et al. (2003). "Neural cell fate analysis in zebrafish using olig2 BAC transgenics." Methods Cell Sci **25**(1-2): 7-14.

Shin, J. T., J. R. Priest, et al. (2005). "Human-zebrafish non-coding conserved elements act in vivo to regulate transcription." Nucleic Acids Res **33**(17): 5437-5445.

Shu, W., H. Chen, et al. "Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains." Nucleic Acids Res **39**(17): 7428-7443.

Sisson, J. C., K. S. Ho, et al. (1997). "Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway." Cell **90**(2): 235-245.

Smelkinson, M. G., Q. Zhou, et al. (2007). "Regulation of Ci-SCFSlimb binding, Ci proteolysis, and hedgehog pathway activity by Ci phosphorylation." Dev Cell **13**(4): 481-495.

Smyth, N., H. S. Vatansever, et al. (1999). "Absence of basement membranes after targeting the LAMC1 gene results in embryonic lethality due to failure of endoderm differentiation." The Journal of cell biology **144**(1): 151-160.

Sommer, L., Q. Ma, et al. (1996). "neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS." Mol Cell Neurosci **8**(4): 221-241.

Song, B. H., S. C. Choi, et al. (2003). "Local activation of protein kinase A inhibits morphogenetic movements during Xenopus gastrulation." Dev Dyn **227**(1): 91-103.

Sorokin, L. and P. Ekblom (1992). "Development of tubular and glomerular cells of the kidney." Kidney Int **41**(3): 657-664.

Sorokin, L. M., S. Conzelmann, et al. (1992). "Monoclonal antibodies against laminin A chain fragment E3 and their effects on binding to cells and proteoglycan and on kidney development." Experimental cell research **201**(1): 137-144.

Sorokin, L. M., F. Pausch, et al. (1997). "Differential expression of five laminin alpha (1-5) chains in developing and adult mouse kidney." Dev Dyn **210**(4): 446-462.

Spicuglia, S. and L. Vanhille "Chromatin signatures of active enhancers." Nucleus **3**(2): 126-131.

Spitz, F. and E. E. Furlong "Transcription factors: from enhancer binding to developmental control." Nat Rev Genet **13**(9): 613-626.

Stamataki, D., F. Ulloa, et al. (2005). "A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube." Genes Dev **19**(5): 626-641.

Stickney, H. L., Barresi, M. J. and Devoto, S. H. (2000). Somite development in zebrafish. Dev Dyn 219, 287-303.

Stock, J. K., S. Giadrossi, et al. (2007). "Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells." Nat Cell Biol **9**(12): 1428-1435.

Stolt, C. C. and M. Wegner "SoxE function in vertebrate nervous system development." Int J Biochem Cell Biol **42**(3): 437-440.

Stuart, G. W., J. V. McMurray, et al. (1988). "Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos." Development **103**(2): 403-412.

Su, Y., W. R. Dostmann, et al. (1995). "Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding domains." Science **269**(5225): 807-813.

Sue S. J., Mortimer J. R., et al. (2005). "oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes." Nucelic Acids Res. Jun 2; **33**(10):3154-64

Sultana, H., S. Verma, et al. "A BEAF dependent chromatin domain boundary separates myoglianin and eyeless genes of Drosophila melanogaster." Nucleic Acids Res **39**(9): 3543-3557.

Sun, Y. M., D. J. Greenway, et al. (2005). "Distinct profiles of REST interactions with its target genes at different stages of neuronal development." Mol Biol Cell **16**(12): 5630-5638.

Swanson CI, Evans NC, Barolo S. (2010). "Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer." Dev Cell. 2010 Mar 16;18(3):359-70.

Sztal, T. E., C. Sonntag, et al. (2012). "Epistatic dissection of laminin-receptor interactions in dystrophic zebrafish muscle." Hum Mol Genet **21**(21): 4718-4731.

Taher, L., D. M. McGaughey, et al. (2011). "Genome-wide identification of conserved regulatory function in diverged sequences." Genome Res **21**(7): 1139-1149.

