uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Noël Ghanem
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Médecine Cellulaire et Moléculaire)
GRADE / DEGREE

Département de Médecine Cellulaire et Moléculaire
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Transcriptional Regulation of Dix Expression in the Subpallial Telencephalon of Vertebrates

TITRE DE LA THÈSE / TITLE OF THESIS

Marc Ekker
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Rashmi Kothary

Carol Schuurmans

John Lewis

Leonard Maler (absent)

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Transcriptional Regulation of *Dlx* Expression in the Subpallial Telencephalon of Vertebrates

Noël Ghanem

Thesis is submitted as a partial fulfillment of the Doctor of Philosophy program in Cellular and Molecular Medicine

University of Ottawa

Department of Cellular and Molecular Medicine

Ottawa, Ontario, Canada

May 2006

Your file Votre référence
ISBN: 978-0-494-25899-6
Our file Notre référence
ISBN: 978-0-494-25899-6

Statement of contribution

The thesis, 'Transcriptional Regulation of *Dlx* Expression in the Subpallial Telencephalon of Vertebrates' comprises three related studies, two of which were collaborative efforts.

I have co-authored with Olga Jarinova the first study entitled: "Regulatory Roles of Conserved Intergenic Domains in Vertebrate *Dlx* Bigene Clusters". This study was published in Genome Research, 2003, 13: 533-543. The following people contributed to the work: I have isolated and characterized the *Dlx* genes from genomic clones of *Spheroides Nephelus*. These clones were screened from a PAC library by Angel Amores. I have sequenced the intergenic regions from the *Dlx1/Dlx2* cluster in mouse and zebrafish whereas Olga Jarinova has done the same with *Dlx5/Dlx6* cluster. Both Olga Jarinova and I have performed the phylogenetic comparisons and generated the constructs for transgenic mice. Olga made also the transgenic zebrafish and prepared the transgenic constructs using a vector made by Dr. Qiaoming Long who is a previous member of the laboratory. Gary Hatch assisted in making the transgenic constructs and maintaining the transgenic mice. These mice were made by Adrianna Gambarotta. Dr. Byung Keon Park who is a previous member of Dr. Ekker's laboratory has identified earlier I12a in mice. Dr. John Rubenstein, from the University of California at San Francisco, reviewed this study. This article was collaboratively written by Dr. Ekker and I. Note that the section entitled 'Material and Methods' of this article was modified from its published version and several experimental details were added to it based on the request of the external examiner.

The second study entitled "Distinct *Dlx*-expressing Progenitors in the Medial and Caudal Ganglionic Eminences Give Rise to Different Subtypes of Adult Cortical Interneurons" was a collaborative work with Dr. John Rubenstein's laboratory. The

following people contributed to the study: Dr. Jason Long and I have performed the DiI labeling experiments in Dr. Rubenstein's laboratory. Transgenic mice were created by Adrianna Gambarotta. Gary Hatch assisted with maintaining the transgenic mice. Dr. Rubenstein helped in data analysis, reviewed thoroughly this study and gave useful suggestions. I have performed the rest of the study and wrote this article which was collaboratively corrected by Dr. Rubenstein and Dr. Ekker.

I have wholly performed the third study entitled "A Dynamic Spatial and Temporal Regulation of *Dlx* Expression Is Conferred by the Activities of Four *Dlx*-specific Enhancers in the Mouse Embryonic Telencephalon". Gary Hatch assisted with maintaining the transgenic mice. Dr. Rubenstein participated in data analysis. I have written this article which was collaboratively corrected and reviewed by Dr. Rubenstein and Dr. Ekker.

Abstract

Vertebrate *Dlx* genes are required for the proper development of an array of tissues and organs including the forebrain, the branchial arches, the limbs and the sensory organs as well as bone and cartilage formation. *Dlx* genes code homeodomain proteins that act as transcription factors. Vertebrates have at least six *Dlx* genes that are organized into three convergently transcribed bigene clusters: *Dlx1/Dlx2*, *Dlx5/Dlx6* and *Dlx3/Dlx7*. *Dlx* genes are often expressed with nested patterns along the proximo-distal axis of organs. Furthermore, *Dlx* paralogs show overlapping expression patterns and exhibit partially redundant functions. Yet, each paralog may also exhibit distinct function(s). *Dlx1*, *Dlx2, Dlx5* and *Dlx6* are expressed in a spatio-temporal manner in the basal ganglia of the developing telencephalon. These four genes are required for normal differentiation and migration of late-born progenitors in the subpallial telencephalon that will, subsequently, give rise to distinct subtypes of cortical GABAergic interneurons. I have identified two novel enhancers that regulate *Dlx* expression in the forebrain of vertebrates: URE2 (Upstream Regulatory Element 2) and I12b (*Dlx1/Dlx2* intergenic enhancer 'b') located in the *Dlx1/Dlx2* locus. Despite little sequence similarities, these enhancers shared overlapping activities in telencephalic domains with the previously described enhancers I56i and I56ii (*Dlx5/Dlx6* intergenic enhancers 'i' and 'ii') from *the Dlx5/Dlx6* locus. All four enhancers function where endogenous *Dlx* genes are expressed between E10.5 and birth. Yet, each enhancer targeted expression of a reporter gene to a different group(s) of cells in the basal ganglia and was characterized by a distinct profile of activity in tangentially migrating cells. These cells followed, between E12.5 and E15.5, migratory routes known to be *Dlx*-positive *in vivo* and gave rise to distinct subtypes of adult cortical interneurons. In addition, the differential activity of the *Dlx* enhancers identified several subdivisions in the medial and caudal ganglionic eminences (MGE and CGE) where distinct *Dlx*-progenitors are born. Consequently, the spatio-temporal regionalization of *Dlx*-progenitors and their acquisition of distinct molecular

properties may explain the diversity of cortical interneuron subtypes derived from distinct ventral subdivisions.

Résumé

Les gènes *Dlx* des vertébrés sont requis pour le développement normal de plusieurs tissus et organes y compris le cerveau antérieur, les arches branchiaux, les membres et les organes sensoriels, ainsi que pour la formation des os et du cartilage. Les gènes *Dlx* codent pour des protéines avec homeoboîtes qui jouent le rôle de facteurs de transcription. Les vertébrés possèdent au moins six gènes *Dlx* qui sont organisés en trois complexes bigéniques où les paralogues sont inversement transcrits: *Dlx1/Dlx2*, *Dlx5/Dlx6* et *Dlx3/Dlx7*. Les gènes *Dlx* sont caractérisés, en général, par des profils d'expression emboîtés le long de l'axe proximo-distal des organes où ils sont exprimés. De plus, les gènes *Dlx* paralogues partagent des profiles d'expression qui se chevauchent et, par conséquent, des fonctions en partie redondantes. Cependant, chaque paralogue peut aussi détenir une ou plusieurs fonction(s) unique(s). Les gènes *Dlx1, Dlx2, Dlx5* et *Dlx6* sont exprimés d'une manière spatio-temporelle dans le télencéphale ventral (cortex ventral) en cours de développement. Ces quatre gènes sont indispensables à la différentiation terminale et la migration successive des cellules nerveuses qui sont nées à partir de cellules précurseurs résidant dans le cortex ventral. Ces dernières vont migrer vers le cortex dorsal et se différencier en plusieurs types de neurones inhibiteurs dont le neurotransmetteur principal est le GABA. J'ai identifié deux nouveaux éléments régulateurs des gènes *Dlx* dont les séquences sont conservées chez les vertébrés: URE2 et I12b qui font partie du locus de *Dlx1/Dlx2*. Ces éléments agissent en tant qu'intensificateurs (enhancers) en *cis*. Malgré le faible degré de similitude entre ces séquences et les séquences régulatrices I56i et I56ii du locus de *Dlx5/Dlx6,* ces quatre éléments manifestent des activités chevauchantes dans le cortex ventral, là ou les gènes *Dlx* sont exprimés *in vivo* chez la souris, et ce, entre le 10ème jour embryonnaire et la naissance. Malgré ces chevauchements, chaque élément régulateur peut induire l'expression d'un gène reporteur dans un ou plusieurs groupes uniques de cellules. De plus, ces éléments ont montré des profils d'activité distincts dans les populations de

cellules nerveuses qui migrent vers le cortex cérébral entre le 12ème et 15ème jour embryonnaire. Ces cellules ont suivi des voies de migration spécifiques aux gènes *Dlx in vivo*, et se sont différenciées en plusieurs types de neurones inhibiteurs. Les différentes activités des éléments régulateurs m'ont aussi permis de distinguer plusieurs subdivisions au sein des régions ganglionnaires médiale et caudale du cortex ventral, là où les gènes *Dlx* sont actifs. Par conséquent, les précurseurs exprimant les gènes *Dlx* acquièrent des propriétés moléculaires distinctes dépendant de leur régulation spatio-temporelle spécifique à la subdivision où ils sont nés. Donc, la diversité des neurones inhibiteurs du cortex dorsal peut être attribuée à la variété des précurseurs exprimant les gènes *Dlx* dans le cortex ventral.

Table of Contents

List of figures and tables

List of abbreviations

AEP: anterior entopeduncular area

A/P: anterior/posterior

bp: base pair

BMP: Bone Morphogenetic Protein

C-: carboxy

°C: degrees Celsius

CB: calbindin

CGE: caudal ganglionic eminence

cDNA: complementary deoxyribonucleic acid

CNS: central nervous system

Cp: caudate putamen

CP: cortical plate

CR: calretinin

CRE: *cis*-regulatory element

dLGE: dorsal lateral ganglionic eminence

dMGE: dorsal medial ganglionic eminence

D/V: dorsal/ventral

Dll: *Distal-less* gene

Dlx: *Distal-less* related mouse homologue

dlx: *distal-less* related zebrafish homologue

DLX: *Distal-less* related human homologue

DIG: digoxigenin

DNA: deoxyribonucleic acid

DP: dorsal pallium

E: embryonic day

EPL: external plexiform layer

FGF: fibroblast growth factor

GABA: γ-amino butyric acid

GAD: glutamic acid decarboxylase

GCL: granule cell layer

GFP: green fluorescent protein

GL: glomerular layer

h: hour

H: Hyoid arch

HB: homeobox

HD: homeodomain

Hi: hippocampus

Hox: homeobox transcription factor

Hy: Hypothalamus

I12a: intergenic enhancer 'a' of *Dlx1/Dlx2* cluster

I12b: intergenic enhancer 'b' of *Dlx1/Dlx2* cluster

I56i: intergenic enhancer 'i' of *Dlx5/Dlx6* cluster

I56ii: intergenic enhancer 'ii' of *Dlx5/Dlx6* cluster

IPL: internal plexiform layer

IZ: intermediate zone

Kb: kilobase

L: limbs

LGE: lateral ganglionic eminence

LV: lateral ventricle

LP: lateral pallium

ML: mitral layer

Md: mandibular component of the first branchial arch

MGE: medial ganglionic eminence

min: minutes

μm: micron

Mz: marginal zone

MZ: mantle zone

N-: amino

nNOS: neuronal nitric oxide synthase

NPY: neuropeptide Y

OP: olfactory placodes

O/N: overnight

OV: olfactory ventricle

P: post-natal day

PV: parvalbumin

PB: phosphate buffered saline

PBS: phosphate buffered solution

PBST: phosphate buffered solution and Tween-20

PCR: polymerase chain reaction

PCx: piriform cortex

PFA: paraformaldehyde

P/D: proximal-distal

PD: pallidum

PLAP: placental-like alkaline phosphatase

POA: preoptic area

PT: prethalamus

URE1: upstream regulatory element 1

URE2: upstream regulatory element 2

RNA: ribonucleic acid

RT: room temperature

s: second

SCB: suprachiasmatic band

SHH: Sonic Hedgehog protein

SOM: somatostatin

Sp: septum

St: somites

ST: subpallial telencephalon

SVZ: subventricular zone

TGF-β: transforming growth factor β

Th: thalamus

TH: tyrosine hydroxylase

vLGE: ventral lateral ganglionic eminence

vMGE: ventral medial ganglionic eminence

VIP: vaso-active intestinal peptide

VP: ventral pallium

VZ: ventricular zone

WT: wild type

III: third ventricle

Acknowledgments

First and foremost, Thanks be to GOD at all times and in all places for everything! I owe a great amount of gratitude and thanking to several ople who have helped me, either directly or indirectly, in completing successfully this doctorate project. Firstly, I thank deeply my supervisor, Dr. Marc Ekker, to whom I hold much recognition and admiration for being an excellent tutor, a kind and generous person. Second, I thank also our collaborator, Dr. John Rubenstein, who reviewed thoroughly this work and gave useful comments and suggestions that helped me analyze better this data. Thirdly, I thank my advisory committee members for their critical review of this research: Dr. Marie-Andrée Akimenko, Dr. Ruth Slack, Dr. Rashmi Kothary and Dr. Kursad Turksen. My thanks go to the members of my examination committee: Dr. Carol Schuurmans, Dr. John Lewis, Dr. Leonard Maler and Dr. Rashmi Kothary.

In Ekker's laboratory, I would like to send many thanks to Gary Hatch and Lucille Joly who taught me the basic techniques in molecular and cell biology and gave me practical advices. I thank also Adrianna Gambarotta for generating and maintaining the transgenic mice. My acknowledgements go to Dr. Luc Poitras and Dr. Fabien Avaron for their help and support in scientific matters. I thank all the present and past co-workers in Ekker's and Akimenko's laboratories. I have truly enjoyed time spent in your midst and I keep great memories of these six years especially with Amanda Smith 'Mandi', Mireille Khacho 'Miro', Danielle Guay 'The Technician from Hell' and Olga Jarinova 'Olja'.

Many thanks go to the OHRI departmental assistant Michelle Leuleu-Evans, the CMM assistant Donna Hooper and the CMM director Dr. Jean-Marc Renaud for their help and orientation. Special thanks go to Mr. Abdo Ghié, member on the board of governors at Ottawa U (ex-international student advisor), for his constant help and guidance with academic matters.

I would like to acknowledge the Lebanese University for supporting my studies with a fellowship for five years.

I would like to acknowledge all my relatives, cousins and friends here in Canada and back in Lebanon: my uncle, Alex El Khoury, and his family, 'my favorite cousin' Soprano Joyce El Khoury and my best friends: Adel Ghié 'Abbout', Nisrine Haddad 'Niso', Dr. Patrick Snaifer 'Sanfour Habibi', Rabih Rached 'President' and Dr. Nadine Karam 'Nanou'. Thank you all for your love, loyalty and encouragement!

My sincere and deepest thanks go also to my precious family back at home; my dedicated dad, Mr. Daoud Ghanem 'Gringo', the sweetest mother, Mrs. Nada El Khoury 'El Hanouné', my loving sister Joëlle 'Jojo', and, my funny and talented brother, Sélim 'Silooo'. You have been a great inspiration to me, thank you for your love, encouragement and support! Nada, thank you for your total dedication and precious prayers too! Not to forget my adorable grandparents in Ottawa, Mr. George El Khoury and Mrs. Helene El Khoury, who have supported me tremendously since I landed in this country. I like to mention also my first godson and most adorable boy, Alex Ghazal 'Kouki' and all my little cousins for they remind me of the kid in me.

Lastly, I remember three great people who passed away while I was studying in Canada and who touched me genuinely: my uncle in law "Mr. Sélim El Khoury", the great Lebanese actor "Mr. Elie Snaifer" and my ex-landlord "Mr. Lyall Graham". May their souls rest in peace!

THANK YOU ALL and MAY GOD ALWAYS BLESS YOU!!!

'Nounou El Wahech',

1. Introduction

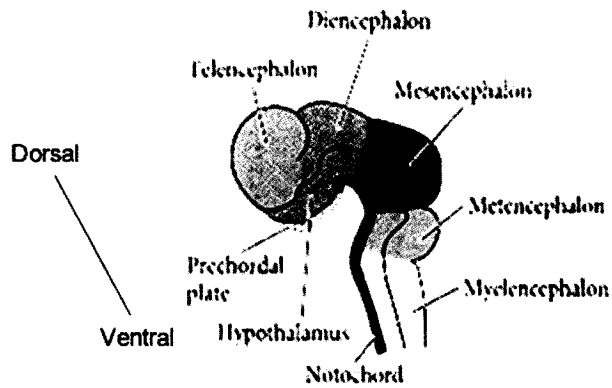
1.1. Overview of central nervous system development

The central nervous system (CNS) is comprised of the brain and the spinal cord, and is the most complex system of the vertebrate body in terms of patterning, regionalization and cellular components. The process by which the CNS is formed early during development, so called neurulation, appears to be similar in amphibians, reptiles, birds, and mammals (Gallera, 1971). It gives rise to the neural tube which is the rudiment of the CNS. The neural tube is formed by the invagination and closure of the neural plate or the portion of the dorsal ectoderm that is specified to become neural ectoderm, and, will ultimately form the brain anteriorly and the spinal cord posteriorly as well as the neural pituitary, the motor neurons and the retina (Figure 1.1. A) (Gilbert, 2000; Rubenstein et al., 1998; Smith and Schoenwolf, 1997). The early mammalian neural tube is comprised of three primary brain vesicles: the forebrain (prosencephalon), the midbrain (mesencephalon) and the hindbrain (rhombencephalon) (Figure 1.1. B). The prosencephalon becomes, in turn, subdivided into the anterior telencephalon and the more caudal diencephalon. The telencephalon will eventually form the cerebral hemispheres (cerebrum), the hippocampus and the olfactory lobes, and, the diencephalon will produce the thalamus and the hypothalamus which receive neural input from the retina. The mesencephalon stays still and gives rise to the cerebral aqueduct connecting the anterior and posterior brain. More posteriorly, the rhombencephalon becomes subdivided into the metencephalon and myelencephalon, which will develop into the cerebellum and medulla oblongata, respectively (Gilbert, 2000; Moore, 1993). Each of the five secondary brain vesicles will thus form several adult derivatives with distinct cellular organizations, boundaries and functions (Figure 1.1. B).

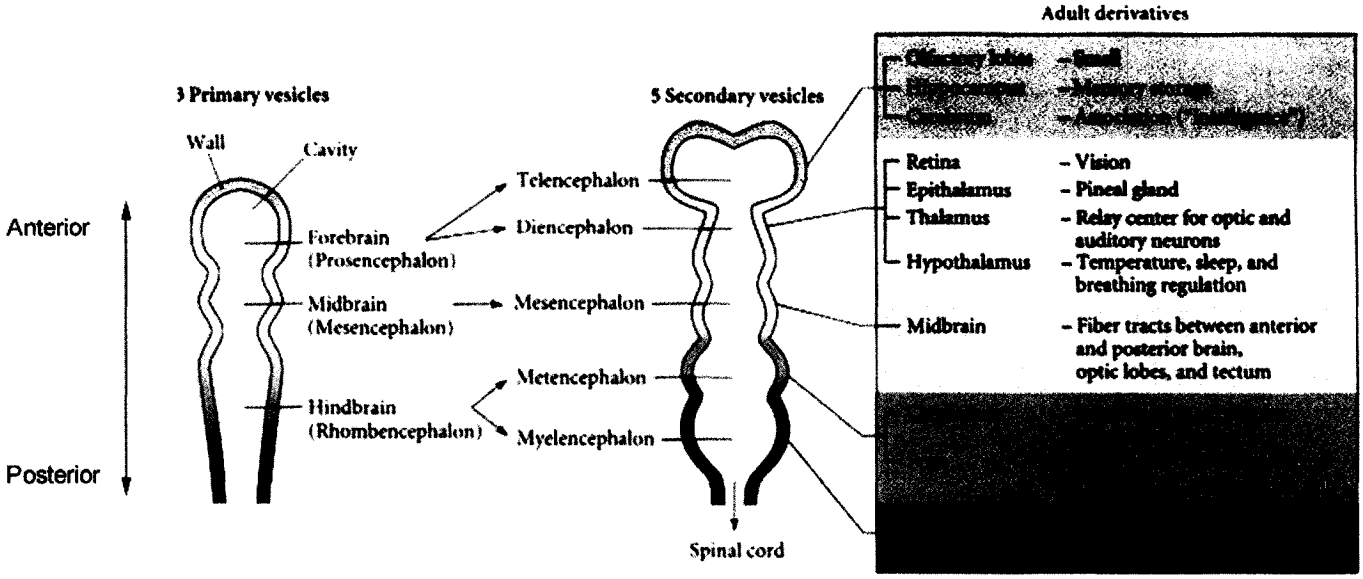
Figure 1.1. Organization and development of the mammalian brain

(A) Drawing of the neural plate and the five secondary vesicles of the brain in the E12 mouse embryo. The prosencephalon (telencephalon in blue and diencephalon in light mauve) derive from the anteriormost portion of the neural plate. The prechordal plate (orange) and notochord (red) form the axial mesendoderm and lie below the neural plate. Signals from the axial mesendoderm (e.g. SHH) and non neural ectoderm (e.g. BMPs) are responsible for the early medio-lateral patterning (dorso-ventral patterning) and formation of longitudinal subdivisions of the neural plate. (B) Early brain development. The early mammalian neural tube is comprised of three primary brain vesicles that are subdivided into five secondary vesicles as development continues. A list of the adult derivatives formed by the walls and cavities of the brain is shown on the right. Modified from Moore and Persaud 1993, and Gilbert 2000.

A


Diencephalon
Telencephalon
Mesencephalon
Dorsal
Metencephalon
Prechordal plate
Myelencephalon
Ventral
Hypothalamus
Notochord

B


Adult derivatives
3 Primary vesicles
5 Secondary vesicles
Wall
Cavity
Anterior
Posterior
Forebrain (Prosencephalon)
Midbrain (Mesencephalon)
Hindbrain (Rhombencephalon)
Telencephalon
Diencephalon
Mesencephalon
Metencephalon
Myelencephalon
Spinal cord
Retina – Vision
Epithalamus – Pineal gland
Thalamus – Relay center for optic and auditory neurons
Hypothalamus – Temperature, sleep, and breathing regulation
Midbrain – Fiber tracts between anterior and posterior brain, optic lobes, and tectum

1.1.1. Early patterning in the neural plate/neural tube

Specifying the brain boundaries and the identity of different structures along the anterior-posterior (A/P) and dorsal-ventral (D/V) axes of the neural tube is controlled, at the molecular level, by several signaling molecules and transcription factors. Such patterning will ultimately generate distinct longitudinal domains (e.g. floor, basal, alar and roof plates), and transverse domains (e.g. prosomeres in the forebrain and rhombomeres in the hindbrain), respectively (Figure 1.2. A) (Rubenstein et al., 1998).

Ventral specification of the neural tube is primarily mediated by: 1) the Sonic Hedgehog protein (SHH), originating from the prechordal plate anteriorly and notochord medially (two embryonic structures lying under the neural tube, see Figure 1.1. A) as well as the floor plate, and genes induced by SHH such as Sonic hedgehog (*Shh*), *HNF3β*, *Nkx2.2* (Briscoe et al., 1999; Chiang et al., 1996; Echelard et al., 1993; Ericson et al., 1996; Hynes et al., 1995; Marti et al., 1995; Roelink et al., 1994; Roelink et al., 1995; Tanabe and Jessell, 1996) and, 2) retinoic acid (RA) which probably comes from the adjacent somites (Pierani et al., 1999). In contrast, the dorsal fates of the neural tube are established by proteins of the transforming growth factor-β (TGF-β) superfamily derived from the dorsal roof plate (embryonic structure lying above the neural tube), especially the bone morphogenetic proteins (BMPs) (Basler et al., 1993; Dickinson et al., 1995; Tanabe and Jessell, 1996) including BMP4, BMP7, dorsalin, activin (Liem et al., 2000; Liem et al., 1997; Liem et al., 1995) and Nodal (Schier et al., 1996; Schier and Shen, 2000).

Early A/P patterning in the neural plate is controlled by several transcription factors including the homeobox genes *Lim1* and *Otx2* (Acampora et al., 1995; Ang et al., 1996; Matsuo et al., 1995; Shawlot and Behringer, 1995). Wnt antagonists, including Cerberus and Dickkopf (Bouwmeester et al., 1996; Glinka et al., 1998), induce neural tissues with anterior properties. Proteins such as Noggin, Follistatin, Chordin and Frizzled related protein Frzb (Doniach, 1993; Harland, 2000; Thomsen, 1997; Wessely and De Robertis, 2002) also participate in the establishment of structures with anterior fate. Candidates for posteriorizing signals include RA (Blumberg et al., 1997), basic

Fibroblast Growth Factor (FGF) (Cox and Hemmati-Brivanlou, 1995; Lamb and Harland, 1995) and Wnt signaling (Altmann and Brivanlou, 2001; Sasai and De Robertis, 1997).

1.2. Organization of the adult telencephalon

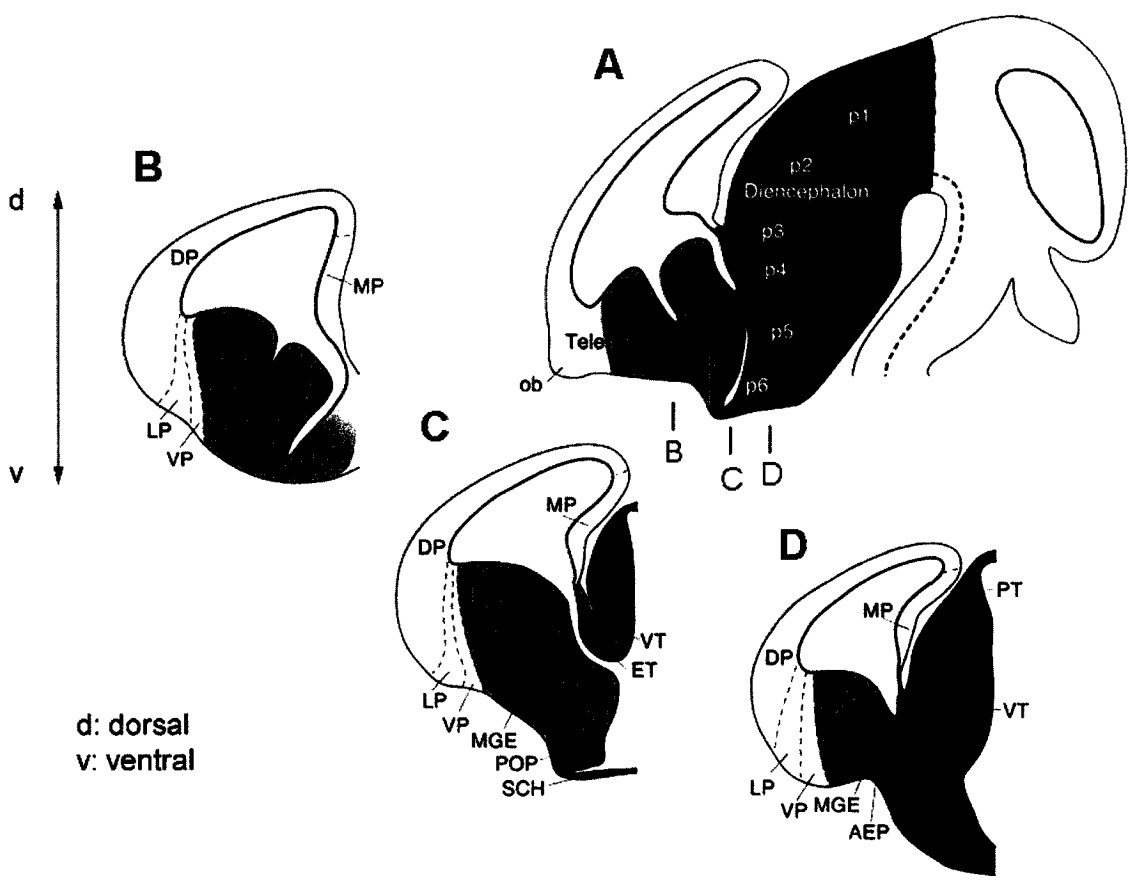
Fate mapping studies have shown that the telencephalon, which is the anterior part of the forebrain, is derived from cells located in the anterior most portion (rostral margin) of the neural plate (Figure 1.1.) (Cobos et al., 2001; Couly and Le Douarin, 1987; Eagleson and Harris, 1990; Inoue et al., 2000; Rubenstein et al., 1998; Varga et al., 1999; Whitlock and Westerfield, 2000). The telencephalon will develop into the cerebrum or the cerebral hemispheres (reviewed in (Rubenstein et al., 1998; Wilson and Rubenstein, 2000). The cerebrum is the most complex structure of the vertebrate brain in terms of structural organization, cellular components and neuronal networks. It is involved in several cognitive and behavioral functions in human such as consciousness, higher cognition, language, motor control and emotions. Despite the highly variable morphologies of the adult telencephalon of diverse vertebrates, the basic organization of telencephalic subdivisions is conserved during embryogenesis (Puelles et al., 2000; Puelles et al., 1999). The two major telencephalic subdivisions are the pallium (roof) and the subpallium (base) (Källen, 1951a; Källen, 1951b; Källen, 1951c; Nieuwenhuys R, 1998; Striedter, 1997).

1.2.1. The pallium

The pallium is organized into four main radial subdivisions: the medial, dorsal, lateral and ventral pallium (Figure 1.2. B-D) (Puelles et al., 2000; Puelles et al., 1999). All of these pallial domains form cortical structures (e.g., superficial laminar neuronal zones) but the lateral and ventral pallium also develop deep-lying nuclear structures, integrated into the claustro-amygdaloid complex. The medial pallium, or limbic cortex, includes the hippocampus and the subicular regions; the dorsal pallium corresponds to the

Figure 1.2. Organization of the prosencephalon

(A) Schema of a sagittal section through the brain of an E13.5 mouse showing the prosencephalic subdivisions proposed by the prosomeric model (Puelles et al., 2000; Puelles et al., 1999). Longitudinal domains include floor plate (fp); alar plate (ap); basal plate (bp) and roof plates (rp) whereas transverse domains comprise prosomeres p1 to p6. (B-D) Transverse sections through the telencephalon and diencephalon showing the different subdivisions of the pallium, subpallium and diencephalon. The pallial subdivisions are: medial pallium (MP); dorsal pallium (DP); lateral pallium (LP) and ventral pallium (VP). The subpallial subdivisions are: lateral ganglionic eminence (LGE); medial ganglionic eminence (MGE); septum (S); anterior entopeduncular area (AEP); anterior preoptic area (POA); posterior preoptic area (POP) and suprachiasmatic nucleus (SCH). Subdivisions in the diencephalon include eminentia thalami (EM); ventral thalamus (VT); dorsal thalamus (DT) and pretectum (PT). Adapted from Rossant and Tam, 2002.


A
p1
p2
Diencephalon
p3
p4
p5
p6
Tele
ob
B
C
D
B
d
v
DP
MP
LP
VP
C
MP
DP
VT
ET
LP
VP
MGE
POP
SCH
D
PT
MP
DP
VT
LP
VP
MGE
AEP
d: dorsal
v: ventral

mesocortex and isocortex (formally known as neocortex), which develops between the medial and lateral pallium; the lateral pallium comprises the primary olfactory cortex, the dorsal claustrum and parts of the amygdalae. The ventral pallium abuts the subpallium and includes the ventromedial claustrum, parts of the amygdala, and parts of the olfactory system and the dorsal most septum (Figure 1.2. B-D) (Puelles et al., 2000; Puelles et al., 1999).

1.2.2. The subpallium

The subpallium includes three primary subdividions: the striatal, pallidal, and telencephalic stalk domains (Figure 1.2. B-D), all of which extend medially into the septum. The striatum abuts the ventral pallium and is derived from a progenitor zone called the lateral ganglionic eminence (LGE). The striatal derivatives include the caudoputamen nucleus, nucleus accumbens, part of the septum and central parts of the amygdala. Under the striatal domain lays the pallidal domain, which is derived from progenitors in the medial ganglionic eminence (MGE) and corresponds to the globus pallidus, ventral pallidum and septum. The subpallial telencepahlic stalk, which is located below the pallidal domain, contains the anterior entopeduncular area (AEP) and preoptic area (POA), and is the site where the major tracts entering and branching out of the telencephalon pass through. Hence, the striatal, pallidal and telencephalic stalk domains converge at different levels of the telencephalon to form embryologically and functionally heterogeneous structures, such as the septum rostrally; the basal ganglia (LGE, MGE, AEP and POA) and piriform cortex medially; and the amygdala, caudally (Figure 1.2. B-D) (Puelles et al., 2000; Puelles et al., 1999; Rossant, 2002).

1.3. Development of the telencephalon

1.3.1. Laminar patterning in the pallium

Histogenesis in the telencephalon varies considerably between the pallium and subpallium. The neocortex, like other dorsal structures such as the tectum of the midbrain, is organized into a precise *laminar architecture* and this is an important

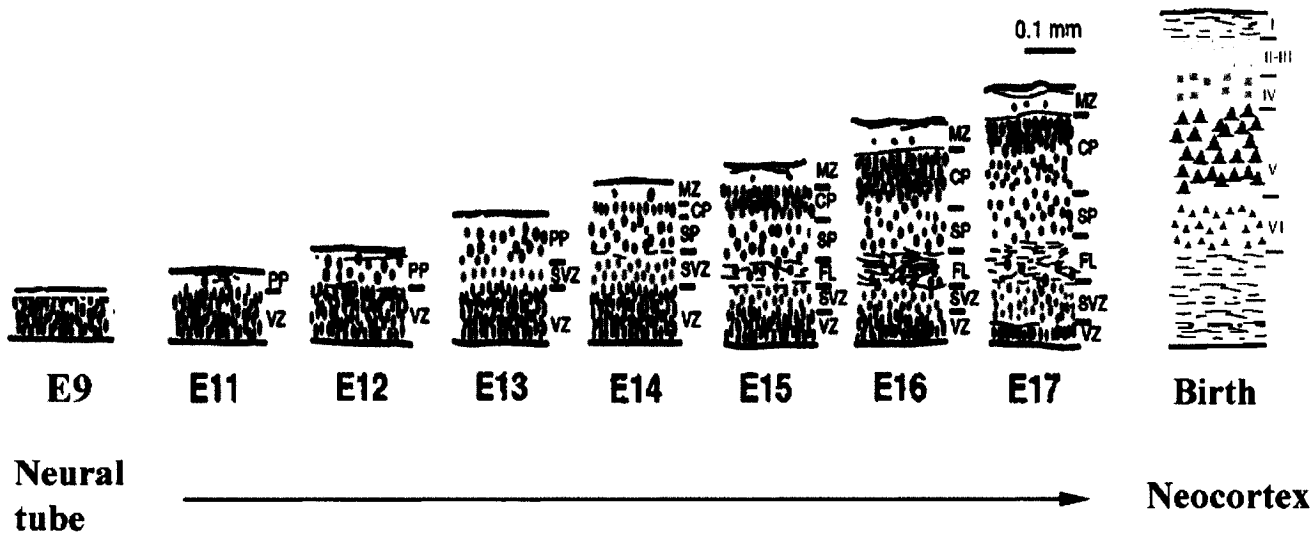
characteristic that distinguishes it from the subpallium and the deep pallial nuclei such as the claustrum, which are organized into *nonlaminar architectures* (Marin, 2002).

The development of cortical lamination takes place in several stages. Around embryonic day 9 (E9), the original neural tube in rodents is a single cell layer thick consisting of a pseudostratified germinal neuroepithelium which is comprised of dividing progenitors. At the preplate stage (E11-E12), two distinctive zones are established: the ventricular zone (VZ) or germinal zone (ependymal) and the preplate (PP) (Figure 1.3.) [reviewed in (Molnar et al., 2006)]. The first population of postmitotic neurons that migrate from the VZ constitute the PP. The PP contains Cajal-Retzius (CR) cells, GABAergic neurons and pioneer neurons (Meyer et al., 2000). At E13, a subventricular (SVZ) is formed between the PP and the VZ. Subsequently, starting around E14, the neuroepithelium is re-organized in two distinct ways; the first one is radial and gives rise to the distinct layers of the mature cortex (1) while the second one is tangential and forms the different columns of the cortex (2).

(1) Upon radial migration of cortical plate neurons, the cortical plate (CP) formation initiates and splits the PP into a marginal zone (MZ) and a subplate (SP) (Figure 1.3.). Layer I neurons, which includes Cajal-Retzius cells, remain near the pial surface in the MZ. Thereafter, successive waves of migration position neurons such that the CP or neocortex now comprises six layers of neuronal cell bodies interacting with one another (Figure 1.3.) (Molnar et al., 2006). These layers are derived from radial glia progenitors located in the VZ where they are born, proliferate, and, after exiting the cell cycle, migrate radially on radial glial processes through the white matter. Migration of young neurons derived from these cortical progenitors toward the CP is largely dependent on radial glia and Cajal-Retzius cells (Rakic, 1971; Rakic, 1972; Rakic, 1974). Radial glia, whose somata reside in the VZ, extend long processes that reach the pial surface. These radial processes, along with signaling (involving the Reelin molecule) from Cajal-Retzius cells located in the MZ, direct the laminar organization. Cajal-Retzius cells (Reelin+ cells) have multiple origins at the border of the developing pallium: the ventral

Figure 1.3. Laminar patterning in the pallium and formation of the neocortex in mouse

The original neural tube is one cell layer thick and is a pseudostratified germinal neuroepithelium (E9). At the preplate stage (E11-E12), two distinctive zones are formed: the ventricular zone (VZ) or germinal zone and the preplate (PP). A subventricular zone (SVZ) separates the PP and the VZ at E13. Upon cortical plate formation, neuronal migration from the VZ forms the cortical plate (CP) and divides the PP into a marginal zone (MZ) and a subplate (SP). Beneath the SP lies a fiber layer (FL) and the SVZ. Successive waves of migration divide the cortical plate or neocortex into six layers (I-VI) of neuronal cell bodies. The layering of the neocortex follows an 'inside-out' gradient of development whereby the deeper layers are formed first, followed by the superficial layers. Modified from Molnar et al. 2006.


0.1 mm
MZ
CP
SP
FL
SVZ
VZ
PP
I
II-III
IV
V
VI
E9
E11
E12
E13
E14
E15
E16
E17
Birth
Neural tube
Neocortex

pallium, the septum and the cortical hem (Bielle et al., 2005), and migrate to their final destinations following different migratory routes (Bielle et al., 2005). Birthdating studies have shown that the layering of the neocortex follows an 'inside-out' gradient of development (Edmondson and Hatten, 1987; Noctor et al., 2001; Rakic, 1972; Rakic, 1974) whereby neuronal precursors that become post-mitotic first (early precursors) form the deeper layers followed later on by the superficial layers. Each neocortical layer differs from the others in its functional properties, its composition of neurons and the synaptic connections that they make (lamina specific pattern). (2) Upon tangential arrangement, the neocortex is further divided into 40 anatomical regions that are highly specialized in distinct behavioral and cognitive functions. Tangential domains include the olfactory cortex, the motor cortices, the somatosensory cortices, and the visual cortex.

1.3.2. Cell formation in the telencephalon

1.3.2.1. Neurogenesis in the telencephalon

Neurons and glial cells (astrocytes and oligodendrocytes) are the two major types of neural cells found in the adult telencephalon. They are derived from three distinct populations of progenitors during neurogenesis: *neuroepithelial cells* that form the wall of the neural tube at the onset of neurogenesis and two other progenitor populations: *radial glial cells* and *basal progenitors*. At early stages, neuroepithelial cells divide symmetrically in the VZ to give rise to two new neuroepithelial cells. Around E12.5, several signaling pathways induce the expression of glial markers by neurepithelial cells, which become radial glial cells. Radial glial progenitors begin to divide asymmetrically to generate the earliest-born neurons in the cortex in mouse. This asymmetric division generates either a neuroepithelial cell and a post-mitotic neuron, which migrates to the PP, or a neuroepithelial cell and a basal progenitor cell, which divide symmetrically away from the VZ surface in the SVZ to generate two post-mitotic neurons (Gotz and Barde, 2005; Haubensak et al., 2004; Miyata et al., 2004; Noctor et al., 2004). Radial glial cells give rise, directly or via the generation of basal progenitors, to most projection neurons of the neocortex. Likewise, the vast majority of neurons in all brain regions of the CNS

derive from radial glia but at distinct time points (Anthony et al., 2004; Gotz and Barde, 2005; Malatesta et al., 2003).

1.3.2.2. Gliogenesis in the telencephalon

Like projection neurons, cortical astrocytes originate from radial glial progenitors located in the cortical VZ. These progenitors are multipotent cells that initially expand via symmetric divisions, but they soon generate more restricted progenitors, including neuronal- and astrocyte-restricted progenitors, which predominate during the late phase of corticogenesis (Davis and Temple, 1994; Luskin et al., 1988; Nieto et al., 2001; Reid et al., 1995; Temple, 2001; Williams and Price, 1995). Of note, astrocytes are born and differentiate after neurons (Altman, 1991). This is due to the sequential generation of neuronal and astrocyte precursors (Qian et al., 2000) as well as to the delayed differentiation of astrocyte precursors, a process that involves multiple signaling, transcriptional, and epigenetic mechanisms that are beginning to be elucidated (Fan et al., 2005; He et al., 2005; Morrow et al., 2001; Sun et al., 2001).

Unlike projection neurons and astrocytes, fate maping experiments showed that oligodendrocytes progenitors (OLPs) are derived from the ventral telencephalon in three successive waves: the first OLPs originate in the MGE and AEP from where they populate the entire embryonic telencephalon including the cerebral cortex before being joined by a second wave of OLPs from the LGE and CGE (Kessaris et al., 2006; Tekki-Kessaris et al., 2001; Thomas et al., 2000). Finally, a third wave that is glial-restricted arises from the cortical SVZ within the postnatal cortex (Gorski et al., 2002; Ivanova et al., 2003; Kessaris et al., 2006; Marshall et al., 2003).

1.3.3. Regionalization in the developing telencephalon

While the morphology of adult pallial and subpallial structures varies considerably between species, very similar regional subdivisions are observed in the telencephalon of all classes of vertebrates during early development. Each subdivision differs from the others by the profile(s) of gene expression that define its boundaries

(Fernandez et al., 1998; Puelles et al., 2000; Puelles et al., 1999). Thus, it seems that the same signaling pathways regulate early regional patterning of the telencephalon, from fish through mammals, by inducing expression of several transcription factors, such as homeodomain proteins, in a spatio-temporal manner.

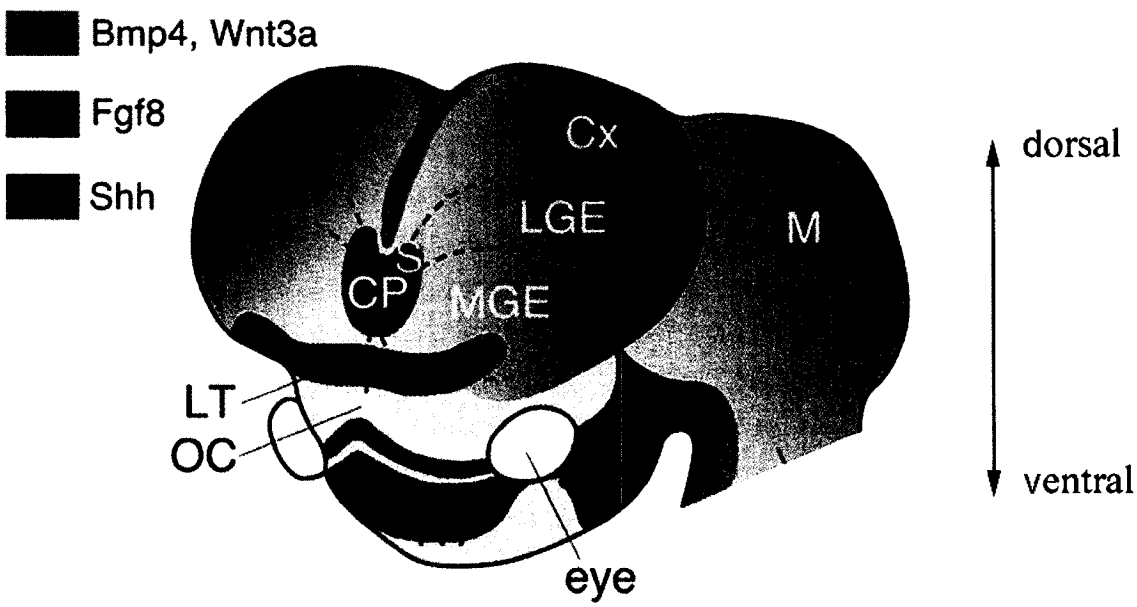
1.3.3.1. Regionalization in the pallium

Besides their role in neural plate formation earlier during development, secreted proteins of the TGF-β (BMP, GDF), and WNT families have been implicated in the dorsal patterning of the telencephalon and its subsequent regionalization. Structures near the dorsal midline, including the cortical hem, choroid plexus and hippocampus, function as signaling centres and express several BMPs (Furuta et al., 1997) and Wnts which are predominantly expressed in the dorsomedial region of the telencephalon (Figure 1.4.) [reviewed in (Campbell, 2003; Zaki et al., 2003)]. For instance, ectopic Bmp expression represses ventral telencephalic markers such as *Nkx2.1* and *Dlx2*, while maintaining dorsal markers, leading to decreased proliferation and increased apoptosis in the basal telencephalon as well as holoprosencephaly (Furuta et al., 1997; Golden et al., 1999). In contrast, dorsal development is disrupted in mice lacking Bmp5 and Bmp7 (Solloway and Robertson, 1999); these mice have delayed closure of the neural tube, hypoplasia of the telencephalic vesicles and reduced apoptosis in the telencephalic roof. Likewise, loss of function of *Wnt3a* and a downstream target of Wnt signaling *Lef1*, leads to loss/reduction of the hippocampus (Galceran et al., 2000; Lee et al., 2000).

The downstream effectors of BMP and/or Wnt signaling in the dorsal telencephalon involves the combinatorial actions of transcription factors such as *Emx1*, *Emx2* and *Lhx2*, which function to specify and expand the medial and dorsal pallium. In fact, mutations in these homeobox genes can cause dorsal to ventral transformations and loss of structures within the pallium. *Lhx2* mutants thus lack most of the hippocampus and neocortex (Monuki et al., 2001). The *Emx* mutants do not show as marked phenotypes in the dorsal pallium as do the *Lhx2* mutants, although hippocampal

Figure 1.4. Signaling centers in the prosencephalon

Schematic representation of the expression of patterning molecules in the prosencephalon of the mouse at midgestation. Members of the BMP and WNT families of signaling molecules such as Bmp4 and Wnt3a are required for dorsal patterning in the prosencephalon whereas Shh and Fgf8 are needed for patterning of ventral structures. Abbreviations: CP, commissural plate; Cx, cortex; HT, hypothalamus; LGE, lateral ganglionic eminence; LT, lamina terminalis; M, mesencephalon; MGE, medial ganglionic eminence; OC, optic chiasm; S, septum. Adapted from Rossant and Tam, 2002.


Bmp4, Wnt3a
Fgf8
Shh
Cx
LGE
M
S
CP
MGE
LT
OC
eye
dorsal
ventral

development is impaired in *Emx2* mutants (Pellegrini M et al. 1996; Yoshida M et al, 1997). Expression of many dorsal markers such as the homeodomain protein *Emx2* is lost in *Gli3* mutant mice, *Gli3* being a repressor of Shh target genes (Rallu et al., 2002; Theil et al., 1999; Tole et al., 2000). Furthermore, *Pax6* and *Emx2* display opposing activities and may generate graded positional identity along the rostro-caudal and D/V axes within the dorsal telencephalon as suggested by their patterns of expression and loss-of-function phenotypes (Figure 1.5.) (Bishop et al., 2000; Mallamaci et al., 2000; Muzio et al., 2002; Pellegrini et al., 1996; Tole et al., 2000; Yoshida et al., 1997).

Besides homeobox proteins, members of the basic-helix-loop-helix (bHLH) family of transcription factors such as *Mash-1* and the two neurogenin genes, *Ngn1* and *Ngn2*, regulate neurogenesis and specify D/V identity in the telencephalon with opposing activities. *Mash-1* is expressed ventrally (in the LGE and MGE) and specifies ventral progenitors (Casarosa et al., 1999; Ma et al., 1997; Parras et al., 2002) while *Ngn1* and *Ngn2* are expressed dorsally and influence neuronal identity in dorsal progenitors (Figure 1.5.) (Fode et al., 2000; Ma et al., 1997; Nieto et al., 2001; Parras et al., 2002; Sun et al., 2001) [reviewed in (Marin, 2002; Schuurmans and Guillemot, 2002; Wilson and Rubenstein, 2000; Zaki et al., 2003)].

1.3.3.2. Regionalization in the subpallium

Each of the embryonic subdivisions in the subpallium, LGE, MGE, and AEP/POA, expresses a different combination of regulatory genes that define its identity and delimit its borders. Some genes including *Otx2*, *Six3* and *Vax1* are expressed by early subpallial progenitors at the neural plate stage, while the expression of other genes including *Dlx1, Dlx2, Gsh1, Gsh2, Mash1, Nkx2.1*, and *Isl1* takes place at later stages following neurulation (Figure 1.5.) (Bulfone et al., 1993a; Bulfone et al., 1993b; Eisenstat et al., 1999; Guillemot and Joyner, 1993; Hallonet et al., 1998; Hsieh-Li et al., 1995; Liu et al., 1997; Olivier et al., 2001; Porteus et al., 1994; Valerius et al., 1995).

Signaling molecules found in the ventral telencephalon include SHH in the MGE, and RA and transforming growth factor-α (TGF-α) in the LGE. Lateral signals are

proposed to participate in patterning the intermediate telencephalon and RA may be involved in this process (Pierani et al. 1999; LaMantia et al. 1993). SHH is required for ventromedial specification (i.e. MGE) (Figure 1.4.) (Chiang et al. 1996; Ericson J et al. 1995; kohtz JD et al. 1998; Machold et al. 2003; Fuccillo et al. 2004) but some aspects of ventrolateral patterning (i.e. LGE) can occur in its absence (Pierani et al. 1999; Rallu et al. 2002). Moreover, D/V patterning can occur in the absence of both SHH and Gli3 (Rallu et al. 2002), suggesting the presence of alternative pathways and/or signals in ventralizing the telencephalon. Mice, fish and humans that have defects in SHH signaling lack ventral telencephalic structures and do not express basal markers (ventral) such as *Nkx2.1* and *Dlx2* (Chiang et al., 1996; Muenke and Beachy, 2000). Alternatively, ectopic expression of Shh in mice and fish induce ventral telencephalic markers such as *Nkx2.1*, *Gsh2*, and *Dlx2* within dorsal telencephalic cells (Corbin et al., 2000; Gaiano et al., 1999), and repress the expression of dorsal markers, such as *Emx1* and *Tbr1* (Kohtz et al., 1998). Many downstream targets of SHH are homeodomain proteins that play a key role in the ventral patterning of the telencephalon including *Nkx2.1*, which is specifically expressed in the MGE (Kimura et al., 1996; Sussel et al., 1999); *Gsh2,* which is expressed in both LGE and MGE (Cajal, 1911; Corbin et al., 2000; Toresson et al., 2000; Yun et al., 2001); and *Pax6,* which is a marker of the dorsal telencephalon and dorsal LGE (Stoykova et al., 2000; Warren et al., 1999) (Figure 1.5.). Thus, *Nkx2.1* mutants display a ventral to dorsal transformation, with cells in the MGE acquiring an LGE identity. Similarly, *Gsh2* mutant mice have ectopic expression of pallial markers in the dorsal LGE, causing dorsalization of this region of the subpallium (Corbin et al., 2000; Szucsik et al., 1997; Toresson et al., 2000; Yun et al., 2001). Complementary to the *Nkx2.1* and *Gsh2* phenotypes, mice lacking *Pax6* express subpallial genes within the ventral pallium (Figure 1.6. B) (Corbin et al., 2000; Stoykova et al., 2000; Toresson et al., 2000; Yun et al., 2001). In addition, *Pax6* and *Gsh2* mutants show milder phenotypes than single mutants (Toresson et al., 2000), confirming that the reciprocal regulatory interactions between these genes mediate D/V patterning on either side of the pallial/subpallial boundary.

Figure 1.5. Regionalization of progenitor domains in the developing telencephalon is defined by the expression patterns of several transcription factors

(A) Coronal hemisections of an E14.5 telencephalon showing the progenitor cell domains in different colors. (B) Transcription factors are expressed with different intensities in the telencephalon and have boundaries of expression that respect pallial and subpallial progenitor domains. Ventral (subpallial) markers include *Isl1, Nkx2.1, Gsh1/2, Mash1* and *Dlx1/2*. Dorsal (pallial) markers include *Pax6, Dbx1, Ngn2, Emx1* and *Lef1*. Abbreviations: AEP, anterior entopeduncular area; Ch, choiroid plexus; dLGE, dorsal lateral ganglionic eminence; DP, dorsal pallium; LGE, lateral ganglionic eminence; LP, lateral pallium; MGE, medial ganglionic eminence; MP, medial pallium; POa, anterior preoptic area; svz, subventricular zone; vLGE, ventral lateral ganglionic eminence; VP, ventral pallium; vz, ventricular zone. Adapted from Rossant and Tam, 2002.