Taipale, J., M. K. Cooper, et al. (2002). "Patched acts catalytically to suppress the activity of Smoothened." Nature **418**(6900): 892-897.

Tajbakhsh, S., D. Rocancourt, et al. (1996). "Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice." Nature **384**(6606): 266-270.

Talbert PB, Henikoff S. (2006) "Spreading of silent chromatin: inaction at a distance." Nat Rev Genet. Oct;7(10):793-803.

Talts, J. F., Z. Andac, et al. (1999). "Binding of the G domains of laminin alpha1 and alpha2 chains and perlecan to heparin, sulfatides, alpha-dystroglycan and several extracellular matrix proteins." The EMBO journal **18**(4): 863-870.

Tamplin, O. J., B. J. Cox, et al. (2011). "Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord." Dev Biol **360**(2): 415-425.

Tanaka, S., Y. Kamachi, et al. (2004). "Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells." Mol Cell Biol **24**(20): 8834-8846.

Tang, Q. Q., T. C. Otto, et al. (2004). "Commitment of C3H10T1/2 pluripotent stem cells to the adipocyte lineage." Proc Natl Acad Sci U S A **101**(26): 9607-9611.

Tapscott, S. J. (2005). "The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription." Development **132**(12): 2685-2695.

Tapscott, S. J., A. B. Lassar, et al. (1992). "A novel myoblast enhancer element mediates MyoD transcription." Mol Cell Biol **12**(11): 4994-5003.

Taylor, S. S., J. Yang, et al. (2004). "PKA: a portrait of protein kinase dynamics." Biochim Biophys Acta **1697**(1-2): 259-269.

Tempe, D., M. Casas, et al. (2006). "Multisite protein kinase A and glycogen synthase kinase 3beta phosphorylation leads to Gli3 ubiquitination by SCFbetaTrCP." Mol Cell Biol **26**(11): 4316-4326.

Tenzen, T., B. L. Allen, et al. (2006). "The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice." Dev Cell **10**(5): 647-656.

Thanos, D. and T. Maniatis (1995). "Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome." Cell **83**(7): 1091-1100.

Thisse, B., V. Heyer, et al. (2004). "Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening." Methods Cell Biol **77**: 505-519.

Thurman, R. E., E. Rynes, et al. "The accessible chromatin landscape of the human genome." Nature **489**(7414): 75-82.

Tidyman, W. E., A. J. Sehnert, et al. (2003). "In vivo regulation of the chicken cardiac troponin T gene promoter in zebrafish embryos." Dev Dyn **227**(4): 484-496.

Timpl, R., H. Rohde, et al. (1979). "Laminin--a glycoprotein from basement membranes." The Journal of biological chemistry **254**(19): 9933-9937.

Tokida, Y., Y. Aratani, et al. (1990). "Production of two variant laminin forms by endothelial cells and shift of their relative levels by angiostatic steroids." The Journal of biological chemistry **265**(30): 18123-18129.

Tosney, K. W., D. B. Dehnbostel, et al. (1994). "Neural crest cells prefer the myotome's basal lamina over the sclerotome as a substratum." Dev Biol **163**(2): 389-406.

True, J. R. and E. S. Haag (2001). "Developmental system drift and flexibility in evolutionary trajectories." Evol Dev **3**(2): 109-119.

Tukachinsky, H., L. V. Lopez, et al. (2010). "A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu-Gli protein complexes." J Cell Biol **191**(2): 415-428.

Tunggal, P., N. Smyth, et al. (2000). "Laminins: structure and genetic regulation." Microsc Res Tech **51**(3): 214-227.

Tyurina, O. V., B. Guner, et al. (2005). "Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling." Dev Biol **277**(2): 537-556.

Tzu, J. and M. P. Marinkovich (2008). "Bridging structure with function: structural, regulatory, and developmental role of laminins." The international journal of biochemistry & cell biology **40**(2): 199-214.

Uchikawa, M., Y. Ishida, et al. (2003). "Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals." Dev Cell **4**(4): 509-519.

Uchikawa, M., Y. Kamachi, et al. (1999). "Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken." Mech Dev **84**(1-2): 103-120.

Ungar, A. R. and R. T. Moon (1996). "Inhibition of protein kinase A phenocopies ectopic expression of hedgehog in the CNS of wild-type and cyclops mutant embryos." Dev Biol **178**(1): 186-191.

Urasaki, A., G. Morvan, et al. (2006). "Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition." Genetics **174**(2): 639-649.

Ureta-Vidal A, Ettwiller L, Birney E. (2003). "Comparative genomics: genome-wide analysis in metazoan eukaryotes." Nat Rev Genet. Apr;4(4):251-62.

Valentine, S. A., G. Chen, et al. (1998). "Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer." Mol Cell Biol **18**(11): 6584-6594.

van den Heuvel, M. and P. W. Ingham (1996). "smoothened encodes a receptor-like serpentine protein required for hedgehog signalling." Nature **382**(6591): 547-551.

van Eeden, F. J., M. Granato, et al. (1996). "Mutations affecting somite formation and patterning in the zebrafish, Danio rerio." Development **123**: 153-164.

Vanderlaan, G., O. V. Tyurina, et al. (2005). "Gli function is essential for motor neuron induction in zebrafish." Dev Biol **282**(2): 550-570.

van Tuinen M. and Hadly, EA. (2004) "Error in estimation of rate and inferred from the early amniote fossil record and avian molecular clocks". J Mol Evol. 59:267–276

Varga, Z. M., A. Amores, et al. (2001). "Zebrafish smoothened functions in ventral neural tube specification and axon tract formation." Development **128**(18): 3497-3509.

Vasios, G., S. Mader, et al. (1991). "The late retinoic acid induction of laminin B1 gene transcription involves RAR binding to the responsive element." Embo J **10**(5): 1149-1158.

Vavouri, T. and B. Lehner (2009). "Conserved noncoding elements and the evolution of animal body plans." Bioessays **31**(7): 727-735.

Vavouri, T., J. I. Semple, et al. (2009). "Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity." Cell **138**(1): 198-208.

Virtanen, I., D. Gullberg, et al. (2000). "Laminin alpha1-chain shows a restricted distribution in epithelial basement membranes of fetal and adult human tissues." Experimental cell research **257**(2): 298-309.

Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." Nature **457**(7231): 854-858.

Visel, A., J. Bristow, et al. (2007). "Enhancer identification through comparative genomics." Semin Cell Dev Biol **18**(1): 140-152.

Visel, A., S. Minovitsky, et al. (2007). "VISTA Enhancer Browser--a database of tissue-specific human enhancers." Nucleic Acids Res **35**(Database issue): D88-92.

Voet D. (2004) "Biochemistry", 3rd ed. Wiley

Vokes, S. A., H. Ji, et al. (2007). "Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning." Development **134**(10): 1977-1989.

Vokes, S. A., H. Ji, et al. (2008). "A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb." Genes Dev **22**(19): 2651-2663.

Wagner, E. J. and P. B. Carpenter "Understanding the language of Lys36 methylation at histone H3." Nat Rev Mol Cell Biol **13**(2): 115-126.

Wallquist, W., J. Zelano, et al. (2004). "Dorsal root ganglion neurons up-regulate the expression of laminin-associated integrins after peripheral but not central axotomy." The Journal of comparative neurology **480**(2): 162-169.

Wang, B. and Y. Li (2006). "Evidence for the direct involvement of {beta}TrCP in Gli3 protein processing." Proc Natl Acad Sci U S A **103**(1): 33-38.

Wang, C., Y. Pan, et al. (2007). "A hypermorphic mouse Gli3 allele results in a polydactylous limb phenotype." Developmental dynamics : an official publication of the American Association of Anatomists **236**(3): 769-776.