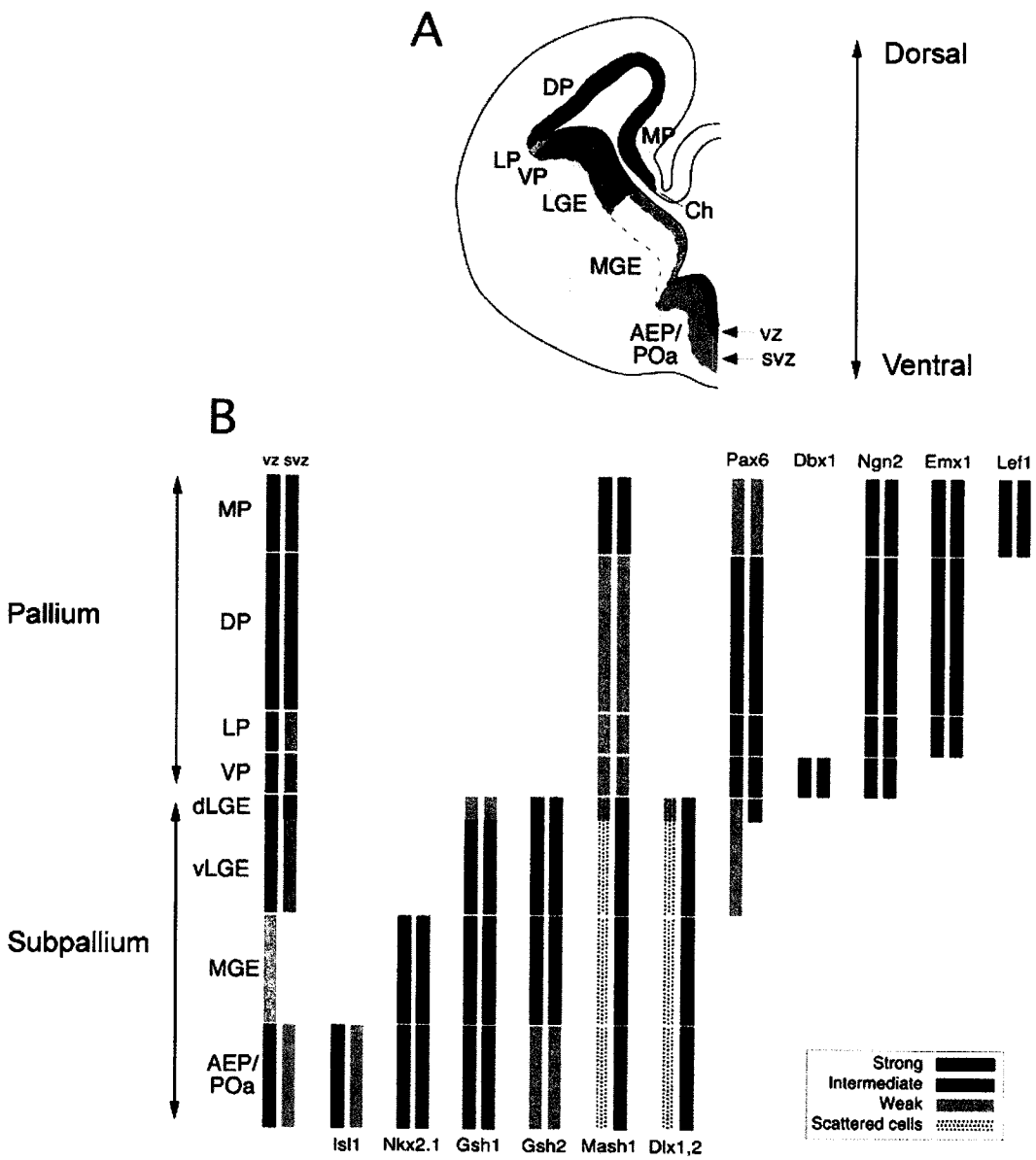
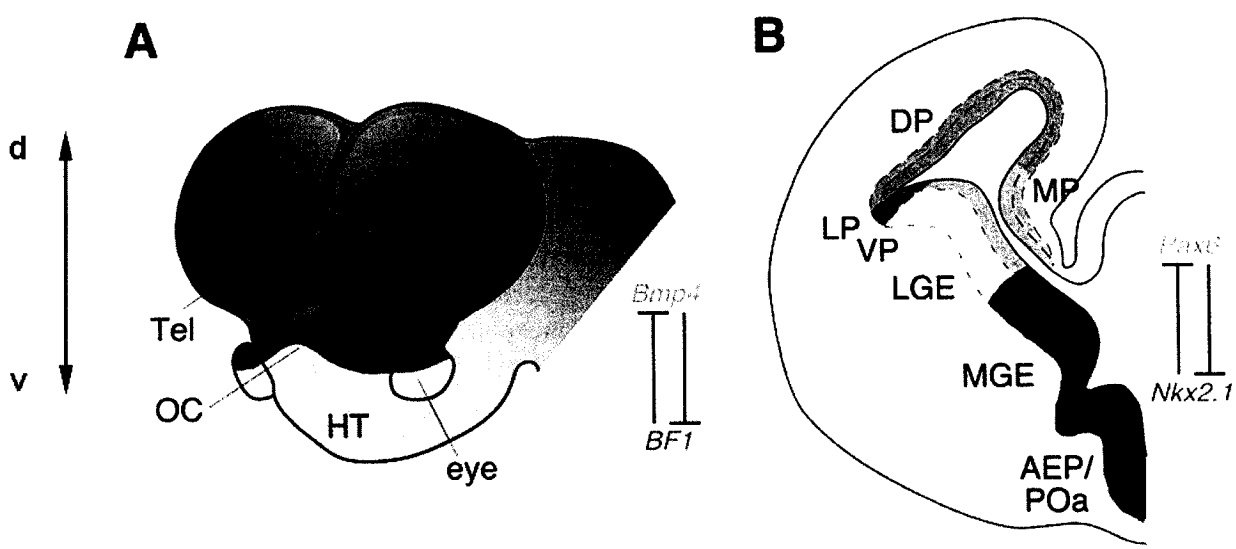

A
DP
MP
LP
VP
LGE
Ch
MGE
AEP/
POa
vz
svz
Dorsal
Ventral
B
vz svz
MP
DP
LP
VP
dLGE
vLGE
MGE
AEP/
POa
Pallium
Subpallium
Isl1
Nkx2.1
Gsh1
Gsh2
Mash1
Dlx1,2
Pax6
Dbx1
Ngn2
Emx1
Lef1
Strong
Intermediate
Weak
Scattered cells

Figure 1.6. Repressive interactions between regulatory genes set the boundaries in the telencephalon

(A) The establishment of appropriate D/V patterning in the telencephalon depends on the reciprocal repression between dorsal and ventral genes such as *Bmp4* and *BF1*, respectively. (B) Independent progenitor cell domains in the LGE and MGE are defined by the interaction between several regional markers such as *Pax6* and *Nkx2.1*. Abbreviations: AEP, anterior entopeduncular area; DP, dorsal pallium; HT, Hypothalmus; LGE, lateral ganglionic eminence; LP, lateral pallium; M, mesencephalon; MGE, medial ganglionic eminence; MP, medial pallium; OC, optic chiasm; POa, anterior preoptic area; Tel, telencephalon; VP, ventral pallium. Adapted from Rossant and Tam, 2002.



d: dorsal
v: ventral

Other factors control the competence of the telecephalon to be ventralized: *BF1* (*Foxg1*) mediates ventralization of the telencephalon by restricting BMP signaling to the dorsal telencephalon (Figure 1.6. A) (Dou et al., 1999).

Fgf signaling pathway is also involved in patterning of the ventral and dorsal telencephalon (Figure 1.4.). Mouse embryos bearing hypomorphic and conditional null *Fgf8* mutations have small and abnormally patterned telencephalon (Fukuchi-Shimogori and Grove, 2001; Meyers et al., 1998; Shanmugalingam et al., 2000; Storm et al., 2006). A major reduction in *Foxg1* expression, a reduced mitotic index, and increased apoptosis contribute to telencephalic hypoplasia in the *Fgf8* null mutants (Storm et al., 2006). *Fgf8* also regulates the expression of *Bmp4*, *Wnt8b* and *Shh*, which in turn affect patterning of both dorsal and ventral structures (Storm et al., 2006).

1.3.4. Cell migration in the telencephalon

Cell migration occurs after patterning and regionalization of progenitors in the forebrain. Once cells are specified, they will migrate to their final destination in the mantle of the forebrain. Like in other CNS regions, there are two types of cell migration in the forebrain based on their orientations: *radial migration*, in which cells migrate from the progenitor zone (place of birth) towards the surface of the brain following the radial disposition of the neural tube and, *tangential migration*, in which cells migrate orthogonally to the direction of radial migration.

1.3.4.1. Radial migration in the telencephalon

As mentioned earlier, radial migration is the major type of migration in the cerebral cortex. Neurons migrate along the radial glial fibers acting as guides to reach their final destination in the cortical plate where they differentiate into glutamatergic neurons (Figure 1.7. B) (Edmondson and Hatten, 1987; Noctor et al., 2001; Rakic, 1971; Rakic, 1972; Rakic, 1974). Radial migration is also found in structures that develop in the absence of lamination (the formation of cortical layers), such as the striatum, although

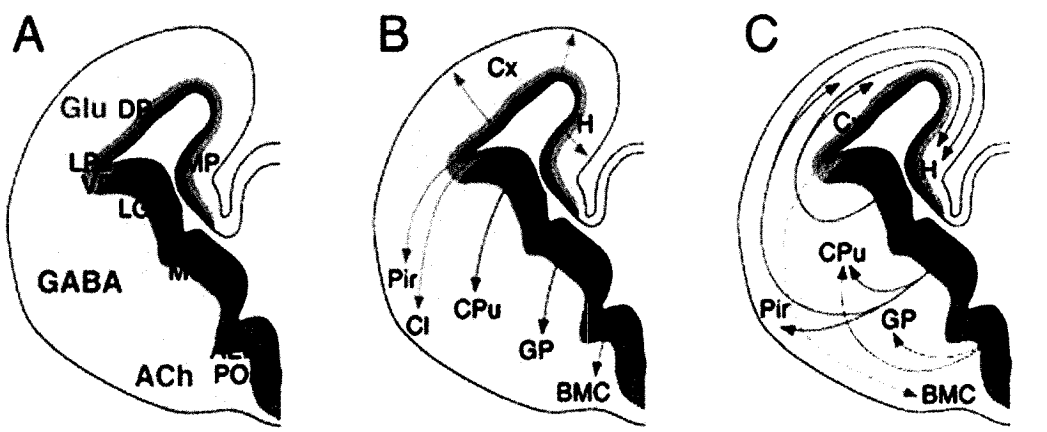
different mechanisms might be involved (Figure 1.7. B) [For review, see (Marin and Rubenstein, 2003a)].

1.3.4.2. Tangential migration in the telencephalon

Different types of neurons are born in distinct subdivisions of the embryonic subpallium and follow restricted routes of migrations before reaching their final destinations in both the pallium and the subpallium (Figure 1.7. C). During embryonic development in mammals, precursors of olfactory interneurons (periglomerular and granule cells) are not intrinsically generated in the olfactory bulb but in the dorsal region of the LGE. They reach their destinations through tangential migration (Corbin et al., 2000; Dellovade et al., 1998; Marin and Rubenstein, 2001b; Marin and Rubenstein, 2003a; Sussel et al., 1999; Wichterle et al., 2001; Yun et al., 2001). This type of migration continues through adulthood and progenitors of olfactory interneurons are constantly born in the SVZ of the lateral ventricle (Lois and Alvarez-Buylla, 1994). They migrate to the olfactory bulb along a highly restricted route called the rostral migratory stream (RMS) (Kornack and Rakic, 2001; Lois and Alvarez-Buylla, 1994; Pencea et al., 2001; Thomas et al., 1996). Likewise, the embryonic subpallium is the origin of a large number of cells that migrate tangentially towards the developing cerebral cortex and hippocampus (Figure 1.7. C) (Anderson et al., 1997a; Corbin et al., 2001; de Carlos et al., 1996; Lavdas et al., 1999; Letinic et al., 2002; Marin and Rubenstein, 2001b; Pleasure et al., 2000; Sussel et al., 1999; Tamamaki et al., 1997; Wichterle et al., 1999). Cells migrating tangentially into the cerebral cortex give rise primarily to GABAergic interneurons (Anderson et al., 2002; Cobos et al., 2001; Stuhmer et al., 2002b; Wichterle et al., 2001), but also generate cortical oligodendrocytes during embryogenesis (He et al., 2001; Olivier et al., 2001; Ross et al., 2003; Santagati et al., 2003; Spassky et al., 1998). These migrating cells originate from multiple sites within the subpallium. Most cortical GABAergic neurons seem to derive from the MGE (Anderson et al., 2002; Anderson et al., 2001; Lavdas et al., 1999; Sussel et al., 1999; Wichterle et al., 1999; Wichterle et al., 2001) although a substantial number of interneurons also originate from other structures

Figure 1.7. Regionalization of the telencephalon defines neural progenitors that will migrate radially and/or tangentially to produce neurons using the three major classes of neurotransmitters in the telencephalon

(A) Neurons using distinct neurotransmitters are derived from different progenitor populations in the telencephalon. Hence, cholinergic (red) and GABAergic neurons (blue) originate from ventral progenitor cells located in the AEP/POa (red), and, LGE and MGE (blue), respectively. In contrast, glutamatergic neurons are specified in the pallium (green). (B, C) Neurons from the three classes mentioned above follow radial (B) and tangential (C) routes of migration in the telencephalon to reach their final destination(s) appropriately. The dotted line indicating a migration of pallial neurons into the BMC in panel C is theoretical. Abbreviations: Ach, acetylcholine; AEP; anterior entopeduncular area; BMC, basal magnocellular complex; Cl, claustrum; Cpu, caudate-putamen nucleus; Cx, cortex; DP, dorsal pallium; GABA, γ-aminobutyric acid; GP, globus pallidus; Glu, glutamate; H, hippocampus; LGE, lateral ganglionic eminence; LP, lateral pallium; MGE, medial ganglionic eminence; MP, medial pallium; Pir, piriform cortex; POa, anterior preoptic area; VP, ventral pallium. Adapted from Rossant and Tam, 2002.



Progenitor Zones **Radial Migration** **Tangential Migration**

such as the CGE (Figure 1.7. C) [(Anderson et al., 2001; Jimenez et al., 2002; Nery et al., 2002), for recent review, see (Wonders and Anderson, 2006)].

1.3.4.3. Routes followed by tangentially migrating neurons and lamninar positioning in the dorsal cortex

GABAergic interneurons migrating to the cortex follow very restricted routes during development. Cell tracing studies identified two different streams of interneurons to round the corticostriatal notch and follow tangentially orientated paths to enter the cortex [reviewed in (Corbin et al., 2001; Marin and Rubenstein, 2001b; Marin and Rubenstein, 2003a; Metin et al., 2006; Nadarajah and Parnavelas, 2002)]. An early stream (E12 in mouse), originating in the MGE, innervates mainly the PP. These cells are tangentially orientated and many show features typical of Cajal–Retzius cells (Figure 1.8. A) (Lavdas et al 1999). A second and more prominent stream, composed also of MGE cells, has been observed to migrate predominantly through the IZ slightly later in development (E13-E15 in mouse) (Figure 1.8. B). At the late stages of corticogenesis, cells originating mostly in the LGE, but also in the MGE, appear in the lower IZ and SVZ (E15-birth) (Figure 1.8. C) (Anderson et al., 2001).

Upon reaching the dorsal cortex, interneurons acquire their laminar position in different ways and through mechanisms that are beginning to be uncovered. Imaging of the surface of the cerebral hemispheres in both explant cultures and brains of living mouse embryos showed interneurons descending into the underlying cortex to assume positions with isochronically generated radially derived neurons (Ang et al., 2003). It seems therefore that once interneurons reach the neocortex, either through the MZ or the IZ/SVZ streams, they migrate radially in the final stage of their journey within the cortex to take up their positions in the appropriate layers (Figure 1.8. D) (Ang et al., 2003; Hevner et al., 2004; Polleux et al., 2002; Tanaka et al., 2003). Interneurons arising in the ventral telencephalon at later stages of corticogenesis actively migrate towards the cortical VZ upon reaching the dorsal telencephalon, a mode of movement called 'ventricle-directed migration' (Nadarajah et al., 2002). After a pause in this proliferative

zone, they migrate radially in the direction of the pial surface to take up positions in the developing CP (Figure 1.8. D) (Nadarajah et al., 2002).

Like pyramidal cells, cortical interneurons are believed to be disposed in an 'inside-out' pattern within the developing cortex. Furthermore, it is believed that contemporaneously born interneurons and pyramidal cells reside within the same layer (Ang et al., 2003; Fairen et al., 1986; Hevner et al., 2004; Miller, 1985; Peduzzi, 1988; Valcanis and Tan, 2003), although this view is still a matter of debate (Yozu et al., 2004).

1.3.4.4. Molecular control of tangential migration

Several factors and guidance molecules control neuronal migration from the subpallium to the pallium. Motogenic factors including the hepatocyte growth factor (HGF) (Powell et al., 2001), the brain derived neurotrophic factor (BNDF) and members of the neurotrophin family such as NT4, strongly stimulate neuronal migration (Polleux et al., 2002), although the specific mechanisms that mediate this function are yet to be elucidated. Guidance molecules/cues play an important role in directing interneuron migration in a ventral to dorsal manner; chemorepulsive and chemoattractive factors produced by the POA and the cortex seem to be involved in this directional guidance, respectively (Marin et al., 2003b; Metin et al., 2006; Wichterle et al., 2003). Furthermore, sorting of interneurons destined to the cerebral cortex or to the striatum appears to be mediated by Neuropilin/Semaphorin interactions. Expression of members of the semaphorin family, including Sema3A and Sema3F, in the mantle of the developing striatum prevents cortical interneurons, which express neuropilin receptors (Neuropilin1 and 2), from entering this structure (Marin et al., 2001a).

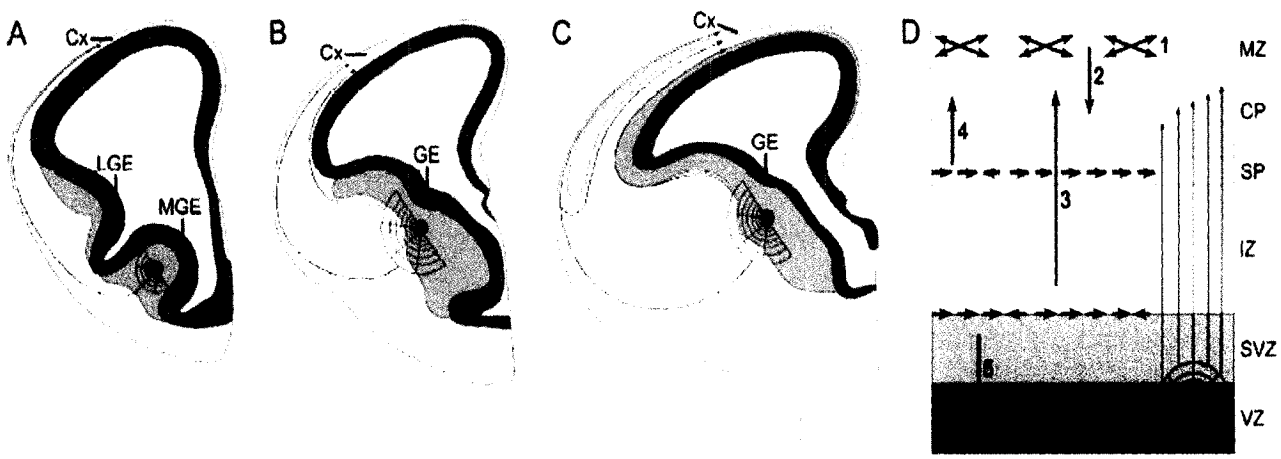
1.4. Inhibitory neurons in the telencephalon - the GABAergic phenotype -

1.4.1. Classes of neurons in the telencephalon

Two general classes of neurons form the complex neuronal network in the telencephalon: 1) the excitatory pyramidal neurons or projection neurons, which send their axonal projections to distinct targets inside and outside the telencephalon (e.g ipsi/

Figure 1.8. Routes of tangential migration followed by immature GABAergic interneurons and laminar positioning of interneurons in the dorsal telencephalon during development in mouse

In mouse, cortical interneurons (red) originate from the ganglionic eminences (GE) in the subpallium. (A) At early stages (E12), these interneurons follow exclusively the MZ while at later stages (B and C) they migrate along routes at the levels of IZ/SVZ and SP. (D) In the MZ, interneurons (1) display varying orientations and (2) enter the developing CP by inward migration. Cells that enter the cortex through the lower streams migrate to the CP (3 and 4) directly or (5) after a short pause in the VZ. Pyramidal cells (blue) are of pallial origin and migrate radially to their positions in the CP. Abbreviations: Cx, cerebral cortex; lateral (LGE) and medial (MGE) ganglionic eminence (GE); MZ, marginal zone, CP, cortical plate; SP, subplate; IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone. Modified from Metin et al. 2006.



E12 **E13-E15** **E15-Birth** **Laminar positioning**

red: cortical interneurons
blue: pyramidal interneurons

contralateral cortex, brain stem, spinal cord, thalamus) and, 2) the inhibitory neurons or interneurons, which form only local synaptic connections between afferent and efferent excitatory neurons and/or other interneurons (DeFelipe and Farinas, 1992; Peters A, 1984). Pyramidal/projection neurons in the neocortex use glutamate as their principle neurotransmitter and comprise 70-80% of neocortical neurons (DeFelipe and Farinas, 1992; Peters A, 1984; Peters and Sethares, 1991; White, 1989). Most striatal neurons (90%) are also projection neurons but use GABA (γ-amino butyric acid) as their main neurotransmitter (Gerfen, 1996; Heimer, 1995; Kawaguchi et al., 1995; Parent and Hazrati, 1995). Inhibitory neurons or interneurons comprise 20-30% of all neocortical neurons (Hendry et al., 1987; Meinecke and Peters, 1987; Parnavelas et al., 1977) and 10% of striatal neurons (Kawaguchi, 1997a; Kawaguchi et al., 1995). They are subdivided into several subtypes based on their neurotransmitter content. The most common type of telencephalic interneurons contains GABA as its main neurotransmitter. However, some interneurons of the olfactory bulb and striatum use dopamine (dopaminergic) or acetylcholine (cholinergic) as their principal neurotransmitter (Gall et al., 1987; Kawaguchi et al., 1995). In mouse, it is suggested that glutamatergic neurons are specified in the pallium, whereas GABAergic and cholinergic neurons are specified in the subpallium (Figure 1.7. A).

1.4.2. Properties of interneurons in the telencephalon

Although there are several types of pyramidal cells with different projection sites, these cells have relatively uniform shapes with apical and basal dendrites and with primary axons that descend towards the white matter (DeFelipe and Farinas, 1992; Feldman, 1984). In contrast, non-pyramidal cells display morphological diversity of dendritic and axonal arborization patterns as revealed by Golgi staining (Cajal, 1911; Fairen, 1984; Peters, 1984), and are anatomically, molecularly, and physiologically heterogenous (Fairen, 1984; Kawaguchi and Kubota, 1997b; Naegele and Barnstable, 1989). Thus, interneurons are subdivided into several subclasses on the basis of: **1)** *their morphological features;* based on their dendritic and axonal morphologies, there are

basket cells, bipolar cells, double bouquet cells, bitufted cells, chandelier cells and Martinotti cells, **2)** *their anatomical features;* based on their axonal arborization, interneurons can be axon- or dendrite-targeting cells or both, and soma-targeting cells, **3)** *their electrophysiological properties;* there are several subtypes of interneurons based on their firing patterns: fast spiking (FS), burst spiking non-pyramidal (BSNP), regular spiking non-pyramidal (RSNP), irregular spiking (IR) and late spiking (LS), **4)** *their molecular properties* primarily their neuro-chemical content. Interneurons can express one or more neurotransmitters including three calcium-binding proteins: parvalbumin (PV), calretinin (CR) and calbindin (CB), and four neuropetides: somatostatin, neuropeptide Y (NPY), vaso-active intestinal peptide (VIP) and cholecystokinin (CCK) [(for review on cortical interneurons, see (Markram et al., 2004)]. Note that one or more combination(s) of the above properties define each subtype of interneurons and it is therefore difficult to systematically distinguish between different subtypes of interneurons based on one specific feature. Nevertheless, several studies have shown that, for instance, PV-, SOM- and CR-positive interneurons constitute distinct, non-overlapping groups of GABAergic neurons in the rat frontal cortex (Kubota et al., 1994), and in the sensory cortex (rat primary visual cortex) (Gonchar and Burkhalter, 1997). These neurons have different electrophysiological characteristics (Cauli et al., 1997; Kawaguchi and Kubota, 1997b). These three subclasses account for more than 80% of all GABAergic neurons in the sensory cortex (Gonchar and Burkhalter, 1997; Tamamaki et al., 2003).

1.4.3. Function of interneurons in the telencephalon

The telencephalon is the seat of language, memory, motor functions, emotions, higher cognition and consciousness in humans. Inhibitory interneurons are indispensable for a normal telencephalic function since they are the regulators of principal neuron activity (pyramidal cells), a role that is primarily mediated through inhibition and establishment of appropriate spatio-temporal balance in the excitation/inhibition ratio depending on the context (Borg-Graham et al., 1998; Kisvarday and Eysel, 1993; Monier

et al., 2003; Murthy and Humphrey, 1999; Santagati et al., 2003; Sato et al., 1996; Sillito, 1984; Wehr and Zador, 2003). The connectivity of the networks and the properties of their intrinsic voltage-gated currents are finely tuned allowing inhibitory interneurons to generate and control the rhythmic output of large populations of both principle cells (pyramidal neurons) and other populations of interneurons (McBain and Fisahn, 2001). GABA itself produces inhibition with different time courses depending on which receptor subtypes ($GABA_A$ or $GABA_B$) are expressed in the post-synaptic membrane (Benardo, 1994; Connors et al., 1988; Kang et al., 1994; McCormick, 1989; Tamas et al., 2003; Thomson and Destexhe, 1999). GABAergic cortical cells are excitatory during development (Ben-Ari, 2002). In addition, the relevance of inhibitory neurons to cortical function is evidenced by the severe consequences of cortical dysfunction related to absence/reduction of inhibition such as in cases of epilepsy (role of chandelier cells) (DeFelipe, 1999), schizophrenia (role of chandelier cells) (Lewis, 2000), working memory (role of GABA in spatial tuning) (Rao et al., 2000), and, perhaps, autism (Rubenstein and Merzenich, 2003).

1.4.4. Origin of interneuron subtypes in the adult cortex

While most GABAergic neurons in the olfactory bulb, cerebral cortex and hippocampus derive from the ventral telencephalon in the anlage of the basal ganglia in rodents (LGE, MGE and CGE) (Anderson et al., 2002; Anderson et al., 2001; Corbin et al., 2001; Jimenez et al., 2002; Marin and Rubenstein, 2001b; Nery et al., 2002; Parnavelas, 2000), it is not known precisely where and when each subtype of interneuron originates during embryogenesis. Recently, *in vitro* cell transplantation experiments (Xu et al., 2004) and *in utero* fate-mapping experiments (Butt et al., 2005) in mice have uncovered the spatio-temporal origin of the major cortical interneuron subtypes. The PV- and SOM-expressing interneuron subgroups seem to originate primarily within the MGE, whereas CR-, VIP- and NPY-expressing interneurons appear to derive mainly from the CGE. Yet, it is still not clear how the diversity of interneuron subtypes is generated and

whether distinct progenitors give rise to different subgroups of interneurons in each of the above regions.

Unlike in rodents, more than half of the cortical interneurons in humans derive from mitoses within the cortical SVZ (Letinic et al., 2002), while the rest originate from progenitors in the subpallium and migrate tangentially to the cortex (Letinic et al., 2002; Rakic and Zecevic, 2003). However, it is also unknown if these cortical and subcortical progenitors give rise to the same subtypes of interneurons or not.

1.5. The Vertebrate *Dlx* family of genes

1.5.1. Origin, genomic organization and evolution of the *Dlx* genes

1.5.1.1. The distal-less gene -*Dll*-

The vertebrate *Dlx* genes are related to the *Drosophila Distal-less* gene -*Dll*- which is required for proximo-distal (P/D) patterning of the leg. As indicated by its name, *Dll* function is required for the specification of distal leg pattern elements in *Drosophila*. *Dll* mutants lack the rudimentary larval limbs and die before birth (Cohen et al., 1989; Cohen and Jurgens, 1989; Gorfinkiel et al., 1997; Wu and Cohen, 1999). *Dll* is also required for development of the P/D axis and specification of the identity of the antenna, which corresponds to both the ear and the nose of the fly (Cohen et al., 1989; Dong et al., 2000). In addition, it is needed in the development of other limb-derived structures including the mouthparts (Cohen and Jurgens, 1989) and the analia (Gorfinkiel et al., 1999). Besides its major role in adult appendage development, *Dll* is expressed in both the optic lobe of the brain (Kaphingst and Kunes, 1994) and the glial cells of the ventral nerve cord (J.B. Skeath and G.P., unpublished observation). However, it remains unknown whether it plays a role in the development of the CNS in the fly.

1.5.1.2. Genomic organization of the *Dlx* genes in Vertebrates

Dlx genes are found in all vertebrates. Mammals including mouse and humans have six *Dlx* genes (Nakamura et al., 1996; Price et al., 1991; Robinson and Mahon,

1994; Robinson et al., 1991; Scherer et al., 1995; Simeone et al., 1994; Stock et al., 1996; Weiss et al., 1994). These genes are organized in three bigene clusters with inverted transcription orientation (*face-to-face*) and are separated by short intergenic regions (<12kb). The three *Dlx* bigene clusters are *Dlx1/Dlx2, Dlx5/Dlx6, and Dlx3/Dlx7.* Each *Dlx* cluster is linked to a *Hox* cluster on a separate chromosome (McGuinness et al., 1996; Nakamura et al., 1996; Simeone et al., 1994; Stock et al., 1996). All *Dlx* genes comprise three exons separated by two relatively short introns (<3kb). They encode proteins of 243-333 amino acids including a highly conserved homeodomain of 61aa (Liu et al., 1997). In some teleost fish (ray-finned fish) such as zebrafish, there are eight known *Dlx* genes (Akimenko et al., 1994; Ekker et al., 1992; Ellies et al., 1997b; Stock et al., 1996). Six of these genes are clustered in bigene clusters similar to those found in mammals, and the remaining two genes (*dlx2b* and *dlx4a*) are not linked to other *Dlx* genes. Based on similarities in sequence motifs within the homeodomain as well as the N- and C-terminal regions of the Dlx proteins, *Dlx* genes are grouped in two general classes: *Dlx2/3/5* and *Dlx1/4/6* (Ellies et al., 1997b; Stock et al., 1996). Yet, it is not known whether the two groups display any functional differences.

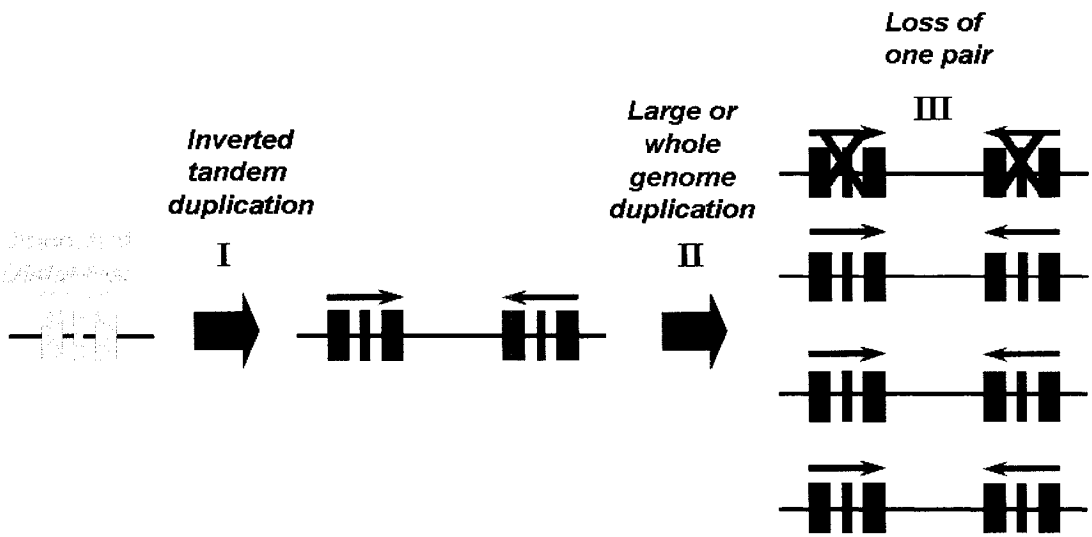
1.5.1.3. Evolution of the Vertebrate *Dlx* genes

As mentioned above, modern vertebrates have six or more *Dlx* genes. Urochordates (primitive vertebrates) such as the tunicate *Ciona intestinalis* possess at least three *Dlx* genes, two of which are linked in one cluster (Caracciolo et al., 2000). Lampreys have at least four *Dlx* genes, two of which are organized in one cluster. The remaining two genes are not linked to another gene (Myojin et al., 2001; Neidert et al., 2001). *Drosophila* and amphioxus have one *-Dll- gene* that is most closely related to *Dlx1* (Holland et al., 1996; Stock et al., 1996) although the support for this relationship is not strong. The latter gene may thus be related to the founding member of the vertebrate family. Given the conserved genomic organization of *Dlx* genes and their linkage to the *Hox* clusters, a scenario has been proposed to explain the evolution of this family; it hypothesizes that an inverted tandem gene duplication of an ancestral *Dlx* gene, followed

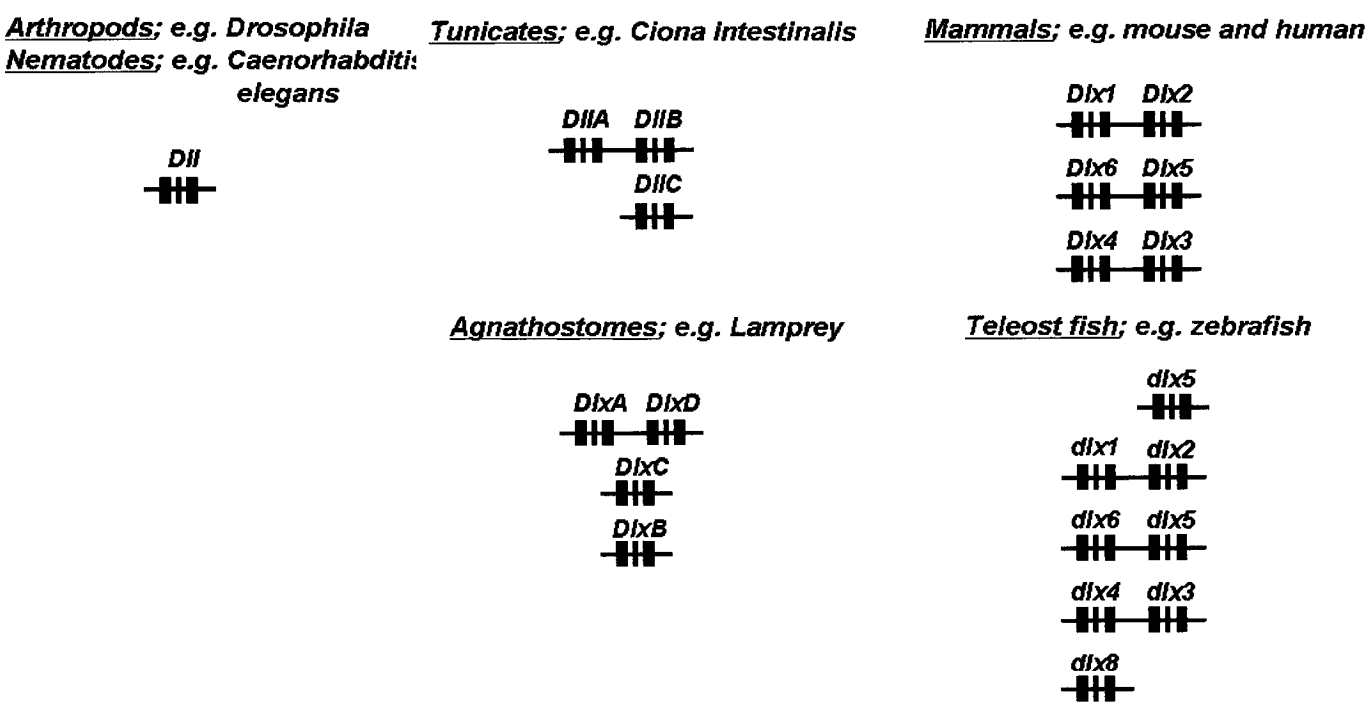
Figure 1.9. A model for *Dll/Dlx* gene family evolution

(A) An initial tandem duplication event (I) from an ancestral *distal-less* gene (orange) gave rise to the first *Dlx* bigene cluster (red and blue). Two rounds of large or whole genome duplication events (II) and a subsequent loss of one *Dlx* pair (III) generated the three *Dlx* bigene clusters of modern vertebrates. Note that, each of these clusters is linked to a *Hox* cluster on a separate chromosome (not shown). (B) Nomenclature and genomic organization of *Dll/Dlx* genes in selected species. Complete genome duplication may have occurred in teleosts after the divergence of this lineage, thus generating additional *Dlx* genes in fish species such as zebrafish. Adapted from Zerucha and Ekker, 2000.

A


Loss of one pair
III
Large or whole genome duplication
II
Inverted tandem duplication
I

B


Arthropods; e.g. Drosophila
Nematodes; e.g. Caenorhabditis elegans
Dll
Tunicates; e.g. Ciona intestinalis
DllA DllB
DllC
Mammals; e.g. mouse and human
Dlx1 Dlx2
Dlx6 Dlx5
Dlx4 Dlx3
Agnathostomes; e.g. Lamprey
DlxA DlxD
DlxC
DlxB
Teleost fish; e.g. zebrafish
dlx5
dlx1 dlx2
dlx6 dlx5
dlx4 dlx3
dlx8

by two rounds of genome duplication and a subsequent loss of one *Dlx* pair gave rise to the mammalian *Dlx* genes (Figure 1.9.) (Ellies et al., 1997b; Neidert et al., 2001). In addition, complete genome duplication may have occurred in teleosts after the divergence of this lineage, thus generating additional *Dlx* genes in fish species such as zebrafish (Figure 1.9.) (Robinson-Rechavi et al., 2001; Stock et al., 1996).

1.6. Functions of the *Dlx* genes in organ development of Vertebrates

Vertebrate *Dlx* genes encode homeodomain transcription factors that regulate transcription of downstream target genes. They are primarily transcriptional activators. They have similar DNA-binding properties *in vitro* and can activate transcription from both artificial (Feledy et al., 1999b; Liu et al., 1997; Masuda et al., 2001; Zhang et al., 1997) and authentic enhancers (Benson et al., 2000; Dodig et al., 1996; Iler et al., 1995; Morasso et al., 1996; Roberson et al., 2001; Stuhmer et al., 2002a; Yu et al., 2001; Zerucha et al., 2000). They can also function as repressors on artificial reporter genes (Ryoo et al., 1997; Yu et al., 2001).

Dlx genes are required for normal development of an array of tissues and organs including the forebrain, the limb/fins, the branchial arches/facial derivatives, sensory organs as well as bone and cartilage formation. Two general characteristics of these genes are: 1) their nested and sequential (spatio-temporal) expression patterns along the proximo-distal axis primarily in the forebrain and the branchial arches (Eisenstat et al., 1999; Liu et al., 1997; Qiu et al., 1997); 2) their functional redundancy/compensation as a result of high overlap in gene expression primarily between paralogs but also among orthologs (Ellies et al., 1997b; Quint et al., 2000). Note that, subtle differences in *Dlx* gene expression also exist and may account for distinct functions among paralogous genes (Akimenko et al., 1994; Ellies et al., 1997b; Liu et al., 1997). Evidence supporting the above observations comes from loss of function studies performed in mice. Even though single *Dlx* mutants die during embryogenesis (*Dlx3*) or shortly after birth (*Dlx1, Dlx2* and *Dlx5*), there is, in general, a mild or lack of obvious phenotype in tissues where

more than one *Dlx* is expressed at any given time (Acampora et al., 1999; Anderson et al., 1997a; Anderson et al., 1997b; Depew et al., 1999; Qiu et al., 1997; Qiu et al., 1995). Furthermore, all heterozygous mice for these mutations appear normal and are fertile. In contrast, when both paralogous genes in one cluster are mutated (double mutants) severe phenotypes are observed, such is the case of *Dlx1/Dlx2* (Qiu et al., 1997) and *Dlx5/Dlx6* double mutants (Robledo et al., 2002).

1.6.1. Limb development

Dll is primarily required for the distal patterning of leg structures in the fly (Cohen and Jurgens, 1989). This seems a common feature of *Dll/Dlx* function in arthropods and mammals for all *Dlx* genes are expressed in the apical ectodermal ridge (AER) of the developing limbs (fore- and hindlimb) (Dolle et al., 1992; Morasso et al., 1995). Their expression regulates the patterning and outgrowth of the limbs in mammals. Of note, no limb phenotype has been reported for any of the *Dlx* single mutants obtained so far (Acampora et al., 1999; Depew et al., 1999; Qiu et al., 1997; Qiu et al., 1995) suggesting that the various *Dlx* genes serve some redundant function in the AER. However, *Dlx5* and *Dlx6* seem to play a more important role than *Dlx1* and *Dlx2* in limb development. This is based on the fact that, unlike in *Dlx1/Dlx2* mutants where the limb is usually normal (Depew M. and Rubenstein J., unpublished observations), *Dlx5/Dlx6* mutants have severe malformations of the distal limb manifested by split distal limb defects similar to ectrodactyly syndromes seen in humans (Robledo et al., 2002). Thus, the *Dlx5/Dlx6* null limb phenotype phenocopied the Split-hand/split-foot malformation (SHFM) with variable penetrance in the forelimbs and complete penetrance in the hindlimbs (Robledo et al., 2002). In humans, this malformation is characterized by missing digits, fusion of the remaining digits (syndactyly), and median clefts (lobster-claw) (Scherer et al., 1994). Since *Dlx5* and *Dlx6* genetically map close to one of four human SHFM disease loci, they are postulated as candidates of SHFM1 (Crackower et al., 1996; Scherer et al., 1994). In contrast, no obvious phenotype was observed in the patterning of limbs and other appendages in the *Dlx5* mutants suggesting a compensatory role by other *Dlx* genes such

as *Dlx6*. All eight zebrafish *dlx* genes are expressed in the median fin fold (primordia of unpaired fins) and the apical ectodermal cells of the developing pectoral fin buds (Akimenko et al., 1994).

1.6.2. Neural crest cells and pharyngial arch development

All *Dlx* genes are expressed in the ectomesenchymal cells derived from the cranial neural crest cells (except zebrafish *dlx2b or dlx5*) (Akimenko et al., 1994; Bulfone et al., 1993a; Davideau et al., 1999; Dolle et al., 1992; Myojin et al., 2001; Neidert et al., 2001; Qiu et al., 1997; Robinson and Mahon, 1994; Simeone et al., 1994; Yang et al., 1998). These cells migrate from the border of the neural plate and give rise to the first and second pharyngeal arches (maxilla and mandible of the first arch and hyoid arch). These arches will later differentiate into the facial skeleton, dental mesenchyme and connective tissue (Depew, 2002b; Depew et al., 2002a). In the pharyngial arches, *Dlx* gene expression follows a temporal sequence of expression that is well correlated with their nested expression patterns along the P/D axis in that region. Thus, *Dlx1* and *Dlx2* are expressed first in more proximal and intermediate domains, followed by *Dlx5* and *Dlx6* in intermediate domains, whereas in distal zones all six genes are expressed (Acampora et al., 1999; Depew et al., 1999; Depew et al., 2002a; Qiu et al., 1997; Qiu et al., 1995). *Dlx1* and *Dlx2* single mutant mice have similar defects in proximal first and second branchial arch derivatives (Qiu et al., 1997; Qiu et al., 1995). However, *Dlx2* mutants also display distinct craniofacial abnormalities such as in the lateral skull dermatocranium. This is probably due to a higher level of expression of *Dlx2* in the first arch ectoderm, from which derives the previous structure (Qiu et al., 1997). *Dlx1/Dlx2* deficient mice display various defects in proximal domains of the first and second pharyngial arches, with a lack of distal defects suggesting a functional redundancy in this region where all of the *Dlx* genes are expressed (Merlo et al., 2000; Qiu et al., 1997; Qiu et al., 1995). These mutants exhibit not only the defects of the single mutants, but also the absence of the upper molar teeth (Qiu et al., 1997). In contrast to *Dlx1* and *Dlx2* single mutants, *Dlx5* mutants exhibit morphological alterations in skeletal elements derived

from both the proximal and distal domains of the first four branchial arches (Acampora et al., 1999; Depew et al., 1999). This implies that redundancy between *Dlx* genes is not generalized. Moreover, *Dlx5/Dlx6* mutants exhibited severe craniofacial defects including the complete absence of the calvaria, resulting in exencephaly, reduction in the size of the eyes, and clefting and dysmorphogenesis of nasal, maxillary and mandibular structures (Robledo et al., 2002).

1.6.3. Sensory organs development

Dll is required for specifying antennal identity in *Drosophila*, the antenna including both the ear and the nose of the fly. This function is evolutionarily conserved since *Dlx* genes have been implicated in the development of ear, nose, and retina in vertebrates. During neurulation, *Dlx3, Dlx2, Dlx5* and *Dlx6* are expressed in the otic placode and, later on, regionally expressed in the otic vesicle (primordium of the inner ear) (Depew et al., 1999; Ekker et al., 1992; Liu et al., 1997; Quint et al., 2000; Robinson and Mahon, 1994; Zhao et al., 1994). In addition, early expression of *Dlx5* and *Dlx6* is detected in the frontonasal ectoderm, olfactory placodes (primordia of nose) and, subsequently, in the olfactory and respiratory epithelium (Acampora et al., 1999; Depew et al., 1999). Mice lacking the function of *Dlx5* exhibit defects in the morphogenesis of the olfactory pit and in the differentiation of the olfactory epithelium (Depew et al., 1999), as well as regional defects in the inner ear, particularly in the semi-circular canals (Acampora et al., 1999). However, the ear defects reported with *Dlx5/Dlx6* mutants are dramatically more severe than those observed in *Dlx5*-deficient mice, again suggesting that *Dlx5* and *Dlx6* have both unique and redundant functions (Robledo et al., 2002).

In the developing and postnatal retina, *Dlx1* and *Dlx2* are expressed in retinal ganglion cells (RGC), amacrine and horizontal cells (de Melo et al., 2003; Dolle et al., 1992; Eisenstat et al., 1999). The retina of *Dlx1/Dlx2* null mice displays a reduced ganglion cell layer (GCL), with specific loss of differentiated RGCs (other subtypes seem less affected) (de Melo et al., 2005). This phenotype suggests that *Dlx1* and *Dlx2* play a

specific role in terminal differentiation and survival of late-born RGCs in the developing mouse retina (de Melo et al., 2005).

1.6.4. Bone and cartilage formation

Unlike all other members of the mammalian *Dlx* family, *Dlx5* and *Dlx6* are expressed in all skeletal elements (developing bone) from the time of initial cartilage formation onwards. However, *Dlx5* displays stronger and wider expression than *Dlx6* and plays a prominent role in osteoblast differentiation and bone formation (Acampora et al., 1999; Depew et al., 1999; Simeone et al., 1994; Zhao et al., 1994). *Dlx5* is expressed in all bones during osteoblast differentiation including dermal bones (intramembranous) but disappears in fully differentiated osteocytes (Acampora et al., 1999; Depew et al., 1999; Zhao et al., 1994). *Dlx5* mutants display a delayed ossification of the roof of the skull and abnormal osteogenesis. They also carry a defect in the endosteal component of the long bone diaphyses and have a reduction in the periosteal lamina (Acampora et al., 1999). *Dlx5/Dlx6* mutants exhibited more severe axial and appendicular skeleton defects including growth retardation and kinked tail vertebrae (Robledo et al., 2002). Furthermore, while early cartilage formation of long bones was normal at E14.5 in *Dlx5/Dlx6* mutants, the axial and appendicular skeletons displayed severe retardation in chondrocyte/cartilage and osteoblast differentiation and subsequent mineralization (Robledo et al., 2002).

1.6.5. Other functions in organ development

Dlx4 (previously *Dlx7)* is implicated in proliferation and survival during erythropoiesis since it is expressed in normal bone marrow cells (Shimamoto et al., 1997; Shimamoto et al., 2000). *Dlx3* is expressed in the surface ectoderm and has been implicated in epidermal development. Humans with a four-base pair deletion in the coding region of *DLX3* have a disorder known as tricho-dento-osseous (TDO) syndrome that affects morphogenesis of epidermal derivatives (hair) and ectodermal derivatives (teeth and craniofacial skeleton) (Price et al., 1998a; Price et al., 1998b). TDO is an

autosomal dominant inheritance of enamel hypoplasia and hypocalcification with associated strikingly curly hair. In addition, *Dlx3* is expressed in ectodermal components of the developing placenta and *Dlx3*-null mice die in midgestation because of a deficiency in the vascularization of the placenta (Morasso et al., 1999). Some *Dlx* genes are also expressed in the genital eminence (Porteus et al., 1994).

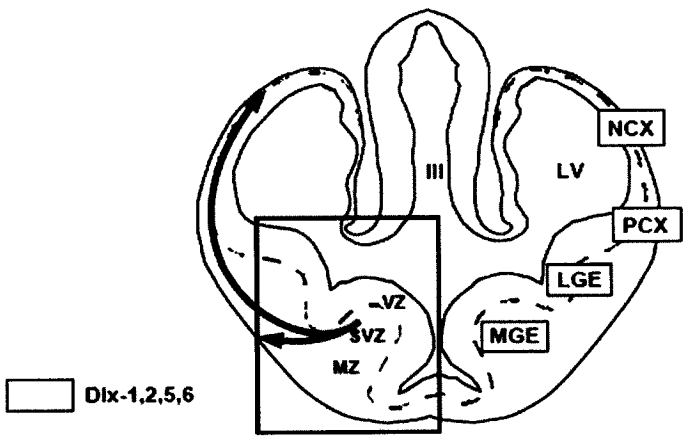
1.7. Function of the *Dlx* Genes in forebrain development

As mentioned earlier, all six mouse *Dlx* genes are expressed in ectodermal derivatives: the nervous system and the surface ectoderm. However, only four *Dlx* genes, *Dlx1, Dlx2, Dlx5* and *Dlx6,* are expressed in the CNS (Bulfone et al., 1993b; Dolle et al., 1992; Eisenstat et al., 1999; Liu et al., 1997; Price et al., 1991; Robinson et al., 1991; Simeone et al., 1994; Yang et al., 1998). This expression is highly restricted to two domains in the forebrain: one telencephalic, which corresponds to the subpallium (basal ganglia), and, one diencephalic, which includes the prethalamus (ventral thalamus) and hypothalamus. Of note, these two domains are also found in other vertebrate species including chicken, frogs, turtles, zebrafish and lampreys (Fernandez et al., 1998; Myojin et al., 2001; Neidert et al., 2001; Puelles et al., 2000; Zerucha et al., 2000). Like in pharyngeal arches, *Dlx* genes display a spatio-temporal sequence of expression in the basal ganglia during development (Figure 1.10.). Thus, *Dlx2* and *Dlx1* (but to a lesser extent) are expressed in early progenitors including some proliferating cells in the VZ (scattered cells). Both *Dlx1* and *Dlx2* are also expressed in most undifferentiated cells in SVZ (uniform expression), and some differentiated cells in the MZ (scattered cells). *Dlx5* is not expressed in VZ progenitor cells but displays very strong expression in the SVZ (proliferating cells) and intermediate expression in the MZ. Finally, *Dlx6* is mainly expressed in differentiated cells in the MZ (Figure 1.10.) (Bulfone et al., 1993b; Eisenstat et al., 1999; Liu et al., 1997; Zerucha et al., 2000).

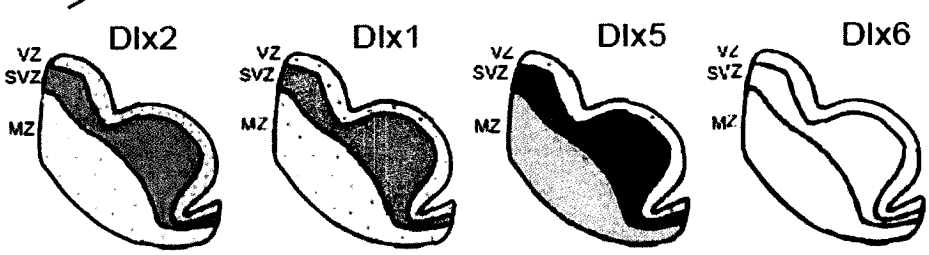
Figure 1.10. Domains of expression of *Dlx* genes in the telencephalon

(A) Schematic representation of an E12.5 coronal section of the mouse telencephalon showing in yellow the combined expression of *Dlx* transcripts. Tangential migration from the subpallium to the pallium is indicated with arrows. (B) *Dlx1, Dlx2, Dlx5* and *Dlx6* expression patterns in the basal ganglia. Most cells in the proliferative zones (VZ and SVZ) in the subpallial telencephalon express one or more *Dlx* gene(s) at some stage of their proliferation and/or differentiation. *Dlx2* is mainly expressed in undifferentiated cells; its expression is found in scattered cells (green dots) in the VZ, in most cells in the SVZ (uniform green) and few cells in the MZ (green dots). *Dlx6* (brown) is primarily expressed in differentiated cells in the MZ (uniform peach). *Dlx1* (red) and *Dlx5* (blue) have intermediate expression patterns in the SVZ and MZ. (C) Proposed model for the spatio-temporal expression of *Dlx* genes and their sequential role at different stages of differentiation. Abbreviations: LGE, lateral ganglionic eminence; LV, lateral ventricle; MGE, medial ganglionic eminence; MZ, mantle zone; NCX, neocortex; PCX, palliocortex; SVZ, subventricular zone; VZ, ventricular zone; III, third ventricle. Adapted from Panganiban and Rubenstein, 2002.

A



B



C

Model:

Dlx2 → Dlx1 → Dlx5 → Dlx6

1.7.1. Control of differentiation and migration of GABAergic neurons in the basal ganglia

Dlx genes are required for the differentiation and migration of most GABAergic neurons in the telencephalon (GABAergic projection neurons and GABAergic interneurons). This is evidenced by their co-localization in most of these neurons (Anderson et al., 1997a; Anderson et al., 1997b; Stuhmer et al., 2002a; Stuhmer et al., 2002b) and from loss-of-function studies in the mouse (Acampora et al., 1999; Anderson et al., 1997a; Anderson et al., 1997b; Qiu et al., 1997; Qiu et al., 1995). The function of *Dlx* genes in the control of GABAergic phenotype is best understood by examining the *Dlx1/Dlx2* mutant phenotype. These mice die within a few hours after birth and show the following defects. **1)** They exhibit a major block in neurogenesis and subsequent differentiation and radial migration of late-born striatal projection neurons that are derived from the LGE after E12.5 (Anderson et al., 1997b; Marin et al., 2000). Late-born cells are unable to migrate out of the proliferative region and tend to accumulate within the LGE where they partially differentiate (Anderson et al., 1997b; Marin et al., 2000). The onset of this abnormality (around E12.5) coincides with the development of the SVZ as the main proliferative zone in the LGE, where *Dlx1, Dlx2* and *Dlx5* are primarily expressed. Furthermore, the failure of the mutant SVZ to mature is also reflected by the lack of *Dlx5, Dlx6* and *Oct6/SCIP* expression, specific markers of this region (Anderson et al., 1997b; Marin et al., 2000). Taken together, these defects demonstrate that: **1)** *Dlx1* and *Dlx2* are upstream regulators of *Dlx5* and *Dlx6* in the LGE and their function is critical for normal striatal development after E12.5. **2)** *Dlx1/Dlx2* null mice display a major block in the development of several types of GABAergic, dopaminergic and cholinergic interneurons derived from the subpallium. Hence, they have reduced numbers of striatal interneurons (GABAergic and cholinergic), olfactory bulb interneurons (>95% of GABAergic and dopaminergic), hippocampal interneurons (>95% GABAergic) and cortical interneurons (>75% of GABAergic) (Anderson et al., 1997a; Anderson et al., 2001; Anderson et al., 1997b; Bulfone et al., 1998; Marin et al., 2000; Marin and Rubenstein, 2001b; Pleasure et al., 2000). The previous defects are caused by a block of

differentiation and a lack of radial and tangential migration of interneurons from the MGE, the AEP/POA, as well as the LGE. *Dlx5* and *Dlx6* expression is also lost in the MGE except in a narrow strip in the deep MZ (Anderson et al., 1997b; Zerucha et al., 2000).

Unlike *Dlx1/Dlx2* double null mutants, *Dlx* single mutants have subtle defects in forebrain development. However, they also die shortly after birth (*Dlx2* mutant) or within the first month of life (*Dlx1* mutant) due to craniofacial and/or enteric nervous system defects (Acampora et al., 1999; Qiu et al., 1997; Qiu et al., 1995). No major abnormalities have been reported in the forebrain of *Dlx1* mutant mice during development (Qiu et al., 1995). In contrast, in the adult brain, *Dlx1* mutant mice display subtype-specific loss (partial loss) of CR+ (bipolar cells) and SOM+ (e.g. bitufted cells) but not PV+ subtypes of interneurons in the hippocampus and the cerebral cortex (Cobos et al., 2005). This partial reduction occurred in a time-dependent manner (in animals older than one month) (Cobos et al., 2005). In addition, these mice showed generalized electrographic seizures and histological evidence of seizure-induced reorganization, thus, associating loss of *Dlx1* function with delayed-onset epilepsy linked to interneuron loss (Cobos et al., 2005). *Dlx2* mutant mice revealed normal histology but a reduction in tyrosine hydroxylase (TH) immunoreactive neurons (dopaminergic neurons) in the olfactory bulb (Qiu et al., 1995). Analysis of these mice at post-natal stages is hampered by their perinatal death. No abnormalities have been reported in the forebrain of *Dlx5* mutant mice (Acampora et al., 1999). Although no dosage effect was reported thus far in *Dlx* mutant mice, Horike and collegues (2005) identified *Dlx5* as a direct target of Mecp2 (gene mutated in individuals with Rett syndrome) and showed that the maternally expressed *DLX5* display a loss of imprinting in lymphoblastoid cells from individuals with Rett Syndrome (Horike et al., 2005). Taken together, these results suggest that *Dlx* genes have considerable redundant functions in forebrain development; yet, each *Dlx* gene may be more or less needed in specific function(s) (with possible dosage effect at least in humans), or, alternatively, can be required for distinct function(s), as suggested by the subtle differences and incomplete penetrance of single *Dlx* mutant phenotypes.

1.8. Regulation of *Dlx* gene function in Vertebrates

The requirement for *Dll* and *Dlx* function in the development of auditory and olfactory system, the mouthparts, appendages, and maybe the brain of both invertebrates and vertebrates suggest that *Dll/Dlx* was somehow involved in the formation of the primitive versions of these structures prior to the divergence of the two lineages. Consequently, some aspects of *Dlx* function may have been inherited from an ancestor that pre-dates the vertebrate radiation. However, to date, there is no direct evidence for this at the molecular level in terms of regulation mechanisms and/or common targets of *Dll/Dlx*. Alternatively, the similarities in *Dll* and *Dlx* functions could be the result of homoplasy, which is the appearance of sameness resulting from independent evolution. In contrast, following the duplication events that led to multiple pairs of physically linked *Dlx* genes (after the divergence of modern vertebrates), the expression patterns of orthologous genes were, in general, conserved. This suggests that selective pressure to maintain the overall regulatory mechanisms that control *Dlx* expression or to preserve specialized functions including novel functions of the *Dlx* paralogues, was acquired after the duplication events (Panganiban and Rubenstein, 2002; Quint et al., 2000; Zerucha and Ekker, 2000). Thus, it is likely that regulatory mechanisms including shared *cis*- and/or *trans*- acting elements, as well as signaling pathways would have likely been included in these duplication events. This may explain at least some of the overlap in expression patterns and redundant functions of *Dlx* genes.

1.8.1. *Trans*- and *cis*-regulation of *Dlx* genes

Several *Dlx* cis-acting elements have been identified, some of which are found in the intergenic regions of *Dlx* bigene clusters and may be shared between the two *Dlx* paralogs. Indeed, two enhancer elements, zI46i (400bp) and zI46ii (300bp) located in the zebrafish *dlx5a/dlx6a* (previously *dlx4/dlx6*) intergenic region were found to drive correct forebrain expression in transgenic mouse at E11.5, although the second element was much less efficient than the first one (Zerucha et al., 2000). *lacZ* expression in these mice

was found in two domains where endogenous *Dlx* genes including *Dlx5* and *Dlx6* are expressed. These domains are the subpallium (basal ganglia) in the telencephalon and the prethalamus and hypothalamus in the diencephalon. The nucleotide sequence of both elements was highly conserved between zebrafish, mouse and human (>85%). The mouse enhancer (mI56i) also targeted *lacZ* expression to the same domains of the forebrain in a constant manner (Zerucha et al., 2000). A 1.6 kb fragment from zebrafish that contains both elements directs expression of a reporter gene that recapitulates endogenous *dlx* expression in the forebrain of zebrafish embryos, suggesting conserved function of these intergenic enhancers (Zerucha et al., 2000). In all of the above experiments, the reporter gene expression in mouse matched *Dlx5* expression more closely than that of *Dlx6.* Nevertheless, the two enhancers may still be shared by the two *Dlx* genes but interact differentially with the promoter of each gene (Zerucha et al., 2000). Furthermore, zI46i (mI56i) appears to be regulated by Dlx1 and Dlx2 *in vivo* and *in vitro* (Stuhmer et al., 2002a; Zerucha et al., 2000). This suggests that Dlx1 and Dlx2 may act as positive regulators of *Dlx5/Dlx6* expression through this intergenic enhancer. Thus, cross-regulatory interactions between *Dlx* genes are important to establish their overall expression in the forebrain.

Similar to the *Dlx5/Dlx6* intergenic region, five conserved non-coding elements (size between ~180 and 358 bp, >82% conservation) were identified in the intergenic region of *Dlx3/Dlx7* in both mouse and human (Sumiyama et al., 2002). One of these elements shows similarity with an element in the zebrafish *dlx3/dlx7* intergenic domain (M. Ekker, unpublished data). Three additional elements were found conserved (~85%) between mouse and human: two were located in the 5'flanking region upstream of *Dlx3* and one upstream of *Dlx7* (Sumiyama et al., 2002). A 79kb reporter construct containing all the above conserved regions, with *lacZ* inserted into the first exon of *Dlx3* gene was tested in transgenic mice. It targeted strong *lacZ* expression in the limb buds and first and second visceral arches at E9.5 and E10.5, consistent with the endogenous *Dlx3* expression pattern. A shorter construct (19kb) with only two elements on each side of the *Dlx3-lacZ* cassette directed expression in the limb buds but not in visceral arches. This

suggests differential regulation of *Dlx3,* and, possibly *Dlx7,* by distinct regulatory element(s).

Transcriptional enhancers of less than 1kb located upstream of the mouse *Dlx3* (Park and Morasso, 1999) and its *Xenopus* ortholog (*Xdll2*) (Morasso et al., 1994; Morasso et al., 1995) have been identified and drive expression in the surface ectoderm (epidermis). A 3.8kb fragment of the mouse *Dlx2* 5'-flanking sequence was shown to confer expression of a reporter transgene to a subset of epithelial cells in both the mandibular and maxillary components of the first branchial arch, with patterns identical to endogenous *Dlx2* expression (Thomas et al., 2000). In addition, *Dlx3* can activate transcription from a 1.7kb fragment of the *dlx5a* 5'flanking region in co-transfection assays suggesting cross-regulatory interactions between *Dlx* genes (Zerucha et al., 1997).

1.8.2. Upstream regulators of *Dlx* genes

Signaling molecules and transcription factors that regulate *Dlx* gene expression are just beginning to be elucidated. Gain of function experiments revealed that SHH can induce *Dlx* expression in the forebrain (Gaiano et al., 1999). Moreover, *Dlx2* expression is greatly reduced in mice lacking Shh (Ohkubo Y, Yun K, Rubenstein J, unpublished data). RA was among the first factors that was implicated in the induction of *Dlx* gene expression although its role is still debated and it is not clear yet whether it is the result of a direct effect or not. Human *DLX5* and *DLX6* expression is transiently induced 2.5 days after treatment of embryonic carcinoma cells (NT2/D1) with RA (Simeone et al., 1994). In contrast, treatment of zebrafish embryos with RA, prior to or during cranial neural crest migration, caused a rapid decrease in transcript levels of several *dlx* genes. This effect was correlated with craniofacial dysmorphologies (Ellies et al., 1997a). Several members of the BMP family of signaling molecules were shown to regulate the expression of *Dlx* genes in different organs. For instance, BMP4 can induce *Dlx5* expression in osteoblasts (Miyama et al., 1999) and *Dlx1/Dlx2* expression in dental mesenchyme (Bei and Maas, 1998), as well as *Dlx3* expression in embryonic ectoderm (Feledy et al., 1999a). BMP induction of *Dlx3* in the embryonic ectoderm is not direct

(Feledy et al., 1999a). It is negatively regulated by β-catenin in early *Xenopus* embryos, either as the result of the downregulation of BMP4 expression or through a more direct interaction (Beanan and Sargent, 2000). BMP2 can induce *Dlx2* expression in chondrocytes (Xu et al., 2001). In the chick, BMP2 and BMP4 play a role, along with Hensen's node and the neural plate, in the induction of *Dlx5* expression in the prospective non-neural epidermis (Pera et al., 1999). Similarly, FGFs can maintain or induce *Dlx* expression. FGF8 induces *Dlx1* and *Dlx2* expression in murine dental mesenchyme (Bei and Maas, 1998) and *Dlx1* expression in mandibular and hyoid branchial arches in chicken (Shigetani et al., 2002). FGF8 negatively regulated the expression, in the first branchial arch epithelium, of a reporter construct carrying a 3.8kb fragment of the mouse *Dlx2* 5'flanking region (Thomas et al., 2000). In contrast, BMP4 had a positive effect (Thomas et al., 2000) on the same construct. In addition, FGF2 maintains *Dlx3* expression in axolotl limb ectoderm (Mullen et al., 1996) and can induce *Dlx5* expression in the nascent chick limb (Ferrari et al., 1999). *Dlx5* expression can also be induced by FGF19 in combination with Wnt8c, in the developing inner ear (Ladher et al., 2000).

A few transcription factors have also been implicated in the regulation of *Dlx* gene expression. *Msx1*, is required to maintain *Dlx2* (but not *Dlx1*) expression in the branchial arch mesenchyme (Bei and Maas, 1998). Ectopic expression of forebrain specific zinc finger *Fez1* in zebrafish can induce *Dlx* expression (Yang et al., 2001). Furthermore, there is evidence that ectopic expression of *Mash1* in the cerebral cortex induces ectopic expression of *Dlx1* and *GAD67* (Fode et al., 2000; Stuhmer et al., 2002a). This suggests that the *Mash1/Dlx/GAD* cascade may be implicated in the regulation of GABA expression.

1.8.3. Downstream targets of *Dlx* genes

Recent studies showed that *Dlx* genes act as transcription factors, mainly transcriptional activators. They can regulate the expression of a variety of genes identified as their targets including the *Dlx* genes themselves. Little is known about the *Dlx* targets in the forebrain; however, there is enough evidence that links *Dlx* function to

the expression of GABA. Ectopic expression of *Dlx2* and *Dlx5* in slice cultures of the mouse embryonic cerebral cortex *in vitro* using an electroporation method induced the expression of glutamic acid decarboxylase (GAD65 and GAD67), the enzymes that synthesize GABA (Stuhmer et al., 2002a). Furthermore, *Dlx* expression vectors can induce expression from an enhancer element isolated from *GAD65* in co-transfection assays (Condie B, unpublished). These results are particularly interesting because they suggest that the *Dlx* genes are important regulators, not only of the differentiation and migration of GABAergic neurons, but also of the expression of *GAD* genes. In addition, as discussed above, *Dlx5* and *Dlx6* are downstream targets of *Dlx1* and/or *Dlx2* in the forebrain (Anderson et al., 1997b; Stuhmer et al., 2002a; Zerucha et al., 2000).

In other organs, several *Dlx* genes have been implicated in the regulation of one or more genes, either directly or indirectly [for review; (Panganiban and Rubenstein, 2002)]. The Dlx2 protein is able to bind a specific sequence (HBS-1) in an enhancer of the *Wnt-1* gene (Iler et al., 1995). Human DLX3 activates directly a human chorionic gonadotropin subunit in the placenta (Roberson et al., 2001) and profilaggrin in differentiating keratinocytes (Morasso et al., 1996). The binding sites that mediate each activation were identified and share both a TAAT core with the recognition sites for other Dlx (and other homeodomain) proteins. Human DLX4 (previously DLX7) activates both *GATA1* and *MYC* in hematopoeitic cells (Shimamoto et al., 1997). Three isoforms of DLX4 have been identified and shown to differentially repress β-globin transcription by binding to identical sequences found in its silencer elements (Chase et al., 2002; Fu et al., 2001). Several targets of *Dlx5* have been identified during bone formation including osteocalcin (Newberry et al., 1998; Ryoo et al., 1997), collagen 1A1 (Dodig et al., 1996) and bone sialoprotein (Benson et al., 2000). The *Dlx5*-binding sites in the regulatory regions of these genes have been characterized (Panganiban and Rubenstein, 2002).

Besides binding to DNA, Dlx proteins may regulate transcription indirectly through other molecular mechanisms. It has been suggested that they are able to form homodimers as well as heterodimers with other Dlx proteins and members of the Msx family of homeodomain proteins (Zhang et al., 1997). These dimerizations are mediated

by the homedomains of both these families. Thus, *Dlx* proteins may act as transcriptional activators to antagonize the repression mediated by *Msx*. This interaction between Dlx and Msx proteins occurs without a requirement for the complex to bind DNA [for review on Dlx and Msx homeoproteins, see (Bendall and Abate-Shen, 2000)].

1.8.4. Consensus sequence and motifs bound by *Dlx* genes

Some of the *Dlx* binding motifs have been characterized. Feledy et al. 1999b identified a consensus DNA sequence for the *Xenopus* Dlx3 protein (Xdll2), (A/C/G) TAATT (G/A) (C/G), that is also recognized by other Dlx proteins *in vitro* (Feledy et al., 1999b). They used binding site selection from a random oligonucleotide pool. Interestingly, this consensus sequence was present twice in the *Dlx* forebrain-enhancer, zI56i (and mI56i), and was bound by Dlx proteins both in *vitro* and *in vivo* (Zerucha et al., 2000). Thus, zebrafish *dlx1, 2, 3, 4* and *6*, and mouse *Dlx1, 2*, and *5* were capable of activating transcription from this intergenic enhancer in at least two different cell lines and in slice culture (Zerucha et al., 2000; Stuhmer et al., 2002a). DNA-binding assays and mutagenesis further demonstrated that the Dlx interaction with zI46i was directly mediated by the two Dlx consensus binding sites (Feledy et al., 1999b; Zerucha et al., 2000). In addition, transgene expression directed by the same enhancer in the forebrain was nearly completely abolished in *Dlx1/Dlx2* mutant mice (Zerucha et al., 2000). This was similar to the marked reduction in *Dlx5* and *Dlx6* expression in these mutants (Anderson et al., 1997b). It suggests that *Dlx1* and/or *Dlx2* regulate *Dlx5/Dlx6* expression through direct interaction with this intergenic enhancer *in vivo*. Indeed, using an optimized Chromatin Immuno-Precipitation procedure (ChIP) and specific polyclonal antibodies to DLX1 and DLX2, Zhou et al. (2004), demonstrated that mI56i was among several DLX1 and DLX2 target genomic sequences that were isolated from E13.5 embryonic forebrain (ganglionic eminences) nuclear extracts. Reporter gene assays demonstrated the functional significance of the binding of the two DLX proteins to this regulatory element, which was further confirmed *in vitro* by electrophoretic mobility shift assays, using tissue extracts or recombinant DLX proteins (Zhou et al., 2004). Besides

the consensus sequence mentioned above, several TAAT/ATTA homeodomain binding motifs were found in the two *Dlx*-forebrain enhancers (Zerucha et al., 2000; Zhou et al., 2004). These motifs may also participate in the recruitment of Dlx proteins and/or other homeodomain proteins.

1.9. Genomic regulatory regions and comparative genomics

1.9.1. Genomic regulatory regions

After sequencing of the human and other vertebrate genomes such as mouse, rat, zebrafish and *Fugu rubripes* (pufferfish) [(Aparicio et al., 2002; Lander et al., 2001; Waterston et al., 2002), Sanger Institute, http://www.sanger.ac.uk/Projects/D_rerio/; Rat Genome Sequencing Project Consortium, 2004)], one of the grand challenges for genomics research in the near future is the comprehensive discovery and the annotation of all functional elements that constitute these genomes including the regulatory elements.

1.9.1.1. Functional elements

The functional elements of the human genome (or any other genome) can be classified, in general, into coding and non-coding regions. These elements contribute to various functions necessary for an organism to progress through life and reproduction. Most mobile elements and pseudogenes do not fit this definition. Although coding regions constitute 1-2% of the human genome (Lander et al., 2001), the minimal amount of the human genome estimated to be under evolutionary constraint or selection pressure is ~5% (Ovcharenko et al., 2004; Waterston et al., 2002). Thus the full complement of protein-coding sequences still remains to be established. Assuming that the presence of constraint is generally correlated with the presence of function (Hardison, 2000), one can conclude that a substantial fraction (~2-3%) of the human genome consists of non-coding functional elements. Although some of these may be non-coding RNA (Eddy, 2002), a larger number are likely to be regulatory elements with roles in controlling gene

expression and chromatin organization, among other functions. Even less is known about the function of roughly half of the genome that consists of highly repetitive sequences or of the remaining non-coding, non-repetitive DNA.

1.9.1.2. Regulatory elements

Regulatory elements can be divided into two main classes: 1) those that control gene expression including promoters, enhancers and silencers; and, 2) those that function in chromatin organization such as insulator elements and matrix attachment regions (MARs) (Arnone and Davidson, 1997; Pennacchio and Rubin, 2001).

A *promoter* is the segment of DNA located immediately proximal to the transcription start site of a gene that is required for initiating transcription. Promoters in in mammals contain different combination of elements, such as TATA boxes, CAAT boxes, GC boxes, an octamer element, that contribute to their function, but none is essential for all promoters. These promoter elements play a role in positioning and/or recruitment of the transcription machinery (RNA polymerase II and upstream transcription factors), and/or initation of transcription as well as strength and efficiency of the promoter (Lewin, 1997). Promoter elements are usually located within ~100bp upstream of the start point (but sometimes more distant), thus, it is relatively easy to predict the approximate locations of proximal promoters. Indeed, a prediction of the promoters for 10,000 human genes was made and nearly 90% of a selected group of predicted promoters (154 segments) were found to function in several cell types (Trinklein et al., 2003).

Enhancers are regulatory elements that upregulate gene expression by recruitment and sequence-specific positioning of transcriptional activators. They can be located within hundreds (sometimes thousands) of kilobases either upstream or downstream of their target genes. They generally function independently of position or orientation (Arnone and Davidson, 1997). Unlike enhancers, *silencers* repress transcription and are generally found near the promoter of their target genes (Ogbourne and Antalis, 1998; Smith et al., 2006).

Insulator elements are barriers that separate domains within chromatin and confine the action of regulatory elements to their appropriate targets. Besides blocking the action of enhancers, they can also prevent the spread of chromatin condensation from nearby regions [(Burgess-Beusse et al., 2002); for recent review, see (Brasset and Vaury, 2005)].

Matrix attachment regions (*MARs*) may mediate binding to the nuclear matrix and may have important roles in the higher-order organization of eukaryotic nuclei (Cremer and Cremer, 2001). A significant fraction of conserved non-coding DNA (11%) in human and mouse consists of predicted MARs (Glazko et al., 2003). Note that some regulatory elements can combine more than one of the above functions.

Unlike promoters, identification and characterization of regulatory elements such as enhancers, silencers and insulators is more challenging because of their unpredictable locations in the genome.

1.9.2. Comparative genomics

Comparative genomics is a powerful approach to uncover functional elements in the human genome. It is based on the assumption that functionally significant parts of genomic sequences evolve more slowly than their non-functional neighborhood, owing to selective pressure (Kimura, 1983). Furthermore, mutations in functional elements are likely to be deleterious and will be selected against (Kimura, 1983), resulting in a reduced rate of evolution (Hardison, 2000; Sidow, 2002; Sumiyama et al., 2001). Conventional methods used to identify functional elements such as classical deletion mapping and functional analyses are time- and money-consuming, and are restricted to short genomic sequences. In contrast, comparative genomics can be used as a large-scale approach that is quick and practical for the annotation of functional elements based on comparing two or more orthologous noncoding sequences.

There are two variants within this approach depending on the set of species analyzed: '*phylogenetic footprinting*' (1) and '*phylogenetic shadowing*' (2). The first variant (1) is used when comparing non-coding sequences from distantly related species e.g. human/mouse/chicken or human/mouse/fish whereas the second (2) is for sequence

comparison among closely related species e.g. primates. '*Phylogenetic shadowing*' was used successfully to identify non-coding RNAs in nematodes (Lim et al., 2003) as well as *cis*-regulatory elements in yeast (Kellis et al., 2003) and primates (Boffelli et al., 2003). Similarly, '*phylogenetic footprinting*' allowed the identification of *cis*-regulatory elements between human, and, primates (Boffelli et al., 2003); mouse (Lee et al., 2004; Loots et al., 2000) as well as chicken (Ahituv et al., 2005; Dodou et al., 2003; Lien et al., 2002; Tumpel et al., 2002) and fish (Bagheri-Fam et al., 2001; Santagati et al., 2003; Shin et al., 2005).

1.9.2.1. *'Phylogenetic footprinting'*

While intra-primate sequence comparisons have allowed the identification of primate-specific regulatory elements including common elements with mammals (Boffelli et al., 2003), human to fish comparisons highlight regulatory elements that are conserved over long evolutionary periods and therefore, are probably shared by all vertebrates. Furthermore, genome sequence alignments between human and rodent species (which diverged ~70 Mya) indicate extensive regions of similarity. Whole genome comparisons between human and mouse have found that 40% of these genomes can be aligned with each other, but only 5% of these genomes are under active selection (Waterston et al., 2002). These data suggest that evolutionary divergence between human and rodents is insufficient to allow resolution of conserved functional DNA from similar, yet non-functional sequences. Thus, the amount of divergence affects the power and resolution of these analyses. As a result, increased sequence divergence can improve the resolution at which features can be discovered and linked to one common ancestor. Indeed, *cis*-regulatory regions were identified after comparing human sequences to more distant species like birds (Dodou et al., 2003; Lien et al., 2002; Tumpel et al., 2002), amphibians (Ahituv et al., 2005) and fish (Bagheri-Fam et al., 2001; Santagati et al., 2003; Shin et al., 2005). Of note, the most remote vertebrate genomes that have been compared to date are those of human and *Fugu rubripes* (pufferfish) (Barton et al., 2001; Nobrega et al., 2003; Ovcharenko et al., 2004; Spitz et al., 2003; Woolfe et al., 2005). In

fact, the latter species offer several advantages compared with other distant vertebrate species: 1) *Fugu rubripes* has the smallest recorded vertebrate genome size (400Mb) (Brenner et al., 1993) which is one-eighth the size of the human genome. This compact size is attributed to the paucity of repetitive sequences, and to short introns and intergenic regions; 2) despite this compactness, the Fugu genome has a similar gene repertoire to mammals (Elgar et al., 1996) and critically, a very similar synteny; 3) the evolutionary distance between mammals and Fugu is estimated to 450 Mya (Kumar and Hedges, 1998); 4) transgenic studies in rodents demonstrated that many of the conserved Fugu non-coding sequences are functional *cis*-regulatory elements involved in tissue- and developmental-stage-specific expression [for review on comparative genomics using Fugu, see (Elgar, 2004; Venkatesh and Yap, 2005)].

In sum, comparative analysis using '*phylogenetic footprinting*' is a very powerful technique to identify genomic regions under evolutionary constraint. However, it should be noted that this method may not be applicable across all evolutionary distances and for all types of genes. Recent results from vertebrate comparisons indicate that strong conservation of *cis*-regulatory regions may occur more frequently in developmental regulator genes [for review on this topic, see (Dickmeis and Muller, 2005); for recent review on the study of vertebrate *cis*-regulatory sequences, see (Gomez-Skarmeta et al., 2006)].

Statement of Inquiry

GABAergic interneurons constitute between 10-30% of telencephalic neurons. They are regulators of principal neuron activity and comprise several subtypes that could be required for common as well as distinct cortical and subcortical functions. Nearly all GABAergic cortical interneurons in mouse and about half of those found in human were shown to derive by tangential migration from distinct ventral progenitors born in the basal ganglia during embryonic development. Yet, little is known about the molecular pathways including the genes and signaling molecules that control the development of these interneuron progenitors. The *Dlx* genes are among few identified genes that were shown to be indispensable for normal differentiation and migration of interneuron progenitors. Consequently, lack of normal *Dlx* function in the basal ganglia causes loss of >80% of cortical, hippocampal and olfactory bulb interneurons in mouse due to premature block of differentiation and lack of migration of mature interneurons. Dlx function in the ventral telencephalon is mediated by the expression of two pairs of *Dlx* genes: *Dlx1/Dlx2 and Dlx5/Dlx6*. These four genes display highly overlapping expression patterns that could explain their partial functional redundancy. Yet, they also exhibit distinct telencephalic function(s) and cross-interactions exist between them.

Our understanding of *Dlx* gene regulation and the relevance of Dlx function to the development of inhibitory interneurons is still limited in many ways: 1) the mechanisms of regulation including regulatory elements that control *Dlx1* and *Dlx2* expression in the forebrain of vertebrates are still unknown, and therefore, 2) it is not clear whether *Dlx1/Dlx2* and *Dlx5/Dlx6* bigene clusters possess evolutionary common and/or distinct mechanisms of regulation that control and refine their nested patterns of expression, 3) little is known about the biological diversity of *Dlx*-progenitors born in different ventral structures such as LGE, MGE and CGE, and how these progenitors give rise to the variety of interneuron subtypes found in the mature cortex, finally, 4) what controls the molecular properties of the *Dlx*-progenitors born in each subdivision of the basal ganglia.

I have addressed in the present doctoral project some of the above questions. In the study done in chapter I, I have specifically tackled the first and second questions. Thus, I have studied and compared, in collaboration with Olga Jarinova, the gene regulation of *Dlx1/Dlx2* and *Dlx5/Dlx6* bigene clusters in vertebrates. We have performed a '*phylogenetic footprinting*' among five vertebrate species in a search for conserved non-coding *cis*-regulatory elements (CREs) in the intergenic region of each cluster, separately. As a result, I have identified a novel *Dlx* enhancer -I12b- located in the intergenic region of *Dlx1/Dlx2* in vertebrates that was active in the forebrain of transgenic mice. We also found the two previously characterized forebrain enhancers, I56i and I56ii, conserved in the *Dlx5/Dlx6* locus of the vertebrate species studied here. In the study performed in chapter II, I tackled partially the third and fourth questions with the purpose of understanding the biological diversity of *Dlx*-progenitors born in the subpallium and the variety of cortical interneurons derived from these progneitors. First, I have described the identification of an additional forebrain-specific enhancer -URE2- located in the *Dlx1* 5' flanking region of four vertebrate species. Second, I have compared the activities of URE2, and, two of the forebrain-specific enhancers described above, I12b and I56i, in the subpallial telencephalon of transgenic mice during midgestation. Consequently, I have suggested the existence of a potential relationship between the distinct spatio-temporal activities of the three CREs in *Dlx*-progenitors during midgestation and the development of various types of cortical interneurons in the adult mouse brain. Finally, in the last study described in chapter III, I have conducted a detailed spatial and temporal comparison of the regulatory activities of all four *Dlx*-CREs in the forebrain of transgenic mice throughout development. The regional and laminar activities of the enhancers were also compared with the endogenous *Dlx* expression during midgestation.

2. Regulatory Roles of Conserved Intergenic Domains in Vertebrate *Dlx* Bigene Clusters

Noël Ghanem[1, 2,*], Olga Jarinova[1, 2,*], Angel Amores[3], Qiaoming Long[1,2,5], Gary Hatch [1], Byung Keon Park[1,6], John L.R. Rubenstein[4], and Marc Ekker[1, 2,7.]

[1]Ottawa Health Research Institute; [2]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, K1Y 4E9; [3]Institute of Neuroscience, University of Oregon, Eugene, OR97403, USA. [4]Nina Ireland Laboratory of Developmental Neurobiology, Center for Neurobiology and Psychiatry, Department of Psychiatry and Programs in Neuroscience, Developmental Biology and Biomedical Sciences, University of California at San Francisco, CA 94143-0984.

[5]Present address: Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN 37232, USA.

[6]Present address: Dept. of Oral anatomy, School of Dentistry, Chonbuk National University, Chonju, Republic of Korea.

[7]Corresponding author.

* These authors contributed equally to this work.

E-MAIL: mekker@uottawa.ca

Running title: Intergenic *Dlx* enhancers.

Keywords: Forebrain, branchial arches, enhancers, gene duplication, *cis*-acting sequences, *Danio rerio*, *Spheroides nephelus*, *Takifugu rubripes*

Abstract

Dlx homeobox genes of vertebrates are generally arranged as three bigene clusters on distinct chromosomes. The *Dlx1/Dlx2*, *Dlx5/Dlx6*, and *Dlx3/Dlx7* clusters likely originate from duplications of an ancestral *Dlx* gene pair. Overlaps in expression are often observed between genes from the different clusters. To determine if the overlaps are due to the conservation of enhancer sequences between paralogous clusters, we compared the *Dlx1/Dlx2* and the *Dlx5/Dlx6* intergenic regions from human, mouse, zebrafish and from two pufferfish, *Spheroides nephelus* and *Takifugu rubripes*. Conservation between all five vertebrates is limited to four sequences, two in *Dlx1/Dlx2* and two in *Dlx5/Dlx6*. These non-coding sequences are more than 75% identical over a few hundred base pairs, even in distant vertebrates. However, when compared to each other, the four intergenic sequences show a much more limited similarity. Each intergenic sequence acts as an enhancer when tested in transgenic animals. Three of them are active in the forebrain with overlapping patterns despite their limited sequence similarity. The lack of sequence similarity between paralogous intergenic regions and the high degree of sequence conservation of orthologous enhancers suggest a rapid divergence of *Dlx* intergenic regions early in chordate/vertebrate evolution followed by fixation of *cis*-acting regulatory elements.

2.1. Introduction

Vertebrates possess anatomical features not seen in their closest living invertebrate relatives, the protochordates such as tunicates or cephalochordates. Genetic changes, such as the evolution of new regulatory pathways, may have permitted the origin of these innovations. Gene duplication followed by functional divergence of paralogs constitutes a major mechanism that permits such changes. An important contribution to the evolutionary divergence of paralogs may be through changes in mechanisms that control gene expression *via cis*-acting regulatory sequences in the non-coding region of genes.

However, the identification of *cis*-acting regulatory elements remains challenging, even after the completion of a few vertebrate genome sequences.

The vertebrate *Dlx* genes, which encode a family of homeobox-containing transcription factors related in sequence to the *Drosophila Distal-less* (*Dll*) gene product, constitute one example of functional diversification of paralogs. All vertebrates investigated thus far have at least six *Dlx* genes that are generally arranged as three bigene clusters: *Dlx1/Dlx2*, *Dlx5/Dlx6*, and *Dlx3/Dlx7* (Ellies et al., 1997b; Liu et al., 1997; McGuinness et al., 1996; Nakamura et al., 1996; Simeone et al., 1994; Stock et al., 1996). Each bigene cluster is localized on a distinct chromosome that also contains one of the *Hox* clusters, suggesting that the duplication events that generated the multiple *Dlx* bigene clusters of vertebrates also involved the *Hox* genes (Amores et al., 1998; Stock et al., 1996). The two linked *Dlx* genes are in an inverted configuration and separated by a short intergenic (3.5- 16 kb) region. Since only one *Dll*-like gene is found in invertebrates such as *Drosophila* and *C. elegans*, the multiple vertebrate *Dlx* genes are thought to have arisen as a result of tandem gene duplication events from one "hypothetical" common ancestor to nematodes, arthropods and vertebrates. The presence, in the tunicate *C. intestinalis*, of a pair of *Dll*-like genes with an organization similar to that of the vertebrate *Dlx* genes (Caracciolo et al., 2000; Di Gregorio et al., 1995) supports the hypothesis that the initial duplication predated the existence of vertebrates.

Gene families such as the *Dlx* family provide attractive models for studying gene regulation and functional divergence between paralogs. The bigene cluster arrangement of *Dlx* genes is conserved amongst distant vertebrates and a direct association is seen between the genomic organization of the genes and their expression patterns in different species (Ellies et al., 1997b; Zerucha et al., 2000) suggesting that the mechanisms of regulation might have been conserved, at least in part. Functional conservation among different orthologs, as inferred from comparative expression patterns seems to be applicable to most vertebrate *Dlx* genes (Quint et al., 2000; Zerucha et al., 2000). Partial functional redundancy between *Dlx* paralogs is suggested by the overlapping gene expression patterns and phenotypes of mice with targeted *Dlx* mutations (Acampora et

al., 1999; Anderson et al., 1997b; Depew et al., 1999; Qiu et al., 1997; Qiu et al., 1995; Robledo et al., 2002). Sharing of *cis*-regulatory elements between members of a *Dlx* bigene cluster may contribute to the overlap in gene expression and to their partial functional redundancy.

Consistent with a model of enhancer-sharing, two highly conserved enhancer elements, I56i and I56ii, were identified in the intergenic region of the *Dlx5/Dlx6* genes of zebrafish, mouse and human and were able to target expression of reporter transgenes to the forebrain of both mouse and zebrafish in patterns that mimic endogenous gene expression (Zerucha et al., 2000). Recently, Sumiyama and collaborators conducted a comparative sequence analysis of the mouse and human *Dlx3/Dlx7* (*Dlx3/Dlx4* was suggested as revised nomenclature by Panganiban and Rubenstein, 2002) bigene cluster (Sumiyama et al., 2002). Conserved sequences were identified both in the coding and non-coding regions of *Dlx3/Dlx7*. Comparisons of the two mammalian loci with the orthologous *dlx3/dl*x7 bigene cluster from zebrafish revealed a much more limited similarity (Sumiyama et al., 2002).

The two genes from the *Dlx1/Dlx2* cluster are expressed in the developing forebrain with patterns that overlap partially with those of *Dlx5* and *Dlx6*. As the *Dlx1/Dlx2* and *Dlx5/Dlx6* bigene clusters probably originate from the duplication of an ancestral cluster, the forebrain expression of *Dlx1* and *Dlx2* could be attributable to enhancer sequences related to I56i and/or I56ii. To address this possibility and to get a comprehensive understanding of *cis*-acting regulatory elements in the *Dlx1/Dlx2* and *Dlx5/Dlx6* intergenic regions, we have performed a homology search ('*phylogenetic footprinting*') between the intergenic regions of the two bigene clusters from five vertebrate species: human, mouse, zebrafish, *Takifugu rubripes* (formerly *Fugu rubripes*) and *Spheroides nephelus*. Sequence conservation between all five species is limited to four distinct sequences of a few hundred base pairs, two in each intergenic region. Each sequence shows enhancer activity in transgenic mice and/or zebrafish. A novel forebrain enhancer, I12b, was identified in the *Dlx1/Dlx2* intergenic region, but surprisingly, it shows almost no sequence similarity to the I56i and I56ii forebrain enhancers, suggesting

that highly overlapping patterns of expression but with subtle differences can be conferred by highly different *cis*-acting regulatory sequences.

2.2. Material and methods

2.2.1. *Dlx* gene nomenclature

In order to help standardize the nomenclature for vertebrate *Dlx* genes, we found it useful to adopt that recently suggested by Panganiban and Rubenstein (Panganiban and Rubenstein, 2002). As the *Dlx* genes are found in regions of conserved synteny that contain the *Hox* clusters, the new nomenclature is aligned with that of the zebrafish *hox* clusters (Amores et al., 1998). Thus, the zebrafish gene we refer to as *dlx5a* in this study is the gene previously named *dlx4* (Akimenko et al., 1994). Similarly, the zebrafish gene previously named *dlx5* is renamed *dlx2b* as it is a *dlx2* duplicate (see Discussion). The previous *dlx1*, *dlx2*, and *dlx6* genes are renamed *dlx1a, dlx2a*, and *dlx6a*, respectively. Finally, the previous *dlx3*, *dlx7*, and *dlx8* genes of zebrafish would be renamed, *dlx3b*, *dlx4b*, and *dlx4a*, respectively. We kept the *Dlx3/Dlx7* nomenclature for the mouse genes throughout the current report for the sake of simplicity but indicated the suggested name change.

2.2.2. Isolation and characterization of *Dlx* genes from *Spheroides nephelus*

Clones from a PAC library (Amemiya et al., 2001) were screened using a PCR approach for a conserved region of *Dlx* genes (Stock et al., 1996). Two degenerate oligonucleotide primers located in the zebrafish *dlx2* homeodomain were used to amplify the homeodomain including intron B of *dlx* positive clones: 5'GGAATTCAA(AG)CC(ACGT)(AC)G(ACGT)AC(ACGT)AT(ACT)TA 3' and 5' CGGGATCC(AG)AACCA(AGT)AT(CT)TT(ACGT)AC(CT)TG 3'. Amplification consisted of five cycles at an annealing temperature of 45°C for 45s with a two minute and 15 seconde ramp to 72°C, followed by 25 step cycles at an annealing temperature of 55°C for 45s. The PCR fragments (300-400bp) were sequenced using the 3100 DNA

Sequencer from Applied Biosystems and preliminary orthology assignment was established. Identity of *dlx1a/dlx2a* positive clones (3 clones) was confirmed by amplifying a 190bp highly conserved fragment located in the I12b sequence using the following oligonucleotide primers (zebrafish): 5' CAGTTGAGCATTCTGGCT 3' and 5' GCTCTAGATGTTATGCTAAAA 3'. Amplification consisted of 30 cycles at an annealing temperature of 45°C for 45s. Identity of *dlx5a/dlx6a* positive clones (5 clones) was confirmed by amplifying a 129bp highly conserved fragment located in the I56i sequence using the following oligonucleotide primers (zebrafish): 5'GGGGTACCATTCTCATAAATG 3' and 5' GGGGTACCAAATAAAGAT 3'. Amplification consisted of 30 cycles at an annealing temperature of 55°C for 45s. Genomic fragments comprising exon 3 of positive *dlx* clones plus the intergenic region between *dlx* genes were sequenced later on using primer walking and the shotgun method. Sequencing was performed using the 3100 DNA Sequencer from Applied Biosystems.

2.2.3. Sequence analysis

The zebrafish, mouse and *Spheroides* intergenic sequences were determined from previously isolated genomic clones (Depew et al., 1999; Ellies et al., 1997b; McGuinness et al., 1996) or from the *Spheroides* clones described in the above paragraph. The sequences are deposited in Genbank under accession numbers: AY168007-AY168012. The sequences from human and *Takifugu rubripes* were obtained from public databases: Human *Dlx1/Dlx2*, Genbank accession number NT_005332.9; Human *Dlx5/Dlx6*, Genbank accession number NT_033964.1; *Takifugu dlx1/dlx2*, scaffold 21, position 120318 to 125668, *Takifugu dlx5/dlx6*, scaffold 3932, position 6627 to 10192. For the Fugu Genome Consortium / JGI (DOE Joint Genome Institute), see http://www.jgi.doe.gov/index.html.

Pairwise sequence alignments are performed with PIPMAKER (available at http://bio.cse.psu.edu/pipmaker/), or with the BestFit, and Mapplot programs of the GCG

Wisconsin package. Multiple sequence alignments are performed with the Pileup and Clustal X programs.

2.2.4. Transgenic animals

For transgenic mice, sequences from the *Dlx* intergenic regions were subcloned into the p1229/p1230 vectors (Yee and Rigby, 1993) that contain a human β-globin minimal promoter and the *lacZ* reporter gene. The constructs that were used to generate the transgenic mice are: 1) Both ends of a 4.8kb -*Bgl*II- mouse intergenic fragment containing I12b were partially filled with nucleotides A and G for 1h at 37°C using Klenow fragment of DNA polymerase I, the fragment was then subcloned into p1230 using *Sal*I restriction sites that were partially filled with C and T nucleotides for site compatibility; 2) a 1.9kb -*Xba*I/*Eco*RI- mouse intergenic fragment containing I12a was subcloned into p1229 using the same restriction sites; 3) two *Eco*RI mouse intergenic fragments containing I56i (4.3kb) and I56ii (3.9kb) were subcloned into PCR2.1, separately, then cut and inserted into p1230 using *Kpn*I/*Hin*dIII restriction sites. Embryos from the mating of a transgenic male with wild type *CD1* females were harvested at E11.5. Pregnant females were sacrificed by cervical dislocation. E11.5 mouse embryos were fixed in 4% paraformaldehyde (PFA) for 1h at room temperature (RT), then washed in 1x PBS for 30 min and stained for β-galactosidase activity overnight at 28°C in a solution of 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, and 0.02% NP-40 in PBS.

For transgenic zebrafish, intergenic enhancer sequences were inserted into a sp72 modified plasmid (sp72-pEGFP-N1) containing the pEGFP-N1 reporter gene (size ~1.1kb inserted in *Sst*I sites) and placed downstream of a 3.5 kb fragment from the immediate 5'-flanking region of zebrafish *dlx6a,* including part of the 5'UTR (inserted in *Xba*I sites). This fragment by itself does not produce any tissue-specific expression in transgenic zebrafish (Figure 2.6. F). Besides the previous construct, two additional constructs were generated and injected into zebrafish fertilized eggs: a 1.4kb zebrafish intergenic fragment containing both I56i and I56ii and a 4kb mouse intergenic fragment

containing I56i were subcloned separately into sp72-pEGFP-N1 using *Kpn*I/*Cla*I restriction sites in a two-step subcloning.

2.3. Results

2.3.1. Genomic organization of *Dlx1/Dlx2* and *Dlx5/Dlx6* bigene clusters in two species of pufferfish

The genomic organization of two loci containing *Dlx* genes was examined in *Spheroides nephelus* and *Takifugu rubripes* and was compared to that of zebrafish, mouse and human. Initial orthology assignment was based on the sequence of the third exon of the genes, which contains part of the homeobox. Orthology was further confirmed by sequence analysis of the intergenic region. As previously described for zebrafish, mouse and human (McGuinness et al., 1996; Simeone et al., 1994; Ellies et al., 1997b; Zerucha et al., 2000), the *dlx1/dlx2* genes and the *dlx5/dlx6* genes of *Spheroides* and *Takifugu* are organized as two pairs of genes, both found in an inverted and convergent configuration (Figures 2.1. A and 2.2. A).

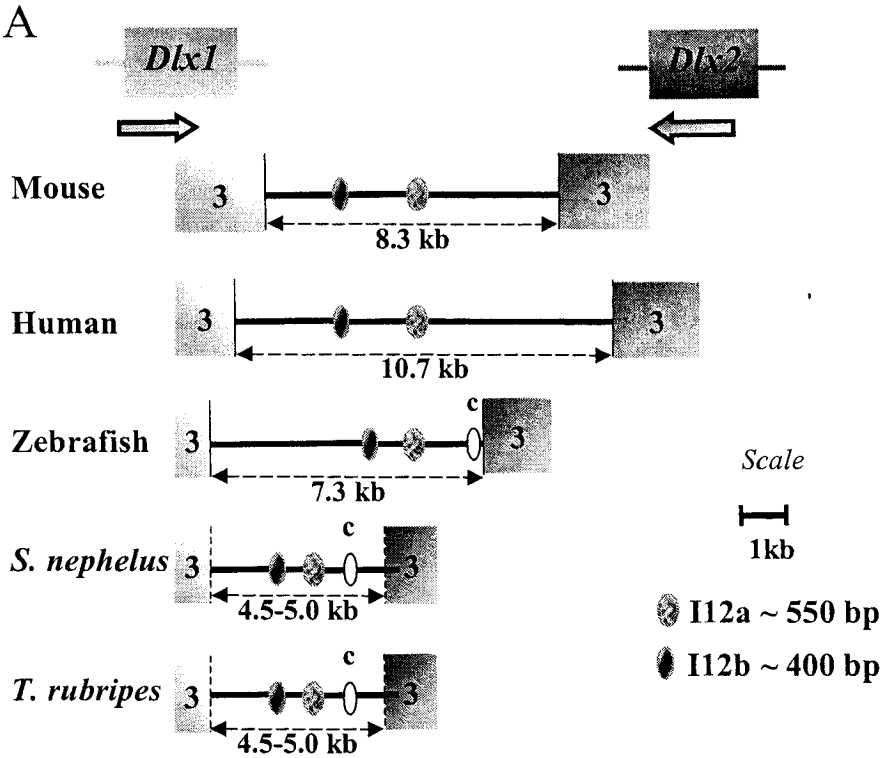
The size of the *Dlx1/Dlx2* intergenic region in the five species varies between about 4.5-5.0 kb for the two pufferfish to 10.7 kb for human (Figure 2.1. A). It was difficult to determine with precision the size of the pufferfish intergenic regions because no cDNA sequences are available for the *Dlx1* and *Dlx2* genes from these species and unequivocal polyadenylation signals were sometimes hard to find in the genomic sequence. The distance that separates the two stop codons is 5.3 kb in both species.

The size of the *Dlx5/Dlx6* intergenic region varied between 10 kb for mouse and human and about 3.0-3.5 kb for the three teleost fish (Figure 2.2. A). Thus despite the fact that the genome size for *Takifugu rubripes* and *Spheroides nephelus* is about 4 and 8 times smaller than those of the zebrafish and mouse/human, respectively, this is not reflected in proportionally smaller intergenic regions.

Figure 2.1. Conserved sequences in the *Dlx1/Dlx2* intergenic region

(A) Schematic representation of the *Dlx1/Dlx2* intergenic region of five vertebrate species. The third exons of the *Dlx* genes are indicated. The position of the polyadenylation sequence in the *Dlx* genes of *Spheroides* and *Takifugu* is an estimate. In addition to the I12a and I12b sequences, ovals labeled c represent a region of sequence conservation between the three teleost fish species. (B) Percentage identity for I12a and I12b in pairwise sequence comparisons.

A



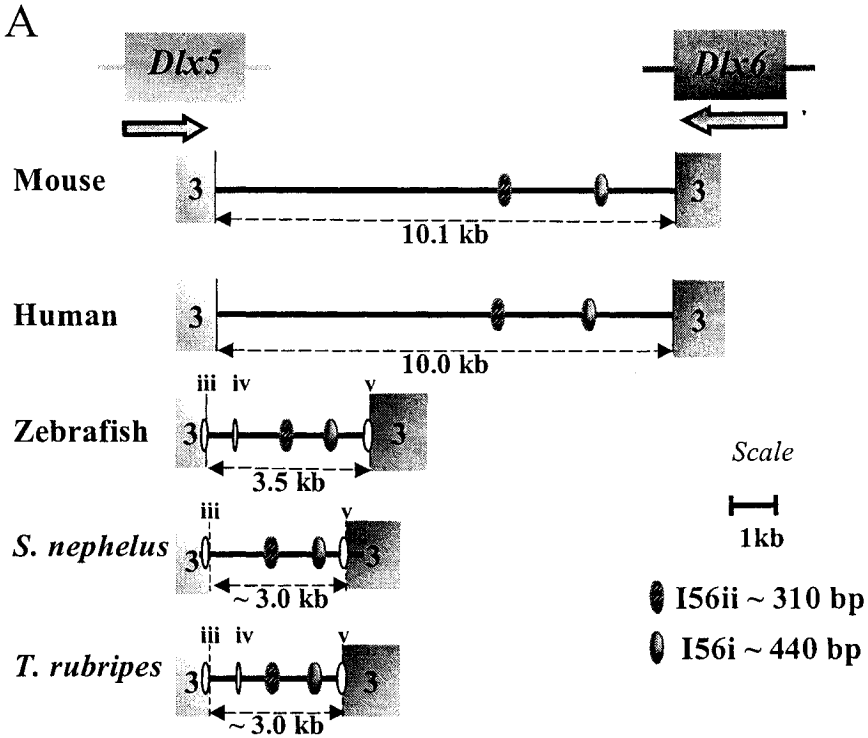
B

I12b% / I12a%	Mouse	Human	Zebrafish	*S. nephelus*	*T. rubripes*
Mouse		97	77	77	75
Human	99		78	77	75
Zebrafish	92	92		76	76
S. nephelus	86	86	87		90
T. rubripes	83	84	86	93	

Figure 2.2. Conserved sequences in the *Dlx5/Dlx6* intergenic region

(A) Schematic representation of the *Dlx5/Dlx6* intergenic region of five vertebrate species. The third exon of each *Dlx* gene is indicated. The position of the polyadenylation sequence in the *Dlx* genes of *Spheroides* and *Takifugu* is an estimate. In addition to the I56i and I56ii sequences, ovals labeled iii, iv, and v represent regions of sequence conservation between a subset of the five species. Sequence alignments can be found in Supplementary Figures. (B) Percentage identity for I56i and I56ii in pairwise sequence comparisons.