Wang, G., K. Amanai, et al. (2000). "Interactions with Costal2 and suppressor of fused regulate nuclear translocation and activity of cubitus interruptus." Genes Dev **14**(22): 2893-2905.

Wang, H., L. Wang, et al. (2004). "Role of histone H2A ubiquitination in Polycomb silencing." Nature **431**(7010): 873-878.

Wang, Q. T. and R. A. Holmgren (2000). "Nuclear import of cubitus interruptus is regulated by hedgehog via a mechanism distinct from Ci stabilization and Ci activation." Development **127**(14): 3131-3139.

Wang X., Zhao Z., Muller J., Iyu A., Khng A. J., Guccione E., Ruan Y., and Ingham, P. W. (2013) "Targeted inactivation and identification of targets of the Gli2a transcription factor in the zebrafish". Biology Open. September 13

Wang, Y., Z. Zhou, et al. (2009). "Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation." Proc Natl Acad Sci U S A **106**(8): 2623-2628.

Watanabe, Y., S. Zaffran, et al. "Fibroblast growth factor 10 gene regulation in the second heart field by Tbx1, Nkx2-5, and Islet1 reveals a genetic switch for down-regulation in the myocardium." Proc Natl Acad Sci U S A **109**(45): 18273-18280.

Weintraub, H., R. Davis, et al. (1990). "MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation." Proc Natl Acad Sci U S A **87**(15): 5623-5627.

Weintraub, H., S. J. Tapscott, et al. (1989). "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD." Proc Natl Acad Sci U S A **86**(14): 5434-5438.

Weirauch, M. T. and T. R. Hughes "Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same." Trends Genet **26**(2): 66-74.

Wendt, K. S., K. Yoshida, et al. (2008). "Cohesin mediates transcriptional insulation by CCCTC-binding factor." Nature **451**(7180): 796-801.

Weth, O., C. Weth, et al. "Modular insulators: genome wide search for composite CTCF/thyroid hormone receptor binding sites." PLoS One **5**(4): e10119.

Willem, M., N. Miosge, et al. (2002). "Specific ablation of the nidogen-binding site in the laminin gamma1 chain interferes with kidney and lung development." Development **129**(11): 2711-2722.

Wilson, C., H. J. Bellen, et al. (1990). "Position effects on eukaryotic gene expression." Annu Rev Cell Biol **6**: 679-714.

Wilson, C. W., M. H. Chen, et al. (2009). "Smoothened adopts multiple active and inactive conformations capable of trafficking to the primary cilium." PLoS One **4**(4): e5182.

Wittkopp PJ, Kalay G. (2011) "Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence." Nat Rev Genet. Dec 6;13(1):59-69.

Wolff, C., S. Roy, et al. (2003). "Multiple muscle cell identities induced by distinct levels and timing of hedgehog activity in the zebrafish embryo." Curr Biol **13**(14): 1169-1181.

Woo, C. J., P. V. Kharchenko, et al. "A region of the human HOXD cluster that confers polycomb-group responsiveness." Cell **140**(1): 99-110.

Wood, A. M., K. Van Bortle, et al. "Regulation of chromatin organization and inducible gene expression by a Drosophila insulator." Mol Cell **44**(1): 29-38.

Woolfe, A., M. Goodson, et al. (2005). "Highly conserved non-coding sequences are associated with vertebrate development." PLoS Biol **3**(1): e7.

Xiao, T., J. Wallace, et al. "Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity." Mol Cell Biol **31**(11): 2174-2183.

Yadon, A. N., B. N. Singh, et al. "DNA looping facilitates targeting of a chromatin remodeling enzyme." Mol Cell **50**(1): 93-103.

Yavari, A., R. Nagaraj, et al. "Role of lipid metabolism in smoothened derepression in hedgehog signaling." Dev Cell **19**(1): 54-65.

Yavari, A., R. Nagaraj, et al. (2010). "Role of lipid metabolism in smoothened derepression in hedgehog signaling." Dev Cell **19**(1): 54-65.

Yoshida, C., F. Tokumasu, et al. (1999). "Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1." Genes Cells **4**(11): 643-655.