A



B

I56i% / I56ii%	Mouse	Human	Zebrafish	*S. nephelus*	*T. rubripes*
Mouse		99	81	82	82
Human	99		81	81	82
Zebrafish	84	84		87	88
S. nephelus	80	79	89		89
T. rubripes	81	82	90	95	

2.3.2. Sequence comparisons and identification of highly conserved non-coding sequence elements in the *Dlx* intergenic regions

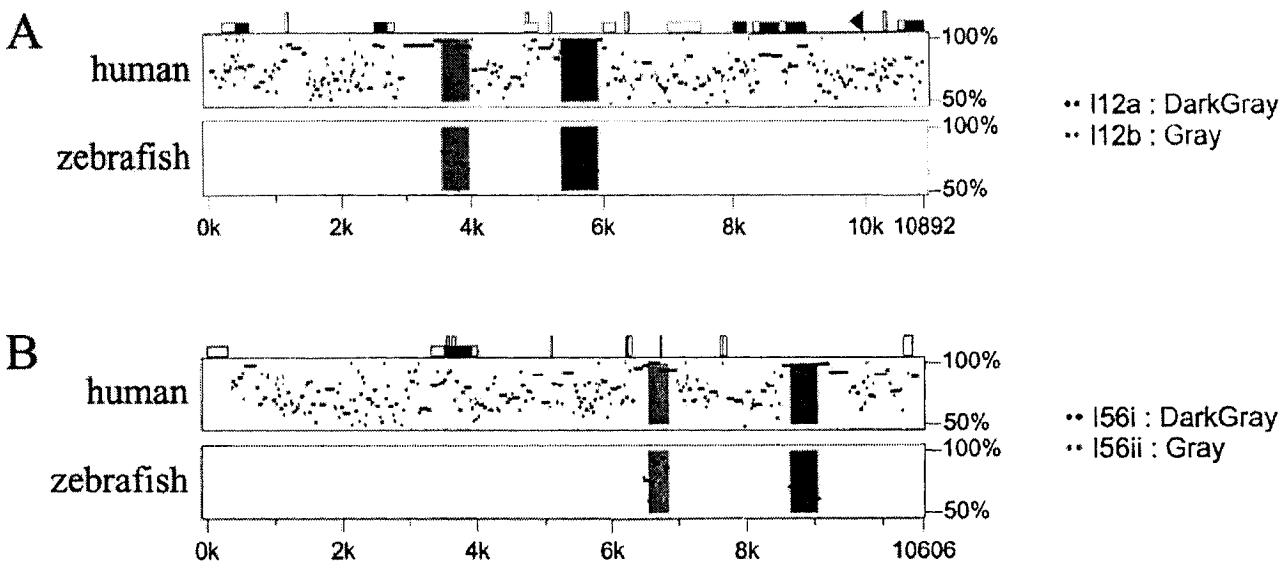
We examined the *Dlx1/Dlx2* and *Dlx5/Dlx6* intergenic regions of the five vertebrate species for conserved sequences. The mouse and human *Dlx1/Dlx2* intergenic regions were highly similar with 80% overall sequence identity (Figure 2.3. A). The same applies for the human *Dlx5/Dlx6* intergenic region (78%; Figure 2.3. B) and for the *dlx1/dlx2* and *dlx5/dlx6* intergenic regions of *Takifugu rubripes* and *Spheroides nephelus* with 85% and 87% sequence identity, respectively (data not shown). This reflects the relatively recent divergence from one common ancestor between mouse and human (approx. 60 million years), on the one hand, and between the two species of pufferfish, on the other hand (between 5-35 million years). Despite the high degree of sequence conservation between orthologous loci, the paralogous intergenic regions, *Dlx1/Dlx2* and *Dlx5/Dlx6,* do not show any striking sequence similarity and no large regions of sequence similarity can be found between the intergenic sequence separating *Dlx3* and *Dlx7* of human, mouse and zebrafish (Sumiyama et al., 2002).

Two highly conserved sequences that were previously identified in the *Dlx5/Dlx6* intergenic region of zebrafish, mouse, and human (Zerucha et al., 2000), I56i and I56ii, were also found in the *dlx5/dlx6* intergenic regions of *Takifugu* and *Spheroides*. They constitute the only two regions of high sequence similarity between all five species (Figures 2.2. A and 2.3. B). The sizes of I56i and I56ii are approximately 440 bp and 310 bp, respectively, and the identity percentages in pairwise comparisons vary between 81 and 99 percent (Figure 2.2. B; 5-species alignment provided as Supplementary Figures 2.1. and 2.2.). The relative positions and orientation of the I56i and I56ii sequences with respect to the flanking genes were identical for all five vertebrates. In both the mouse/human (Figure 2.3. B) and the *Takifugu/Spheroides* (data not shown) alignments, I56i and I56ii reside in a region of overall stronger sequence conservation.

In addition to I56i and I56ii, we found two sequences of 150-200 bp with higher than 80% identity between zebrafish, *Takifugu*, and *Spheroides* (Figure 2.2. A;

Figure 2.3. Percentage identity plot (PIP) of the (A) *Dlx1/2,* and (B) *Dlx5/6* intergenic regions between mouse, human and zebrafish

The mouse sequence is shown on the horizontal axis and the percentage identity to the human (top plot) and zebrafish sequences (lower plot) are shown on the vertical axis. Sequences used for comparison include the intergenic regions and the 3'UTRs of both flanking genes. In (A), *Dlx1* is to the left and in (B), *Dlx5* is to the left. Shaded dark and light gray areas indicate the positions of enhancers. Repetitive sequences are shown as follows: black triangles, MIR, mammalian interspersed repeats; vertical rectangles, simple sequence repeats; CpG islands: white horizontal rectangle, CpG ratio >0.60; gray rectangles, CpG ratio >0.75. For further details on PIP analyses see http://globin.cse.psu.edu/pipmaker.


A
human
zebrafish
100%
50%
100%
50%
0k
2k
4k
6k
8k
10k
10892
•• I12a : DarkGray
•• I12b : Gray
B
human
zebrafish
100%
50%
100%
50%
0k
2k
4k
6k
8k
10606
•• I56i : DarkGray
•• I56ii : Gray

Supplementary Figure 2.1. Multiple sequence alignment of I56i in five vertebrate species

M: mouse; H: human, T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. The consensus sequence represents identity in four out of five species.

```
          1                                                        50
H_I56_i   GCATTATAAT TTTGGTTAAT TTTCAATTTT AG.G.TC.CT ACGTCTCT..
M_I56_i   GCATTATAAT TTTGGTTAAT TTTCAATTTT AG.G.TC.CT ACGTCTCT..
T_I56_i   GCGTCGTAAT TTCAGATAAT ATCCCG.... AGCGCTCACT .C..CTCTCC
S_I56_i   GCGTCGTAAT TTCAGATAAT ATCCCG.... AGCGCTCACT .C..CTCTCC
Z_I56_i   ACATTGTAAT TTTAGATAAT ATCCCA.... AGCGTTCACT .C.TC.CT.C
Consensus GC-T--TAAT TT--G-TAAT -T-C------ AG-G-TC-CT -C--CTCT--

          51                                                      100
H_I56_i   .GCAATTTGT GTATGAATAA CAGAATAATT TCCCTCTTTT GTTTCGCCTT
M_I56_i   .GCAATTTGT GTATGAATAA CAGAATAATT TCCCTCTTTT GTTTCGCCTT
T_I56_i   GGCAATTTGT ATATGAATAA CCGAATAATT TCCCCCTTTT GTTCCA....
S_I56_i   GGCAATTTGT ATATGAATAA CCGAATAATT TCCCCCCTTT GTTCCA....
Z_I56_i   GGCAATTTGT ACATGAATAA CCGAATAATT T.CATCTTTT GTTTCG....
Consensus -GCAATTTGT -TATGAATAA C-GAATAATT TCCC-CTTTT GTT-C-----

          101                                                     150
H_I56_i   TCCTGT..T. CCT.GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT
M_I56_i   TCCTGT..T. CCT.GAATCT AAATAAAGAT GGCTTTTTAG TATTAAAAGT
T_I56_i   T.CTGTGCTA CCTCAAATCC AAATAAAGAT .GCCTTTTAG TATTAAAAGT
S_I56_i   T.CTGTGCTA CCTCAAATCC AAATAAAGAT .GCCTTTTAG TATTAAAAGT
Z_I56_i   T.CTTTGCCA CTTCAAATCC AAATAAAGAT .GCCTTTTAG TATTAAAAGT
Consensus T-CTGT--T- CCT--AATC- AAATAAAGAT -GC-TTTTAG TATTAAAAGT

          151                                                     200
H_I56_i   GGAAGAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG
M_I56_i   .G..GAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG
T_I56_i   GGTAGAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG
S_I56_i   GGTAGAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG
Z_I56_i   GGTAGAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG
Consensus GG-AGAAAAT TACAGGTAAT TATCTTTGAC GGTAAAAACG CTGTAATCAG

          201                                                     250
H_I56_i   CGGGCTACAT GAAAAATTAC TCTAATTATG GCTGCATTTA AGAGAAT...
M_I56_i   CGGGCTACAT GAAAAATTAC TCTAATTATG GCTGCATTTA AGAGAAT...
T_I56_i   CGGGCTACAT CAAAAATTAC CCTAATTATG CCTGCATTTA TGAGAAT.GG
S_I56_i   CGGGCTACAT CAAAAATTAC CCTAATTATG CCTGCATTTA TGAGAATGGG
Z_I56_i   CGGGCTACAT CAAAAATTAC CCTAATTATG TCTGCATTTA TGAGAAT...
Consensus CGGGCTACAT -AAAAATTAC -CTAATTATG -CTGCATTTA -GAGAAT---

          251                                                     300
H_I56_i   ........G. ...G.....A AAAAAACCTT CT.TGTGGAT AAAAA.CCTT
M_I56_i   ........G. ...G.....A AAAAAACCTT CT.TGTGGAT AAAAA.CCTT
T_I56_i   .......GGA ...G..A.AA AAAAAAGCTT CACT.TGGAT AAAAACCCAC
S_I56_i   CCCTATAGGA GTCGATACAA AAAAAAGCTT CACT.TGGAT AAAAACCCAC
Z_I56_i   ........G. ...G.....A AAAAAACCCT CTCT.TGGAT .AAAACCCAT
Consensus --------G- ---G-----A AAAAAA-CTT C--T-TGGAT AAAAA-CC-

          301                                                     350
H_I56_i   AAATTGTCCC CAATGTCTGC TTCAAATTGG ATG..GCACT GCAGCTGGAG
M_I56_i   AAATTGTCCC CAATGTCTGC TTCAAATTGG ATG..GCACT GCAGCTGGAG
T_I56_i   AAATTGTCCC AAATATCTCC TTCATATTGA ACGCCG..CT GCAGCTGGCT
S_I56_i   AAATTGTCCC AAATATCTCC TTCATATTGA ACGCCG..CT GCAGCTGGCT
Z_I56_i   AAATTGTCCC AAATATCTCC TTCATATTGA ACG.AG.ACT GCAGCTGGAT
Consensus AAATTGTCCC -AAT-TCT-C TTCA-ATTG- A-G--G--CT GCAGCTGG-
```

```
          351                                                      400
H_I56_i   GCTTTGTT.C AGAATTGATC CTGGGGAGCT ACGAACCCAA AGTTTCA.CA
M_I56_i   GCTTTGTT.C AGAATTGATC CTGGGGAGCT ACGAACCCAA AGTTTCA.CA
T_I56_i   GCTTTGTTCC A.CATCGATC CCGCGGAG.T TTG...GC.. A.TTT.AAC.
S_I56_i   GCTTTGTTCC A.CATCGATC CCGCGGAG.T TTG...GC.. A.TTT.AAC.
Z_I56_i   GCTTTGTT.C AGGATCGATC CTGTGGCG.A AAG...GC.. A.TTC.AGC.
Consensus GCTTTGTT-C A--AT-GATC C-G-GGAG-T --G----C-- A-TTT-A-C-

          401                                                  447
H_I56_i   G...TAG..G A...AGG.GG GA.AAAAA.G AAAAGAAAAC ATTTTTC
M_I56_i   G...TAG..G A...AGG.GG GA.AAAAA.G AAAAGAAAAC ATTTTTC
T_I56_i   GGCCT.GC.. AT.TAGG.AC GA.AGAGACG .GGAGAAAAC A.TTTT~
S_I56_i   GGCCT.GC.. AT.TAGG.AC GA.ANAGACG .GGAGAAAAC A.TTTT~
Z_I56_i   GTCTT.GCAG ATCAAGGCA. GAGAGAAAGG .GGA.AAAAC A.TTTT~
Consensus G---T-G--- A---AGG--- GA-A-A-A-G ---AGAAAAC A-TTTT-
```

alignments provided as Supplementary Figures 2.3. and 2.4.). The first is found in the 3'UTR sequence of zebrafish *dlx5a* (see note concerning the nomenclature of zebrafish *dlx* genes in the Methods section) and at a corresponding position, with respect to the predicted stop codons of the *Takifugu* and *Spheroides* orthologs (Figure 2.2. A). The second is found just downstream of the 3'UTR of zebrafish *dlx6a* and at a similar position in the pufferfish orthologs. Finally, a fragment of about 100 bp with 83% sequence identity was found between the end of *dlx5a* and I56ii in zebrafish and *Takifugu* but was not found in *Spheroides* (alignment provided as Supplementary Figure 2.5.). None of the three shorter conserved sequences could be identified in the two mammalian loci.

We identified two highly conserved sequences in the *Dlx1/Dlx2* intergenic regions of the five vertebrates. The first, I12a, is approximately 550 bp in length and the percentages in sequence identity in pairwise comparisons vary between 83% and 99% (Figures 2.1. B and 2.4.). The second, I12b, is about 400 bp in length and shows percentages of identity that vary between 75% and 97% (Figures 2.1. B and 2.5.). The relative positions and orientations of I12a and I12b with respect to the *Dlx1* and *Dlx2* genes were identical in all five species. As for I56i and I56ii, the I12a and I12b sequences reside in a region of overall stronger sequence conservation in mouse/human (Figure 2.3. A) and in *Takifugu/Spheroides* (data not shown) pairwise comparisons.

In addition to I12a and I12b, we identified a sequence of about 320 bp, I12c, that was conserved between *Takifugu*, *Spheroides*, and zebrafish. This sequence is located between the end of *dlx2* and I12a (Figure 2.1.A; alignment provided as Supplementary Figure 2.6.). Finally, a sequence of about 110bp was found in or near the 3'UTR of *Dlx1* of mouse and human and the in the zebrafish *dlx1/dlx2* locus, between the 3'end of *dlx1* and I12b (alignment provided as Supplementary Figure 2.7.). This sequence contains a TTA tri-nucleotide repeat but sequence conservation extends beyond this repeat.

Supplementary Figure 2.2. Multiple sequence alignment of I56ii in five vertebrate species

M: mouse; H: human, T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. The consensus sequence represents identity in four out of five species.

```
           1                                                    50
H_I56_ii   AAATAGGGTC TAAATAATGG CTGAATTAGC CCTACTGGAC A.GTTTCAGA
M_I56_ii   AAATAGGGTC TAAATAATGG CTGAATTAGC CCTACTGGAC A.GTTTCAGA
T_I56_ii   AAATCGAGTC TAAATAATGA C..AGTTAGC TTTACTGGAC T.GCTCCG.A
S_I56_ii   AATTCGAGTC TAAATAATGA C..AGTTAGC TTTACTGGAC A.GCTTCA.A
Z_I56_ii   AAACCGCGTA TAAATAATGA C..AGTTAGC TTTACTGGAC ACCCT.CA.A
Consensus  AAAT-G-GTC TAAATAATG- C--A-TTAGC --TACTGGAC A-G-T-CA-A

          51                                                   100
H_I56_ii   .TGTAACACT CTGTAATAAT TATA.TTGCA GGCTGGATT. AGGATGCTAT
M_I56_ii   .TGTAACACT CTGT.ATAAT TATA.TTGCA GGCTGGATT. AGGATGCTAT
T_I56_ii   CTGTAACACT CCACAATAA. T.TA.TT.CT .CCAGGATTA AGGCCGCCAT
S_I56_ii   CTGTAACACT CCACAATAA. T.TA.TT.CT .CCAGGATTA AGGCCGCCNG
Z_I56_ii   CTGTAACACT CAACAATAAT TATACTTCCA .CCAGGATT. A.G.CTCCAT
Consensus  -TGTAACACT C---AATAA- T-TA-TT-C- --C-GGATT- AGG--GC-AT

           101                                                  150
H_I56_ii   TATCATAA.T CTGGACGTTT ACAATTATCT GTAATTTGCA AAGATGCGCC
M_I56_ii   TATCATAA.T CTGGACGTTT ACAATTATCT GTAATTTGCA AAGATGCGCC
T_I56_ii   TATCATAA.T CTGGACATTG ACAATTACCT ATAATTTGCA AAGATGCGCT
S_I56_ii   TATCATAATT CTGGACATTG ACAA.TACCT ATAATTTGCA AAGATGCGCT
Z_I56_ii   TATCATAA.T CTGGACATTT ACAATTATCT ATAATTTGCA AAGATGCGCC
Consensus  TATCATAA-T CTGGAC-TT- ACAATTA-CT -TAATTTGCA AAGATGCGC-

           151                                                  200
H_I56_ii   AGG.TCTTGA TTACAG.CA. G...TTTTTT TTTTTTTTTT TTTTTTTTTT
M_I56_ii   AGG.TCTTGA TTACAG.CA. GCTTTTTTTT TTTTTTTTTT TTTTTTTTTT
T_I56_ii   TGGTTCTTGA TTGCAAAGAG GC........ .......... ....TTTTTC
S_I56_ii   TGGTTCTTGA TTGCAGAGA. GC........ .......... ...ATTCCCC
Z_I56_ii   TGGTTCTTGA TTGCAGA.AG G......... .......... ....TTTTT.
Consensus  -GG-TCTTGA TT-CAG--A- G--------- ---------- ----TTTTT-

           201                                                  250
H_I56_ii   GTACCAC..G CTAACCATCA CTAAACAGTG ACAGTAATAA CAGCTAATTT
M_I56_ii   GTATCAC..G CTAACCATCA CTAAACAGTG ACAGTAATAA CAGCTAATTT
T_I56_ii   CC.CC.CAAG CTCGCCATCA CTAAACAGTG ACAGTAATAA CAGCTAATTT
S_I56_ii   CC.CC.CAAG CTCGCCATCA CTAAACAGTG ACAGTAATAA CAGCTAATTT
Z_I56_ii   .T.CC.CTGG CTCACCATCA CTAAACAGTG ACAGTAATAA CGGCTAATTT
Consensus  ---CC-C--G CT--CCATCA CTAAACAGTG ACAGTAATAA CAGCTAATTT

           251                                                  300
H_I56_ii   TGCTGGC.A. ATATAAGAGG TG...CTGGG GTGTGCAAAC AATTT.CACA
M_I56_ii   TGCTGGC.A. ATATAAGAGG TG...CTGGG GTGTGCAAAC AATTT.CACA
T_I56_ii   TGCAGGCGAT ATATAAGAGC T.TACC.AGG GTGTGCAAAC ACTTTCC.CA
S_I56_ii   TGCAGGCAAT ATATAAGAGC T.TACC.AGG GTGTGCAAAC ACTTTTC.CA
Z_I56_ii   TGCAGGCAAT ATATAAGAGG TGT..C.GGG GTGTGCAAAC ACTTT.CACA
Consensus  TGC-GGC-A- ATATAAGAG- T----C--GG GTGTGCAAAC A-TTT-C-CA

           301          317
H_I56_ii   CCTGGATGTG CTCA..C
M_I56_ii   CCTGGATGTG CTCA..C
T_I56_ii   CCTGGATTTG CTTATTC
S_I56_ii   CCTGGATTTG CTTATTC
Z_I56_ii   CCTGGATTTG CTTA..C
Consensus  CCTGGAT-TG CT-A--C
```

Supplementary Figure 2.3. Multiple sequence alignment of I56iii in three fish species

T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. The consensus sequence represents identity in three out of three species.

```
           1                                                    50
T_I56_iii  ATGTTGATTT ..GACTTTTT .TTTTT.TCC T.TCTATGGG ACTCG.TG.T
S_I56_iii  ATGTTGATTT CAGA....TT .TTTTT.TCC T.TCTATGGG ACTCG.TG.T
Z_I56_iii  A..TTG.TTT ...A....TT ATTTTTGT.. TGTATATTGG ACTGGTTGTT
Consensus  A--TTG-TTT ---A----TT -TTTTT-T-- T-T-TAT-GG ACT-G-TG-T

           51                                                   100
T_I56_iii  .ACAA.ATTT CACAGGAATA TGCAATTGTA T...T.TGAC AGTCATAGAA
S_I56_iii  .ACAA.TTTT CACAGGAATA TGCAATTGTA ....T.TGGC AGTCATAGAA
Z_I56_iii  AACAATTTTT TTGAGGAATA TGCAA.TGTA TCGATATGGC AGTCCTAGAA
Consensus  .ACAA--TTT ---AGGAATA TGCAA-TGTA ----T-TG-C AGTC-TAGAA

           101                                                  150
T_I56_iii  GAACCGTGTA AAATGTGTAA A.TGTGTGCA TGTAATTTAT TGCA.TTTTG
S_I56_iii  GAACCGTGTA AAATGTGTAA A.TGTGTGCA TGTAATTTAT TGCATTTTTG
Z_I56_iii  GAA.CGTGTA TAATGTGTAA ATTGTGTGCA TGTAATTTAT TGCA..TTTG
Consensus  GAA-CGTGTA -AATGTGTAA A-TGTGTGCA TGTAATTTAT TGCA--TTTG

           151          167
T_I56_iii  GAAGACTATT AAACGTT
S_I56_iii  GAAGACTATT AAACGTT
Z_I56_iii  GAAG...... .AA..TT
Consensus  GAAG------ -AA--TT
```

Supplementary Figure 2.4. Multiple sequence alignment of I56v in three fish species

T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. The consensus sequence represents identity in three out of three species.

```
          1                                                        50
 T_I56_v  ACGGGAA.AC GGGCCTGTTA TGGTGCGACG TGTTGGCCCA ATACGGAGCA
 S_I56_v  ACGGGAA.AC GGGCCTGTTA TGGTGCGACG TGTTGGCCCA ATATGGAGCA
 Z_I56_v  AAAGTAATA. ..GCCTG... TGTTGC.A.. TAAT...CTA AT....A.CA
Consensus A--G-AA-A- --GCCTG--- TG-TGC-A-- T--T---C-A AT----A-CA

          51                                                      100
 T_I56_v  CAA...A..A A.GGGGGC.T .TCTTTTGC. .CTACTTTCC CTTTGAAATC
 S_I56_v  CAA...A..A A.GGGGAC.T .TCTTTTGCG TTTTTTTTCC CTTTG.AATC
 Z_I56_v  TAATTGATTA ATGGGG.CAT GTC.TTT.C. ATTA.TTTCT TTTTTAAA.A
Consensus -AA---A--A A-GGGG-C-T -TC-TTT-C- --T--TTTC- -TTT--AA--

          101                                                     150
 T_I56_v  ACGCAACA.G TCAGCATTTG CTTTG.AGTC ACGGTTTACT .CA.ATCTGA
 S_I56_v  ACNCAACA.G TCAGCGTCTG GTTTG.AGTC ACGGGTTACT .CA.ATCTGG
 Z_I56_v  ATG.AACACG ACA.C.TTTA C.TTGAAGTC ATGGTTTACT CCATTTATGA
Consensus A-G-AACA-G -CA-C-T-T- --TTG-AGTC A-GG-TTACT -CA--T-TG-

          151                                                     200
 T_I56_v  CAAAACATAA AAAGTGTATG T.ACAACTAA CATC.G.CAG ATTATATTTA
 S_I56_v  CAAAACATAA AAAGGGTTTG T.TC.ACTAA CTTC...CAG ATTATATTTA
 Z_I56_v  CAAAACATAA AAA..GTTT. TAACAACTAA AATCGGACA. A.CATATTTA
Consensus CAAAACATAA AAA--GT-T- T--C-ACTAA --TC---CA- A--ATATTTA

          201                          229
 T_I56_v  CAGA.CAGAC AATTTGCTTG GC.TTTGCA
 S_I56_v  CAAA.CAGAC AGTTTGCTTG GC.TTTGCA
 Z_I56_v  CA.ATCA.A. A..TTGCTT. GCTTTTGCA
Consensus CA-A-CA-A- A--TTGCTT- GC-TTTGCA
```

Supplementary Figure 2.5. Multiple sequence alignment of I56iv in two fish species T: *Takifugu rubripes*; and Z: zebrafish.

```
          1                                                        50
T_I56_iv  CCCTACAATT AGCCGCTTTG AATGGCAGAA AAGAGGCGAG GGGGGCGGCA
Z_I56_iv  CCTTACAATT AGCAGCTTTG AATTGTAGTG AGGAGGCGA. AGGGACGGCG
Consensus CC-TACAATT AGC-GCTTTG AAT-G-AG-- A-GAGGCGA- -GGG-CGGC-

          51                                                       99
T_I56_iv  CGCACATCTG AGACAATCCG CATTCCTGTG GGGGTAGACC GGCTCCAGA
Z_I56_iv  CGCAGATCTG AGGCAGTCCG CATGCCTGTG GGGGCAAACC GGCTCCAGC
Consensus CGCA-ATCTG AG-CA-TCCG CAT-CCTGTG GGGG-A-ACC GGCTCCAG-
```

2.3.3. The sequences conserved between all five vertebrate species contain enhancers

To determine that the conserved *Dlx* intergenic sequences, I56i, I56ii, I12a and I12b, constitute *cis*-acting regulatory sequences, they were tested in reporter constructs that were injected to produce transgenic mice and zebrafish. As previously reported, I56i and I56ii target expression of *lacZ* reporter constructs to the forebrain of transgenic mice and zebrafish starting at E10 and persisting in adult mice (Zerucha et al., 2000). The mouse I56i sequence can efficiently target expression to the forebrain by itself in 100% of primary transgenic mice expressing the transgene and in 3 out of 4 transgenic lines (Figure 2.6. A, Table 2.1.; Zerucha et al., 2000). The zebrafish I56i sequence also targeted expression to the forebrain of 12 out of 12 primary transgenic mouse embryos (Zerucha et al., 2000). In both cases, reporter gene expression precisely mimics that of the endogenous *Dlx5* gene and highly overlaps with that of *Dlx6* (Zerucha et al., 2000).

Three primary transgenic mice and 2 established lines containing a mouse I56ii reporter construct expressed *lacZ* in the forebrain (Figure 2.6. B) although the intensity of the ß-galactosidase staining was more variable between the telencephalic and diencephalic expression domains and staining seemed often weaker than that observed with I56i constructs. However, the mouse I56ii (this work) was more efficient at targeting transgene expression to the forebrain than its zebrafish counterpart (Zerucha et al., 2000).

When tested in transgenic zebrafish, a construct containing both zebrafish I56i and I56ii targeted expression of the green fluorescent protein (GFP) reporter transgene to the domains of *dlx* expression in the telencephalon and diencephalon (Figure 2.6. G and H). In this transgene construct, GFP is placed immediately downstream of a 3.5 kb fragment of the *dlx6a* 5'-flanking region including the promoter and part of the 5'UTR. This 5'-flanking fragment does not, by itself, target expression of GFP in a specific manner (Figure 2.6. F; no reproducible pattern in more than 150 embryos injected). However, in the presence of the zebrafish enhancers, 75-80% of injected embryos (n>400) had forebrain expression starting at 18 hours post-fertilization (hpf) and lasting until at least 96 hpf. Three transgenic lines could be produced all with comparable expression patterns

and intensity. An embryo from one line is shown in Figure 2.6. G and H. In contrast, the same intergenic fragment coupled to the ß-globin minimal promoter, which was used for transgenic mouse constructs, showed forebrain expression in only 8% of injected embryos and only 0.5% of them had more than 10 GFP-positive cells (Zerucha et al., 2000). The difference between efficiency of the human ß-globin minimal promoter fragment between human and zebrafish is, at present, unclear.

Similar transgene constructs containing the mouse I56i sequence (Figure 2.6. I) or a combination of I56i and I56ii (Figure 2.6. J), inserted in the *5'-dlx6a-GFP* plasmid, expressed GFP in the forebrain of transgenic zebrafish although the proportions of transgenic embryos were smaller than those observed with the corresponding construct containing zebrafish sequences. Thus, for both constructs, 35-40% embryos showed forebrain expression (n>150 for each construct) with most of the GFP-positive cells in the telencephalic domain of *dlx* expression (Figure 2.6. I and J).

The mouse I12b conserved sequence targeted reporter transgene expression to the forebrain of transgenic mice, starting at E10 and lasting until E16, the latest time point examined (Figure 2.6. C and Table 2.1.; 3/3 primary embryos and 5/5 transgenic lines). This construct also produced expression in the apical ectodermal ridge, another site of endogenous *Dlx* expression although expression was more variable in intensity (Table 2.1.) compared to that observed in the forebrain.

Preliminary examination of sections of brains from lines of transgenic mice expressing the *I12b-lacZ* construct indicates that the constructs faithfully mimic expression of *Dlx1/Dlx2* in the telencephalon and diencephalon (data not shown). Thus, despite the fact that their sequences are highly divergent (see below), the three intergenic sequences, I56i, I56ii, and I12b, act as *cis*-acting forebrain enhancers with highly overlapping patterns of activity.

A 1.9 kb *Xba*1-*Eco*R1 fragment containing the I12a conserved sequence targeted *lacZ* expression to a subset of *Dlx*-expressing cells in the mesenchyme of the mandibular component of the first branchial arch and in the hyoid arch starting at E9.5 and lasting until at least E16, when expression gradually diminishes (Figure 2.6. D, E and Table 2.1.;

Figure 2.4. Multiple sequence alignment of I12a in five vertebrate species

M: mouse; H: human, T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. The consensus sequence represents identity in four out of five species.

```
          1                                                   50
  H_I12_a CAGAATAGGA CTGAAATTTT CGGCTATATA GCCTCTGTCT TTATAGCTGA
  M_I12_a CAGAATAGAA CTGAAAATTT CGGCTATATA G.CTCTGTCT TTATAGCTGA
  Z_I12_a CCGAGCAGGA ATGAAAAATC CTGCT.T.TA TTTTCCCTCT TTATAGCTGA
  T_I12_a CGGAATATGA CTGAAAATGT CTGCT.T.TA TTTCCCCTCT TTATAGCTGA
  S_I12_a CGGAATATGA CTGAAAATCT CTGCT.T.TA TTTCCCCTTT TTATAGCTGA
Consensus C-GAATA-GA CTGAAAAT-T C-GCT-T-TA ----C--TCT TTATAGCTGA

          51                                                  100
  H_I12_a T.GAAAAAAA TTGCAGATTA TTAGCATAAA TGTTTACTCT TCATTACGCT
  M_I12_a T.GAAAAAAA TTGCAGATTA TTAGCATAAA TGTTTACTCT TCATTACGCT
  Z_I12_a T.GGAAAAAA TTGCTGATTA TTAGCATAAA TGTTTACTCC TCATTACGCT
  T_I12_a T.GGGGAAAA ATGTAGAATA TTAGCAT.AA .G....GT.C .C...AGGCT
  S_I12_a TGGGGGGAAA ATGTAGACTA TTAGCATAAA TGTTTACTCC TCATTACGCT
Consensus T-G---AAAA -TG-AGA-TA TTAGCATAAA TGTTTACTC- TCATTACGCT

          101                                                 150
  H_I12_a GATGACATTG TGCACTCGAG ATCTTGGTAA TCTTT.GGGA AAATTATGAG
  M_I12_a GATGACATTG TGCACTCGAG ATCTTGGTAA TCTTT.GGGA AAATTATGAG
  Z_I12_a GATGACATTG TGCACTCGAG ATCTTGGTAA TCTTT.GGGA AAATTATCAG
  T_I12_a .A.GTTATTG TGCACTCGAG ATCTTGGTAA TCTTTGGGGG AAATTATGAG
  S_I12_a GATGACATTG TGCACTCGAG ATCTTGGTAA TCTTTTGGGG AAATTATGAG
Consensus GATGACATTG TGCACTCGAG ATCTTGGTAA TCTTT-GGG- AAATTATGAG

          151                                                 200
  H_I12_a TAA..TTTT. ...T...AAA AATTTTAA.. AGCAGCAGCA GTTACCATGC
  M_I12_a TAA..TTTT. ...T...AAA AATTTTAA.. AGCAGCAGCA GTTACCATGC
  Z_I12_a TAA..TTTT. ...T...AAA AA.TTT.A.. A..AGCAGCA GTTACCATGC
  T_I12_a TAATTTTTTC C.CTC..TCT TTTTTTTATT TAAAGCAGCA GTTGCCATGC
  S_I12_a TAATTTTTTC CTCTCTTTTT TTTTTTTATT TAAAGCAGCA GTTGCCATGC
Consensus TAA--TTTT- ---T------ --TTTT-A-- ---AGCAGCA GTT-CCATGC

          201                                                 250
  H_I12_a AATTCAGGCT AATTCTGCGT AGGCTCCGAA CGGATATAAT TATCGAGGAG
  M_I12_a AATTCAGGCT AATTCTGCGT AGGCTCCGAA CGGATATAAT TATCGAGGAG
  Z_I12_a AATTCAGGCT AATTTTGCGT AGGCTCCAAA CGGATATAAT TATCGCGGAG
  T_I12_a AATTCAGCCT AATTCTGCGT GGGCTCCACA CGGATATAAT TATCGTTGAG
  S_I12_a AATTCAGCCT GATTCTGCGT GGGCTCCACA CGGATATAAT TATCGTTGAG
Consensus AATTCAG-CT AATTCTGCGT -GGCTCC--A CGGATATAAT TATCG--GAG

          251                                                 300
  H_I12_a TCAAGATGTT ATGCTAAAAA CTATGCATTG ATATTCCCAT TTATTATGTA
  M_I12_a TCAAGATGTT ATGCTAAAAA CTATGCATTG ATATTCCCAT TTATTATGTA
  Z_I12_a TCAAGATGTT ATGCTAAAAA CTATGCATTG ATATTACCAT TTATTATGTA
  T_I12_a TCAAGATGCT AGGCTAAAAA CTATGCATTG ATATTCCCAT TTATTATGTA
  S_I12_a TCAAGATGCT AGGCTAAAAA CTATGCATTG ATATTCCCAT TTATTATGTA
Consensus TCAAGATG-T A-GCTAAAAA CTATGCATTG ATATTCCCAT TTATTATGTA


          301                                                 350
  H_I12_a CATACAACTT TGACAATGTT GATGCTTGAA ATAGCAGGAA ATTGTCTTGC
  M_I12_a CATACAACTT TGACAATGTT GATGCTTGAA ATAGCAGGAA ATTGTCTTGC
  Z_I12_a CATACAACTT TGACAATGTT GATGCTTGAA ATAGCTGGAA ATTGTCTTGC
  T_I12_a CATAGGACTC TAACATTGTT GATGCTTGAA ATAGCGGGAA ATTGTCTTGT
  S_I12_a CCTAGGACTC AGACAATGTT GATGCTTGAA ATAGCGGGAA ATTGTCTTGT
Consensus CATA--ACT- TGACAATGTT GATGCTTGAA ATAGC-GGAA ATTGTCTTG-
```

```
          351                                                 400
  H_I12_a G.CAAAAATT ACAGTCCATA TTAGGACATC TGAAATTGCG AAGAATTATA
  M_I12_a G.CAAAAATT ACAGTCCATA TTAGGACATC TGAAATTGCG AAGAATTATA
  Z_I12_a G.CAAAAATT ACAGTCCATA TTAGGACATC TAAAATTGCG AAGAATTATA
  T_I12_a GT.AAAAATA ACACCCCATA TTAGGACATC TAAAATTGGG AAGAATTATG
  S_I12_a GTAAAAAATA ACACCCCATA TTAGGACATC TAAAATTGCG CAGAATTATG
Consensus G--AAAAAT- ACA--CCATA TTAGGACATC T-AAATTGCG AAGAATTAT-

          401                                                 450
  H_I12_a TAGTAATTGC AGGCTTTCAG ATGGGAGAGC CAGAATGCTC AAACTAGTGC
  M_I12_a TAGTAATTGC AGGCTTTCAG ATGGGAGAGC CAGAATGCTC AAACTAGTGC
  Z_I12_a TAGTAATTGC AGGCTTTCAG ATCGGGGAGC CAGAATGCTC .AACTGGAGC
  T_I12_a TAGTAATTGC AGGCTTTCAG ATTTGAAAGC CAGAATGCTC AAACTAGATC
  S_I12_a TAGTAATTGC AGGCTTTCAG ATTTGAAAGC CAGAATGCTT AAACTAGATC
Consensus TAGTAATTGC AGGCTTTCAG AT--GA-AGC CAGAATGCTC AAACTAG--C

          451                                                 500
  H_I12_a AAATGTGGAT AAAATTACAG ATCAGAGCAA AAATTAGGGA AAAAGGGGGT
  M_I12_a AAATGTGGAT AAAATTACAG ATCAGAGCAA AAATTAGGGG AAAAGGGGGT
  Z_I12_a AATTGTGGAT AAAATTACAG ATCAGAGTAT AAATTAGGG. ...ATGGGGT
  T_I12_a AAATGTGGAT GGAATTACAG ATGAGGGTAT AAATTA.GG. AGAGGGG.GT
  S_I12_a AAATGTGGAT AAAATTACAG ATCAGGGTAT AAATTA.GG. ATAGGGGAGT
Consensus AAATGTGGAT AAAATTACAG ATCAG-G-A- AAATTA-GG- .A-A-GGG-GT

          501                                                 550
  H_I12_a CTGGGATTTT TCAGGTTGGC TATAGGATTC CGGAAGTGTC TAATGCCACA
  M_I12_a CTGGGATTTT TCAGGTTGGC TATAGGATTC CGGAAGTGTC TAAAGCCACA
  Z_I12_a ATGGCATTTT TCAAGTTGTC TACAGGATTC CGGAAGTGTC TAAAACCACA
  T_I12_a ATAGGCTGTG ACA..TT.TA AAT.CCCCTC GTCGAATG.C .AAAACAGCA
  S_I12_a ATAGGCTATG GCA..TT.TA AAT.CTTCTC GACAATTG.C .AAAACCACA
Consensus -T-GG-T-T- -CA--TT--- -AT-----TC ---AA-TG-C -AAA-CCACA
```

Figure 2.5. Multiple sequence alignment of I12b in five vertebrate species

The consensus sequence represents identity in four out of five species. Sequences similar to the binding site for Dlx protein ([A/C/G]TAATT[G/A][C/G], (Feledy et al., 1999b) are shown in bold. See also Figure 2.7. Abbreviations as in Figure 2.4.

```
          1                                                   50
  T_I12_b CAGCTGCACA TCCAAGAGGG TCAGTGTTAT TTCGCTGTAT TCTCCTCTTG
  S_I12_b CAGCTGCACC TCCAAGAGGG TCAGTG.TA. TT..C..... TCT.C.CTTG
  H_I12_b CAGCTGCAAA CCCAAGAGGG TCAGCATCAT TTCACTGTAT TCTCTTCTTG
  M_I12_b CAGCTGCAAA CCCAAGAGGG TCAGCATCAT TTCACTGTAT TCTCTTCTTG
  Z_I12_b CAGCTGTGGA GCAGAGAGGG TCAGCGTTAT TTCACTCTAT TCTCCTCTTG
Consensus CAGCTGCA-A -CCAAGAGGG TCAG--T-AT TTC-CT-TAT TCTC-TCTTG

          51                                                  100
  T_I12_b ATTATACAAG CTGAGCCCAT CAAAACCACC ATAATTACAG TCATTTCGCC
  S_I12_b ATTATACAAG CTGAGCCCAT CAAAACCACC ATAATTACAG TCATTTCGCC
  H_I12_b A.T.TACAAG CCGGGCCCAT CAAACACAAC ATAATTACAG TAATTTCAGG
  M_I12_b A.T.TACAAG CCGGGCCCAT CAAACACAAC ATAATTACAG TAATTTCAGG
  Z_I12_b A.T.TACGCG CCGGGCCCAT CAAACCAGAC ATAATTACAG TCATTCCCTC
Consensus A-T-TACAAG C-G-GCCCAT CAAA--CA-C ATAATTACAG T-ATTTC---

          101                                                 150
  T_I12_b CTTATTTATT CTAATGCCGT TT.CCTATCT GTGAGGTAAT TATGAGC.A.
  S_I12_b CTTATTTATT CTAATGCCGT TT.CCTATCT CTGAGGTAAT TATGAAC.AT
  H_I12_b TTTATTTATT CTAATGCAGT TTCCCCATCT CTCTGGTAAT TATGAGCAA.
  M_I12_b TTTATTTATT CTAATGCAGT TTCCCCATCT CTCTGGTAAT TATGAGCAA.
  Z_I12_b CTTACTTATT CTAATGCAGC TT.CCCATCT ACGGGATAAT TATGAGCAAT
Consensus -TTATTTATT CTAATGC-GT TT-CC-ATCT -T--GGTAAT TATGAGC-A-

          151                                                 200
  T_I12_b ..TTTTT.GC AAGGAGAATC TTTCTGTGTT GACAAAGGAG ATA..G.TCA
  S_I12_b TTTTTTT.GC AAGGAGAATC TTTCTGTGTT GACAAAGGAG ATA..G.TCT
  H_I12_b .TTTTTTCGC CCAGGGAATC TTTTTGCATT AACAAAAGAG ATAACGCACT
  M_I12_b .TTTTTTCGC CCAGGGAATC TTTTTGCATT AACAAAAGAG ATAACGCACT
  Z_I12_b TTTTTTTCGC ACGTGGAATC .TTTTG..TT AACAAAAGAG ATAGCGCGCT
Consensus -TTTTTT-GC ---G-GAATC TTT-TG--TT -ACAAA-GAG ATA--G--CT

          201                                                 250
  T_I12_b GAAAG..AAA TCTGCTGCCG CAAGACAAAA AAAGAAGAGA AAATAGGTGA
  S_I12_b AAGAG..AAA T.T..T.... ...G.C...A ATAGAA.A.A AAATAGGTGA
  H_I12_b GAAAGCCAAA TTTGCTGTGC .A.TTGAGAA AAGGAAAAAA AAA.AAATCA
  M_I12_b GAAAGCCAAA TTTGCTGCGC .A.TTGAGAA AAGG..GAAA AAA.AAATCA
  Z_I12_b GAAAG.AAAA TTTGCTGTGC GACGGGGGAA AAAGAAAAGG GGG.AAAAAA
Consensus GAAAG--AAA T-TGCTG--- -A------AA AA-GAA-A-A AAA-A--T-A

          251                                                 300
  T_I12_b GCCTCCCTTT AGTCACTG.T TCCTTTAGAC TTCGCCTGGA CTATAAGGCT
  S_I12_b GCCTCCTTTT CGTCACTG.T TCCTTTAGAC TTAGGCTGGA CTATGAGGCT
  H_I12_b A.AT.AGGTG CG.AGCTGCC ATCTCTGCAA TTCTCTGGTA CCG.GAGCCG
  M_I12_b A.AT.AGGTG CG.AGCTGCC ATCTCTGCAA TTCTCCGGTA CCG.GAGCCC
  Z_I12_b G.AG.A.GAG CG.AAATAGG TGCGCAGCGG TAGT.TATAA GTA.AGGCTG
Consensus ---T----T- CG---CTG-- --CT-T--A- TT-----G-A C----AG-C-

          301                                                 350
  T_I12_b TCTAAATTGC TCGCAGGTGT AGTCTCCATC AC.TGTCACT GCGAAGAGTC
  S_I12_b GCTAAATTGC TCGCAGGTGT AGTCTGCATC AC.TGTCACT GCGAAGAGCC
  H_I12_b GC.AAATTGC TTGCAGGTGT ATGGAGCAAG .CTTGTCAAT GG..CCAGGC
  M_I12_b GC.AAATTGC TTGCAGGTGT ATGGAGCAAG .CTTGTCAAT GG..CCGGGC
  Z_I12_b GC.AAATTGC TTGCAGGTGT ATACCGAGAG ATTTGTCAAT GGAAAGAGGG
Consensus GC-AAATTGC T-GCAGGTGT A----GCA-- -C-TGTCA-T G-----AG-C
```

```
         351                                                  393
  T_I12_b TTCGAAATTA GCAAA.GCAC AATGTAAGAG TGATGACAGA CGC
  S_I12_b TTCAAAATTA GCAAA.GCAC AATATGAGAA CGATCGCAGC CAA
  H_I12_b CTCCAAATTA GCAAATGCAC AGCAGCAAAG TAATGA.AGA CAG
  M_I12_b CTCCAAATTA GCAAATGCAC AGCAGCGAAG TAATGA.AGA CAG
  Z_I12_b ATCAAAATGA GCAAAAGCAC AATGTCGGAG TGATGA.AGG CAG
Consensus -TC-AAATTA GCAAA-GCAC A-------AG T-ATGA-AG- CA-
```

Supplementary Figure 2.6. Multiple sequence alignment of I12c in three fish species

T: *Takifugu rubripes*; S: *Spheroides nephelus*; and Z: zebrafish. A sequence located at position 291-642 downstream of zebrafish *dlx2b* (*dlx5*) polyadenylation signal shows similarity to I12c. The consensus sequence represents identity in three out of four species.

```
          1                                                        50
   S_I12c CCATGG.ATT GCTGTGGTAN G.CTCCGGGG CTGTCTTGGG AACCCAGTTC
   T_I12c CCAT.G.A.T GCTGT.G.CA G.CTCC.GGG CTGTCTT.GG AA.CGAGTTC
   Z_I12c CGGTGGAAGT TGTGAGAAAT GTTTCCATGT TGTTTAGCTC GAGGCATTTG
Z_dlx2b-3'~CGAGG.GCT TGAGGAGAAT GTTTCCAGGG CGGTCTGCTT GAGGCGGTTT
Consensus CC-TGG-A-T --TG--G-AT G--TCC-GGG C-GTCT---- -A--CAGTT-

          51                                                      100
   S_I12c CGCTTCC.CT TTCTCTTGTT AGACTGAGTT CGTAAAACGT CTGG.AAAGC
   T_I12c CGCTCCT.CT TT.TCTTGTT AGACTGAGTT CATAAAACGT CTGG.AGAGC
   Z_I12c TGAGCCTGGT TT.T.TGCGT GCGCTGACTT CATAAAAAGT CTGG.AGCGC
Z_dlx2b-3'TATGTATTTT TTCTCTGGTT GAGTGGCGTT CATAAAACGT CTGGAAGAGG
Consensus -G---CT--T TT-TCT-GTT ---CTGAGTT CATAAAACGT CTGG-AGAGC

          101                                                     150
   S_I12c AGCTCACNGA CTCCACGGAG TCTG.AGTGT G.AG.GCGCT TGT.ATA.TC
   T_I12c AGCTCCCCGA CTCCACGGAG TC.......T G.AG.GCGCT TGT.ATA.TC
   Z_I12c AGATCTCTGA CTCCAGAGAG TCTGAAGCCA GCCGAACGC. CTTCATGC.C
Z_dlx2b-3'AGCGAGCTGA CTCCAGAGAG TCTGGAGTGG GAAGAGCCCT CGCCGCACGC
Consensus AGCTC-CTGA CTCCA--GAG TCTG-AG--- G-AG-GCGCT -GT-ATA--C

          151                                                     200
   S_I12c .CACCCTGCC T.C.TGCCCN TCT.CACATG CCAAGGCGTG TTTTTGTGTN
   T_I12c .CACCCT.CA T.C.TTCCCC TCT.CACATG CCAAGGCCTG TTTTGGCGTC
   Z_I12c GGGGCGTGTT TGC..GCGCT CATTCCCA.G CC.CGAGGC. TCTGTA.G.C
Z_dlx2b-3'GGAACGCCCT CGCATGTGCT TCTGCGCATG CCAAGACGCG TCCGAATGCC
Consensus --A-C-T-C- T-C-TGC-CT TCT-C-CATG CCAAG-CG-G T-T----G-C

          201                                                     250
   S_I12c .GACTTCCCC TCTGCAGAGC CTGCGC.CCC GGGCTTTTTG CCTGTCCG.A
   T_I12c .GACTCCCCC TCCGCTGAGC CTGCGC.CCC GGGCCTTTTG CCTGCCTG.C
   Z_I12c ..GCG.TCTC TCTGCT.GTC G.GA.CTCCT CGAGTTTTTG TCTTTCAGTT
Z_dlx2b-3'AGGCGTTCCC CCTCAT.ACC GAGAGGTACA GGATAGTACC TC.TTCACCA
Consensus -G-C---CCC TCTGCT-A-C --G-GC-CC- GG---TTTTG -CT-TC-G--

          251                                                     300
   S_I12c ACCCTG.GTT CTTTCCGCTA A.AT..GT.C TG..CTG.AC GAA.ACTGTG
   T_I12c ACCCCG.GTT CTTTCCGC.C A.A...GT.C TG..CTG.AC GGA.ACTGTG
   Z_I12c TTAC.GTCTT CAAAAATACA AGATGCGTGC TG.TTTGTTC CCCTGCTTTT
Z_dlx2b-3'AGACAATTTA CAAAGGGATA AAAAG.GAGT AGAGTTGTT. GCATAATCTC
Consensus A--C-G--TT C-----G--A A-A---GT-C TG---TG--C G-A-ACT-T-

          301                                                     350
   S_I12c TAA.AACAGT TGATTTGCTT GTTCTAGTAC TGACGCCTAA TAAGNACACT
   T_I12c TAA.AACAGT TGATTTGCCT GTTCTAGTAC TGACGCCTAA TAA.ATCACT
   Z_I12c TAATAACAAA GCTTCTGCTA ATTTTTGAAC TCGAG..TAA AAAGTTAAAC
Z_dlx2b-3'TAAAGATAAA AAATAT.ATA CTTTATGCAT TGTAG..GAA TTAGAGGAAG
Consensus TAA-AACA-- --AT-TGCT- -TT-T-G-AC TG--G--TAA TAAGA--A--

          351      361
   S_I12c GCATCTTCCT T
   T_I12c GCATCTTCCT T
   Z_I12c NTCTCTCTTC T
Z_dlx2b-3'ATATGAACAG A
Consensus G-ATCT-C-- T
```

Supplementary Figure 2.7. Multiple sequence alignment of I12

The consensus sequence represents identity in three out of three species. H: human; M: mouse and Z: zebrafish.

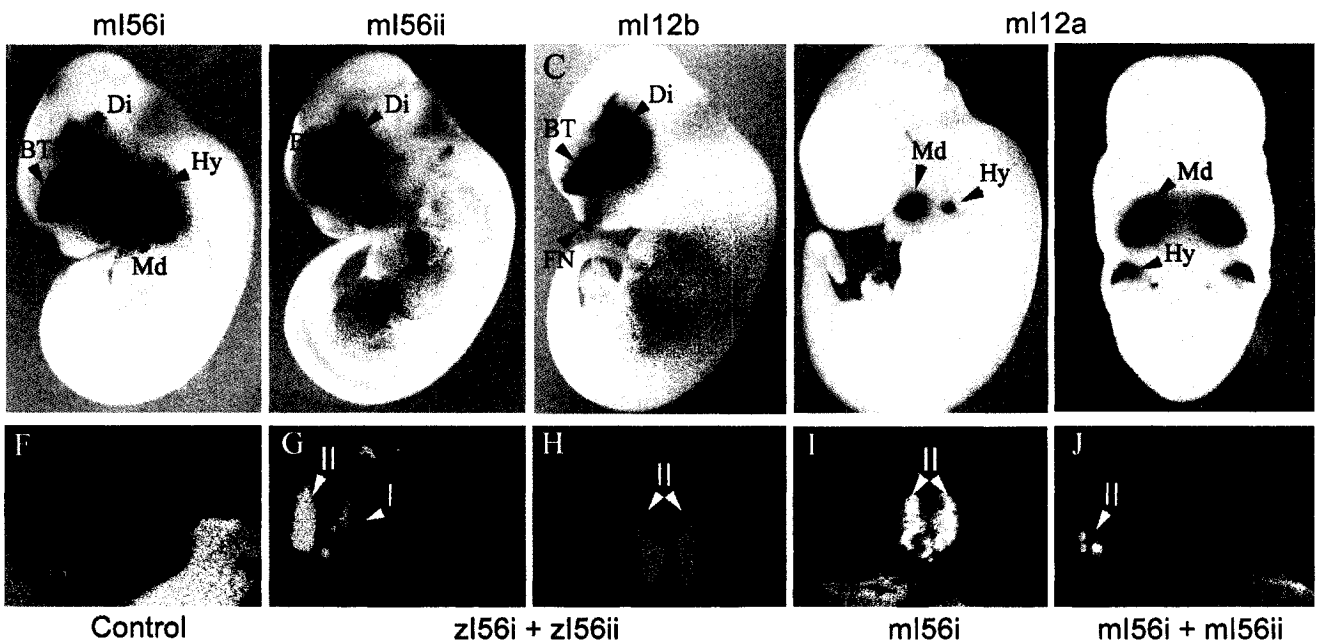
```
          1                                                        50
 M-I12    GATCTAATTG CTGTA.CAT. ATTATTGTTG TTATTATTAT TAATTATTAT
 H-I12    ~~TATATTTT TTTAATTATG ATTATTATTA TTATTATTAT T.ATTATTAT
 Z-I12    GATCTAATGA CTGTA.CAT. ATTA...TTG TTATTATTAT T.ATTGTTAT
Consensus --T-TA-T-- -T--A--AT- ATTA---TT- TTATTATTAT T-ATT-TTAT

          51                                                      100
 M-I12    TATTATTGTT .CTGTAAACA TGTTG..CA. CAAGCTTAG. CCTCTTTC.C
 H-I12    TATTATTATT ATTATAAAGA TGATGATTAT TATTATTAGT GGTATTAGTA
 Z-I12    TATTGTTGTT .CTGTAAACA TGTTG..CA. CAAGCTTAG. CCTTTTTG.C
Consensus TATT-TT-TT --T-TAAA-A TG-TG---A- -A---TTAG- --T-TT----

          101          116
 M-I12    GTTCTGTTG. TGTGTG
 H-I12    GTAGTATCGT TGCGTT
 Z-I12    GTTCTGTTG. TGTG~~
Consensus GT--T-T-G- TG-G--
```

Figure 2.6. Enhancer activity of conserved *Dlx* intergenic sequences in transgenic mice (A-E) and zebrafish (F-J)

(A) Mouse I56i, (B) mouse I56ii, and (C) mouse I12b each drive reporter gene expression to the basal telencephalon (BT) and diencephalon (Di) of transgenic mice, as shown here in mouse embryos. (D, E) Mouse I12a drives reporter gene expression to a subset of mesenchymal cells in the mandibular (Md) component of the first branchial arch and in the second branchial arch (Hy) of an E11.5 embryo. (A-D) are sagittal views and (E) is a frontal view of the embryo shown in (D). All embryos are at stage E11-12. FN, frontonasal prominence. (F) Head of 48 hpf primary transgenic zebrafish embryo, dorso-lateral view, injected with the control *dlx6a-GFP* reporter plasmid. Injection of this construct results in very few GFP-positive cells with no tissue specificity (n>150). (G, H) Lateral and frontal views, respectively, of a 48 hpf zebrafish embryo from a transgenic line produced with a construct made with the *dlx6a-GFP* reporter plasmid that also contained a 1.4 kb *dlx5a/dlx6a* intergenic fragment containing I56i and I56ii. I and II indicate the diencephalic and telencephalic domains of transgene expression that also correspond to endogenous *dlx* expression patterns in the zebrafish forebrain. (I) Frontal view of a 48 hpf primary transgenic zebrafish embryo injected with a *dlx6a-GFP* that also contained a 4.0 kb mouse *Dlx5/Dlx6* intergenic fragment that comprises I56i. The transgene is expressed predominantly in the telencephalic domain II. (J) Lateral view of a 48 hpf primary transgenic zebrafish embryo injected with a *dlx6a-GFP* that also contained a 2.8 kb mouse *Dlx5/Dlx6* intergenic fragment that comprises both I56i and I56ii. GFP-positive cells are seen only in the telencephalic domain, II.


ml56i
ml56ii
ml12b
ml12a
Di
BT
Hy
Md
C
FN
F
G
H
I
J
II
I
Control
zl56i + zl56ii
ml56i
ml56i + ml56ii

B.K. Park, S. Sperber, B. Thomas, G. Hatch and M. Ekker, unpublished observations). Reporter transgene expression was observed in 6 out of 7 transgenic lines (Table 2.1.). A 1.6 kb *Xho*1 fragment containing zebrafish I12a targeted expression in 1 out of 2 lines of transgenic mice (Table 2.1.).

As the *Dlx1/Dlx2* intergenic regions of mouse and human showed sequence conservation that extended beyond the above two enhancers (Figure 2.3. A), we produced transgenic mice with reporter constructs containing mouse intergenic fragments outside I12a and I12b. Thus, a construct containing a 1.5 kb DNA fragment located between I12a and I12b, with 80% identity between mouse and human (Figures 2.1. and 2.3. A), did not show enhancer activity in mouse embryos (0 out of 3 primary transgenic embryos, as determined by detection of the transgene using PCR). Transgenic analysis of combinations of fragments from the mouse *Dlx1/Dlx2* intergenic region failed to indicate any enhancer activity that could be assessed to sequences outside I12a and I12b. Notably, some of these constructs included I12c (0 out of 6 PCR-positive embryos) suggesting that this sequence has no enhancer activity by itself, although it cannot be ruled out that it may cooperate with either I12a or I12 b in a quantitative manner.

2.3.4. The three forebrain enhancers show limited sequence similarity

The similar activity of the I12b, I56i, and I56ii enhancers in transgenic mice led us to investigate whether there could be sequence similarities between them. We made pairwise and dot matrix alignments of the three forebrain enhancers in both orientations. We also compared the forebrain enhancers with I12a. We did not find long stretches of sequence similarity among the four enhancers. The best dot matrix alignment was obtained by comparing I12b with I56i (Figure 2.7. A). A short fragment that extended between 60-80 bp, depending on individual pairwise alignments, was present in all three forebrain enhancers but not in I12a. The two enhancers from the *Dlx5/Dlx6* locus are in opposite orientations in this alignment (shown for the zebrafish sequences in Figure 2.7. B). The overall similarity over the short region is between 50-60%, thus smaller than the similarity between orthologous enhancer sequences (Figures 2.1. B and 2.2. B).

Interestingly, this region of similarity was also found downstream of the zebrafish *dlx2b* gene, a gene thought to be a duplicate of *dlx2a*, but that is not part of a bigene cluster (A. Amores and M. Ekker, unpublished observations).

The sequences shown in Figure 2.7. B include a putative *Dlx* binding site [(A/C/G/) TAATT (G/A) (C/G); (Feledy et al., 1999B)], near both ends of the similarity region. The core binding site for many homeodomain proteins, (TAAT/ATTA) was also found beween the two putative *Dlx* binding sites, in many of the enhancers (Figure 2.7. B). The spacing between the *Dlx* binding sites was similar in all three enhancers. We previously showed that mutagenesis of both *Dlx* binding sites in I56i abolished almost completely reporter gene expression in the forebrain of transgenic mice, suggesting that these sites are essential for activation or maintenance of enhancer activity, possibly through a cross-regulatory or auto-regulatory mechanism (Zerucha et al., 2000). The *Dlx* binding sites and surrounding nucleotides are less conserved in I56ii than those in I12b and I56i. The I56ii sequence is not activated by Dlx proteins in transfection assays, contrary to I56i and I12b (Zerucha et al., 2000 and N. Ghanem and M.Ekker, data not shown). This may also explain why it is less efficient than the other two enhancers in targeting a strong and consistent forebrain expression.

We also looked for additional protein binding sites within the four enhancers (using Genomatix, Matinspector professional software; www.genomatix.de) and could not find any that were consistently found in all of them or in the three forebrain enhancers except for the homeodomain protein binding sites TAAT/ATTA. Interestingly, the *Dlx* binding site is also a low-affinity binding site (Chen and Schwartz, 1995) for members of the *Nkx* family, that are known to be expressed in the forebrain. *Nkx*2.1, for instance, regulates regionalization in a subset of cells in the basal ganglia (Sussel et al., 1999) where the *Dlx* genes are also expressed.

In summary, the similarity between enhancers from paralogous bigene clusters occurs only in a small region of the total enhancer sequence which, in turn, is highly conserved and over a much longer distance between orthologous, but not paralogous loci.

Table 2.1. Expression of reporter constructs in primary transgenic mouse embryos and transgenic mouse lines.

Enhancer element	Primary (P)	Lines (L)	Ventral Forebrain	Frontonasal Prominence	Apical Ectodermal Ridge	Visceral Arches	Ectopic Expression
MI12a	N.D.	7	0L	0L	1L	6L	2L
ZI12a	N.D.	2	0L	0L	0L	1L	0L
MI12b	3	5	3P, 5L	3P, 4L	3P, 5L	1P•,0L	2P,1L
ZI12b	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
MI56i *	13	4	13P, 3L	2P, 1L	1P, 0L	7P, 2L	N.D.
ZI56i *	12	N.D.	12P	0P	0P	3P	N.D.
MI56ii	3	2	3P♦, 2L♦	1P, 0L	2P, 1L	0P, 0L	1P, 0L
ZI56ii *	10	2	1P, 0L	1P,1L	1P, 0L.	0P, 2L	N.D.

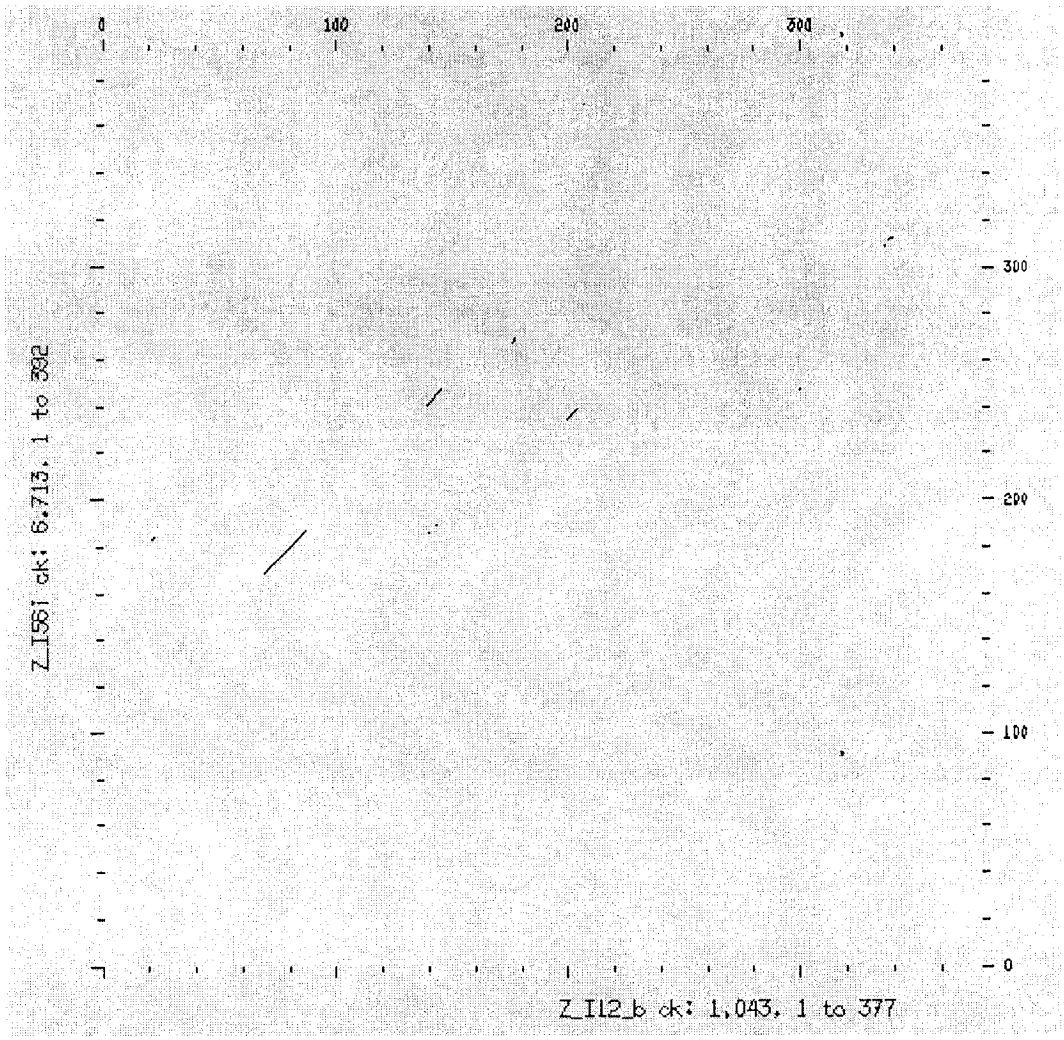
All constructs were made by inserting the enhancer fragments into the p1229/1230 plasmids that contain the LacZ reporter gene under the control of a β-globin minimal promoter. M: mouse; Z: zebrafish.

- • Weak expression in the second arch
- ♦ Both founders and lines showed variable expression patterns in the two domains of the forebrain; thus, 2P showed a weaker expression pattern in the telencephalon (domain II) compared to the diencephalon (domain I), whereas, 2L showed expression only in the diencephalon (domain I), and, 1P (shown in Fig. 6B) showed equally strong expression in both domains (I, II).
- * From Zerucha et al., 2000

Figure 2.7. Limited similarity between intergenic forebrain enhancer sequences

(A) Dot matrix comparison of the zebrafish I12b and I56i. The main two regions of sequence similarity are shown in (B) as multiple sequence alignments between I12b, I56i and I56ii and a sequence downstream of the zebrafish *dlx2b*. A three out of four consensus is shown. Putative *Dlx* binding sites, (A/C/G/) TAATT (G/A) (C/G), are indicated in bold, with mismatches highlighted. Additional TAAT/ATTA core homeodomain protein binding sites are also highlighted.

A



B

```
 Z_I12b      CCATCA.AAC CAGACATAAT TACAGTCATT CCCTCCTTAC TTATTCTAAT
 Z_dlx2b-3'  CATTGA.ATC GAATCATAAT TGCAGCCATT TGGT.TTTGT GCATCCTCGC
 Z_I56i      TCATAA.ATG CAGACATAAT TAGGGTAATT TTTGATGTAG CCCGCTATTA
 Z_I56ii     CATTTACAAT TATCTATAAT TTG.CAAAGA TGCGCCTGGT TCTTGATTGC
Consensus    C--T-A-A-- -A--CATAAT T---G--ATT T-----TT-- -C-T--T---

 Z_I12b      GCAGC.TTCC C.ATCTACGG GATAATTATG AGCAATTTT
 Z_dlx2b-3'  GCATT.TTCC T.CTGTCTGC CGCAATTACG AGCAATTTC
 Z_I56i      .CAGCGTTTT TACCGTCAAA GATAATTACC TGTAATTTT
 Z_I56ii     AGAAGGTTTT TTCCCTGGCT CACCATCACT AAACAGTGA
Consensus    -CA---TT-- T-C--T---- -A-AATTAC- AG-AATTT-
```

Supplementary Figure 2.8. Multiple sequence alignment of I12a and Z_*dlx2b*-3'in zebrafish

The latter sequence is located at position 1505-1605 from the polyadenylation signal of *dlx2b* (*dlx5*).

```
          351                                                  400
Z_I12_a   TTAGGACATC TAAAATTGCG AAGAATTATA TAGTAATTGC AGGCTTTCAG
Z_dlx2b-3'~~~~~~~AT. TGAAA.TGCC AGGGCTTAT. TTGCAA..CC A..C.ACCA.
Consensus -------AT- T-AAA-TGC- A-G--TTAT- T-G-AA---C A--C---CA-

          401                                                  450
Z_I12_a   ATCGGGGAGC CAGAATGCTC AACTGGAGCA ATTGTGGATA AAATTACAGA
Z_dlx2b-3'CTATGAG.TC CA.AATG... .ACTGCTTCC GTT.TGGATT CAAAGCTACA
Consensus -T--G-G--C CA-AATG--- -ACTG---C- -TT-TGGAT- -AA----A-A

          451                       477
Z_I12_a   TCAGAGTATA AATTAGGGAT GGGGTATG
Z_dlx2b-3'TC..TGCATG AA..AGAGAG GGTCTCTG
Consensus TC---G-AT- AA--AG-GA- GG--T-TG
```

2.4. Discussion

2.4.1. Conserved organization of the intergenic region of orthologous *Dlx* bigene clusters

We have performed a search for homologies in the intergenic region separating the two *Dlx* genes of bigene clusters in five different vertebrate species. Our analysis further illustrates the usefulness of "*phylogenetic footprinting*" (Muller et al., 2002), to identify *cis*-acting regulatory sequences. Examination of the region that separates the two *Dlx* genes that constitute the *Dlx1/Dlx2* or the *Dlx5/Dlx6* bigene clusters reveals regions of high sequence conservation as well as conserved organization of the intergenic region for orthologous loci of distantly-related vertebrates. Each of the two bigene clusters contains two regions of high sequence conservation that extends over a few hundred base pairs as well as a few shorter regions of sequence similarity. For both bigene clusters, the relative position and orientation of the conserved intergenic sequences are identical in all five species (Figures 2.1. and 2.2., and deposited sequence data).

The use of compact genomes found in tetraodontid species, such as the two pufferfish *Takifugu rubripes* and *Spheroides nephelus,* was initiated to facilitate the search for regulatory elements. This is mainly due to the fact that large regions of neutral DNA were lost in the course of genome reduction in these species, leaving the non-coding DNA regions enriched for *cis*-acting regulatory elements. We found that the presence of highly conserved sequences in *Dlx* intergenic regions probably contributes to maintain its size even in species with a compact genome. Thus, the size of the *Dlx1/Dlx2* and of the *Dlx5/Dlx6* intergenic regions in the two pufferfish, although smaller than their mammalian counterparts does not follow, proportionally, the smaller size of the genome of the two species.

Orthology assignment for the vertebrate *Dlx* genes was sometimes made difficult by the high degree of sequence similarity in the coding region of *Dlx* genes and by their highly overlapping patterns of expression. Conserved synteny, particularly with the *Hox* clusters was useful in establishing orthology relationships, as the *Dlx* bigene clusters have

been found consistently on the same chromosome as one of the *Hox* clusters (Stock et al., 1996; Amores et al., 1998). Here, we propose that the sequence of the intergenic region is also a reliable predictor of orthology as the paralogous intergenic sequences are quite different while orthologous bigene clusters contain highly conserved sequences.

We examined whether or not the above prediction also applies to a duplicate gene in zebrafish: *dlx2b* (previously, *dlx5;* see comments about nomeclature in Methods). This gene shows high sequence similarity with members of the *Dlx2* and *Dlx5* orthology groups. Mapping of *dlx2b* indicates that it is found in a group of genes with conserved synteny and that are a duplicate of a chromosome region that includes *dlx2* (Amores et al., 1998). We examined about 8 kb of DNA downstream of *dlx2b* and found some sequence similarity with the non-coding sequence elements located in the *Dlx1/Dlx2* intergenic region. Thus, sequences similar to I12a, I12b, and I12c were found (Figure 2.7. B and Supplementary Figures 2.6. and 2.8.) although similarity was generally lower than when comparing individual elements between species. No sequence was found that resembled the conserved elements from the *Dlx5/Dlx6* intergenic region except for the short sequence shown in Figure 2.7. B. Thus, in addition to synteny analysis, conservation of non-coding sequence elements can be useful in establishing relationships between duplicate genes.

2.4.2. Highly conserved *cis*-acting regulatory sequences in the intergenic region of *Dlx* bigene clusters

The largest conserved sequences found in the *Dlx1/Dlx2* and *Dlx5/Dlx6* intergenic regions are also the only ones conserved in all five species that were examined in the present study. The role of each of these sequences as *cis*-acting regulatory elements is demonstrated by their ability, once coupled to a promoter to drive expression of a reporter transgene in a tissue- and stage-specific manner. Sequence comparisons between mouse and human, or between *Takifugu* and *Spheroides*, reveals an overall high degree of sequence similarity and are therefore of less predictive value in the identification of regulatory elements. This may be due to the small evolutionary distance between the two

mammals (~50-60 MY) as well as the two pufferfish (~5-35 MY), and to the slow rate of divergence for neutrally evolving regions among vertebrates in general (0.1% to 0.5% per million years) (Tautz, 2000). Intergenic fragments outside the enhancers with 75-80% overall conservation between mouse and human failed to act, by themselves, as enhancers when tested in transgenic mice. Therefore, caution should be exerted when identifying putative *cis*-acting sequences based on comparisons between vertebrates of the same order. Comparisons that include multiple species with some that are distantly related might be a more efficient approach to identify non-coding sequence elements of functional importance, while keeping in mind that absence of sequence conservation does not necessarily indicate absence of functional conservation (Flint et al., 2001).

The relatively high degree of sequence conservation between the mouse and human *Dlx1/Dlx2* intergenic region (80%) or *Dlx5/Dlx6* intergenic region (78%) contrasts with the *Dlx3/Dlx7* intergenic region that is only 69% identical, overall, between the two species (Sumiyama et al., 2002) despite the presence of sequences with higher percentage identity that may have a regulatory function (Sumiyama et al., 2002). However, comparisons of the mammalian *Dlx3/Dlx7* intergenic region with those of zebrafish (Sumiyama et al., 2002), or *Takifugu rubripes* (N. Ghanem and M. Ekker, unpublished observations) did not show conserved sequences comparable in length or percent identity to the four enhancers that we identified in the *Dlx1/Dlx2* or in the *Dlx5/Dlx6* bigene clusters. Therefore, the *Dlx3/Dlx7* bigene cluster may differ from its two paralogous *Dlx* clusters by a relatively low importance of the intergenic region in the mechanisms that control gene expression or by a higher divergence in regulatory mechanisms between the different vertebrate lineages. Consistent with this latter hypothesis is the observation that zebrafish *dlx3/dlx7* have marked differences in their early patterns of expression compared to their mammalian orthologs (Quint et al., 2000).

2.4.3. Function of intergenic elements in *Dlx* regulation and evolution

The organization of *distal-less*-related genes in bigene clusters may have preceded the evolution of vertebrates as two of the three characterized *Dll* genes of the ascidian

Ciona intestinalis, Dll-A, and *Dll-B* are organized similarly with a short intergenic region (Di Gregorio et al., 1995). Recently, an enhancer located upstream of *Dll-A* was identified and shown to recapitulate most aspects of the endogenous expression pattern (Harafuji et al., 2002). Enhancers have yet to be found in the intergenic region that separates the *Ciona Dll-A* and *Dll-B* genes and preliminary sequence comparisons did not reveal similarities in sequence between this region and the four *cis*-acting regulatory sequence found in vertebrate *Dlx* genes (M. Ekker, unpublished observations).

Although the three *Dlx* bigene clusters of vertebrates are likely the result of duplication of an ancestral bigene cluster, we did not observe a high degree of conservation between paralogs, regardless of the species. This extends the observation, previously made by Sumiyama and collaborators who compared the three human bigene clusters (Sumiyama et al., 2002). This lack of sequence similarity between paralogs is surprising, considering the similarities in expression patterns of genes found in paralogous bigene clusters.

Enhancers with overlapping patterns of activity (Figure 2.6.) show only a limited conservation in sequence (Figure 2.7.) that contrasts sharply with the high degree of conservation between orthologous sequences. Furthermore, enhancer sequences found in one *Dlx* bigene cluster are not found in the two paralogous clusters. Although one or several *Dlx* intergenic enhancers could originate from a sequence found in the ancestral *Dlx* bigene cluster, they would have diverged following the duplication events that took place early in vertebrate evolution, and that led to the three *Dlx* bigene clusters of modern vertebrates. This divergence happened before the separation of the lineages leading to modern-day teleost and tetrapods. Since then, purifying selection maintained most, if not all, regulatory mechanisms that involve these intergenic sequences, at least for the *Dlx1/Dlx2* and *Dlx5/Dlx6* bigene clusters. The region of limited similarity found between the three forebrain enhancers may suggest that they resulted from a tandem duplication (I56i and I56ii) that also predated the split between the ray-finned fish lineages, and/or represent what subsists from a sequence present in the ancestral *Dlx* bigene cluster.

Although the current study suggests that *cis*-acting regulatory elements of diverse sequence may exert similar enhancer function, the converse may also be true. Thus, I56i from mouse targets expression of a reporter transgene to the forebrain and mesenchymal cells of the branchial arches (Figure 2.6. A) whereas the orthologous sequence from zebrafish only directs expression to the forebrain, in either transgenic mice or zebrafish (Zerucha et al., 2000) despite the fact that the two sequences are more than 80% identical (Figure 2.2. B). Thus, the small differences in sequence between the enhancers from the two species may have a profound effect on enhancer function.

Evidence has been previously presented for cross-regulatory interactions between *Dlx* genes. Thus, the *Dlx1* and *Dlx2* genes are expressed earlier in the forebrain and are involved in either the activation or maintenance of *Dlx5* and *Dlx6* expression through the enhancer(s) found in the *Dlx5/Dlx6* intergenic region (Zerucha et al., 2000). In contrast, there is, at present, no evidence that *Dlx5/6* regulate *Dlx1/2* in the brain. In the branchial arch mesenchyme, *Dlx5/6* regulate *Dlx3*, but not *Dlx1/2* (Depew et al., 2002a). Thus, the divergence of the intergenic enhancer sequences may have contributed to the specificity of cross-regulation between *Dlx* genes, allowing for sequential expression of paralogs.

The present study indicates an important role for the intergenic region in the *cis*-regulatory mechanisms that are responsible for many aspects of the expression of genes from two *Dlx* bigene clusters. Intergenic regulatory elements are not solely responsible for *Dlx* regulation. Thus, a fragment of the 5'-flanking region of mouse *Dlx2* was shown to recapitulate expression in the epithelial cells of the branchial arches (Thomas et al., 2000). A targeted mutation, that inactivates the function of mouse *Dlx1* and *Dlx2,* eliminates the entire intergenic region (Anderson et al., 1997b). Intriguingly, homozygous mutants expressed truncated *Dlx1* transcripts in the forebrain despite the absence of the I12b sequence (Zerucha et al., 2000). Although our results indicate that I12b is sufficient to confer expression of a reporter transgene to the forebrain (Figure 2.6. C), distinct sequences located upstream of *Dlx1* also share this property (N. Ghanem and M. Ekker, unpublished observations), suggesting a cooperative or synergistic effect between multiple and distinct enhancers in forebrain regulation of *Dlx1* and/or *Dlx2*.

Distinct mechanisms may take place at the *Dlx5/Dlx6* locus. The *lacZ* reporter gene, introduced in a targeted mutation of *Dlx5/Dlx6* that also removes the intergenic sequence (including I56i and I56ii), is only weakly expressed in the forebrain (Robledo et al., 2002). This suggests that enhancers outside the intergenic region may exist but that the intergenic enhancers play an essential role in conferring proper levels of gene expression, in as much as detection of transcripts by *in situ* hybridization can be considered quantitative. Taken together, these observations suggest complex mechanisms of *Dlx* expression control. These mechanisms involve multiple enhancers with overlapping but not necessarily redundant activity and a high degree of conservation in distant vertebrates for at least some of these enhancers.

Acknowledgments

We thank Luc Poitras and Fabien Avaron for useful discussions and Adrianna Gambarotta and Lucille Joly for technical assistance. N.G. was supported in part by a scholarship from the Lebanese University, Beyrouth. This work is supported by grants from the Canadian Institutes of Health Research (MOP14460) and the March of Dimes Birth Defects Foundation (FY01-207). M.E. is an Investigator of the CIHR.

3. Distinct *Dlx*-expressing Progenitors in the Medial and Caudal Ganglionic Eminences Give Rise to Different Subtypes of Adult Cortical Interneurons

Noël Ghanem[1], Jason Long[2], Gary Hatch[1], John L. R. Rubenstein[2], Marc Ekker[1, 3]

1 Centre for Advanced Research in Environmental Genomics, Dept. of Biology, University of Ottawa, Canada

2 Nina Ireland laboratory of Developmental Neurobiology, Centre for Neurobiology and Psychiatry Genetics, Department of Psychiatry, University of California at San Francisco, USA

3 Corresponding author: mekker@uottawa.ca

Short title: Distinct *Dlx*-progenitors in the subpallial telencephalon

Key Words: ganglionic eminence; regionalization; migration; progenitors; enhancers; cerebral cortex; GABAergic neurons; homeobox; mice.

Abstract

Distinct subtypes of cortical GABAergic interneurons provide inhibitory signals that are indispensable for neural network function. The *Dlx* homeobox genes have a central role in regulating their development and function. We have characterized the activity of three *cis*-regulatory sequences involved in forebrain expression of vertebrate *Dlx* genes: URE2, I12b, and I56i. The three regulatory elements display regional and temporal differences in their activities within the MGE and CGE, and label distinct populations of tangentially migrating neurons at E12.5 and E13.5. We provide evidence that the dorsal and ventral MGE (dMGE and vMGE) are distinct sources of tangentially migrating neurons during midgestation. In the adult cortex, URE2 and I12b are differentially expressed in parvalbumin-, calretinin-, NPY- and nNOS-positive interneurons; I12b was specifically active in somatostatin- and VIP-positive interneurons. These data suggest that interneuron subtypes utilize distinct combinations of *Dlx1/Dlx2* enhancers from the time they are specified through adulthood.

3.1. Introduction

Two mains classes of neocortical neurons constitute the basic functional units upon which complex neuronal networks are established in the cerebral cortex during development: 1) glutamatergic neurons, which are largely projection neurons; and, 2) GABAergic neurons, which are local circuit neurons (DeFelipe and Farinas, 1992; Peters, 1984). GABAergic neurons in the adult brain are inhibitory and comprise 20-30% of all neocortical neurons (Hendry et al., 1987; Meinecke and Peters, 1987; Parnavelas et al., 1977). They are subdivided into several subtypes based on molecular, morphological and electrophysiological features (Markram et al., 2004). For instance, subtypes of interneurons express one or more of three calcium-binding proteins; parvalbumin (PV), calretinin (CR) and calbindin (CB), and four neuropetides; somatostatin (SOM),

neuropeptide Y (NPY), vaso-intestinal peptide (VIP) and cholecystokinin (CCK) (Markram et al., 2004).

Inhibitory interneurons are indispensable for a normal cortical function through establishment of the appropriate spatio-temporal balance of the excitation/inhibition and through the generation of circuit oscillations (Borg-Graham et al., 1998; Hensch, 2005; Kisvarday and Eysel, 1993; Monier et al., 2003; Murthy and Humphrey, 1999; Sato et al., 1996; Sillito, 1984; Wehr and Zador, 2003).

In rodents, most interneurons in the neocortex, hippocampus and olfactory bulb are derived from the embryonic basal ganglia [for review, see (Corbin et al., 2001; Marin and Rubenstein, 2001b; Parnavelas, 2000; Wonders and Anderson, 2006)]. Immature interneurons migrate along several distinct routes towards their destinations where they differentiate into different subtypes. Moreover, these progenitors derive from several ventral structures: **1)** the dorsal LGE (dLGE) is the site of origin of olfactory bulb interneurons (Corbin et al., 2000; Dellovade et al., 1998; Stenman et al., 2003; Sussel et al., 1999; Wichterle et al., 2001; Yun et al., 2001); **2)** the MGE and CGE are major sources of cortical and hippocampal interneurons (Anderson et al., 2002; Anderson et al., 2001; Lavdas et al., 1999; Sussel et al., 1999; Wichterle et al., 1999; Wichterle et al., 2001). Distinct subtypes of interneurons have different spatio-temporal origins (Anderson et al., 2002; Butt et al., 2005; Valcanis and Tan, 2003; Wichterle et al., 2001; Xu et al., 2004). For example, the PV- and SOM-expressing interneurons seem to derive primarily from progenitors located in the MGE between E12.5 and E16.5 (Butt et al., 2005; Xu et al., 2004). In contrast, interneurons expressing CR derive exclusively from progenitors found in the CGE between E14.5 and E16.5 (Butt et al., 2005; Xu et al., 2004).

The mechanisms that specify cortical interneuron subtypes produced within the MGE are unknown. The existence of microenvironments within a given progenitor domain may explain the diversity of interneuron subtypes derived from this domain. Regional specification of progenitor cells may be directly associated with the establishment of such microenvironments. Alternatively, the environment where the interneurons migrate and settle may also participate in their fate determination and final

maturation. Thus, progenitor cells may be under the control of different spatio-temporal cues and/or mechanisms of regulation that are primarily dictated by the expression profiles of several genes including transcription factors.

Dlx genes are among factors that are required for the proper differentiation and subsequent migration of interneurons in the telencephalon. Mutants lacking *Dlx1/Dlx2* function display a block in the differentiation of progenitors in the basal ganglia and lack of migrations to the cerebral cortex, olfactory bulb and hippocampus (Anderson et al., 1997a; Anderson et al., 1997b). This is evidenced by an 80% reduction in neocortical GABAergic neurons and >95% reduction in hippocampal and olfactory bulb interneurons.

The *Dlx* family comprises six genes organized into three bigene clusters on separate chromosomes: *Dlx1/Dlx2*, *Dlx5/Dlx6* and *Dlx3/Dlx4*. Four of them, *Dlx1, Dlx2, Dlx5* and *Dlx6* are sequentially expressed, albeit in an overlapping fashion, in the subcortical telencephalon starting around E8.5-E9 (Eisenstat et al., 1999; Liu et al., 1997). They display highly overlapping expression patterns in the basal ganglia (subpallium) but with subtle differences [for review, see (Panganiban and Rubenstein, 2002)]. We previously identified *cis*-regulatory elements (CRE) located in the intergenic regions of the *Dlx1/Dlx2* and *Dlx5/Dlx6* loci of vertebrates These CREs, I56i, I56ii and I12b, are highly conserved between vertebrates and were able to target reporter transgene expression to the subpallial telencephalon (subpallium) and the diencephalon (prethalamus and hypothalamus) of transgenic animals (Ghanem et al., 2003; Zerucha et al., 2000).

Here, we describe the characterization of a novel forebrain-specific CRE, URE2 (Upstream Regulatory Element 2) located in the 5'flanking region of *Dlx1* in several vertebrate species. By comparing the spectrum of activity of URE2 with those of I12b and I56i in the telencephalon of transgenic mice, we show that the three CREs displayed differential activities in the dorsal and ventral parts of the MGE and CGE, and labeled distinct populations of tangentially migrating neurons at E12.5 and E13.5. We propose the existence of microenvironments within dorsal and ventral parts of the MGE and CGE

that produce different cell-types. Furthermore, we demonstrate that URE2 and I12b are differentially active in several subtypes of adult cortical interneurons.

3.2. Material and methods

3.2.1. Transgenic mice

To generate transgenic mice, the enhancer sequences of URE2, I12b and I56i were subcloned into the p1229/p1230 vectors (Yee and Rigby, 1993) that contain a human β-globin minimal promoter and the *lacZ* reporter gene. A 1.5kb mouse fragment containing the URE2 sequence was amplified by PCR using the following oligonucleotide primers: 5'AAAGCTTCCTTTGAATTCATCCCGCTTG · 3' and 5'CAAGCTTCATGCAAATTCATACTCTGCC 3' (*Hin*dIII restriction site added). Amplification consisted of 30 cycles at an annealing temperature of 62°C for 45s. This was followed by cloning of the PCR product into PCR 2.1 and, then, into p1229 using the *Hin*dIII restriction site. For subcloning details and PCR conditions used in the making of I12b and I56i transgenic constructs refer to chapter 2 of this thesis. The p1230-I12b-alkaline phosphatase construct was generated by replacing the *lacZ* reporter gene with the human placental alkaline phosphatase gene (PLAP) in the p1230-I12b-*lacZ* vector using *Bam*HI restriction sites. The human PLAP sequence was amplified by PCR from the pGT0-PFS plasmid using the following oligonucleotides primers to which *Bgl*II restriction sites were added: 5'AGATCTGCTGCTGCTGCTGCTGCTGCTG 3' and 5' AGATCTGATGAGTTTGGACAAACCAC 3'. Amplification was done in 30 cycles using an annealing temperature of 60°C for 45s. Correct subcloning of the PLAP gene was confirmed by sequencing using M13 reverse universal primer.

3.2.2. Morphological analysis of transgenic mice

Embryos from the mating of a transgenic male with normal *CD1* females were harvested at various embryonic stages. Pregnant females were sacrificed by cervical dislocation. E11.5 and E12.5 mouse embryos were fixed in 4% PFA for 1h at RT, then

washed in 1x PBS for 30 minutes and stained for β-galactosidase activity overnight (O/N) at 28°C in a solution of 1 mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, and 0.02% NP-40 in PBS. Mouse brains were dissected out after staining and soaked in 20% sucrose solution at 4°C O/N. The following day, brains were cryoprotected and frozen using cold isopentane (kept at -80°C). 30-50μm sections were then cut using a cryostat (Leica CM3050 S).

3.2.3. Sectioning and tissue preparation of adult brains

Postnatal day 35 (P35) adult mice were screened by PCR using genomic DNA extracted from ear tissue and specific oligonucleotide primers for *lacZ* (5' ACGGCAGAGCCATCTATTGC 3' and 5'CGCTCATCCGCCACATATCC3') and/or PLAP (5' ACGGCAGAGCCATCTATTGC 3' and 5'CTTGGACAGAGCCACATATGG 3'). Amplification consisted of 30 cycles at an annealing temperature of 60°C for 45s. Positive mice were anesthetized with isoflurane and were subjected to cardiac perfusion in 1x PBS followed by 4% cold PFA. Brains were dissected out afterwards and post-fixed in 4% PFA for 4h at RT. Sections of 30-40μm thickness were cut into cold 0.1M phosphate buffer saline (PB) using a vibratome (Vibratome series 1000, Technical Products International).

3.2.4. Double immunohistochemistry

E12.5 and E13.5 mouse brains were dissected from AP+/*lacZ*+ embryos and sectioned as described above. Frozen sections were dried for 1h, then washed for 3 x 5 min each in 0.1M PB, pH=7.4, to eliminate residues from tissue protection medium. Sections were incubated in blocking solution: 1% bovine serum albumin (BSA); 5% goat serum and 0.3% triton-X in 0.1M PB, for at least 1h at RT. This was followed by incubation with the primary antibody at 4°C O/N. 3 x 10 min washes in 0.1M PB were performed the next day. This was followed by incubation with the secondary antibody for 2h at RT away from light. Sections were washed for 3 x 10 min, processed for a second

round of immunohistolabeling and, finally, mounted in Immuno Floure mounting media (Fisher).

Immunohistochemistry on floating sections of adult mouse tissue was performed as described earlier (Cobos et al., 2005). We used the following antibodies: guinea pig anti-β-Gal (1:1000, generous gift from Thomas Sargent laboratory), rabbit anti-PLAP (1:100 Serotec), rabbit anti-parvalbumin (1:1000, Swant Swiss Abs), rat anti-somatostatin (1:100, Chemicon), rabbit anti-neuronal nitric oxide synthase enzyme (1:150, Zymed-Invitrogen), rabbit anti-calretinin (1:4000, Swant Swiss Abs), rabbit anti-neuropeptide Y (1:4000, Immunostar), rabbit anti-tyrosine hydroxylase (1:350, Chemicon), and rabbit anti-GABA (1:4000, Sigma); rabbit anti-vaso-active intestinal peptide (1:200; ImmunoStar). Secondary antibodies were purchased from Molecular Probes (Invitrogen): goat anti-rabbit Alexa fluor 488, goat anti-guinea pig Alexa fluor 594 and goat anti-rat Alexa fluor 488.

3.2.5. Organotypic culture

Organotypic slice cultures of embryonic mouse forebrain were prepared as previously described (Tobet et al., 1994). Shortly, E12.5 and E13.5 mouse embryos were placed in ice cold Krebs buffer containing 126 mM NaCl, 2.5 mM KCl, 1.2 mM NaH_2PO_4, 1.2 mM $MgCl_2$, 2.5 mM $CaCl_2$, 11 mM glucose, and 25 mM $NaH0_3$. Brains were then removed and embedded in 5% low-melt agarose (Sigma). 250 μm thick coronal sections were then cut on a vibrotome into cold Krebs buffer, and the sections were transferred to sterile Krebs buffer (pH 7.4; filtered Krebs with 10 mM HEPES, penicillin, streptomycin, and gentamicin) on ice. After 15 min, the sections were transferred to polycarbonate culture membranes (diameter, 13 mm; pore size, 8 μm; Whatman) in Falcon organ tissue culture dishes containing 1 ml of medium (Gibco MEM with glutamine, 10% fetal calf serum, and pen/strep antibiotics). They were subsequently placed in a sterile incubator (5% CO2, 37°C) for 1 hr, after which the medium was changed to Neurobasal/B-27 (Gibco). DiI crystals (Molecular Probes, about 100 μm) were placed using an insect pin and a dissection microscope. Transplanted tissue was cut

using fine dissecting scissors and forceps. After incubation for various times, the slices were fixed in 4% PFA in PBS at 4°C for at least 4 hr, and mounted on glass slides under Vectashield mounting medium (Vector). Slices were visualized on a Zeiss microscope using rhodamine or fluorescein fluorescence filters. In each experiment, 6-8 mouse brains were cut and 4 sections on average per brain were transferred into culture and labeled with DiI. E12.5 and E13.5 sections were cultured for 48-72h. Four independent DiI labeling experiments and two transplantation experiments were performed.

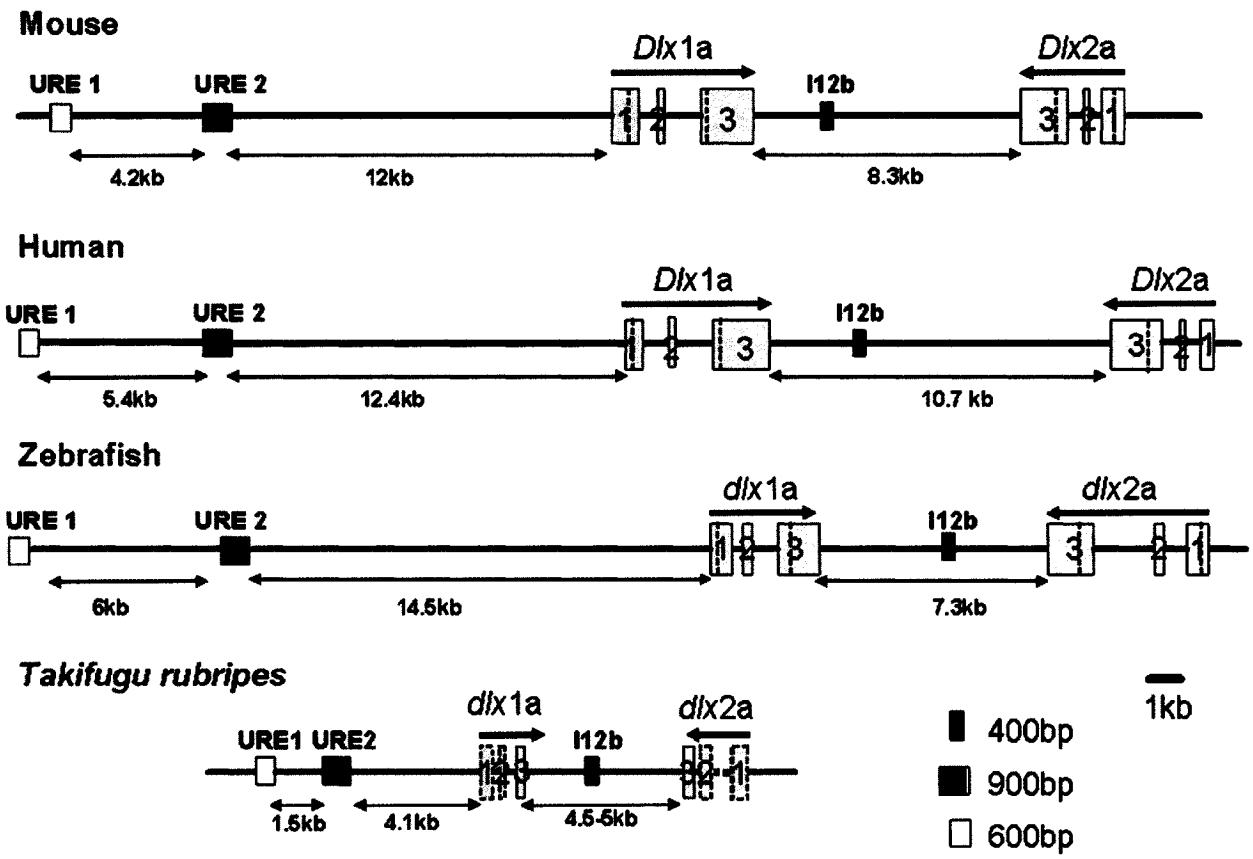
3.3. Results

3.3.1. Regulation of *Dlx* gene expression in the forebrain by tissue-specific enhancers

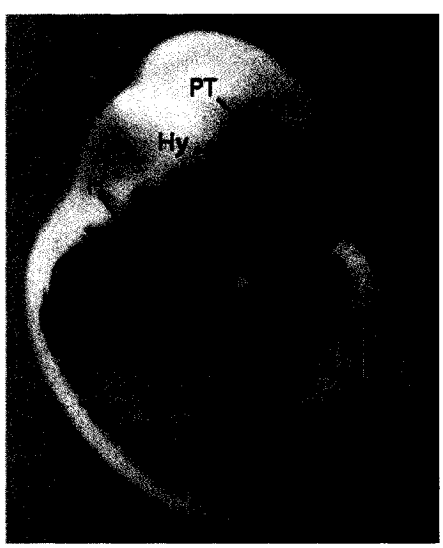
The previously identified CREs, I56i and I12b, located in the intergenic regions of *Dlx5/Dx6* (Zerucha et al., 2000) and *Dlx1/Dlx2* (Ghanem et al., 2003), respectively, show comparable regulatory activities in the subpallial telencephalon and diencephalon when examined in whole mount transgenic mouse embryos (Ghanem et al., 2003). In a search for additional regulatory elements controlling *Dlx1* and *Dlx2* gene expression in the forebrain, we examined, by phylogenetic footprinting, the 5' flanking regions of each gene for conserved non-coding DNA sequences among mouse, human, *Takifugu rubripes*, and zebrafish. We identified two conserved sequences in the 5'flanking region of *Dlx1* that we named "Upstream Regulatory Element 1 and 2", URE1 and URE2. The two sequences were found in all the vertebrate species listed above as well as others. The percentage of sequence conservation varied between 72-97% for URE2 and 72-92% for URE1 depending on the species compared. In the mouse, they were located ~17kb (URE1; ~900bp) and ~12kb (URE2; ~600bp) upstream of the *Dlx1* ATG start site (Figure 3.1. A and data not shown). The URE1 and URE2 sequences were 92% and 97% conserved between mouse and human, respectively.

Page 123 missing.

A


Mouse
Dlx1a
Dlx2a
URE 1
URE 2
I12b
4.2kb
12kb
8.3kb
Human
5.4kb
12.4kb
10.7 kb
Zebrafish
dlx1a
dlx2a
6kb
14.5kb
7.3kb
Takifugu rubripes
URE1 URE2
1.5kb
4.1kb
4.5-5kb
1kb
400bp
900bp
600bp

B


PT
Hy

Supplementary Sequence 3.1. Multiple sequence alignment of URE2 in five vertebrate species