Young, M. D., T. A. Willson, et al. "ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity." Nucleic Acids Res **39**(17): 7415-7427.

Young, T., Y. Poobalan, et al. "Mutated in colorectal cancer (Mcc), a candidate tumor suppressor, is dynamically expressed during mouse embryogenesis." Dev Dyn **240**(9): 2166-2174.

Yu, M., Y. Xi, et al. "Activity of dlx5a/dlx6a regulatory elements during zebrafish GABAergic neuron development." Int J Dev Neurosci **29**(7): 681-691.

Yurchenco, P. D. (2011). "Basement membranes: cell scaffoldings and signaling platforms." Cold Spring Harbor perspectives in biology **3**(2).

Yurchenco, P. D., P. S. Amenta, et al. (2004). "Basement membrane assembly, stability and activities observed through a developmental lens." Matrix biology : journal of the International Society for Matrix Biology **22**(7): 521-538.

Yurchenco, P. D., Y. Quan, et al. (1997). "The alpha chain of laminin-1 is independently secreted and drives secretion of its beta- and gamma-chain partners." Proceedings of the National Academy of Sciences of the United States of America **94**(19): 10189-10194.

Yusuf, F. and B. Brand-Saberi (2006). "The eventful somite: patterning, fate determination and cell division in the somite." Anat Embryol (Berl) **211 Suppl 1**: 21-30.

Zagris, N. and A. E. Chung (1990). "Distribution and functional role of laminin during induction of the embryonic axis in the chick embryo." Differentiation **43**(2): 81-86.

Zagris, N., A. E. Chung, et al. (2000). "Differential expression of laminin genes in early chick embryo." Int J Dev Biol **44**(7): 815-818.

Zarnegar, M. A., J. Chen, et al. "Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements." Mol Cell Biol **30**(20): 4922-4939.

Zelenchuk, T. A. and J. L. Bruses (2011). "In vivo labeling of zebrafish motor neurons using an mnx1 enhancer and Gal4/UAS." Genesis **49**(7): 546-554.

Zent, R., K. T. Bush, et al. (2001). "Involvement of laminin binding integrins and laminin-5 in branching morphogenesis of the ureteric bud during kidney development." <u>Dev Biol</u> **238**(2): 289-302.

Zhang Z, Gerstein M. (2003). "Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements." J Biol. (2):11. Epub 2003 Jun 6.

Zhang, C., Z. Xuan, et al. (2006). "A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome." <u>Nucleic Acids Res</u> **34**(8): 2238-2246.

Zhang, X. M., M. Ramalho-Santos, et al. (2001). "Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R asymmetry by the mouse node." <u>Cell</u> **105**(6): 781-792.

Zhang, X. M., M. Ramalho-Santos, et al. (2001). "Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node." <u>Cell</u> **106**(2): 781-792.

Zhao, Y. Y., F. J. Zhang, et al. (2011). "The association of a single nucleotide polymorphism in the promoter region of the LAMA1 gene with susceptibility to Chinese high myopia." <u>Molecular vision</u> **17**: 1003-1010.

Zhao, Z., G. Tavoosidana, et al. (2006). "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions." <u>Nat Genet</u> **38**(11): 1341-1347.

Zheng, X., R. K. Mann, et al. "Genetic and biochemical definition of the Hedgehog receptor." <u>Genes Dev</u> **24**(1): 57-71.

Zhu, A. J., L. Zheng, et al. (2003). "Altered localization of Drosophila Smoothened protein activates Hedgehog signal transduction." <u>Genes Dev</u> **17**(10): 1240-1252.

Zhu, M., W. Zhao, et al. (2009). "The oldest articulated osteichthyan reveals mosaic gnathostome characters." <u>Nature</u> **458**(7237): 469-474.

Zinkevich, N. S., D. V. Bosenko, et al. (2006). "laminin alpha 1 gene is essential for normal lens development in zebrafish." <u>BMC Dev Biol</u> **6**: 13.