The consensus sequence represents identity in five out of five species. Sequences similar to the binding site for *Dlx* protein ([A/C/G]TAATT[G/A][C/G]) (Feledy et al., 1999b) are highlighted in grey. TAAT/ATTA putative homeodomain binding motifs are underlined. M, mouse; H, human; T, *Takifugu rubripes*; S, *Spheroides nephelus*; and Z, Zebrafish.

```
          1                                                   50
  H_URE_2 ~GAGAACACT AGATTCTTAT TCAAGCATTC TATCGAGCT. CTGCATTCAT
  M_URE_2 ~GAGAACACT GGATTCTTAT TCAGGCATTC TGTGGAGTT. CTGCATTCAT
  S_URE_2 GCAAAGAACA AAATTATTCT TCAAACTTTC TATCAAGCTT CTGCATTCAC
  T_URE_2 GCAAAGAACA AAATTATTCC TCAAGCTTTC TATCAAGCTT CTGCATTCAC
  Z_URE_2 GCAAAGCACA GAATTATTCT CCCTGCCTTC TATCAAGCTC CTGCATTCAG
Consensus ~-A-A--AC- --ATT-TT-- -C---C-TTC T-T--AG-T- CTGCATTCA-

          51                                                  100
  H_URE_2 GGCTGTGTCT AAAGGGCATG TCAG.CCT.T TGATTCTCTC TGAGAGGTAA
  M_URE_2 GGCTGTGTCT AAAGGGCATG TCAG.CCT.T TGATTCTCTC TGAGAGGTAA
  S_URE_2 GGCCAAGTCT AAAGGGCATG TCAGTCCTCC TG.C.CTCTC ACAGAGGTAA
  T_URE_2 GGCCAAGTCT AAAGGGCATG TCAGTCCTCC TG.C.CTCTG AGAGAGGTAA
  Z_URE_2 CAGCGCGTCT GAAGGGCGCG TCAGTCCGCC TGCCTCTCTC AGAGAGGTAA
Consensus ------GTCT -AAGGGC--G TCAG-CC--- TG---CTCT- --AGAGGTAA

          101                                                 150
  H_URE_2 TTATCC.TTT TCCTGTCACG GAACAACAAA TGATAGCTAA CTACAGAGGC
  M_URE_2 TTATCC.TTT TCCTGTCACG GAACAACAAA TGATAGCTAA CTACAGAGGC
  S_URE_2 TTATCC.TTT TCCTGTCAGA GAACAACAAA TGATTGCCGA GTA.TGGGGT
  T_URE_2 TTATCC.TTT TCCTGTCAGA GACCAACAAA TGATCGCCAA CTA.T.GGG.
  Z_URE_2 TTATCCTTTT TCCTGTCACA GGGCTCCAAA TGATCTCTAA CCACTGGGGA
Consensus TTATCC-TTT TCCTGTCA-- G--C--CAAA TGAT--C--A --A----GG-

          151                                                 200
  H_URE_2 ACATTTGCAG TAG.TCACAT TCATCAACT. GCAG.AAAAA AAAATTCAA.
  M_URE_2 ACATTTGCAG TAG.TCACAT TCATCAACT. GCAGAAAAAA AAAATTCAA.
  S_URE_2 GCATTTTCTG GAGCTGAAAT CTGTCAGTTG GGAGAAAAAG GGG.TTTAAT
  T_URE_2 GCATTTTCTG GAGCTG.AAT TTGTCAGTTG GGAGAAAAAG GGGATTTAAT
  Z_URE_2 GCTTTGTTAG GGGCTCAGCT TTGGCAATAG GAAGTCGAAA GAGCTTTAAT
Consensus -C-TT----G --G-T----T ----CA---- G-AG---AA- ----TT-AA-

          201                                                 250
  H_URE_2 TTT...A.A. TTGTGCAAC. A....CA.G. .C.T.GC.AC A.TGGGC.TT
  M_URE_2 TTT...A.A. TTGTACAAC. A....CA.G. .C.T.GC.AC A.TGGGC.TT
  S_URE_2 TTTCT...A. CT.T...TC. A.TGT.AT.T TTTT..CCCC AATGGTCATT
  T_URE_2 TTTCTGGCAG CTGTACATCT ACTTTCATGT TTTTCCCCCC ACTGGTCATT
  Z_URE_2 TTTCAGACAG CTG..CA.C. A.TGT.ACTT TCATCGCGGC A.TTAT.AAA
Consensus TTT-----A- -T------C- A-----A--- ---T--C--C A-T-------

          251                                                 300
  H_URE_2 TT..GAGCAT TTCTGT..T. G.TTC.TCC. CTGTCTCGC. TATTCCTCCC
  M_URE_2 TC..GAGC.. TTCTGT..T. G.TTC.TCC. CTGCCTTGC. TAGTCCTCCC
  S_URE_2 TC.T.TTCTT T.CTTTTTTT TATT..TCC. C...CCTGCA CAGAGAGTGC
  T_URE_2 TCTTGTTTTT TGGGGGGGGG GGTTCCCCCA CTATCCTGCA CAGAGAGCTC
  Z_URE_2 NC..CAT.AT AGCGGTCATC NCTACGAGCG C.A..C.GCG C.G.CACTTT
Consensus ---------- ---------- --T-----C- C------GC- ----------

          301                                                 350
  H_URE_2 T.CCAGATCT ATTTTTTAAA C...TTTTTT TCTG...GT. TATTT.T.TT
  M_URE_2 T.CCAGATCT ATTTTTTAAA CTTTTTTTTT TCTG...GT. TATTT.T.TT
  S_URE_2 TGGCGGAGGT .TTTTTCTGC CTGTGAAGCC TTTGAGGGTG TATTTATGAG
  T_URE_2 TGCCGGGGCT GTTTGTGCGC CTGTGAAGCC TCTGAGAGTG TATTTATGTG
  Z_URE_2 TGTTTGTGAC GTGGGGCCGC GCGCG.CGCG CCTGCCTGTG CA.CGGCCT.
Consensus T----G---- -T-------- ---------- --TG---GT- -A--------
```

```
          351                                                        400
  H_URE_2 CCC..C..T. ...T.TTTGT CTCT..TCTT .C.CATTT.T TAC..TC.TC
  M_URE_2 CCC..CTTT. ...T.TTTGT CTCT..TCTT .C.CATTT.T TAC..TC.TC
  S_URE_2 CCCTACATTA ACATACCTAT TTATAGTCCT ACAC.TTTCT TACAGTCATC
  T_URE_2 CCCTACATTA ACATACCTAT TTATAGTCCT ACAC.TTTCT TACAGTCATC
  Z_URE_2 CCCCGCATTA GCATATTTAT TTATAGACTT ACACTTTTCT TACAGTCATT
Consensus CCC--C--T- ---T---T-T -T-T---C-T -C-C-TTT-T TAC--TC-T-

          401                                                        450
  H_URE_2 TGTACTTT.C ..TTGT.... ...TA.A.AG ..TA.AT.TT .TCC...T..
  M_URE_2 TGTACTTT.C ..TTGT.... ...TA.A.AG ..TA.AT.TT .TCC...T..
  S_URE_2 TTTTTTTTCC TGTTGTGAC. CATTTCACAG ATTATATATT GTCCAAGT.A
  T_URE_2 TTTTTTTT.C TGTTGTGAC. CATTTCACAG ATTATATATT GTCCAAGT.A
  Z_URE_2 TTTGCGCAAC AATTTTCACA CTCGACATTG TCTGTGTAAT TTCCCTTTCA
Consensus T-T------C --TT-T---- ------A--G --T---T--T -TCC---T--

          451                                                        500
  H_URE_2 .T.TGTGGC. .TC.TCAT.T CT.TTT.TCC CCCATT..GA .AG.GCTATG
  M_URE_2 .T.TGTGGC. .TC.TCGT.T CT.TTT.TTC CCCATT..GA .AG.GCTATG
  S_URE_2 AT.T.TCCCT TTCCTCATCT GTGTTT.TTT ACCTTTCAGA GAAAGGAGTG
  T_URE_2 AT.T.TCCCT TTCAGCATCT GTGTTT.TTT ACCTTTCAGC GAAAGGAGTG
  Z_URE_2 GTGTCTCTCT GTTTTCTCCT CTCTTTCCTT CCCTTTCAGA AAGAGCCATG
Consensus -T-T-T--C- -T---C---T -T-TTT---- -CC-TT--G- -A--G---TG

          501                                                        550
  H_URE_2 AATGTA..GA AAATTATCAC AATTACTCAT ATAATTGAGC C.TCTTTGTA
  M_URE_2 AATGTA..GA AAATTATCAC AATTACTCAT ATAATTGAGC C.TCTTTGTA
  S_URE_2 AATGGAGGGA AAATTATCAC AATTACTCAT ATAATTGAGC CGTCTTTGTA
  T_URE_2 AATGGAGGGA AAATTATCAC AATTACTCAT ATAATTGAGC CGTCTTTGTA
  Z_URE_2 AATGGATGGA AAATTATCAC AATTACTCAC ATAATTGAGC CGTCTTTGTG
Consensus AATG-A--GA AAATTATCAC AATTACTCA- ATAATTGAGC C-TCTTTGT-

          551                                                        600
  H_URE_2 GCAAGTGCAA CTCCAGTAG. ..CCTTTCTC CATCA.TGAA AATGGTTTCA
  M_URE_2 GCAAGTACGA CTCCAGTAG. ..CCTTTCTC CATCA.TGAA AATGGTTTCA
  S_URE_2 GCAAGTGCAG CTTCAGTAGC CCCCTTTTTC CATCAGACAA AATGGTTTCA
  T_URE_2 GCAAGTGCAG CTTCAGTAGC CCCCTTTTTC CATCAGCCAA AATGGTTTCA
  Z_URE_2 GCAAGTGCAG CTCGAGGAG. .CCCTTTTTC CATCAGCCAA AATGGTTTCA
Consensus GCAAGT-C-- CT--AG-AG- --CCTTT-TC CATCA---AA AATGGTTTCA

          601                                                        650
  H_URE_2 TTATA.GGGT TTTTCATATT CTCTGACAC. CATCTACACA GAGGAACAGG
  M_URE_2 TTATA.GGGT TTTTCATATT CCCTGACAC. CATCTACACA GAGGAGCAAG
  S_URE_2 TTATATGGGT TTTTCATATT CCCTGACACA GGGCTCTGCT GAGGGCCTGG
  T_URE_2 TTATATGGGT TTTTCATATT CCCTGACACA GGGCTCTGCT GAGGGCCTGG
  Z_URE_2 TTATA.GGGG TTTTCATATT CCCTGACACG GGGCGCAGGG GGGGCCGCGG
Consensus TTATA-GGG- TTTTCATATT C-CTGACAC- ---C------ G-GG-----G

          651                                                        700
  H_URE_2 CGTGCAGATG AGATGTGCTA GGAACAGGCT AGATCAGTAA GGTCACAGTA
  M_URE_2 CGTGCAGATG AGATGTGCTG GGAACAGGCT AGATCAGTAA GGTCACAGTA
  S_URE_2 CGCGTGGATG AGATGCCGGA TGAACAAAGT GAATCATTAA GGTCACAGTA
  T_URE_2 CGCGTGGATG AGATGCAGGA TGAACAGAGT GAATCATTAA GGTCACAGTA
  Z_URE_2 TGTGCGCAGG AGTCGCTGGG GGAGAAGGGT GGATCATTAA GGTCACAGTG
Consensus -G-G---A-G AG--G----- -GA--A---T --ATCA-TAA GGTCACAGT-
```

```
          701                                                       750
  H_URE_2 GGAATAATTA GCTCTGCTAT GGAAAGAGCA TCTAG.GCCT TT.TACT...
  M_URE_2 GGAATAATTA GCTCTGCTAT GGAAAGAGCA TCCAG.GCCT TT.TACT...
  S_URE_2 GGAATAATTA TCCCAGCTAT GGAAAGAGCA T.CAG..CCT TTATTCTC..
  T_URE_2 GGAATAATTA TCCCAGCTAT GGAAAGAGCA T.CAG..CCT TTATTCTC..
  Z_URE_2 GGGATAATTA TCTTGACTAT GGAAAGAGCA TCCAGCACCT TTTTTCTCTT
Consensus GG-ATAATTA -C----CTAT GGAAAGAGCA T--AG--CCT TT-T-CT---

          751                                                       800
  H_URE_2 .GCTA.CATA AATGTACTGT CCATGGCTTT TAGTCACAAA AAAAACTTAC
  M_URE_2 .GCTA.CATA AATGTACTGT CCGTGGCTTT TAGCCAC.AA AAAAACTTAC
  S_URE_2 CCCCGGCACA AATGCAATGT CCATGG.C.T T......... .AGAACTTTG
  T_URE_2 CCCCAGCACA AATGCAATGT CCATGG.C.T T......... .AGAACTTTG
  Z_URE_2 CCCCGCCACA AAAGCACCGT CCATGGCCTT TCGGCAC..A AAGAACTTGG
Consensus --C---CA-A AA-G-A--GT CC-TGG---T T--------- -A-AACTT--

          801                                                       850
  H_URE_2 TAACAAATGG AGCTCCCGCC TACTACTT.T .....GA..A .A.....A.A
  M_URE_2 TAACAAATGG AGCTCCCGCC TACTACTT.T .....GA..A .A.....A.A
  S_URE_2 TTGCAAATGG AGTTTCCACC TTCCACTTAT ACAAAGACGA GAGGGGGAGA
  T_URE_2 TTACAAATGG AGTTTCCACC TTCCACTTAT ACAAAGACGA GAGGGGGAGA
  Z_URE_2 AAATAAATGG AGCTTCCACC TACAACTTAT TCACA.A..A GA....GAGA
Consensus ----AAATGG AG-T-CC-CC T-C-ACTT-T ------A--A -A-----A-A

          851                                                       900
  H_URE_2 ..AAGATTTG TATCAACACT ACAATTTTCC ATCATTAAGA CTAATAACA.
  M_URE_2 ..AAGATTTG TATCAACACT ACAATTTTCC ATCATTAAGA CTAATAACA.
  S_URE_2 GCAAGACTTG TATCAACATG ACAATTTTCC ATCATTAAGA CTAATGACAG
  T_URE_2 GCAAGACTTG TATCAACATG ACAATTTTCC ATCATTAAGA CTAATGACAG
  Z_URE_2 G..AGACTTG TATCAACATG ACAATTTTCC ATTATTAGGA CTAATGGC.G
Consensus ---AGA-TTTATCAACA-- ACAATTTTCC AT-ATTA-GA CTAAT--C--

          901                                                       950
  H_URE_2 .CAGAGCCTA GTATACATCA AGGGGAATA. AAAAGAAAAA TCTCACATTC
  M_URE_2 .CAGAGCCTA GTATACATCA AGGGGAATA. AAAAGAAAAA TCTCACATTC
  S_URE_2 GCACAGACT. GTGGGGATCA AGGGGAGCAC AAAAAAAGAG AGGCGCATTC
  T_URE_2 GCACAGACT. GTGGGGATCA AGGGGAGCAC AAAAAAAGAG AGGCGCATTC
  Z_URE_2 G.GCTGCCT. TAGGAGATCA AGGGGAG... ......AGGG GGACGCGTTC
Consensus -----G-CT- ------ATCA AGGGGA---- ------A--- ---C-C-TTC

          951                                                      1000
  H_URE_2 AAGTGGCGGC TGGGTGCTGA CCTTTGTTCC CTTTTTTTGT GTACGACTTA
  M_URE_2 AAGTGGTGGC TGGGCGCTGA CCTTTGTTCC CCCTTTTTGT ATACGACTTA
  S_URE_2 AGGTGTTGGC TGAGTGCTGA CCTCC.TT.C CTGTCTTTGT GTGCGACTTA
  T_URE_2 AGGTGTTGGC TGAGTGCTGA CCTCC.TT.C CTGTCTTTGT TGGAGACTTA
  Z_URE_2 AGGTG.TGCT GGAGTACTGA CCTCT.CC.C CCGTCTTTGT TTGGGACTTA
Consensus A-GTG--G-- -G-G--CTGA CCT------C C--T-TTTGT ----GACTTA

          1001                 1024
  H_URE_2 ACTCTTTACA AAAAAGAGCC ACAC
  M_URE_2 ACTCTTTACA AAAAAGAGCC ACAC
  S_URE_2 ACTCTTTACA AAAA.TCTAG AG~~
  T_URE_2 ACTCTTAACA AAAACTAAAA AG~~
  Z_URE_2 ACTCCGAACA AAAACCCTAA AG~~
Consensus ACTC---ACA AAAA------ A---
```

We looked for a potential regulatory role played by URE1 and URE2 by generating transgenic mice using a construct comprised of a *lacZ* reporter gene under the control of the human β-globin minimal promoter. Analysis of the β-galactosidase activity in the *URE2-lacZ* transgenic lines at E11.5 showed that URE2 is active in the forebrain of these mice as well as the hyoid arch, the somites and the apical ectodermal ridge of the limbs (Figure 3.1. B). Expression of *lacZ* in the forebrain is restricted to the subpallial telencephalon, the prethalamus and the hypothalamus and was consistent among independently generated transgenic lines (n=4). In whole-mount preparations, *URE2-lacZ* reporter activity in the forebrain is reminiscent of the activities of *I12b-lacZ* and *I56i-lacZ* (Ghanem et al., 2003). The other CRE, URE1, showed no activity in the forebrain but was active in other regions of the embryo, notably in the retina, and will not be discussed further here. We also examined the 5'flanking region of *Dlx2* but did not find sequences that were conserved among the five species.

3.3.2. *Dlx* enhancers identify divisions in the MGE and CGE

We compared the expression of *lacZ* driven by URE2, I12b and I56i in coronal sections of E11.5 and E12.5 telencephalon. Results were confirmed using at least two independent transgenic lines. The *I12b-lacZ* and *I56i-lacZ* transgenic embryos display nearly indistinguishable reporter expression at E11.5 and E12.5 in the subventricular zone (SVZ) and mantle zone (MZ) of the subpallium: the LGE, the MGE, the anterior entopeduncular area (AEP), the preoptic area (POA) and the CGE (Figure 3.2.). However, *URE2-lacZ* transgenic lines display distinct reporter gene expression in two main regions:

1) *URE2-lacZ* was the only CRE directing expression to the ventricular zone (VZ), where it is expressed in the LGE, dorsal MGE (dMGE) and CGE at E11.5 (Figure 3.2. A and D). *URE2-lacZ* expression in the VZ weakens by E12.5 (Figure 3.2. G and J), and, in the CGE, it becomes restricted to a medial domain or mCGE (Figure 3.2. J). Cells expressing *URE2-lacZ* in the VZ are arranged in radial chains suggesting that they may be clonally related (Figure 3.3. B and Supplementary Figure 3.1. B). This URE2 activity

in the VZ is weak before E11.5 (data not shown) and is maintained at least until E15.5 although with weaker intensity (data not shown). Thus, contrarily to I12b and I56i, URE2 is active in subpopulation(s) of VZ progenitor cells between E11.5 and E15.5.

2) *URE2-lacZ* is differentially expressed in subpallial SVZ domains, where it is largely excluded from the ventral MGE (vMGE), from the AEP, and somewhat excluded from the vCGE (Figure 3.2. A, G and J), whereas expression of *I12b-lacZ* and of *I56i-lacZ* is stronger in those regions (Figure 3.2. B, C, H, I, K and L). Thus, the distinct spatio-temporal activities of URE2, compared to I12b and I56i, provide genetic evidence for two subdivisions of the MGE: dMGE and vMGE, and, three subdivisions of the CGE: dCGE, mCGE and vCGE.

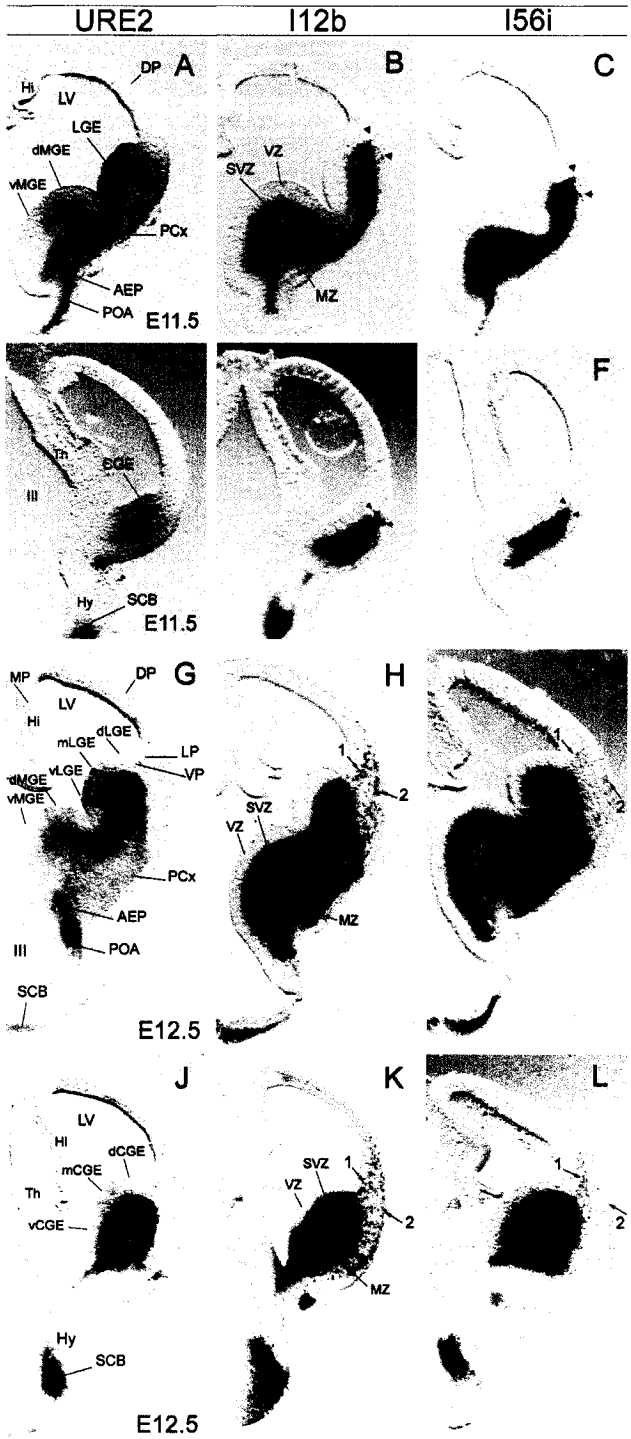
3.3.3. URE2 and I12b label distinct populations of tangentially migrating interneurons

To address whether the three CREs, URE2, I12b, and I56i, are active in distinct and/or overlapping subsets of cells of the developing telencephalon, we generated a transgenic mouse line in which the human alkaline phosphatase gene, *AP*, is driven by the I12b sequence (*I12b-AP* transgene). We crossed *I12b-AP* mice with *I56i-lacZ* and *URE2-lacZ* lines, separately, and screened the F1 generations for *lacZ*+ and *AP*+ double hemizygote embryos. Then, we compared the expression patterns of the two reporter genes, at E12.5 and E13.5, by double immunohistochemistry on brain sections using antibodies against β-galactosidase and AP.

First, we compared the activities of URE2 and I12b at E12.5. In the dMGE, *URE2-lacZ* expression is observed in a subset of cells in the VZ that are arranged in radial chains (Figure 3.3. B) whereas *I12b-AP* expression is not detected except in few columns of cells at the VZ/SVZ boundary (Figure 3.3. A). Both transgenes are expressed in larger subsets of cells in the SVZ, many of which are double-labeled but some are single-labeled (Figure 3.3. C). Similar differences in the profiles of expression of the two

Figure 3.2. Comparative enhancer activities of URE2, I12b and I56i in the subpallial telencephalon of transgenic mice

Coronal hemisections showing *lacZ* expression under the control of each enhancer at (A-F) E11.5 and, (G-L) E12.5 at medial and caudal levels. (A, D, G, J) *URE2-lacZ*; (B, E, H, K) *I12b-lacZ*; (C, F, I, L) *I56i-lacZ*. Arrowheads in B, C, E and F mark the beginning of tangential migrations at the level of the dLGE at E11.5. Arrows in H, I, K and L, show two streams of migratory cells budding from the LGE and merging into one stream at the level of the cortico-striatal boundary at E12.5: a superficial stream/prospective intermediate zone (stream 1) and a deep stream/prospective marginal zone (stream 2). There is no tangential migration of *URE2-lacZ*-positive cells at these stages. III: third ventricle; AEP: anterior entopeduncular area; CGE: caudal ganglionic eminence; CP: cortical plate; dCGE: dorsal caudal ganglionic eminence; dLGE: dorsal lateral ganglionic eminence; dMGE: dorsal medial ganglionic eminence; DP: dorsal pallium; Hi: hippocampus; IZ: intermediate zone; LGE: lateral ganglionic eminence; LP: lateral pallium; LV: lateral ventricle; mCGE: medial caudal ganglionic eminence; mLGE: medial lateral ganglionic eminence; MGE: medial ganglionic eminence; MZ: mantle zone; Mz: marginal zone; PCx: piriform cortex; POA: preoptic area; SCB: suprachiasmatic band; SVZ: subventricular zone; Th: thalamus; vCGE: ventral caudal ganglionic eminence; vLGE: ventral lateral ganglionic eminence; vMGE: ventral medial ganglionic eminence; VP: ventral pallium;VZ: ventricular zone.


URE2
I12b
I56i
A
B
C
DP
Hi
LV
LGE
dMGE
vMGE
PCx
AEP
POA
E11.5
VZ
SVZ
MZ
F
Th
SGE
III
Hy
SCB
E11.5
G
H
MP
DP
LV
Hi
dLGE
mLGE
LP
VP
vLGE
dMGE
vMGE
PCx
AEP
III
POA
SCB
E12.5
1
2
SVZ
VZ
MZ
J
K
L
LV
HI
dCGE
mCGE
Th
vCGE
Hy
SCB
E12.5
SVZ
VZ
1
2
MZ

transgenes are detected in the dLGE and dCGE (Supplementary Figure 3.1. A-C and data not shown).

In the vMGE and AEP, *URE2-lacZ* expression is weaker than in the dMGE; it was absent in the VZ and restricted to some radial columns of cells in the SVZ (Figures 3.2. G, arrows in 3.3. B and data not shown). By contrast, *I12b-lacZ* (or *I12b-AP*) expression, although not detected in the VZ, appears to be homogenously found in the SVZ of the vMGE and AEP (Figures 3.2. H, 3.3. A and data not shown).

The *URE2-* and *I12b*-containing transgenes also show differences in CGE expression (Figure 3.2. J-K and Supplementary Figure 3.1. A-C). *URE2-lacZ* expression is present in the VZ of the middle domain of the CGE (mCGE), but not in the dorsal or ventral domains (dCGE, vCGE) (Figures 3.2. J and Supplementary Figure 3.1. B). Furthermore, *URE2-lacZ* expression is weaker in the vCGE (SVZ and MZ) compared with the dCGE and mCGE (Figures 3.2. J and Supplementary 3.1. B).

The distinct expression of the *Dlx* CREs within the MGE prompted us to investigate whether the enhancers are differentially expressed in interneurons that tangentially migrate to the cortex. Most cortical GABAergic interneurons in mouse are derived from the MGE and CGE (Butt et al., 2005; Marin and Rubenstein, 2001b; Marin and Rubenstein, 2003a; Nery et al., 2002; Xu et al., 2004) and express *Dlx* genes (Anderson et al., 1997a; Cobos I et al., 2006; Cobos et al., 2005; Stuhmer et al., 2002b).

Cells in early tangential migrations from the subpallium express *I12b-lacZ* much more than they express *URE2-lacZ.* Of the two transgenes, *I12b-lacZ* is the only one expressed in migrating cells at E11.5 (Figure 3.2. B, E, arrowheads). By E12.5, there are two streams of tangentially migrating cells entering the pallium from the dLGE and dCGE, corresponding two migratory routes known to be *Dlx+* (Anderson et al., 2001; Lavdas et al., 1999; Marin and Rubenstein, 2001b; Wichterle et al., 2001). One is a deep stream that traverses the SVZ of the ventral and lateral pallium (VP, LP) and enters the SVZ/intermediate zone of the dorsal pallium (DP) (stream 1 in Figure 3.2. H and K); the second, is a superficial stream that traverses the mantle of the VP and LP, and continues in the marginal zone (Mz) of the DP (stream 2 in Figures 3.2. H, K, 3.3. G and H). There

are ~3-4 times more *I12b-AP* than *URE2-lacZ* cells tangentially migrating from the MGE and CGE at this age (Figure 3.3. G, H, Table 3.1. A, and data not shown).

By E13.5, a larger number of tangentially migrating neurons express *URE2-lacZ* in the cortical intermediate and marginal zones (Figures 3.4. B, H, and Supplementary Figure 3.2. B and H). Yet, there are still are ~2-2.5-fold fewer *URE2-lacZ* expressing cells than *I12b-AP-* or *I56i-lacZ-* expressing cells (Table 3.1. A).

We compared URE2 and I12b activities in tangentially migrating cells using double-labeling. About half of the tangentially migrating cells expressing *URE2-lacZ* also express the *I12b-AP* transgene at E12.5 or E13.5 (Figure 3.3., Supplementary Figure 3.1., and Table 3.1. A). URE2+/I12b+ cells are primarily found in stream 2 (Mz; Figures 3.3. I, 3.4. I, Supplementary Figures 3.1. I and 3.2. I). By contrast, URE2+/I12b- migrating cells are largely found in stream 1 (IZ, Figures 3.3. I, 3.4. I, Supplementary Figure 3.1. I, 3.2. I and Table 3.1. B). In summary, differential activity of CREs from the *Dlx1/Dlx2* locus in tangentially migrating cells suggest the existence of several populations (or subpopulations) in both the Mz and IZ.

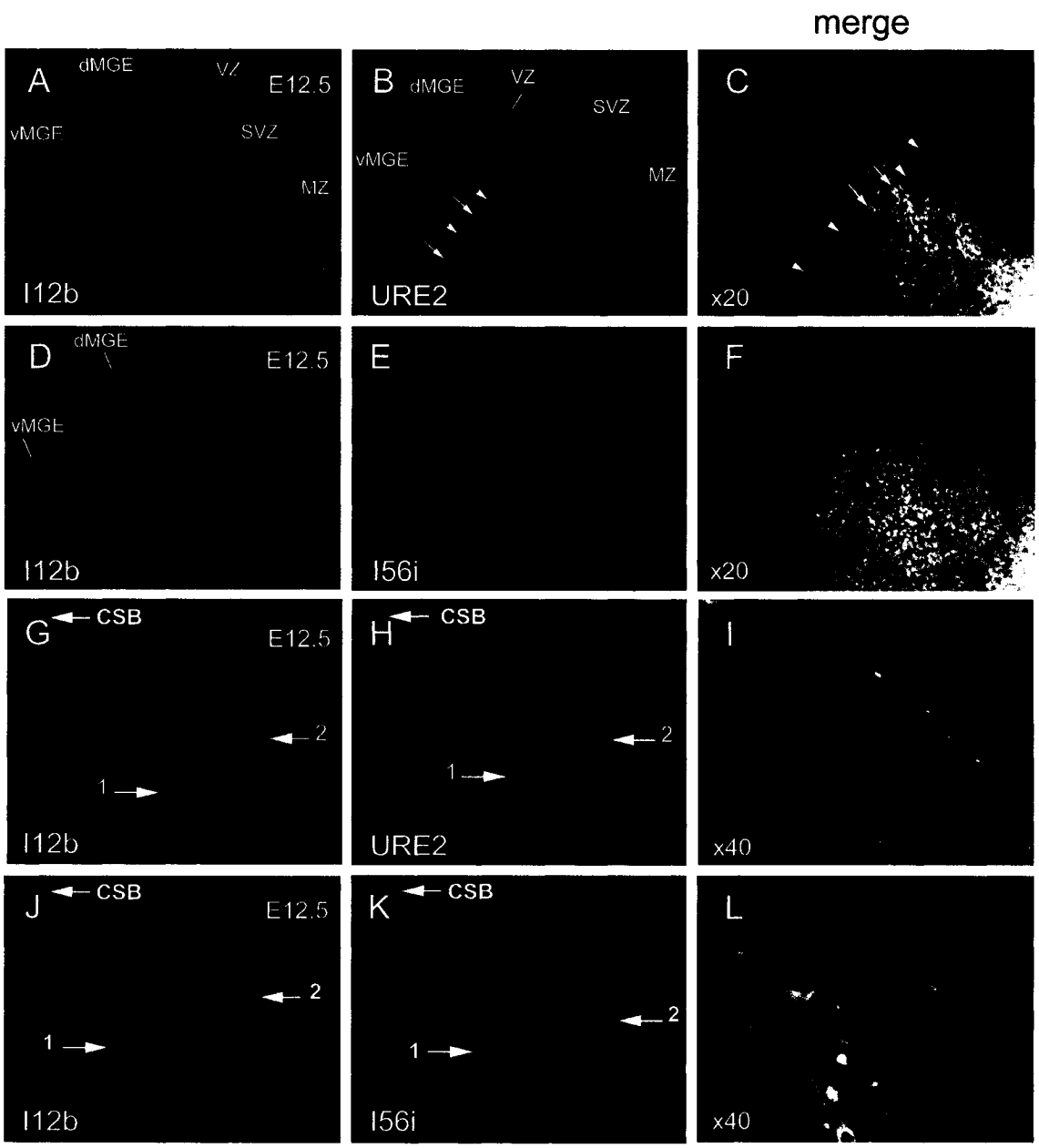
3.3.4. Highly similar activities of I12b and I56i

We then compared the activities of I12b and I56i at E12.5. The majority of cells in the SVZ and MZ are double-labeled in the MGE and AEP (Figure 3.3. D-F and data not shown) and in the CGE (Supplementary Figure 3.1. D-F) confirming X-gal staining (Figure 3.2.) that suggests activity of the two enhancers in overlapping population(s).

We next compared the expression of I12b and I56i in tangentially migrating cells at E12.5 and E13.5 as we did for URE2/I12b comparisons. The *I12b-AP* and *I56i-lacZ* transgenes are co-expressed in tangentially migrating cells (Table 3.1. A and Figures 3.3. J-L, 3.4. D-F, J-L, Supplementary Figures 3.1. J-L, 3.2. D-F and J-L). We also detect two populations of single-labeled migrating cells at the levels of the MGE and the CGE. The I56i+/I12b- and the I56i-/I12b+ populations accounted for less than 20% and 12.5% of migrating cells, respectively (Table 3.1. A). All three cell populations (I12b+/I56i+, I12b+/I56i- and I12b-/I56i+) are equally distributed between migration streams 1 and 2

Figure 3.3. Activities of three *Dlx* CREs in the MGE and cortico-striatal boundary (CSB) at E12.5

(A-F) Double immunohistochemistry showing the expression of (A, D) *I12b-AP*, green; (B) *URE2-lacZ*, red; and (E) *I56i-lacZ*, red. (C) and (F) are merged pictures of (A, B) and (D, E), respectively. Compared with the dMGE, only few radial columns of cells in (B) express *URE2-lacZ* in the SVZ of vMGE (arrowheads) whereas others do not (arrows). In (C), arrows and arrowheads show radial columns of double labeled cells (yellow) and single labeled ones (red or green), respectively. (G-L) Labeling of tangentially migrating cells towards the CSB of (G, J) *I12b-AP* mice (green), (H) *URE2-lacZ* mice (red) and (K) *I56i-lacZ* mice (red) at E12.5. (I) and (L) are merged pictures of (G, H) and (J, K), respectively. As described in Figure 3.2. H, I, K and L, migratory cells (arrows 1 and 2 in G-L) follow superficial (stream 2) and deep (stream 1) streams of migrations at E12.5 that will merge into one along the Mz of the CSB. There is little tangential migration of *URE2-lacZ*-positive cells at this age (H). (I, L) Single-labeled and double-labeled cells are indicated with arrows and arrowheads, respectively. Symbols as in Figure 3.2.


merge
A
dMGE
VZ
E12.5
vMGE
SVZ
MZ
I12b
B
dMGE
VZ
SVZ
vMGE
MZ
URE2
C
x20
D
dMGE
E12.5
vMGE
I12b
E
I56i
F
x20
G
CSB
E12.5
2
1
I12b
H
CSB
2
1
URE2
I
x40
J
CSB
E12.5
2
1
I12b
K
CSB
2
1
I56i
L
x40

Supplementary Figure 3.1. Activities of three *Dlx* CREs in the CGE and CSB at E12.5

(A-F) Double immunohistochemistry showing (A, D) the expression of *I12b-AP*, green; (B) *URE2-lacZ*, red; and (E) *I56i-lacZ*, red. (C, F) are merged pictures of (A, B), and, (D, E), respectively. *URE2-lacZ* expression is absent from the VZ and weaker in the SVZ of the vCGE compared with the mCGE. (G-L) Labeling of tangentially migrating cells at the level of the CSB at E12.5 in (G, J) *I12b-AP*, green, (H) *URE2-lacZ*, red and (K) *I56i-lacZ*, red, mouse embryos. (I, L) are merged pictures of (G, H) and (J, K), respectively. Similar to the MGE, migratory cells follow two streams of migrations at this age (arrows 1 and 2 in G-L), and (H) there are few tangentially migrating cells that express *URE2-lacZ* at this age. Single-labeled and double-labeled cells are indicated with arrows and arrowheads, respectively. Symbols as in Figure 3.2.

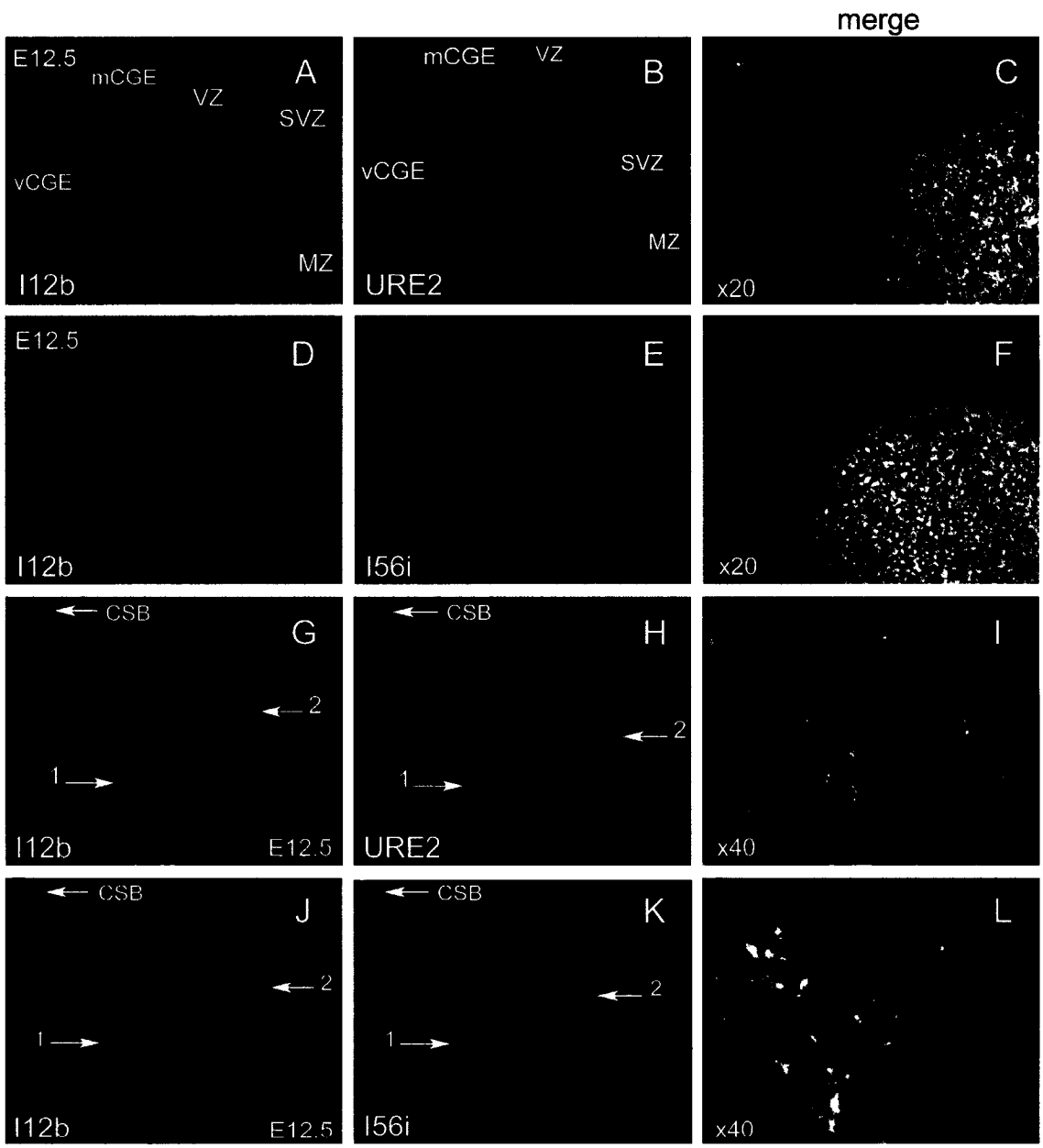

merge
E12.5
mCGE
VZ
SVZ
vCGE
MZ
A
I12b
mCGE
VZ
B
vCGE
SVZ
MZ
URE2
C
x20
E12.5
D
I12b
E
I56i
F
x20
CSB
G
2
1
I12b
E12.5
CSB
H
2
1
URE2
I
x40
CSB
J
2
1
I12b
E12.5
CSB
K
2
1
I56i
L
x40

Table 3.1. Percentage of migrating cells derived from the MGE and CGE in CRE reporter lines

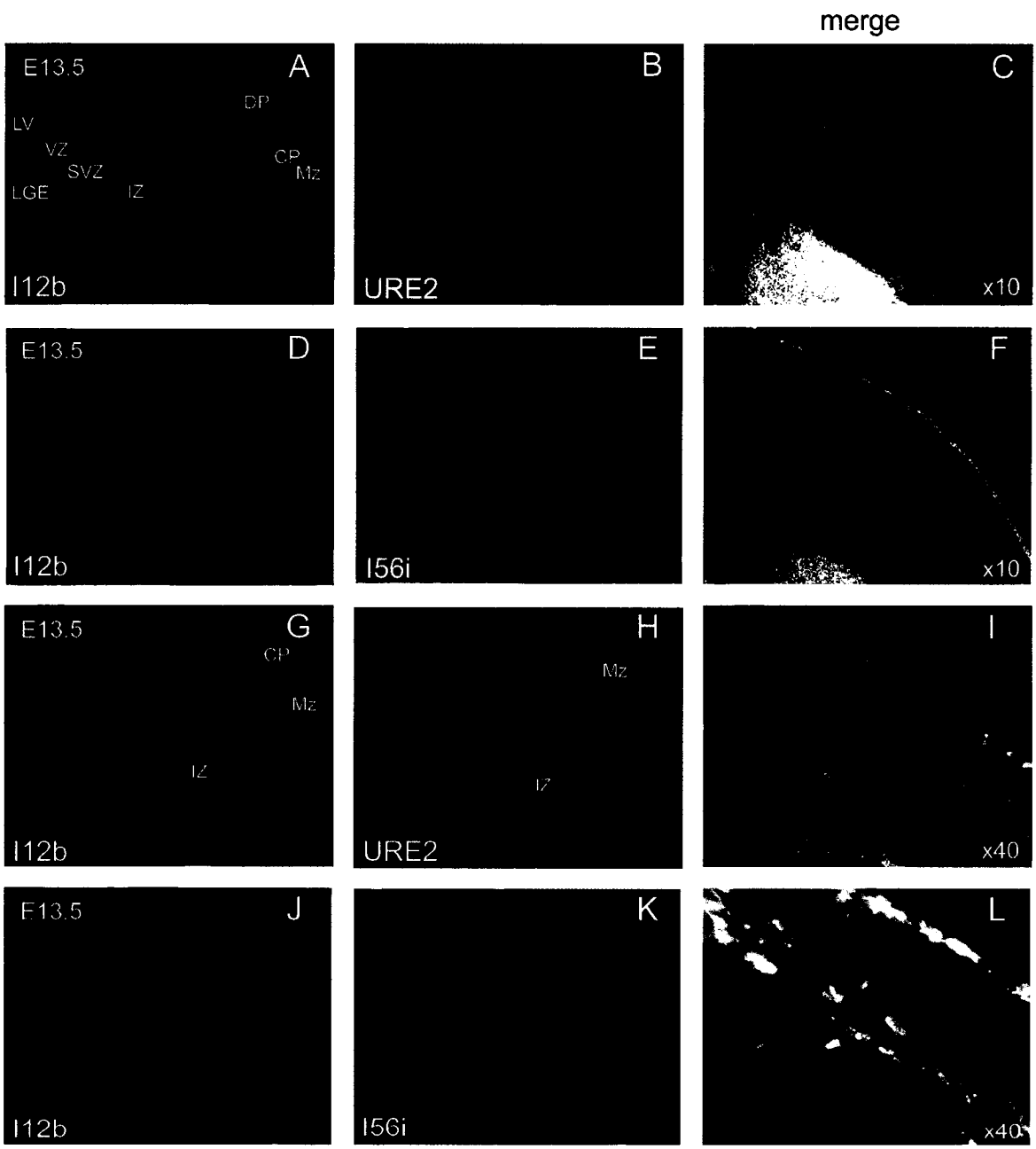
Table 3.1. A		E12.5 MGE/CSB %	E12.5 CGE/CSB %	E13.5 MGE/DP %	E13.5 CGE/DP %
URE2 / I12b	URE2+/I12b-	9.3 ± 0.6	10.9 ± 1.5	15.4 ± 0.7	18.9 ± 0.5
	URE2+/I12b+	10.3 ± 0.5	14.1 ± 1.6	13.2 ± 0.7	13.3 ± 0.4
	URE2-/I12b+	81.4 ± 2.0	75.0 ± 2.5	71.4 ± 1.5	67.8 ± 2.5
I56i / I12b	I56i+/I12b-	14.1 ± 0.9	10.7 ± 2	19.9 ± 1	13.3 ± 1.7
	I56i-/I12b+	8.7 ± 2.1	7.5 ± 2.5	9.8 ± 1	12.5 ± 2.5
	I56i+/I12b+	77.1 ± 3.0	81.7 ± 3.5	70.2 ± 1	74.2 ± 4.2

Table 3.1. B		E12.5 MGE/CSB %	E12.5 CGE/CSB %	E13.5 MGE/DP %	E13.5 CGE/DP %
URE2+/I12b-	Marginal zone (stream 2)	18.4 ± 1.6	18.8 ± 7.0	31.6 ± 5.9	29.2 ± 1
	Intermediate zone (stream 1)	80.0 ± 5.0	68.8 ± 6.2	70.7 ± 3.4	64.8 ± 4.8

(Table 3.1. A) Percentage of single and double labeled migrating cells at the level of the CSB and DP at E12.5 and E13.5. (Table 3.1. B) Distribution of the URE2+/I12b- cells along the two main streams of migration in the CSB at the level of the MGE and CGE. The number of *lacZ* and/or AP positive cells in the CRE lines were determined from 2-3 sections in the CSB or DP at the level of the MGE or CGE per mouse and averaged (n=2). Data are presented as mean ± SD.

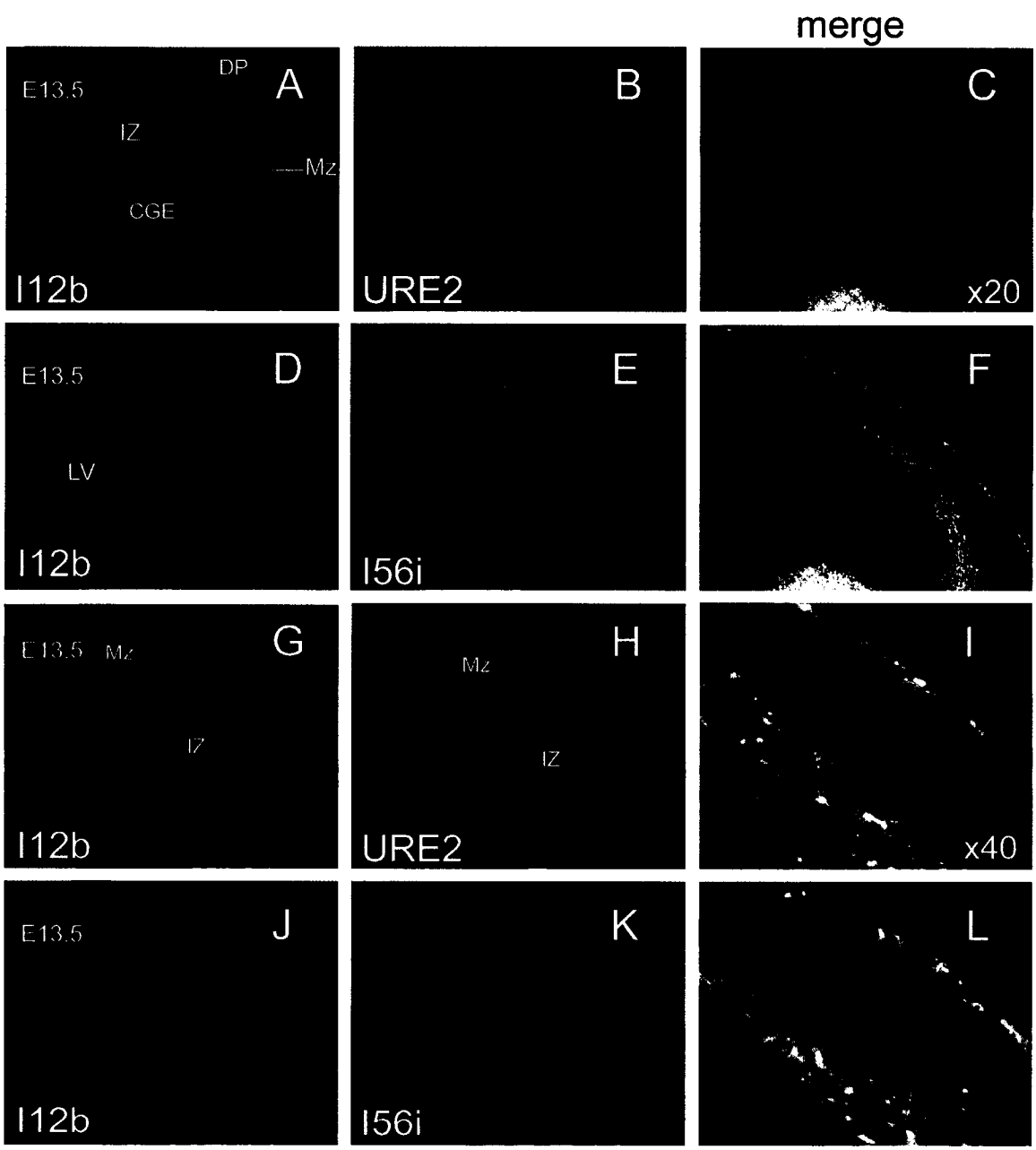
Figure 3.4. Activities of three *Dlx* CREs in tangentially migrating cells derived from the MGE at E13.5

(A-F) Double immunohistochemistry showing the expression of (A, D) *I12b-AP*, green; (B) *URE2-lacZ*, red; and (E) *I56i-lacZ*, red, in the DP. (C) and (F) are merged pictures of (A, B) and (D, E), respectively. Tangentially migrating cells follow the Mz and Iz of the DP at E13.5. (G-L) Higher magnification pictures of the boxes shown in (A-F), respectively. (I) and (L) are merged pictures of (G, H) and (J, K), respectively. There are more migrating cells in *URE2-lacZ* mice at this age (B, H) compared with E12.5 (Figure 3.3. H). Single-labeled and double-labeled cells in I and L are indicated with arrows and arrowheads, respectively. Most migrating cells in *I12b-AP/I56i-lacZ* mice are double labeled (arrowheads in L) whereas a large number of cells in *I12b-AP/URE2-lacZ* mice are single labeled (arrows in I). Symbols as in Figure 3.2.


merge
E13.5
A
DP
LV
VZ
SVZ
LGE
IZ
CP
Mz
I12b
B
URE2
C
x10
E13.5
D
I12b
E
I56i
F
x10
E13.5
G
CP
Mz
IZ
I12b
H
Mz
IZ
URE2
I
x40
E13.5
J
I12b
K
I56i
L
x40

Supplementary Figure 3.2. Activities of three *Dlx* CREs in tangentially migrating cells derived from the CGE at E13.5

(A-F) Double immunohistochemistry showing the expression of (A, D) *I12b-AP*, green; (B) *URE2-lacZ*, red; (E) *I56i-lacZ*, red, in the DP. (C, F) are merged pictures of (A, B) and (D, E), respectively. (G-L) Tangentially migrating cells follow the Mz and Iz of the DP at E13.5. (G-L) Higher magnification pictures of the boxes shown in (A-F), respectively. Panels (I, L) are merged pictures of (G and H), and, (J and K), respectively. As observed in the MGE, there are more migrating cells derived from the CGE that express *URE2-lacZ* at this age (B, H) compared with E12.5 (Supplementary Figure 3.1. H). Most migrating cells in *I12b-AP/I56i-lacZ* double transgenic mice are double labeled (arrowheads in L) whereas a large number of cells in *I12b-AP/URE2-lacZ* mice are single-labeled (arrows in I). Single-labeled and double-labeled cells in I and L are indicated with arrows and arrowheads, respectively. Symbols as in Figure 3.2.


merge
E13.5
DP
A
IZ
Mz
CGE
I12b
B
URE2
C
x20
E13.5
D
LV
I12b
E
I56i
F
E13.5
Mz
G
IZ
I12b
Mz
H
IZ
URE2
I
x40
E13.5
J
I12b
K
I56i
L

(Mz and Iz) (Figures 3.3. L, 3.4. L, Supplementary Figures 3.1. L and 3.2. L). These results suggest that the I12b and I56i CRE are mainly active in the same cells and could be responding to similar mechanisms.

3.3.5. The dMGE and the vMGE produce cortical interneurons

Since the I12b and URE2 CREs are differentially active in the MGE and AEP, and may label distinct subpopulations of tangentially migrating neurons derived from these regions, we sought to investigate whether this could be linked to intrinsic properties of the progenitor cells found in these regions in terms of regional specification and migration potential. For this purpose, we performed *in vitro* DiI labeling experiments on brain slice culture taken from wild type litters at E12.5 and E13.5. First, we compared the migration potential of neurons derived from progenitors residing in the MGE versus those derived from progenitors found in the AEP and POA, separately. We found no evidence of tangential migration that reached the cortex from either the AEP or POA (Supplementary Figures 3.3. B-C, F-H, 3.4. B, D, J-L, N, and P) but we observed robust tangential migration of neurons from the MGE that enter the cortex (Supplementary Figures 3.3. B, D, 3.4. B-C and E-H, n=15) consistent with previous slice studies (Anderson et al., 2001; Anderson et al., 1997b).

Next, we compared the migration potential of neurons derived from progenitors found in the dMGE versus vMGE at E12.5 and E13.5. We observe robust tangential migration to the cortex from both regions (Figures 3.5. A-D, Supplementary Figure 3.4. E-H, N-O and data not shown, n=12). We also performed tissue transplantation assays coupled to DiI experiments at E12.5; the vMGE was dissected out and transplanted into a rostral or a caudal section to replace either the LGE or CGE, respectively. Then, the tissue transplanted from the vMGE and the one from the remaining dMGE were labeled separately with DiI (Figure 3.5. E, I and M). The vMGE transplants show robust tangential migration to the cortex regardless of their new environment (Figure 3.5. M-P and data not shown, n=4) indicating that E12.5 vMGE progenitor cells are specified to produce cells that tangentially migrate to the cortex. In addition, progenitors located in

the remaining dMGE retain their ability to migrate to the cortex (Figure 3.5. F-G and J-K). These experiments demonstrate that both the vMGE and dMGE are sources of cells that tangentially migrate to the cortex.

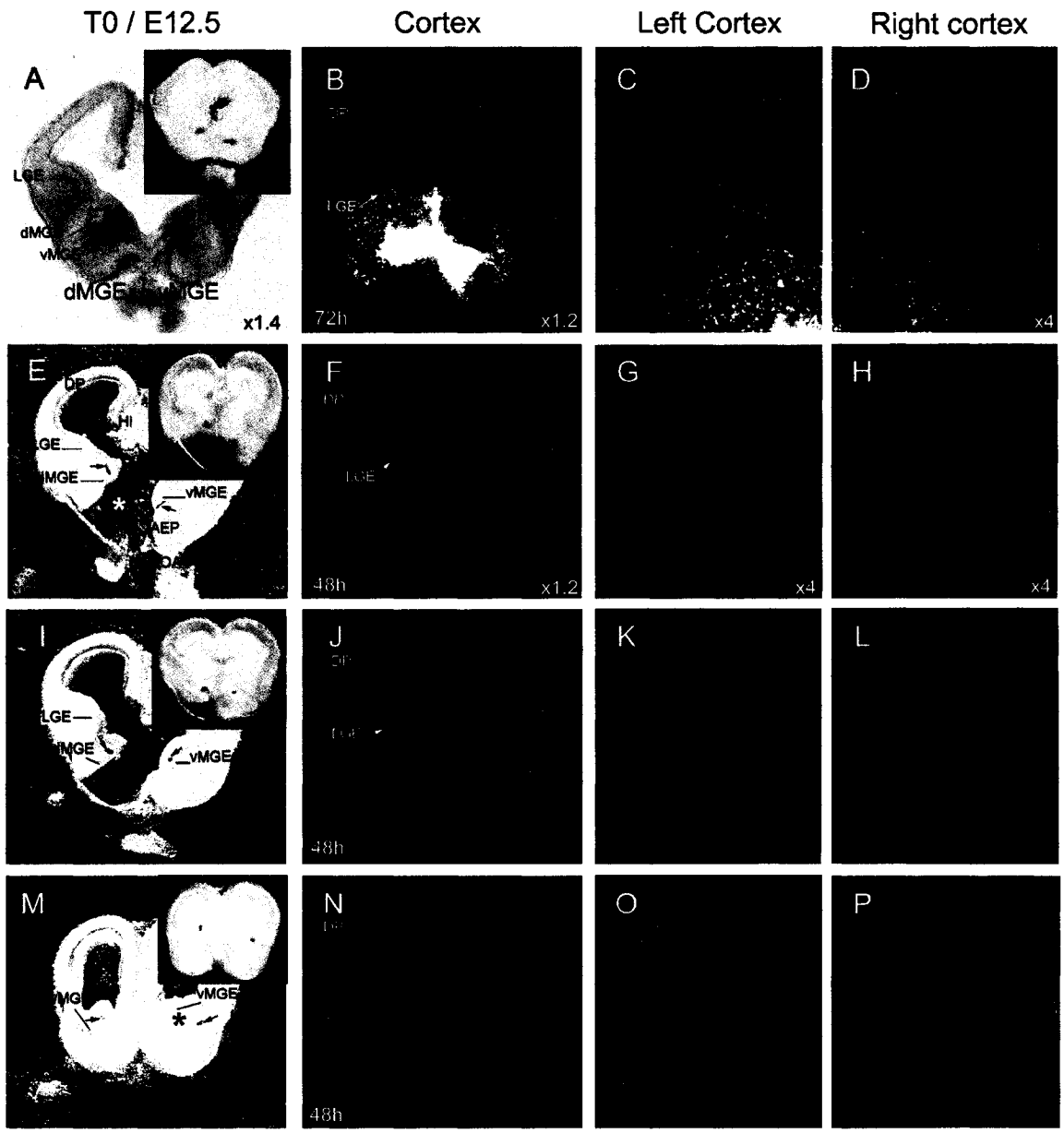
3.3.6. URE2 and I12b mark overlapping and distinct subtypes of adult cortical interneurons

Expression of URE2 and I12b reporter transgenes persists postnatally in subsets of forebrain neurons. We focused on their expression in interneurons of the somatosensory and motor cortex of 5 week old mice using double immunohistochemistry. We co-labeled either *URE2-lacZ+* or *I12b-lacZ+* cells with GABA. We find that approximately two-thirds of GABA+ cells (~66.1 $\pm$ 2.8%, n=2) were *URE2-lacZ+* and nearly all of the *URE2-lacZ+* cells express GABA (~92.9 $\pm$2.1%, n=2). By contrast, nearly all of GABA+ cells are *I12b-lacZ+* (~93.8 $\pm$ 2.1%, n=2) and vice-versa (93.5 $\pm$ 3.4%, n=2) (Table 3.2. and data not shown).

Next, we determined whether *URE2-lacZ* and *I12b-lacZ* were differentially expressed in subtypes of cortical interneurons. Thus, we co-labeled *URE2-lacZ+* and *I12b-lacZ+* cells with the following subtype markers: calcium-binding proteins, CR and PV; neuropeptides, SOM, NPY, nNOS, and VIP. We find that *URE2-lacZ* is expressed in almost all PV+ (99 $\pm$ 1%), CR+ (99.2 $\pm$ 0.8%), NPY+ (97.8 $\pm$ 3.2), and nNOS+ (99.9 $\pm$ 0.1%) interneurons (n=2). On the other hand, *I12b-lacZ* is expressed in a subset of these cells: PV (85.5 $\pm$ 3.3%), CR (79.4 $\pm$ 5.1%), NPY (75.7 $\pm$ 5.0%) and nNOS (23.7 $\pm$ 7.1%) (n=2) (Figures 3.6. I-X, 3.7. A-P and Table 3.2.). Importantly, *URE2-lacZ* is expressed in only a very small number of SOM- (6.4 $\pm$ 1.2%) or VIP- (11.8 $\pm$ 5.4%) interneurons (n=2) (Figures 3.6. E-H, 3.7. Q-T and Table 3.2.). By contrast, *I12b-lacZ* is expressed in most of these two interneuron subtypes (SOM- 87.5 $\pm$ 4.7%; VIP- 96.1 $\pm$ 2.6%) (n=2) (Figures 3.6. A-D, 3.7. U-X and Table 3.2.). Thus, we conclude that URE2 and I12b mark overalpping as well as distinct population(s) or subpopulation(s) of cortical interneurons in the adult mouse brain.

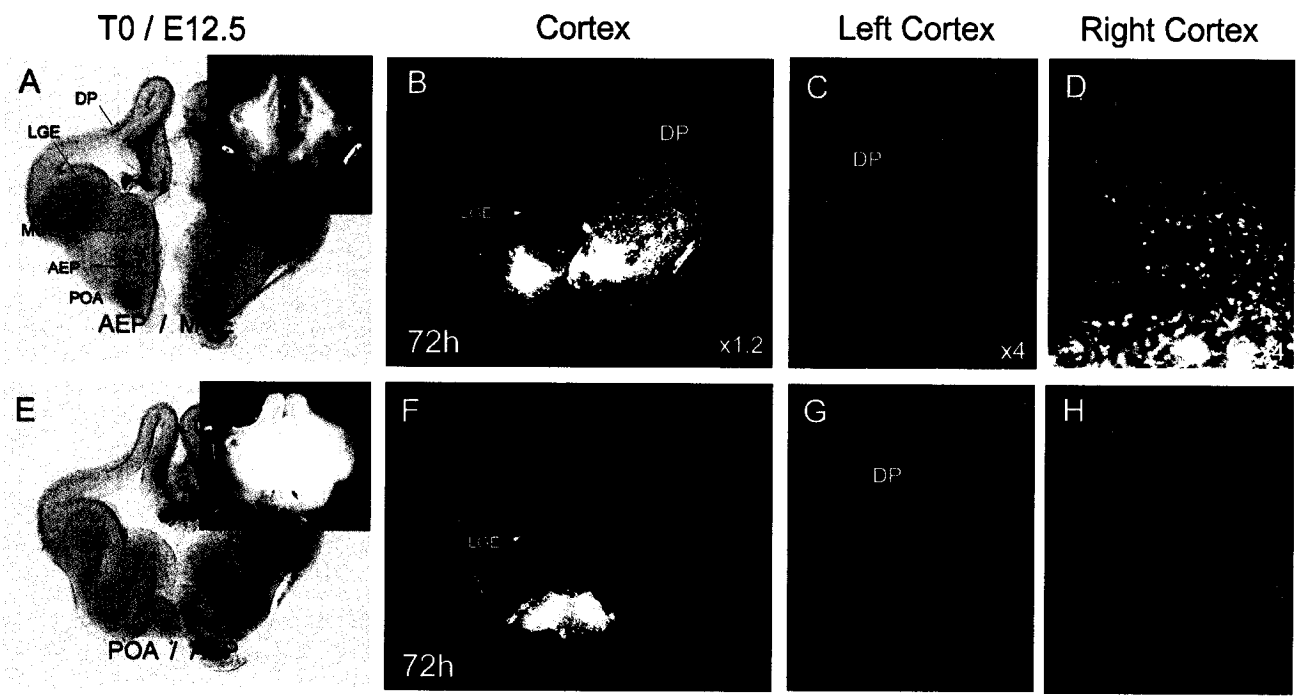
Figure 3.5. The dMGE and vMGE are two distinct sources of tangentially migrating cells to the DP at E12.5

(A) DiI labeling (arrows) of the vMGE and dMGE. Inset in A is a picture taken after 72h in culture. (B-D) Robust migrations towards the cortex are derived from both subdivisions. In panels (E) and (I), the left vMGE was dissected and transplanted into a rostral level to replace the LGE (* in E, I and M), respectively. The left dMGE and right vMGE in (E, I) as well as the transplanted vMGEs in (M) were labeled with DiI. Insets in (E, I, M) are pictures taken after 48h in culture. (F, J, N) Fluorescent pictures of (E, I, M), respectively. (G, K, O) and (H, L, P) are higher magnifications of the left and right hemispheres in (E, I, M), respectively. Robust tangential migration to the DP is observed from progenitors located in the dMGE (G, K), control vMGE (H, L) as well as transplanted vMGE (O, P). Symbols as in Figure 3.2.


T0 / E12.5
Cortex
Left Cortex
Right cortex
A
B
C
D
LGE
dMGE
vMGE
x1.4
72h
x1.2
x4
E
F
G
H
LGE
vMGE
AEP
48h
x1.2
x4
I
J
K
L
LGE
vMGE
48h
M
N
O
P
vMGE
48h

Supplementary Figure 3.3. Cell migration from ventral structures to the DP at E12.5

(A, E) DiI labeling on coronal sections at medial levels. The location of the DiI crystal is indicated with an arrow. Labeled structures are indicated at the bottom of each section. Insets in A and E are pictures taken after 72h in culture. (B, F) Fluorescent pictures of (A, E) taken after 72h in culture, respectively. Higher magnifications of the left and right hemispheres in panels (B) and (F) are shown in (C, G) and (D, H), respectively. Little or no tangential migration towards the cortex is observed from progenitors in (C, H) the AEP; and (G) POA. (D) Strong tangential migrations take place from the MGE. Symbols as in Figure 3.2.


T0 / E12.5
Cortex
Left Cortex
Right Cortex
A
DP
LGE
AEP
POA
B
DP
72h
x1.2
C
DP
x4
D
E
F
72h
G
DP
H

Supplementary Figure 3.4. Cell migration from ventral structures to the DP at E13.5

(A, E, I, M) DiI labeling on coronal sections at medial levels. The location of the DiI crystal is indicated with an arrow. Labeled structures are indicated at the bottom of each section. Insets in (A, E, I, M) are pictures taken after 48h in culture. (B, F, J, N) Fluorescent pictures of (A, E, I, M) taken after 48h in culture, respectively. Higher magnifications of the left and right hemispheres in panels (B, F, J, N) are shown in (C, G, K, O) and (D, H, L, P), respectively. Robust tangential migration of cells to the DP are observed in sections where (O) the dMGE and (C, G, H) the vMGE are labeled. Little or no tangential migration is derived from progenitors in (D, K, P) the AEP and (L) POA. Symbols as in Figure 3.2.

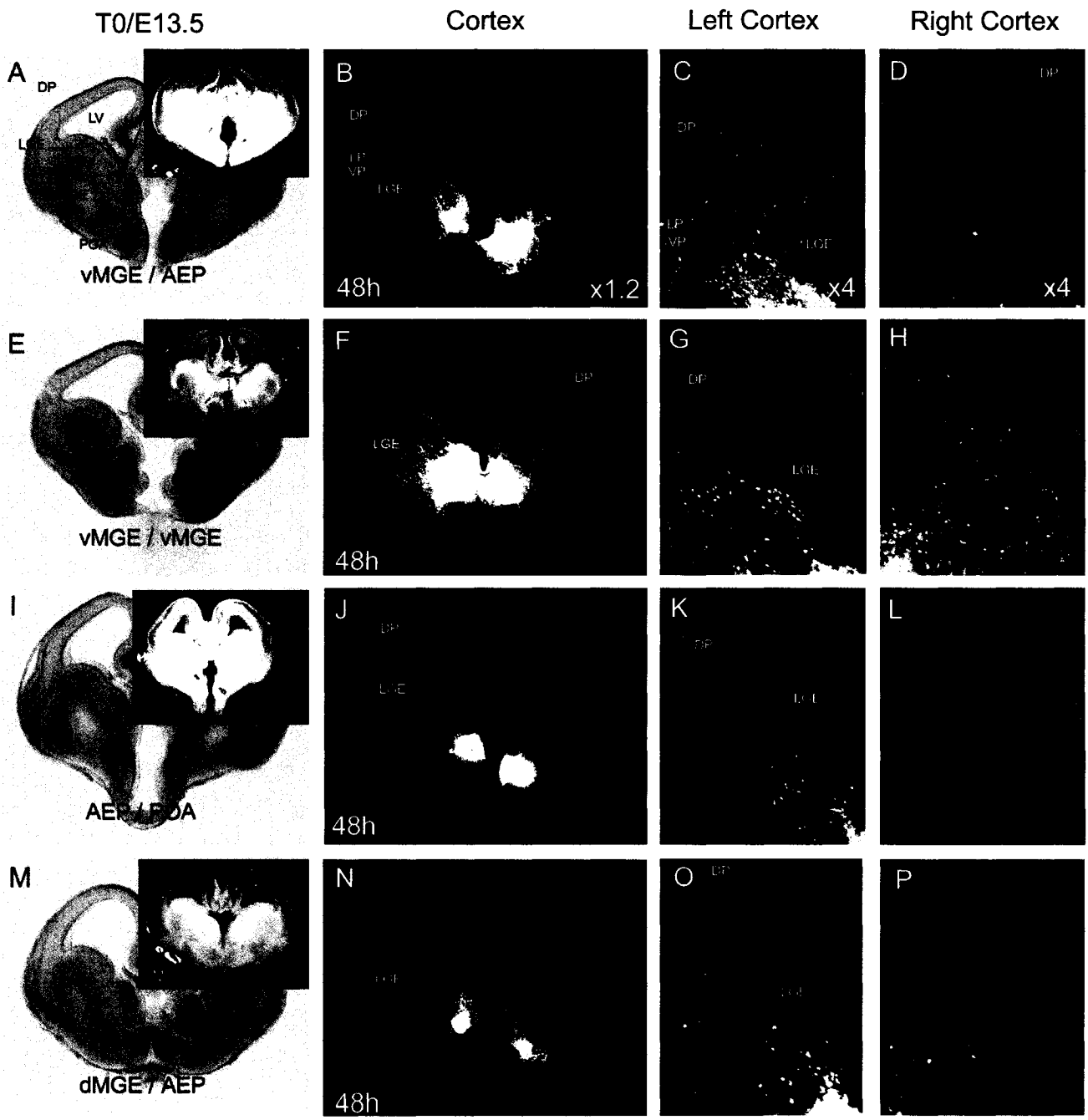

T0/E13.5
Cortex
Left Cortex
Right Cortex
A
DP
LV
vMGE / AEP
B
48h
x1.2
C
x4
D
x4
E
vMGE / vMGE
F
48h
G
H
I
J
48h
K
L
M
dMGE / AEP
N
48h
O
P

Figure 3.6. Co-labeling of URE2- and I12b-expressing interneurons with subtype markers in the mouse adult somatosensory cortex at P35

(A, I, Q) Cells labeled with *I12b-lacZ* or (E, M, U) *URE2-lacZ* are shown in red. Interneurons-expressing (B, F) somatostatin; (J, N) parvalbumin; or (R, V) calretinin, are shown in green. (D, H, L, P, T, X) Higher magnifications of boxes shown in (C, G, K, O, S, W), respectively. (O, P) Almost all PV+, (W, X) CR+ neurons express *URE2-lacZ*. In contrast, SOM-expressing interneurons (G, H) do not express *URE2-lacZ*. The *I12b-lacZ* transgene is expressed in a large number of (K, L) PV+; (S, T), CR+, as well as (C, D) SOM+ interneurons. Double and single-labeled interneurons are indicated with arrows and arrowheads.

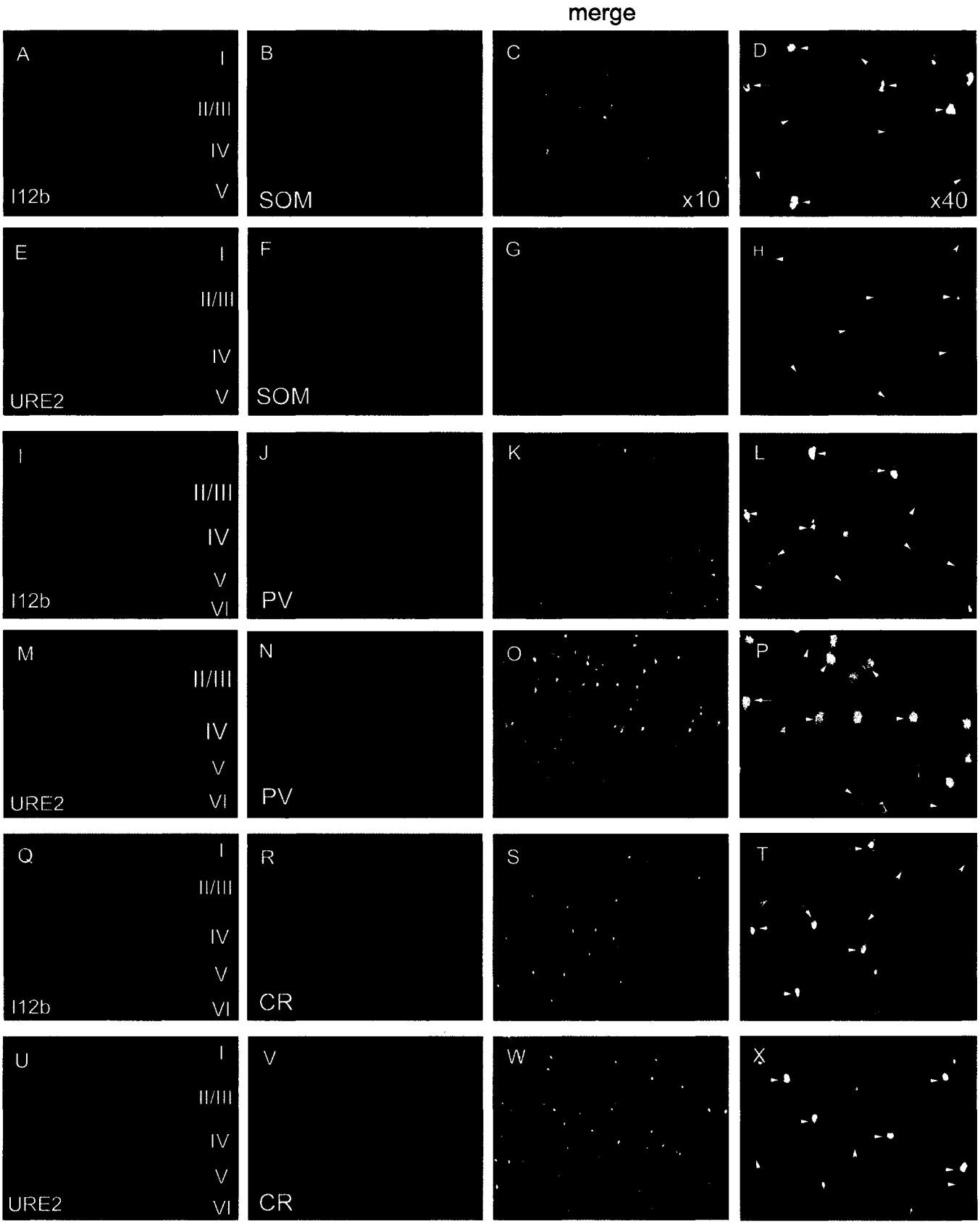

merge
A
I
II/III
IV
I12b
V
B
SOM
C
x10
D
x40
E
I
II/III
IV
URE2
V
F
SOM
G
H
I
II/III
IV
V
I12b
VI
J
PV
K
L
M
II/III
IV
V
URE2
VI
N
PV
O
P
Q
I
II/III
IV
V
I12b
VI
R
CR
S
T
U
I
II/III
IV
V
URE2
VI
V
CR
W
X

Figure 3.7. Co-labeling of URE2- and I12b-expressing interneurons with subtype markers in the mouse adult somatosensory cortex at P35

(A, I, Q) Cells expressing *URE2-lacZ* or (E, M, U) *I12b-lacZ* are shown in red. Interneurons-expressing (B, F) neuropeptide Y; (J, N) neuronal nitric oxide synthase; or (R, V) vaso-active intestinal peptide are shown in green. (D, H, L, P, T, X) Higher magnifications of boxes shown in (C, G, K, O, S, W), respectively. (C, D) Almost all NPY+; (K, L) nNOS+ intereneurons express *URE2-lacZ* but (S, T) VIP+ interneurons do not express this transgene. (G, H) A large number of NPY+; or (W, X) VIP+, but (O, P) few nNOS+ interneurons express *I12b-lacZ*. Double and single-labeled interneurons are indicated with arrows and arrowheads.

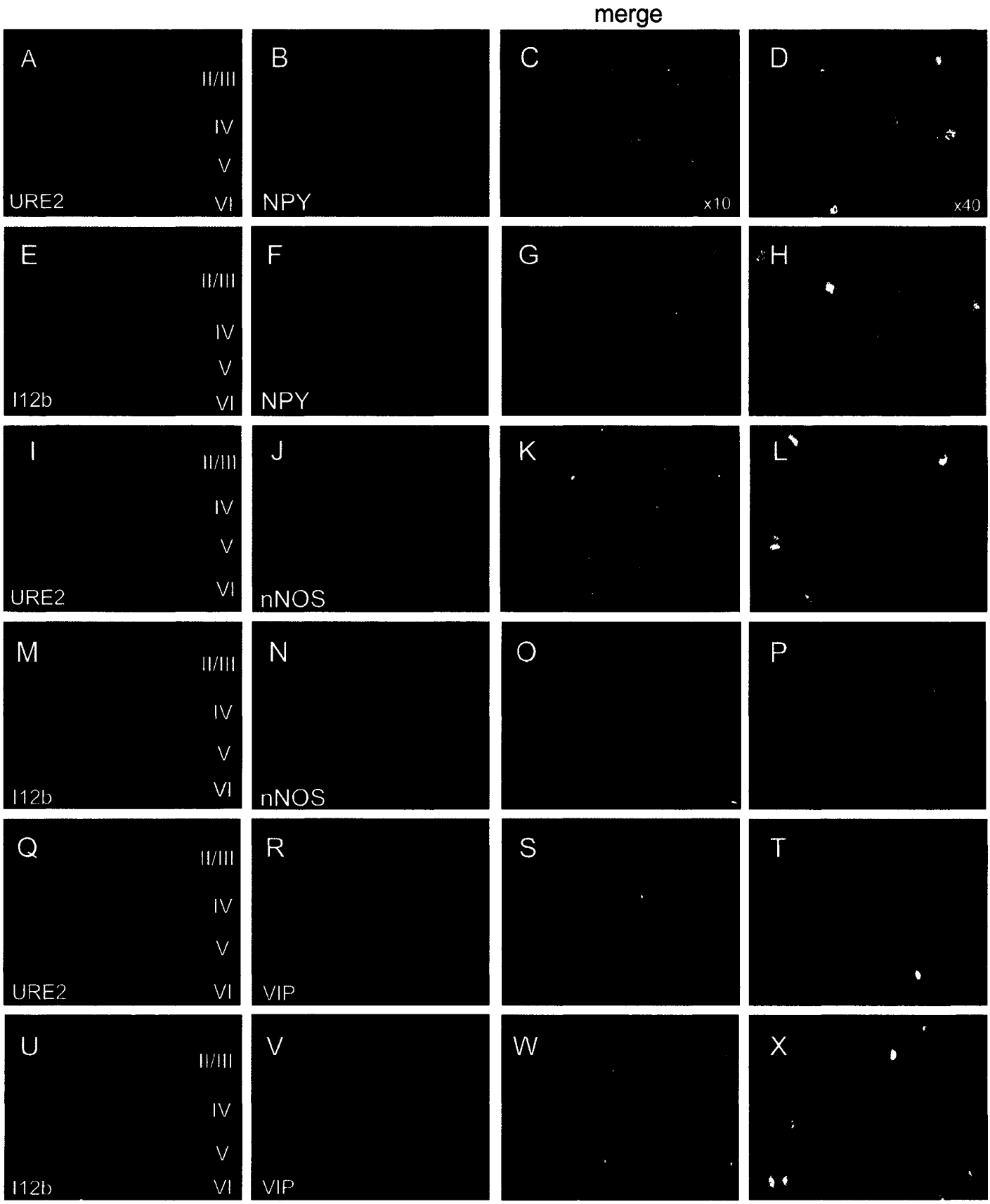

merge
A
II/III
IV
V
URE2
VI
B
NPY
C
x10
D
x40
E
II/III
IV
V
I12b
VI
F
NPY
G
H
I
II/III
IV
V
URE2
VI
J
nNOS
K
L
M
II/III
IV
V
I12b
VI
N
nNOS
O
P
Q
II/III
IV
V
URE2
VI
R
VIP
S
T
U
II/III
IV
V
I12b
VI
V
VIP
W
X

Table 3.2. Percentage of cortical interneurons subtypes labeled with URE2 and I12b in the somatosensory and motor cortices at P35

	URE2	**I12b**
GABA %	66.1 ± 2.8	93.8 ± 2.1
Parvalbumin (PV) %	99 ± 1.0	85.5 ± 3.3
Calretinin (CR) %	99.2 ± 0.8	79.4 ± 5.1
Somatostatin (SOM) %	6.4 ± 1.2	87.5 ± 4.7
Neuropeptide Y (NPY) %	97.8 ± 3.2	75.7 ± 5.0
Neuronal Nitric Oxide Synthase (nNOS) %	99.9 ± 0.1	27.3 ± 7.1
Vaso-active Intestinal Peptide (VIP) %	11.8 ± 5.4	96.1 ± 2.6
Calbindin (CB) %	6.7 ± 1.1	N.D.*

N.D*. Not Determined

The numbers of positive cells were determined from 8-10 coronal sections of the somatosensory and motor cortices per mouse and averaged (n = 2). Sections were selected at different levels with respect to the rostro-caudal axis. In each section, a rectangular area of neocortex from the white matter to the pial surface was analyzed. Data are presented as mean ± SD.

3.4. Discussion

Herein, we provide evidence that the overall patterns of *Dlx* expression in the developing and mature brain represent the sum of patterns generated by distinct enhancer elements that may be regulated by distinct mechanisms. This conclusion is based on the detailed examination of *cis*-regulatory elements (CREs) that control *Dlx* gene expression in the developing telencephalon. We report the identification of URE2, a novel *Dlx* forebrain-specific CRE that is highly conserved in vertebrates (Figure 3.1. and Supplementary Figure 3.1.). URE2 and two previously characterized *Dlx* CREs active in the forebrain, I56i and I12b, display partially overlapping activities in the basal ganglia (Figures 3.2., 3.3. A-F and data not shown). Importantly, their activities are distinct in two subdivisions of the MGE; dMGE and vMGE (Figures 3.2. A-C, G-I and 3.3. A-C), as well as three subdivisions in the CGE; dCGE, mCGE and vCGE (Figure 3.2. J-L and Supplementary Figure 3.1. A-C). Furthermore, I12b/I56i and URE2 are active in overlapping, but distinct population(s) of tangentially migrating neurons to the cortex derived from the MGE and CGE (Figures 3.3., 3.4., Supplementary Figures 3.1. and 3.2.). In the adult mouse cortex, *URE2-lacZ* and *I12b-lacZ* are expressed in common as well as distinct subtypes of cortical interneurons (Figures 3.6. and 3.7.). The differential expression of *URE2-lacZ* and *I12b-lacZ* suggests that: 1) the vMGE and dMGE may be distinct sources of cortical interneurons. The same conclusion may also be true for the dCGE, mCGE and vCGE; 2) *Dlx*-progenitors born in these subpallial subdivisions have different molecular properties and may subsequently give rise to distinct subtypes of adult cortical interneurons, yet, direct evidence of such a relationship is still lacking.

3.4.1. Multiple enhancers regulate *Dlx1/2* and *Dlx5/6* expression in the MGE and CGE

We have identified thus far four CREs that are involved in expression of four *Dlx* genes in the forebrain of vertebrates. URE2 and I12b are located in the *Dlx1/Dlx2* locus, and, I56i and I56ii are located in the *Dlx5/Dlx6* locus (Ghanem et al., 2003; Zerucha et

al., 2000). Preliminary analysis of I56ii activity shows that this CRE is active in a distinct distribution at E11.5-E13.5 when compared to URE2, I12b, and I56i (Ghanem N. and Ekker M., unpublished data).

URE2 is the only CRE that is strongly active in subset(s) of radially arranged cells in the VZ of the LGE, dMGE and mCGE at E11.5 and E12.5 (Figures 3.2. A, D, G, J, 3.3. B, Supplementary Figure 3.1. B and data not shown). Likewise, *Dlx2* is expressed in discrete radial clusters of mitotically active VZ cells (Eisenstat et al., 1999; Porteus et al., 1994). *Dlx1* is expressed in a smaller subset of VZ cells (Eisenstat et al., 1999). Hence, URE2 activity is similar to the endogenous expression of *Dlx1* and *Dlx2* in this region. Ongoing studies are aimed at establishing whether URE2 is regulating the expression of one *Dlx* gene or both.

Three different types of cells are defined by differential expression of URE2 and I12b in the SVZ and MZ of the dMGE, dCGE and mCGE at E11.5-E13.5: *URE2-lacZ+/I12b-AP+*, *URE2-lacZ+/I12b-AP-* and *URE2-lacZ-/I12b-AP+*. Some of the *URE2-lacZ+/I12b-AP-* cells are radially arranged in the VZ (Figure 3.3. C, Supplementary Figure 3.1. C and data not shown). The above cell types are also found in the vMGE/AEP and vCGE (Figure 3.3. C, Supplementary Figure 3.1. C and data not shown) although the activity of URE2 is weaker in these regions and the majority of labeled cells are *URE2-lacZ-/I12b-AP+*. These results suggest that there are microenvironments within MGE and CGE where distinct groups of progenitor cells express *URE2-lacZ* and/or *I12b-lacZ*.

I56i and I12b are both active in the MGE and CGE in most SVZ cells, and a substantial number MZ cells, (Figures 3.2., 3.3. D-F and Supplementary Figure 3.1. D-F). Thus, these two CREs are likely to be active in common cell populations at E11.5 - E13.5.

The CREs' overlapping activities recapitulate similarities in the endogenous expression of *Dlx* genes suggesting that they may be involved in common regulatory mechanisms such as cross-interactions between *Dlx* genes. In fact, *Dlx2*, *Dlx1*, and *Dlx5* share strong and uniform expression patterns in the SVZ and, intermediate ones in the MZ, where most cells express more than one *Dlx* gene (Eisenstat et al., 1999; Liu et al.,

1997). Previous work provided evidence that cross-regulatory interactions between *Dlx* genes explain some of the spatio-temporal similarities in *Dlx* patterns of expression. *Dlx1* and/or *Dlx2* are upstream regulators of *Dlx5* and *Dlx6*. In *Dlx1/2* null mice, *Dlx5* and *Dlx6* expression is lost in the most of the telencephalon except in cells derived from the septal region (Anderson et al., 1997b; Zerucha et al., 2000). Furthermore, Zerucha et al., 2000 demonstrated that cross-interaction between these four *Dlx* genes is mediated, at least in part, by the I56i enhancer. This enhancer is regulated by *Dlx1* and/or *Dlx2* both *in vitro* and *in vivo* (Stuhmer et al., 2002a; Zerucha et al., 2000; Zhou et al., 2004).

3.4.2. Distinct cell populations tangentially migrate to the cortex between E11.5 and E13.5

Differential spatio-temporal activities of URE2, I12b and I56i are detected in tangentially migrating cells that are derived from the MGE and CGE between E11.5 and E13.5. Based on activity of *Dlx* CREs, there are at least three major cell populations that tangentially migrate to the cortex at these ages: *URE2-lacZ+/I12b-AP+* (*I56i-lacZ+*), *URE2-lacZ+/I12b-AP-* (*I56i-lacZ-*), *URE2-lacZ-/I12b-AP+* (*I56i-lacZ+*). *URE2-lacZ* expression in tangentially migrating cells begins at least one day later than most *I12b/I56i-lacZ+* cells.

There are important spatio-temporal differences in the nature of *URE2-lacZ*-negative and *URE2-lacZ*-positive cells that migrate to the cortex: **1)** *URE2-lacZ*-positive cells appear to be primarily derived from the dMGE whereas *URE2-lacZ*-negative cells can derive from the dMGE and/or vMGE. The vMGE may be the major source of E11.5 migrations since URE2 is weakly active in this region and no tangential migration is detected in *URE2-lacZ* embryos at this age. Later, production of interneurons (between E12.5 and E15.5) may derive from both the dMGE and vMGE (Figures 3.2., 3.3., 3.4. and data not shown). The same conclusion may apply to the dCGE and/or mCGE versus vCGE (Supplementary Figures 3.1., 3.2. and data not shown). Alternatively, the increase in migrating cells expressing *URE2-lacZ* at later ages could also be due to increased URE2 enhancer activity in pre-existing cells that did not express the transgene

previously. **2)** Expression of the *URE2-lacZ* transgene is detected in a larger number of migrating interneurons at later stages (E13.5 and E14.5) (Figure 3.4., Supplementary Figure 3.2 and data not shown). **3)** The majority of cells expressing *URE2-lacZ* but not *I12b-AP* (*URE2-lacZ+/I12b-AP-*) follow the IZ as their major route of migration between E12.5 and E13.5 (Figures 3.3. I, 3.4. I and Table 3.1. B). In contrast, *URE2-lacZ-* cells migrate following both the Mz and IZ at these ages (Figures 3.3. I and 3.4. I).

Taken together, our data suggest that migrating cells expressing various combinations of URE2-*lacZ*+, I12b-*AP*+ and I56i-*lacZ*+ may correspond to distinct types of immature cortical interneurons derived from distinct microenvironments in the MGE and CGE. However, direct evidence for the existence of such lineage(s) is still lacking at this point.

3.4.3. Characterization of distinct sources of migrating cells within the MGE

Most mouse cortical GABAergic neurons are derived from the MGE and CGE (Anderson et al., 2002; Anderson et al., 2001; Gorski et al., 2002; Lavdas et al., 1999; Sussel et al., 1999; Wichterle et al., 1999; Wichterle et al., 2001). Using tissue transplantation coupled to DiI labeling in slice cultures, we provide evidence that the dMGE and vMGE are distinct sources of tangentially migrating cells, consistent with the regional expression of *URE2-lacZ* in the MGE. The DiI experiments seem to reveal more tangentially migrating cells from the vMGE versus dMGE between E12.5 and E13.5, consistent with the smaller size of tangential migration marked by *URE2-lacZ* compared with *I12b-lacZ* or *I56i-lacZ* (Figures 3.2. G-I, 3.3. G-L, 3.4., and Table 3.1. A). Alternatively, some of the dMGE migrating cells could derive from vMGE and are thus labeled at different positions along their migratory routes.

3.4.4. *Dlx* CREs are active in the majority of cortical interneurons in the adult cortex

Dlx genes are expressed in most mouse neocortical GABAergic interneurons (Stuhmer et al., 2001; Cobos et al., 2005, 2006). URE2 and I12b appear to be major

regulators of *Dlx1* and *Dlx2* in adult cortical GABAeric neurons, as they are expressed in 66% and 93% of these cells, respectively (Table 3.2. and data not shown). Most *URE2-lacZ*+ (~92%) and *I12b-lacZ*+ (~93%) cells in the cortex contain GABA (data not shown).

PV-, SOM-, and CR-positive interneurons constitute, molecularly and physiologically, distinct subtypes of cortical GABAergic neurons (Markram et al., 2004). *I12b-lacZ* but not *URE2-lacZ* labels the majority of interneurons expressing SOM (87% vs 7%) and VIP (96% vs 12%) (Figures 3.6. A-H, 3.7. Q-X and Table 3.2.). Moreover, *URE2-lacZ* marks only a few CB-positive interneurons (6.7%) (data not shown and Table 3.2.). In contrast, *URE2-lacZ* is expressed in nearly all PV-, CR-, NPY- and nNOS-expressing interneurons (>97% of each subtype) (Figures 3.6., 3.7. and Table 3.2.). *I12b-lacZ* is expressed in large numbers of interneurons belonging to these subtypes but not all (75-87% except for nNOS ~27%) (Figures 3.6., 3.7. and Table 3.2.).

We did not analyze co-expression of *I56i-lacZ* with markers of interneuron subtypes but we anticipate that it will label the same subtypes as *I12b-lacZ.* This conclusion is based on the fact that both enhancers were co-active in 70-80% of tangentially migrating cells between E12.5 and E13.5. Furthermore, Stuhmer et al., 2002b showed that zI46i, which is the zebrafish ortholog of I56i, marked >90% of adult cortical interneurons and showed no evidence for subtype specificity (Stuhmer et al., 2002b).

3.4.5. Distinct subtypes of cortical interneurons may derive from *Dlx*-progenitors born in subdivisions of the MGE and CGE

Previous studies showed that PV- and SOM-expressing interneurons derive primarily from progenitors located in the MGE between E12.5 and E16.5 (Butt et al., 2005; Xu et al., 2004). Our data are in agreement with these results but further propose that these subtypes may derive from distinct *Dlx*-progenitors born in subdivisions within the MGE such as dMGE and/or vMGE. Since nearly all PV- neurons express *URE2-lacZ*, it is likely that this subtype may derive from progenitors in which the URE2 CRE is

active and which are found in the dMGE rather than in the vMGE (Figure 3.8. A). Furthermore, this could correspond to the *URE2-lacZ+* cells that undergo robust tangential migrations to the cortex between at E13.5 and E14.5. In contrast, SOM-interneurons that exclusively express *I12b-lacZ* may derive from progenitors born primarily in the vMGE and, possibly, in the dMGE. In fact, as discussed earlier, *I12b-AP+/URE2-lacZ-* cells are mainly located in the vMGE but a substantial number of them are also found in the dMGE (Figure 3.3. C). Furthermore, SOM progenitor cells are observed in both subdivisions at E13.5 (Cobos I and Rubenstien J., unpublished), suggesting that both regions could contribute to this subtype (Figure 3.8. A). Interneurons expressing NPY and nNOS represent subpopulations of the SOM population (Freund and Buzsaki, 1996; Gonchar and Burkhalter, 1997; Kubota et al., 1994). Thus, they could derive from subgroups of progenitors found in the MGE and give rise to the SOM-expressing interneurons (Figure 3.8. A).

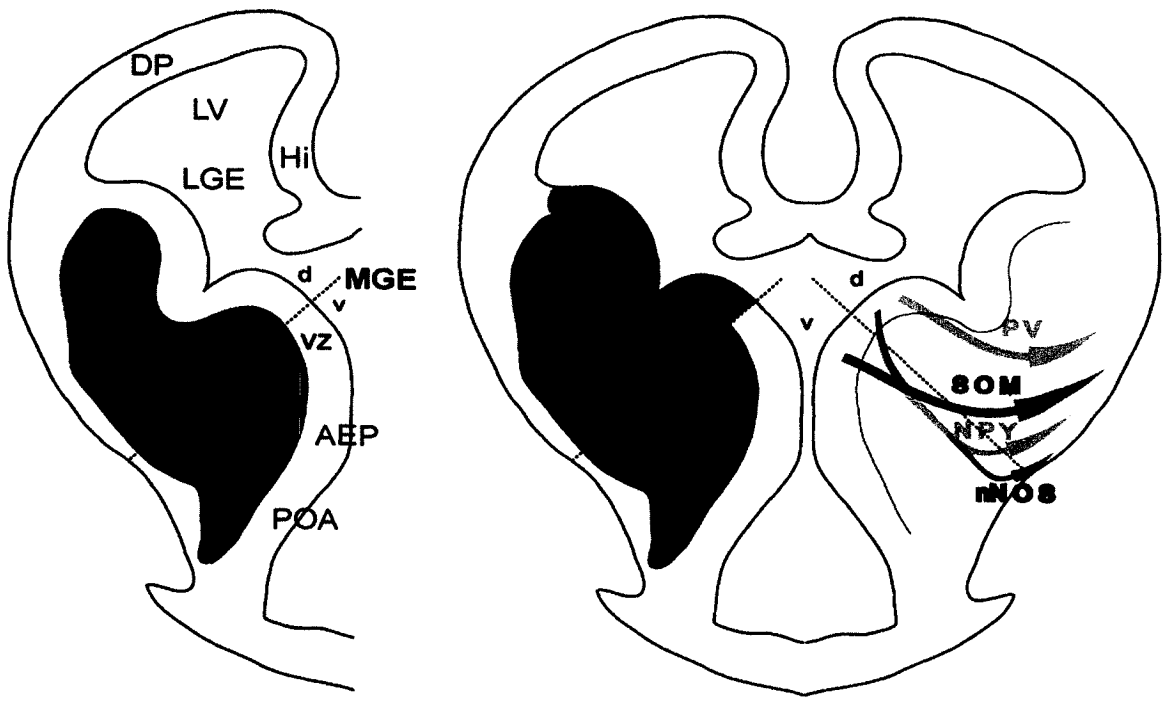
CR-expressing cortical interneurons express *URE2-lacZ* (~99%) and *I12b-lacZ* (~80%) suggesting that they may derive from *URE2-lacZ+/I12b-AP+* cells found in the dCGE and/or mCGE between E13.5 and E14.5 (Figure 3.8. B). Although direct evidence for the existence of such lineage(s) is still lacking, two arguments support this conclusion and favor the dCGE/mCGE over vCGE include: 1) both enhancers are strongly active in the dCGE and mCGE. However, URE2 is less active in the vCGE and is only active in the VZ of the mCGE (Figure 3.2. J and Supplementary Figure 3.1. A-C). 2) Previous studies have shown that interneurons expressing CR are derived exclusively from progenitors found in the dCGE between E13.5 and E15.5. (Butt et al., 2005; Xu et al., 2004).

Unlike CR+ interneurons, VIP+ interneurons mainly express *I12b-lacZ* (96%) and can derive from *I12b-lacZ+* progenitors located in any of the three subdivisions of the CGE. Arguments supporting a dCGE/mCGE origin (Figure 3.8. B) include: 1) *In utero* transplantation studies performed by Butt et al., 2005 showed that most VIP-progenitors are located in the dCGE. 2) CR and VIP populations are largely overlapping populations of interneurons in the rat frontal cortex (Kubota et al., 1994). Thus, one would expect a

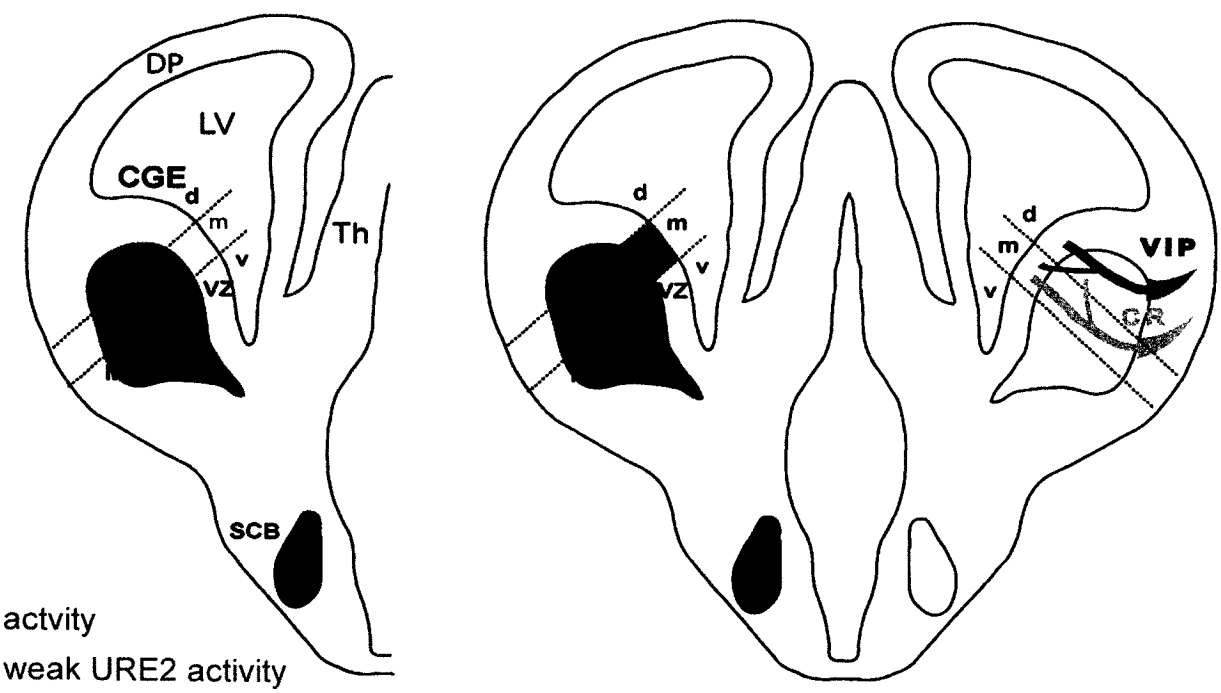
Figure 3.8. Proposed model for the origin of adult cortical interneurons subtypes in the embryonic mouse brain

(A, B) Schematic representation of E12.5 coronal hemisections/sections at the level of (A) the MGE and (B) the CGE. Enhancer activities of I12b (red) and URE2 (blue) in the MGE and CGE are depicted in the left hemispheres in (A, B). We propose the existence of microenvironments within dorsal and ventral parts of the MGE and CGE that produce different *Dlx*-positive progenitor cells. Although a direct lineage relationship is still lacking, these *Dlx*-progenitors may give rise later on to various subtypes of adult cortical GABAergic interneurons as shown in colored arrows in the right hemispheres in (A, B). (See discussion for altenative explanations). *Dlx*-progenitors that express *URE2-lacZ* and *I12b-AP* are represented by blue and red arrows, respectively. Progenitors that express both transgenes are shown in orange. Symbols as in Figure 3.2.

A



B



red : I12b actvity

light blue: weak URE2 activity

dark blue: strong URE2 activity

common origin for these neurons namely, the dCGE and/or mCGE. However, *URE2-lacZ* is expressed in only 11% of VIP-expressing cells in the somatosensory cortex and motor cortices. This is unexpected since all CR-expressing interneurons are labeled by this enhancer. One possible explanation for this discrepancy is that the distribution of CR+ interneurons is not identical between species. Alternatively, we could have underestimated the number of VIP+ cells that are marked by URE2.

In sum, these studies demonstrate that distinct *Dlx1/2* enhancer elements define molecularly distinct subdivisions of the MGE and CGE, and distinguish some adult cortical interneuron subtypes. It is noteworthy that, due to the limitation of the techniques used in this study, the existence of a direct lineage relationship between tangentially migrating cells expressing one or more transgenes/enhancers and the generation of specific subtype(s) of adult cortical interneurons discussed above is still lacking. Alternatively, *Dlx* enhancer activity in distinct progenitors and interneuron population(s) may not be reflective of a lineage relationship but rather be a sign of enhancer activity unrelated to specific lineage(s) that changes in time and space (switched on and off). Ongoing studies are aimed at using loss of function methods to establish the *in vivo* functions of these enhancers and lineage analyses to investigate whether their differential expression in the basal ganglia anlagae leads to the differential production of cortical interneuron subtypes or not.

Acknowledgments

We thank Dr Thomas Sargent laboratory for providing the guinea-pig β-gal antibody and Dr. Ruth Slack for useful comments on the manuscript. We thank Dr Luc Poitras for his assistance in drawing the model shown in Figure 3.8., Adrianna Gambarotta for creating the transgenic mice and Elaine Coderre for technical assistance with sectioning. N.G. was supported in part by a scholarship from the Lebanese

University, Beyrouth. This work is supported by a grant from the Canadian Institutes of Health Research (CIHR, #MOP14460) to M.E. and research grants to J.L.R.R. from Nina Ireland and NIH RO1 MH49428 and K02 MH01046. M.E. was supported by a CIHR Investigator Award during part of this work.

4. A Dynamic Spatial and Temporal Regulation of *Dlx* Expression Is Conferred by the Activities of Four *Dlx*-specific Enhancers in the Mouse Embryonic Telencephalon

Noël Ghanem[1], Gary Hatch[1], John L. R. Rubenstein[2], Marc Ekker[1, 3]

1 Centre for Advanced Research in Environmental Genomics, Dept. of Biology, University of Ottawa, Canada

2 Nina Ireland laboratory of Developmental Neurobiology, Centre for Neurobiology and Psychiatry Genetics, University of California at San Francisco, USA

3 Corresponding author: mekker@uottawa.ca

Short title: Four *Dlx* forebrain-specific enhancers

Keywords: regulation, homeobox, forebrain, enhancer activity, ß-galactosidase, transgenic mice, migration, differentiation

Abstract

Regulation of region-specific differentiation and migration in the embryonic forebrain is a complex mechanism that involves a variety of transcription factors such as the *Dlx* genes. Four forebrain-specific enhancers -CREs- (*cis*-acting regulatory elements) contribute to the *Dlx* transcriptional regulation in the subcortical telencephalon and the rostral diencephalon. Despite high sequence conservation among distant vertebrates, the *Dlx1/Dlx2* forebrain enhancers, URE2 and I12b, showed little sequence similarities with two previously characterized *Dlx* forebrain enhancers, I56i and I56ii, from the *Dlx5/Dlx6* locus. We carried out a detailed spatial and temporal analysis of the reporter transgene expression driven by each of the four enhancers. Despite significant overlaps, each enhancer targets expression to distinct groups of cells at all ages examined. I56ii is exclusively active in group(s) of cells in the mantle zone (MZ) of the basal ganglia between E12.5 and E13.5. Interestingly, the order of *lacZ* expression among the CRE lines differs with respect to the distance from the lateral ventricle to the MZ and changes during development from *URE2-lacZ / I12b-lacZ / I56i-lacZ / I56ii-lacZ* at E11.5-E12.5, to *URE2-lacZ / I56i-lacZ / I12b-lacZ / I56ii-lacZ* at E13.5-PO. This sequence of enhancer activity generally correlates with endogenous *Dlx* expression which largely follows this temporal pattern: *Dlx2/Dlx1/Dlx5/Dlx6*. Several regional differences between the activities of the four CREs were also highlighted in the primordia of the septum, the striatum and the pallidum throughout development. *lacZ* staining in all CRE lines except I156ii lines is found in *Dlx*-positive domains of the adult mouse brain including the neocortex one month after birth. This data reflect a complex and dynamic regulation of *Dlx* expression during embryonic development implicating several CREs with potentially identical and distinct role(s).

4.1. Introduction

Over the past fifteen years, genetics studies have focused on elucidating the mechanisms underlying the development of the telencephalon, one of the most complex

and diverse regions of the CNS. Several models of telencephalic organization and development have emerged implicating a diverse array of signaling molecules and transcription factors that control cell fate specification and regional identity as well as proliferation and differentiation [for reviews; (Marin, 2002; Puelles and Rubenstein, 2003; Schuurmans and Guillemot, 2002; Zaki et al., 2003)]. These studies have established that the embryonic telencephalon is subdivided into distinct progenitor domains within two major regions: the pallium (primordium of neocortex, hippocampus, piriform cortex and amygdala) and the subpallium (primordium of basal ganglia). Programs of regional identity defined by the transcription factors expressed in the progenitor cells, control most aspects of histogenesis within these domains; however, in some cases, particular cell types are generated in one region and then tangentially migrate to another domain, where that cell type apparently is not made, at least in large numbers. For instance, there is tangential migration of GABAergic interneurons from the subpallium to the cortex (Anderson et al., 1997a; de Carlos et al., 1996; Tamamaki et al., 1997). These neurons express *Dlx* genes (Porteus et al. 1994) and contribute the majority of interneurons to the mouse and chicken pallium (Anderson et al., 1997a; Anderson et al., 2001; Cobos et al., 2001; Gorski et al., 2002; He et al., 2001; Lavdas et al., 1999; Nery et al., 2002; Wichterle et al., 2001). Newly differentiating cells arising in the proliferative zones of the subpallium migrate tangentially through the MGE, LGE and CGE into the cerebral cortex, and rostro-dorsally from the dLGE along the rostral migratory stream into the olfactory bulb to give rise to several types of GABAergic interneurons and oligodendrocytes [reviewed in (Marin and Rubenstein, 2003a)]. Using several methods such as DiI labeling, retroviral transduction, transgene expression and transplantation, previous studies investigated the properties of subpallial interneurons in terms of timing of birth, origins, routes of migration, final destination and phenotype (Anderson et al., 2001; Marin and Rubenstein, 2001b; Nery et al., 2003; Stuhmer et al., 2002b; Wichterle et al., 2001).

Analyses of *Arx*, *Dlx1/Dlx2* and *Mash1* mutant mice have begun to define some of the genetic cascades underlying differentiation of forebrain GABAergic neurons

(Anderson et al., 1997b; Casarosa et al., 1999; Kitamura et al., 2002). There is evidence that *Mash1* is acting upstream of the *Dlx* genes (Fode et al., 2000; Yun et al., 2002) whereas *Arx*, *Dlx5/Dlx6* and *GAD67* are acting downstream of *Dlx1/Dlx2* [(Anderson et al., 1997a; Anderson et al., 1997b; Cobos et al., 2005; Stuhmer et al., 2002b; Zerucha et al., 2000); Cobos and Rubenstein, unpublished data)]. However, direct evidence for the mechanisms interconnecting the expression of the *Mash1*, *Dlx1/2/5/6*, *Arx* and *GAD67* genes is just beginning to be established, and will require identifying the *cis*-acting regulatory elements of these genes.

Towards elucidating the biochemical mechanisms that control development of telencephalic GABAergic neurons, we have investigated the regulatory elements of the *Dlx* genes. The vertebrate *Dlx* gene family is mainly comprised of three bigene clusters located on separate chromosomes; *Dlx1/Dlx2*, *Dlx5/Dlx6*, and *Dlx3/Dlx4* (additional genes exist in certain species, such as zebrafish). A short intergenic region (3.5kb-16kb) separates the two *Dlx* genes of each cluster (Ellies et al., 1997b; Ghanem et al., 2003; Liu et al., 1997; McGuinness et al., 1996; Nakamura et al., 1996; Simeone et al., 1994; Stock et al., 1996). During forebrain development, four *Dlx* genes -*Dlx1/2/5/6*- are expressed in the telencephalon and diencephalon, in overlapping cell populations, suggesting potential redundant functions especially between the two linked paralogs. Hence in the subpallial telencephalon, *Dlx1* and *Dlx2* expression is first induced in subpopulations of undifferentiated neural progenitors undergoing proliferation in VZ and SVZ. Shortly after, *Dlx5* and *Dlx6* expression is triggered in differentiating cells in the SVZ and MZ (Bulfone et al., 1993b; Eisenstat et al., 1999; Liu et al., 1997). Upon terminal differentiation, most *Dlx*-positive cells migrate towards their final destinations that include the striatum, the olfactory bulb, the hippocampus and the cerebral cortex, where they differentiate into GABAergic neurons, as well as others types of neurons (dopaminergic and cholinergic). While *Dlx1, 2* and *5* single null mutants exhibit subtle defects in olfactory bulb and cortical interneuron development (Acampora et al., 1999; Cobos et al., 2005; Long et al., 2003; Qiu et al., 1995), *Dlx1/Dlx2* double mutants have a

severe block in the differentiation of neurons derived from the LGE and MGE at around E12.5, with no detectable expression of *Dlx5* or *Dlx6* in these regions. A four-fold decrease in the number of cortical GABAergic interneurons was also observed (Anderson et al., 1997b; Bulfone et al., 1998; Marin et al., 2000; Yun et al., 2002).

We previously hypothesized that shared regulatory mechanisms may underlie the functional redundancy among *Dlx* paralogs. Furthermore, similarities in *Dlx* expression patterns led us to suggest that control mechanisms might have been conserved during evolution, persisting through the several rounds of duplication events that gave rise to multiple *Dlx* clusters. We and others have shown that the genomic organization of the *Dlx* genes is conserved among distant vertebrates and the intergenic regions are sites enriched for highly conserved *cis*-acting regulatory elements (CREs) (Zerucha et al., 2000; Sumiyama et al., 2002; Ghanem et al., 2003). Hence, we have earlier identified several CREs acting as forebrain specific enhancers for the *Dlx* genes in vertebrates; I56i and I56ii in the *Dlx5/Dlx6* intergenic region (Zerucha et al., 2000) and, URE2 and I12b in the *Dlx1/Dlx2* locus (Ghanem et al., 2003; Ghanem N and Ekker M, unpublished). Unexpectedly, despite high sequence conservation among distant vertebrates, the four enhancers displayed little sequence conservation when compared to each other, suggesting a rapid divergence of *Dlx* regulatory elements early in chordate/vertebrate evolution (Ghanem et al., 2003). To investigate potentially distinct role(s) among these enhancers, we carried out a detailed spatial and temporal analysis of reporter transgene expression driven by each enhancer between E10.5 and P25. We report a high degree of overlap in activity patterns among URE2, I12b and I56i in the telencephalon. However, each CRE also targets to distinct group(s) of cells at all ages examined. We also show that, unlike the previous enhancers, I56ii is active in a separate group(s) of cells in the deep mantle of the basal ganglia between E12.5 and E13.5 and is not active in tangentially migrating cells. The regional and laminar activities of the four CREs were also compared with the endogenous expression of the four *Dlx* genes expressed in the subpallial telencephalon. Our data reflect a complex and dynamic regulation of *Dlx*

expression during embryonic development implicating several regulatory elements with overlapping and distinct role(s).

4.2. Material and methods

4.2.1. Transgenic animals

For transgenic mice, sequences containing the four mouse enhancers (URE2, I12b, I56i and I56ii) were subcloned separately into the p1229/p1230 vectors (Yee and Rigby, 1993) that contain a human β-globin minimal promoter and the *l*acZ reporter gene. Subclonings were done using a PCR-based approach or using convenient restriction sites and followed by sequencing to verify the integrity of each insert (see 'Material and Methods' in chapters 2 and 3 of this thesis for PCR conditions, primers used and subcloning details). Transgenic animals were produced and analyzed as previously described (Zerucha et al., 2000). At least two independent transgenic lines that show *lacZ* staining were generated with each construct. Staining results on whole mount embryos and on coronal sections of the forebrain were replicated from independent lines.

4.2.2. Histology

E9.5 to E12.5 mouse embryos were fixed for 45min-1h in 4% cold PFA in PBS at RT, then, washed and stained with a *lacZ* substrate solution as described in 'Material and Methods' in chapter 2 of this thesis. The stained brains were dissected and equilibrated in 20% sucrose solution O/N at 4°C prior to sectioning. Brains were embedded in Tissue-Tek media the following day, cryoprotected and processed for frozen sectioning at 30-50µm using a cryostat (Leica CM3050 S). Sections were mounted with Aquatex (EM Science, VWR). E13.5-P0 mouse brains were fixed for 1-4h in 4% cold PFA in PBS at 4°C, dissected and sectioned prior to staining. *lacZ* staining was performed as described in 'Material and Methods' of chapter 2 of this thesis. One month old animals were anaesthetized and perfused with cold 4% PFA in PBS. Subsequently, the brains were removed and post-fixed for 2-4 hours, then equilibrated in 20% sucrose at 4°C O/N. The

next day, the brains were cryoprotected and sectioned as described above. Sections were post-fixed for 20min at RT in 10% formalin (3.7% formaldehyde/PBS), washed and stained with the *lacZ* substrate solution O/N as described earlier (chapter 2; 'Material and Methods'). After washing, slides were mounted with Aquatex the next day.

4.2.3. Double immunohistochemistry

E12.5 and E13.5 coronal sections of the mouse telencephalon were generated as described above. Frozen sections were dried for 1h, then, washed for 3 x 5min each in 0.1M PB to eliminate residues from tissue protection medium. Sections were incubated in blocking solution: 5% goat serum, 1% bovine serum albumin (BSA) and 0.2% gelatin in 0.5% PBST (1x PBS with 0.5% Tween-20), for at least 1h at RT. This was followed by incubation with the primary antibody at 4°C O/N. 3 x 10 min washes in 0.1M PB was performed the next day before incubating in the secondary antibody for 2h at RT away from light. Sections were washed for 3 x 10 min in 0.1M PB, processed for a second round of immunohistolabeling and, finally, mounted in Immuno Floure mounting media (Fisher).

We used the following antibodies: guinea pig anti-β-gal (1:1000, generous gift from Thomas Sargent lab), rabbit anti-PLAP (1:100 Serotec). Secondary antibodies were purchased from molecular probes –Invitrogen-; goat anti-rabbit Alexa fluor 488 and goat anti-guinea pig Alexa fluor 594.

4.2.4. *In situ* hybridization

In situ hybridization on mouse coronal sections was performed as follows: embryos were fixed for 2h at RT in 4% PFA in PBS (or O/N at 4°C), and then washed for 2 x 5min in PBS. Subsequently, embryos were dehydrated for 2 x 5 min in 100% methanol (MeOH) at RT and stored in the same solution at -20°C for at least 60 min (up to 6 months). Embryos were rehydrated progressively in three steps of 1 x 5 min each: 75% MeOH/25% PBS, 50% MeOH/50% PBS, 25% MeOH/75 % PBS, and then washed for 3 x 5 min in 0.1% PBST. After rehydration, brains were dissected, equilibrated in

20% sucrose solution at 4°C O/N and sectioned as described above. Sections were either processed for hybridization or stored at -20°C until required (up to one year). Prior to hybridization, the RNA anti-sense DIG-labeled probe is diluted (1:200-1:1000) in hybridization buffer [50% formamide, 1x salt, 10% dextran sulfate, 1mg/ml yeast tRNA, 1x Denhardt's, 1x double distilled H_2O (d.d.H_2O)]. To prepare 1 liter of 10x salt used in hybridization buffer we add the following: 114g of NaCl, 14.04g of Tris HCl, 1.34g of Tris base, 7.8g of $NaH_2PO_4.H_2O$, 7.1g of Na_2HPO_4, 100ml of 0.5M EDTA, in d.d.H_2O (pH=7.5). The probe mix is denatured for 5-10 min at 70°C, and then 200µl of the mix are added to each slide and covered with a coverslip. Hybridization is performed O/N at 65°C in a humidified chamber (1x salt/50% formamide). After hybridization, slides are transferred to a coplin jar and washed as follows: 3 x 30 min at 65°C in solution A (1x SSC, 50% formamide, 0.1% triton-X or tween 20 in d.d. H_2O) and 2 x 30 min at RT in 1x TBST (1.4M NaCl, 27mM KCl, 0.25M Tris HCl pH=7.5, 1% Tween 20 in d.d. H_2O). After washing, slides are blocked in blocking solution (10% heat-inactivated sheep serum in 1x TBST) for at least 1h at RT. Antibody staining consists of adding 100µl of 1:2000-5000 anti-digoxigenin (anti-DIG) Fab IgG fragment (Roche) in blocking solution and incubating O/N in a humidified chamber at 4°C (the slides are covered with coverslips). The next day, washes are performed at RT as follows: 4-5 x 20 min washes in 1x TBST, 2 x 10 min in fresh 1x NTMT [(100mM NaCl, 100mM Tris HCl pH 9.5, 50 mM $MgCl_2$, 0.1% Tween 20 in d.d.H_2O), the solution is filtered)]. Slides are stained for 10 min-2h in the dark at RT with staining solution [4ml of 1x NTMT, 14ul of BCIP (50mg/ml; Sigma) and 27ul of NBT (50mg/ml; Sigma)]. Stop the reaction with 3-5 x 10 min washes in d.d.H_2O followed by 2 x 10 min in TBST + 10mM EDTA. Fix slides in 4% PFA in PBS for 20 min and mount in aquamount (EMD science, VWR).

The digoxigenin-labeled *Dlx* probes were generated using the following cDNAs: *Dlx1*, 240bp fragment linearized with BamHI / T7 polymerase (Bulfone et al., 1997b); *Dlx2*, 560bp fragment linearized with EcoRI / T3 polymerase (Porteus et al., 1991; Bulfone et al., 1997b); *Dlx5*, 1.6kb fragment linearized with SmaI / T3 polymerase) (Liu et al., 1997) and *Dlx6*, 1.3kb fragment linearized with NotI / T3 polymerase (Liu et al.,

1997). *Dlx* radioactive probes were generated as described in Zerucha et al. 2000. Briefly, 10µg of each cDNA is linearized in 50µl of DEPC-treated water (add 0.1% DEPC to 1 liter of water in the fume hood O/N and autoclave). After digestion, the mix is extracted with 1 volume of phenol:chloroform (1:1) by vortexing for 30s and spinning for 3 min. An additional extraction with 1 volume of chloroform/AIA (24:1) is performed. After the second extraction, the linearized cDNA is precipitated with 2x ethanol and 1/10 of 3M sodium acetate (pH=5.2) at -20°C O/N or at -80°C for 30 min. This is followed by a 15 min. spin at 4°C. The cDNA is then washed with ice cold 70% ethanol, dried and resuspended in 10µl DEPC-water. Probe synthesis is performed at 37°C for 2h using: 1µl (1µg) of linearized cDNA, 2µl of NTP-labeling mix (Roche), 2µl of 10x transcription buffer (Roche), 1µl of RNAsin (20units/µl; Fermentas), 2µl of RNA polymerase (Roche) in d.d.H_2O (final volume is 20 µl). Following probe synthesis, the RNA/cDNA mix is precipitated with 2.5µl of 4M LiCl and 75 µl of 100% cold ethanol for 30 min at -80°C. The mix is spun for 30min in a Biofuge pico centrifuge (Kendro Laboratory Products) and washed twice with 70% ethanol and resuspended in RNase-free water. The size and amount of each probe were checked on a RNA gel in RNase-free conditions.

4.3. Results

4.3.1. Spatio-temporal analyses of the reporter gene expression -*lacZ*- during the development of the telencephalon

Divergence in the sequences of the four *Dlx* forebrain CREs namely, URE2, I12b, I56i and I56ii, suggests that each of them is involved in distinct aspects of *Dlx* regulation either spatially, temporally, or both. To test this possibility, we conducted a detailed spatial and temporal analysis, in the telencephalon, of *lacZ* reporter gene expression driven by each of the four CREs. At least, two independent transgenic lines were generated for *URE2-lacZ, I12b-lacZ, I56i-lacZ*, and *I56ii-lacZ*. We then compared, on coronal sections, the reporter gene expression conferred by all four enhancers in the telencephalon at E10.5, E11.5, E12.5, E13.5, E14.5, E15.5, P0 and P25. The activities of

the four enhancers were also compared with the endogenous expression of *Dlx* genes between E12.5 and E15.5.

4.3.2. E10-10.5

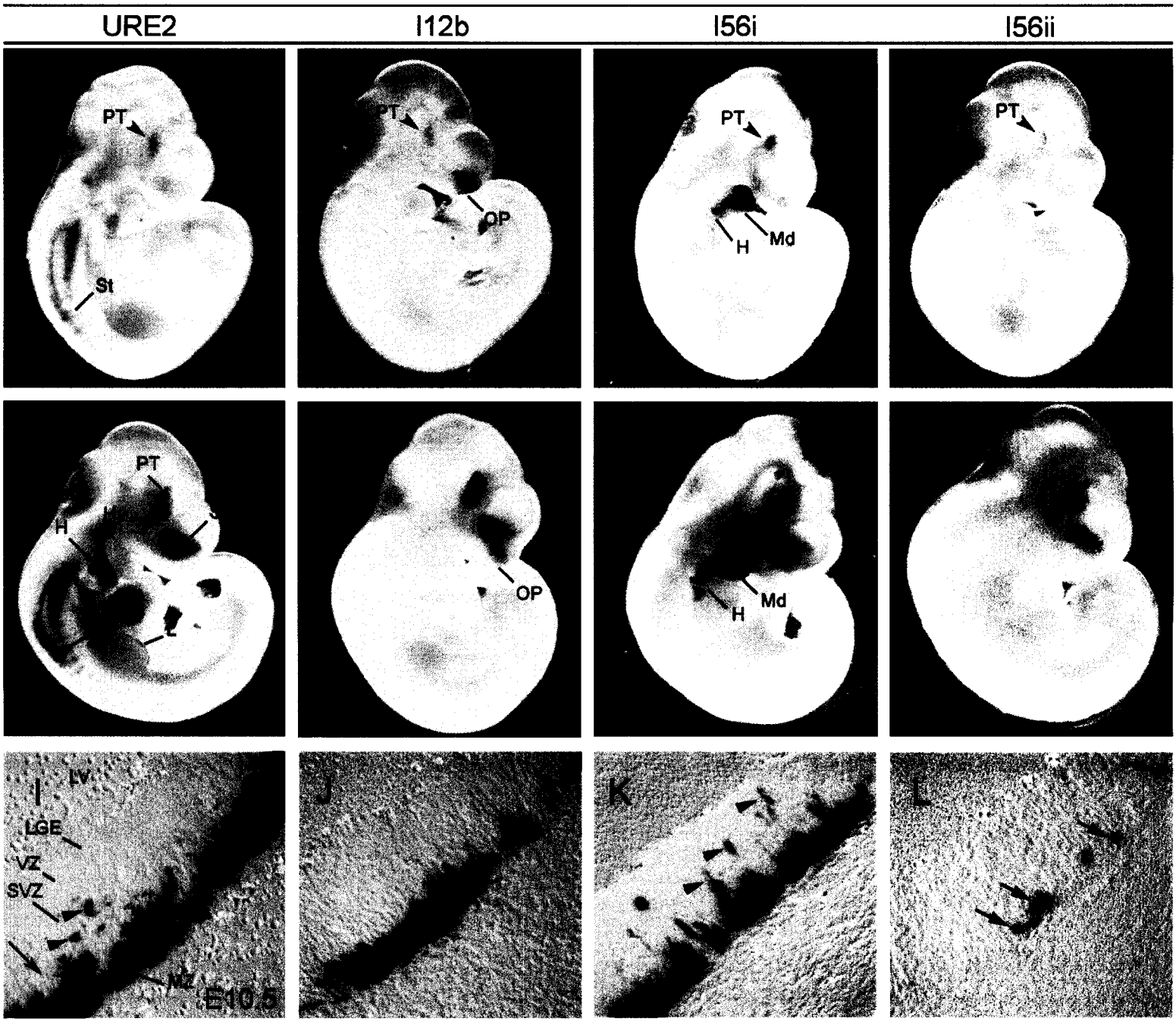
We determined the onset of *lacZ* expression conferred by all four enhancers is around E10 in a small cluster of cells located in the region of the primordia of the prethalamus (previously known as the ventral thalamus) (Figure 4.1. A-D; arrowheads). This is close to the onset of endogenous expression of *Dlx1* and *Dlx2* in the prethalamus at E9.0 and E9.5, respectively, as determined by *in situ* RNA hybridization (Bulfone et al., 1993a; Price et al., 1991) and by RNA blot analysis (Northern blot) (McGuinness et al., 1996). The onset of *Dlx5* and *Dlx6* expression occurs around E9.5 (Simeone et al., 1994).

Shortly after their onset, expression of all four transgenes becomes visible in the two domains where endogenous *Dlx* genes are expressed. Domain I or the diencephalon includes expression in the prethalamus and a Dlx^+ longitudinal domain in the hypothalamus. Domain II is the subpallial telencephalon (Figure 4.1. E-H) (Bulfone et al., 1993a; Bulfone et al., 1993b; Eisenstat et al., 1999; Ghanem et al., 2003; Liu et al., 1997; Porteus et al., 1994). Ectopic expression is seen in the dorsal diencephalon (mesencephalon) in the *I56i-lacZ* line shown here (Figure 4.1. C and G) but is not consistent among lines. In *URE2-lacZ*, *I12b-lacZ*, and *I56i-lacZ* transgenics, β-galactosidase-positive cells are largely absent from the VZ at E10.5 except for scattered cells in the URE2 and I56i lines (Figure 4.1. I and K; arrowheads), and are concentrated as a thick subpial layer (Figure 4.1. I-K). The subpial tissue consists of the emerging SVZ and MZ (Yun et al., 2002). Unlike expression from the other three enhancers, *I56ii-lacZ* transgenics produced only a few $lacZ^+$ cells that are present in the subpial region of the LGE (Figure 4.1. L; arrows).

In all four transgenic lines, the β-galactosidase-positive cells almost certainly are *Dlx*-positive since the patterns of enhancer activity at this age overlap with the endogenous patterns of *Dlx* expression. However, the transgene expression in all lines

Figure 4.1. Reporter gene expression in *Dlx*-CRE transgenic mice

(A-H) *lacZ* expression in whole mount transgenic embryos carrying *lacZ* reporter under the control of URE2 (A, E), I12b (B, F), I56i (C, G) and I56ii (D, H) *Dlx* enhancers at E10 (A-D) and E10.5 (E-H). Arrowheads in A-D indicate the onset of *lacZ* expression in the prethalamus. At E10.5, all four CREs target *lacZ* expression to the subpallial telencephalon and two domains in the diencephalon: prethalamus (PT) and hypothalamus (Hy). *lacZ* staining is also detected in other *Dlx*-positive domains such as the limbs (*URE2-lacZ* lines), the branchial archs (*URE2-* and *I56i-lacZ* lines) and the olfactory placodes (*I12b-lacZ* lines). (I-L) Comparison of *lacZ* expression on coronal sections in the neuroepithelium of transgenic mice at E10.5. (I-K) A uniform and thick layer of *lacZ* staining is seen in the subpial region of the LGE in (I) *URE2-lacZ,* (J) *I12b-lacZ* and (K) *I56i-lacZ* transgenic lines. Only a few stained cells were found in the previous region in *I56ii-lacZ lines* (L; arrows). The arrow in (I) indicates the boundary between MGE and LGE. Arrowheads in (I) and (K) mark stained cells found in the VZ and/or SVZ of LGE in *I56i-lacZ* and *URE2-lacZ* lines. Abbreviations: H, Hyoid arch; Hy, Hypothalamus; L, limbs; LGE, lateral ganglionic eminence; LV, lateral ventricle; Md, mandibular component of the first branchial arch; MZ, mantle zone; OP, olfactory placodes; PT, prethalamus; St, somites; ST, subpallial telencephalon; SVZ, subventricular zone; VZ, ventricular zone.


URE2
I12b
I56i
I56ii
PT
St
OP
H
Md
I
LV
LGE
VZ
SVZ
MZ
E10.5
J
K
L

examined lack one feature of telencephalic *Dlx* expression at E10.5; previous studies showed that the dorsal LGE (dLGE) has a high concentration of $DLX2^+$ cells in the VZ (Eisenstat et al., 1999; Yun et al., 2002). As none of the transgenic lines studied here show this property, it suggests that this expression is conferred by an unknown enhancer element.

4.3.3. E11.5

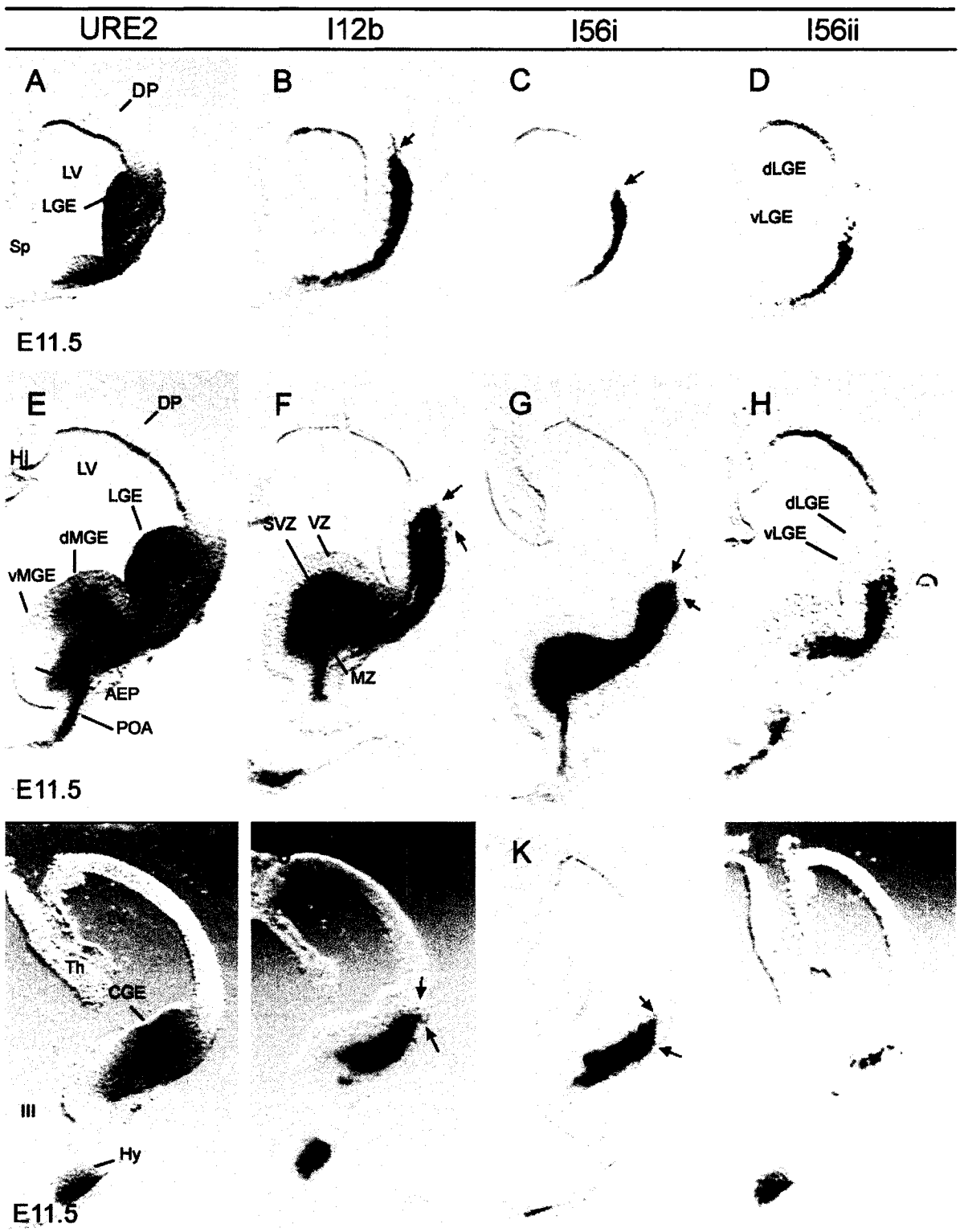
Transgenic animals produced with all four reporter constructs show stronger *lacZ* expression in the subpallial telencephalon at E11.5 compared to E10.5. Transgene expression in the forebrain is restricted to regions where endogenous *Dlx* genes are expressed. Expression patterns of the four transgenes overlap extensively although discrete and reproducible differences in enhancer activity were observed. *URE2-lacZ* expression is characterized by its nearly homogeneous expression in the VZ, SVZ and MZ in most areas of the subpallium [rostroventral pole (part of the olfactory bulb anlage) dLGE, vLGE, dMGE, CGE] (Figure 4.2. A, E, I and data not shown), except in the vMGE, AEP and POA, where it is excluded from the VZ (Figure 4.2. E).

Unlike the URE2 transgenics, generally no staining was detected in the telencephalic VZ in transgenics produced with the other three enhancers. *I12b-lacZ* and *I56i-lacZ* lines show nearly identical expression in the subpallial SVZ and MZ (Figure 4.2. B, C, F, G, J and K), except in the rostral pole, where *I12b-lacZ* may be in the VZ (data not shown). As at E10.5, expression from the *I56ii-lacZ* construct is in a subset of MZ cells, that are largely limited to the vLGE, dMGE, part of the vMGE, CGE and POA (Figure 4.2. D, H and L). Expression of *lacZ* from all four lines is sparse in the septum at this age (Figure 4.2. A-D and Table 4.1. A), as is the case for endogenous *Dlx* expression.

Thus, within the entire subpallium at E11.5, enhancer activity followed this sequence with respect to the radial dimension of the neuroepithlium (VZ-SVZ-MZ): *URE2-lacZ* > *I12b-lacZ* > *I56i-lacZ* > *I56ii-lacZ*.

Figure 4.2. Comparison of the activities of four *Dlx* CREs in the subpallial telencephalon of transgenic mice at E11.5

(A-L) Coronal hemisections showing *lacZ* expression when under the control of URE2 (A, E, I), I12b (B, F, J), I56i (C, G, K) and I56ii (D, H, L) enhancers at rostral (A-D), medial (E-H) and caudal (I-L) levels. Arrows in (B, C, F, G, J and K) mark the onset of tangential migration to the dorsal pallium in *I12b-lacZ* and *I56i-lacZ* lines. Note that no migrating *lacZ*-positive cells are detected with the other two enhancer lines. Abbreviations: III, third ventricle; AEP, anterior entopendular area; CGE, caudal ganglionic eminence; dLGE, dorsal lateral ganglionic eminence; dMGE, dorsal medial ganglionic eminence; DP, dorsal pallium; Hi, hippocampus; Hy, Hypothalamus; LGE, lateral ganglionic eminence; LV, lateral ventricle; MGE, medial ganglionic eminence; MZ, mantle zone; POA, preoptic area; Sp, septum; SVZ, subventricular zone; Th, thalamus; vLGE, ventral lateral ganglionic eminence; vMGE, ventral medial ganglionic eminence; VZ, ventricular zone.


URE2
I12b
I56i
I56ii
A
DP
LV
LGE
Sp
B
C
D
dLGE
vLGE
E11.5
E
DP
HI
LV
LGE
dMGE
vMGE
AEP
POA
F
SVZ
VZ
MZ
G
H
dLGE
vLGE
E11.5
Th
CGE
III
Hy
E11.5
K

4.3.4. E12.5

Regional differences in the expression of the four transgenes, and their relative activities with respect to the radial dimension of the neuroepithlium, persist at E12.5. Salient modifications in the expression patterns compared to E11.5 are as follows: 1) there is definitive expression in the septum at this age, which persists until P25 in all CRE lines except *I56ii-lacZ* lines (Figures 4.3., 4.4., 4.6. and 4.7.; panels A-D; and Table 4.1. A); 2) tangentially migrating cells are seen entering the cortex in the I12b and I56i lines (Figure 4.3. F, G, J and K; arrowheads), note that, a few tangentially migrating cells are also seen at E11.5 in the I12b and I56i lines (Figure 4.2. B, C, F, G, J and K; arrows); 3) the intensity of *URE2-lacZ* expression in the neuroepithelium of all eminences and POA varied as follows: SVZ > MZ > VZ (Figure 4.3. A, E and I); 4) the intensity of I12b and I56i expression is MZ > SVZ (Figure 4.3. B, C, F, G, J and K).

We compared transgene expression with the distribution of endogenous *Dlx* transcripts. Firstly, *Dlx2*, *Dlx1* and *URE2-lacZ* are expressed in the VZ of the ganglionic eminences (except vMGE in *URE2-lacZ* lines), and, thus, distinguish themselves from the other *Dlx* genes and enhancers, respectively (Figure 4.3. E, M and N). Secondly, the *Dlx* genes (except *Dlx6*), and, the transgenes (except *I56ii-lacZ*), when compared, are expressed with apparently equal intensity in the SVZ (Figure 4.3. E-G and M-O). Thirdly, in the MZ, the intensity of *Dlx* expression is as follows: *Dlx5* > *Dlx1* > *Dlx2* (Figure 4.4. M-O) (Eisenstat et al. 1999) whereas the intensity of *lacZ* expression in the same zone in the CRE lines follows this order: *I56i-lacZ* = *I12b-lacZ* > *URE2-lacZ* (Figure 4.4. E-G). Taken together, we conclude that URE2 activity is similar to *Dlx2* expression whereas I12b and I56i activities resemble mainly *Dlx5* expression. Finally, some similarities exist between the expression of *Dlx6* and the activity of *I56ii-lacZ* transgenics; both are considered the weakest in the SVZ and are basically restricted to the deep MZ where cells have reached an advanced stage of differentiation (Figure 4.3. H and P) (Simeone et al., 1994; Liu et al., 1997).

In sum, *lacZ* expression, with respect to the lateral ventricle as reported here with the four CREs -URE2/I12b/I65i/I56ii- was comparable to the endogenous order of expression of the four *Dlx* genes *Dlx2/Dlx1/Dlx5/Dlx6*- at E12.5.

4.3.5. Tangential migration towards the dorsal pallium

Two major paths of tangential migration from the subpallium have been described: a 'dorsal' migration to the pallium and a "rostral" migration to the olfactory bulb. We have described earlier the migration to the cortex, between E12.5 and E13.5, from three CREs lines: URE2, I12b and I56i (Chapter 3 of this thesis). In addition, we show here that: 1) there is no tangential migration in *I56ii-lacZ* lines between E11.5 and E15.5 (Figures 4.2., 4.3., 4.4. and 4.6.; panels D, H and L), 2) starting at E14.5, the *URE2-lacZ*, *I12b-lacZ* and *I56i-lacZ* transgenes are expressed in a third migratory pathway that runs through the cortical SVZ, deep to the Mz and IZ (Figure 4.6. A-C, E-G, I-K and data not shown), consistent with previous reports showing that later born *Dlx*-positive cells derived from both the MGE and LGE migrate largely within the cortical proliferative zones between E14.5 and E16.5 (Figure 4.6. M-P) (Anderson et al., 2001; Stuhmer et al., 2002b); 3) β-galactosidase-positive neurons are still found in the mouse neocortex at birth (Figure 4.7.) as well as at one month after birth (data not shown) in all CRE lines except I56ii lines.

4.3.6. I56ii marks distinct group(s) of cells at E12.5 and E13.5

We noticed that a major regional modification in the activities of the four CREs occurs between E12.5 and E13.5 in part of the MZ of the MGE and CGE. Whereas at E12.5, *URE2-lacZ*, *I12b-lacZ* and *I56i-lacZ* lines show strong and homogenous expression throughout the MZ of the MGE and CGE (Figure 4.3. E-G and I-K, Table 4.1. C), by E13.5, mantle expression is greatly reduced in the MGE and CGE in all three lines (Figure 4.4. E-G and I-K; arrows, Table 4.1. C). In parallel, *I56ii-lacZ* expression seems to be largely specific to a group of cells located in the MZ of the MGE and CGE (Figure

Figure 4.3. Comparison of the activities of four *Dlx* CREs in the subpallial telencephalon of transgenic mice at E12.5

(A-L) Coronal hemisections showing *lacZ* expression when under the control of URE2 (A, E, I), I12b (B, F, J), I56i (C, G, K) and I56ii (D, H, L) enhancers at rostral (A-D), medial (E-H) and caudal (I-L) levels. Arrowheads in (F, G, J and K) show two streams of tangentially migratory cells that are budding from the LGE (F, G) and CGE (J, K) and merging into one stream at the level of the cortico-striatal boundary. Double arrowheads in (E-H) show the distance of *lacZ* expression with respect to the lateral ventricle in the CRE lines. The order of activity of the four CREs at E12.5 is as follows: URE2/I12b/I56i/I56ii. (M-P) *In situ* hybridization on E12.5 coronal hemisections showing the expression of *Dlx2* (M), *Dlx1* (N), *Dlx5* (O) and *Dlx6* (P) at medial levels. Abbreviations: III, third ventricle; AEP, anterior entopendular area; CGE, caudal ganglionic eminence; dLGE, dorsal lateral ganglionic eminence; dMGE, dorsal medial ganglionic eminence; DP, dorsal pallium; Hi, hippocampus; Hy, Hypothalamus; LGE, lateral ganglionic eminence; LP, lateral pallium; LV, lateral ventricle; MGE, medial ganglionic eminence; MZ, mantle zone; POA, preoptic area; Sp, septum; SVZ, subventricular zone; Th, thalamus; vLGE, ventral lateral ganglionic eminence; vMGE, ventral medial ganglionic eminence; VP, ventral pallium; VZ, ventricular zone.

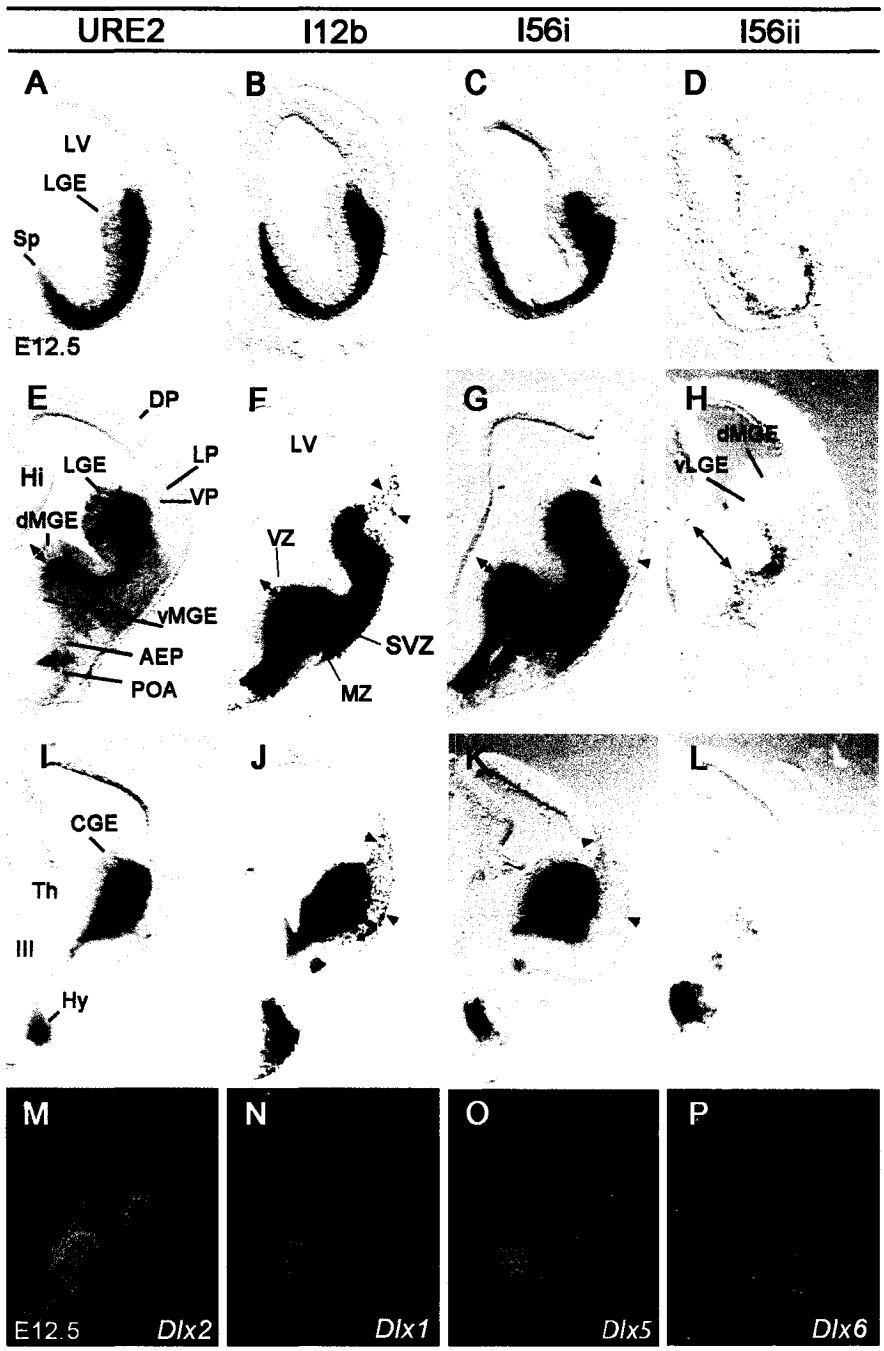

URE2
I12b
I56i
I56ii
A
B
C
D
LV
LGE
Sp
E12.5
E
DP
LP
Hi
LGE
VP
dMGE
vMGE
AEP
POA
F
LV
VZ
SVZ
MZ
G
H
dMGE
vLGE
I
CGE
Th
III
Hy
J
K
L
M
E12.5
Dlx2
N
Dlx1
O
Dlx5
P
Dlx6

Table 4.1. Summary of *lacZ* expression in various telencephalic domains of the *Dlx*-CRE lines between E10.5 and P25

A

Septum	10.5	11.5	12.5	13.5	15.5	P0	P25
URE2	-	+	+++	+++	+++	++	++
I12b	-	+	+++	+++	++	+	+
I56i	-	+	++	+++	+++	++	++
I56ii	-	-	+	++	-	-	-

B

Striatum	10.5	11.5	12.5	13.5	15.5	P0	P25
URE2	+++	+++	+++	+++	+++	+++	++
I12b	++	+++	+++	++	+	+	+
I56i	++	+++	+++	+++	+++	++	++
I56ii	+	++	+	+	-	-	-

C

Pallidum	10.5	11.5	12.5	13.5	15.5	P0	P25
URE2	+++	+++	+++	+	-	-	-
I12b	++	+++	+++	+	++	++	++
I56i	++	+++	+++	-	+++	+	+
I56ii	+	+	+	++	-	-	-

D

Olfactory Bulb	P0					P25			
	SVZ	GCL	ML	EPL	GL	GCL	ML	EPL	GL
URE2	+++	++	++	-	+	++	+	-	++
I12b	+	++	++	-	++	++	+	-	++
I56i	++	++	+	-	+	+++	++	-	++
I56ii	-	-	-	-		-	-	-	-

(-), no expression; (+), weak; (++), intermediate and (+++), strong.

4.3. H, L; 4.4. H, L and Table 4.1. C). In order to investigate whether I56ii labels distinct group(s) of cells, we generated an alkaline phosphatase (AP) reporter line under the control of I12b (*AP-I12b*) and mated it with the *I56ii-lacZ* line. We performed double immunohistochemistry on brain coronal sections from double hemizygote embryos (*AP+/lacZ+*) at E12.5 and E13.5 using antibodies against β-gal and AP. We found that I56ii, but not I12b, is active in a transverse column of cells lining the MZ of the vLGE and MGE at E12.5 and E13.5 (Figure 4.5. A-F; arrowheads in F and data not shown). We did not check whether this is also true at E10 and E11.5. Nevertheless, our data suggests that I56ii is likely active in distinct subset(s) of cells precisely between E10.5 and E13.5 in comparison with I12b, and its activity is not detectable after E14.5 except in some hypothalamic cells (Figure 4.6. H, Tables 4.1. B and C, and data not shown).

4.3.7. Enhancer activity in the telencephalon from E13.5 to P25

Subpallial expression of the four CREs at E13.5 shows several interesting regional and laminar differences compared to earlier stages. We will first address changes in the laminar pattern.

URE2 activity in the VZ is reduced at E13.5 compared to younger ages (Figures 4.4 A, E, I and 4.5. K). Next, while at earlier ages I12b activity precedes I56i, by E13.5, I56i activity is closer to the ventricle than I12b activity; *I56i-lacZ* is still strongly expressed in the SVZ (Figures 4.4. C, G, K and 4.5. H) whereas *I12b-lacZ* becomes weaker in this region in most of the subpallium, except in a stripe of cells located at the dorsal tip of the dLGE (Figures 4.4. B, F, J and 4.5. G, J; arrows). Thus, at E13.5, the order of expression with respect to the lateral ventricle is as follows: URE2 / I56i / I12b / I56ii (Figures 4.4. E-H and 4.5. G, H, J, K; double arrows). This order stays the same until P25 (Figures 4.6., 4.7., Table 4.1. and data not shown). In contrast, the order of expression of the *Dlx* genes at this age is identical to E12.5; *Dlx2* > *Dlx1* > *Dlx5* > *Dlx6* (Figure 4.4. M-P). Also, at E13.5, *lacZ* expression in the SVZ of the LGE in URE2, I12b and I56i lines is continuous with a radial stream of cells that are aligned along the pallial/subpallial limit in medial and caudal regions (Figures 4.4. E-G, I-K and 4.5. G, H,

J, K). This is particularly clear for I12b in the CGE, where its expression forms a boundary around the subpallium (Figure 4.4. J).

At E15.5, *URE2-lacZ* and *I56i-lacZ* expression remain very strong in the proliferative zones of the LGE, MGE and CGE whereas I12b expression greatly decreases (Figure 4.6. and Table 4.1.). At P0, URE2 and I56i activities remain robust in the striatum whereas I12b activity is barely detectable in this region (Figure 4.7. A-C; E-G and Table 4.1. B). At P25, URE2 and I56i activities persist in small periventricular domains adjacent to the striatum that are continuous with the rostral migratory stream (Table 4.1. B and data not shown). Robust expression from the URE2, I56i and I12b transgenes is observed in all layers of the olfactory bulb (OB) containing GABAergic neurons at P0 and P25 but with different intensities: the SVZ of the OB, the granular cell layer (GCL), the mitral layer (ML) and the glomerular layer (GL) (Figure 4.8. A-C, E-G; Table 4.1. D and data not shown). *I56ii-lacZ* is not expressed in the OB (Figure 4.8. D, H and Table 4.1. D).

While some aspects of the regional pattern of subpallial *lacZ* expression are stable, other aspects are highly dynamic. URE2 is actived in the striatum but not in the pallidum at E15.5-P0 (Figures 4.6. E; 4.7. E, arrow, Tables 4.1. B, C and data not shown), and, by P25, there is only trace expression in this region (Table 4.1. C and data not shown).

I12b activity strikingly decreases in the striatum between E13.5 and E15.5 (Figures 4.4. B, F and 4.6. B, F). Striatal expression remains very low at P0 (Figure 4.7. B, F; Table 4.1. C and data not shown). At P25, striatal expression persists in scattered cells, particularly in the caudal striatum (Table 4.1. C and data not shown). While I12b expression was not detected in the pallidum at E13.5, expression is moderate in caudal parts of this region at E15.5 and P0 and weak at P25 (Figures 4.4. F, 4.6. F, 4.7. F and data not shown).

Figure 4.4. Comparison of the activities of four *Dlx* CREs in the subpallial telencephalon of transgenic mice at E13.5

(A-L) Coronal hemisections showing *lacZ* expression when under the control of URE2 (A, E, I), I12b (B, F, J), I56i (C, G, K) and I56ii (D, H, L) enhancers at rostral (A-D), medial (E-H) and caudal (I-L) levels. Arrowheads in A-C, E-G and I-K indicate the two major routes, IZ and Mz, followed by tangentially migrating cells to the DP at E13.5. The same migratory routes are followed by *Dlx*-expressing cells (M-P; arrowheads). Double arrowheads in (E-H) show the distance of *lacZ* expression with respect to the lateral ventricle in each CRE line. The order of enhancer activity at E13.5 is URE2/I56i/I12b/I56ii. Arrows in H and L indicate a restricted domain of expression of *I56i-lacZ* in the MZ of the vLGE and MGE (H) as well as the CGE (L). Of note, no *lacZ* staining is detected in the *I56ii-lacZ*+ domain with the other three CRE lines (E-G, I-K; arrows). (M-P) *In situ* hybridization on E13.5 coronal hemisections showing the expression of *Dlx2* (M), *Dlx1* (N), *Dlx5* (O) and *Dlx6* (P) at medial levels. Abbreviations: III, third ventricle; AEP, anterior entopendular area; CGE, caudal ganglionic eminence; Cp, caudate putamen; CP, cortical plate; dLGE, dorsal lateral ganglionic eminence; dMGE, dorsal medial ganglionic eminence; DP, dorsal pallium; Hi, hippocampus; Hy, Hypothalamus; IZ, intermediate zone; LGE, lateral ganglionic eminence; LV, lateral ventricle; MGE, medial ganglionic eminence; Mz, marginal zone; MZ, mantle zone; PD, pallidum; POA, preoptic area; Sp, septum; ST, subpallial telencephalon; SVZ, subventricular zone; Th, thalamus; vLGE, ventral lateral ganglionic eminence; vMGE, ventral medial ganglionic eminence; VZ, ventricular zone.

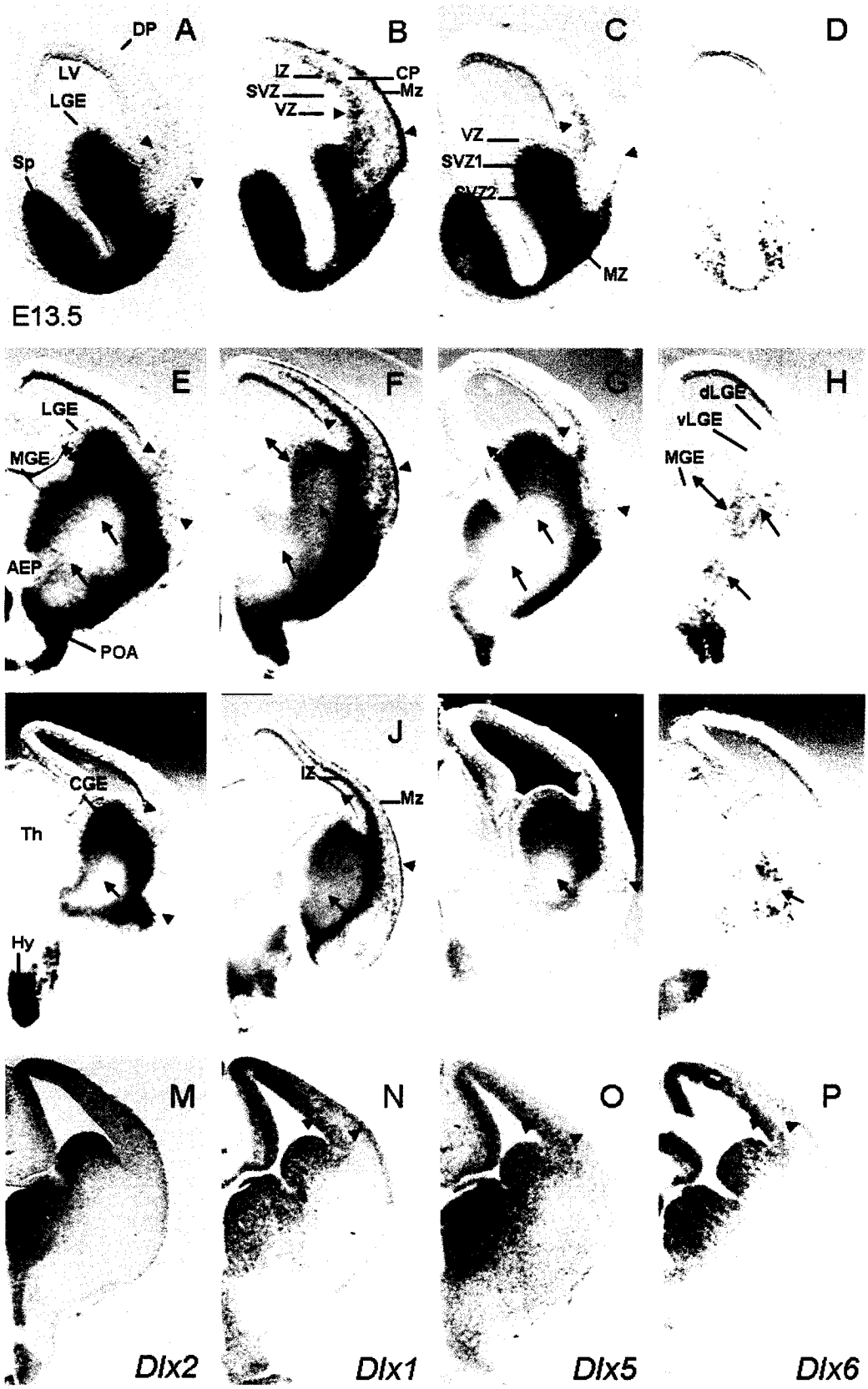

URE2
I12b
I56i
I56ii
A
B
C
D
DP
LV
LGE
Sp
IZ
CP
SVZ
Mz
VZ
VZ
SVZ1
SVZ2
MZ
E13.5
E
F
G
H
LGE
MGE
AEP
POA
dLGE
vLGE
MGE
J
CGE
Th
Hy
IZ
Mz
M
N
O
P
Dlx2
Dlx1
Dlx5
Dlx6

Figure 4.5. I56ii labels distinct group(s) of cells in the subpallial telencephalon at E13.5

(A-F) Double immunohistochemistry showing the activity of I12b in green (A, D) and I56ii in red (B, E) in the MZ of the LGE and MGE. (C, F) Merged pictures of (A, B) and (D, E), respectively. (D, E, F) Higher magnification pictures of the domains shown in insets in (A, B, C), respectively. I56ii-positive cells are not co-labeled with I12b (F; arrowheads). (G-L) Double immunohistochemistry showing the activity of I12b (G, J) in green, and, I56i (H) and URE2 (K) in red, in the LGE. (I, L) Merged pictures of (G, H) and (J, K), respectively. Note that the activity of I12b in the SVZ of the LGE is weaker compared with the activities of I56i and URE2, and is particularly strong in a group of cells located at the dorsal tip of the LGE (G, J; arrows). The distance of *lacZ* or AP expression with respect to the lateral ventricle in the CRE lines is indicated with double arrowheads in G, H, J and K. Legend as in Figure 4.4.

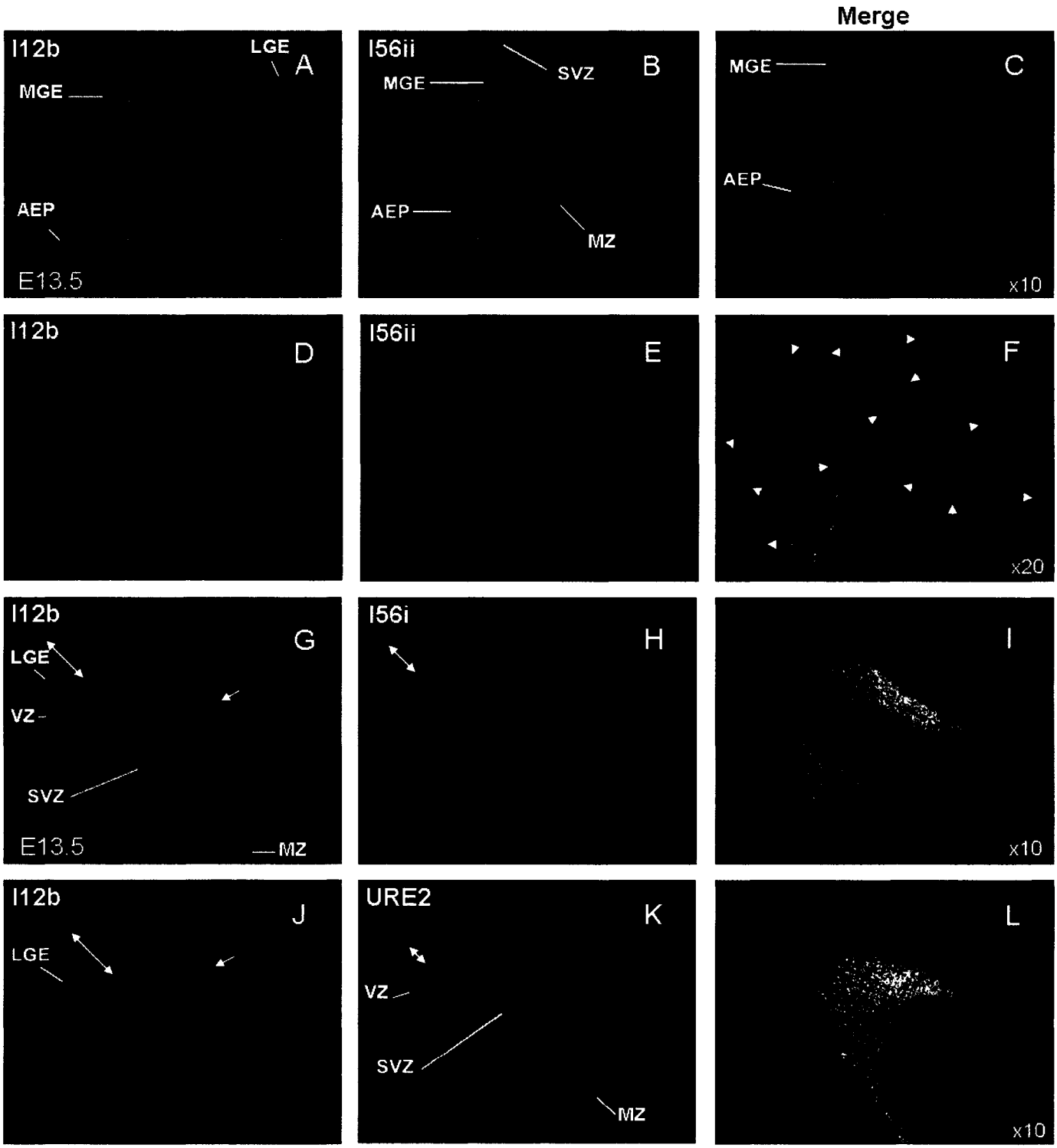

Merge
I12b
LGE
A
MGE
AEP
E13.5
I56ii
SVZ
B
MGE
AEP
MZ
MGE
C
AEP
x10
I12b
D
I56ii
E
F
x20
I12b
G
LGE
VZ
SVZ
E13.5
MZ
I56i
H
I
x10
I12b
J
LGE
URE2
K
VZ
SVZ
MZ
L
x10

Figure 4.6. Comparison of the activities of four *Dlx* CREs in the subpallial telencephalon of transgenic mice at E15.5

(A-L) Coronal hemisections showing *lacZ* expression when under the control of URE2 (A, E, I), I12b (B, F, J), I56i (C, G, K) and I56ii (D, H, L) enhancers at rostral (A-D), medial (E-H) and caudal (I-L) levels. Arrowheads in A-C, E-G and I-K indicate the three routes followed by tangentially migrating cells to the DP at E15.5: SVZ, CP and Mz. The same migratory routes are followed by *Dlx*-expressing cells (M-P; arrowheads). Arrows in E-G mark the pallidum (PD) where I12b (F) and I56i (G) but not URE2 (E) are active. At this age, *I12b-lacZ* expression regresses considerably in the LGE and MGE (F) whereas expression from *URE2-lacZ* (E) and *I56i-lacZ* (G) lines remains strong. No *lacZ* staining is detected with *I56ii-lacZ* lines except in the hypothalamus (L; arrow). (M-P) *In situ* hybridization on E15.5 coronal hemisections showing the expression of *Dlx2* (M), *Dlx1* (N), *Dlx5* (O) and *Dlx6* (P) at medial levels. Legend as in Figure 4.4.

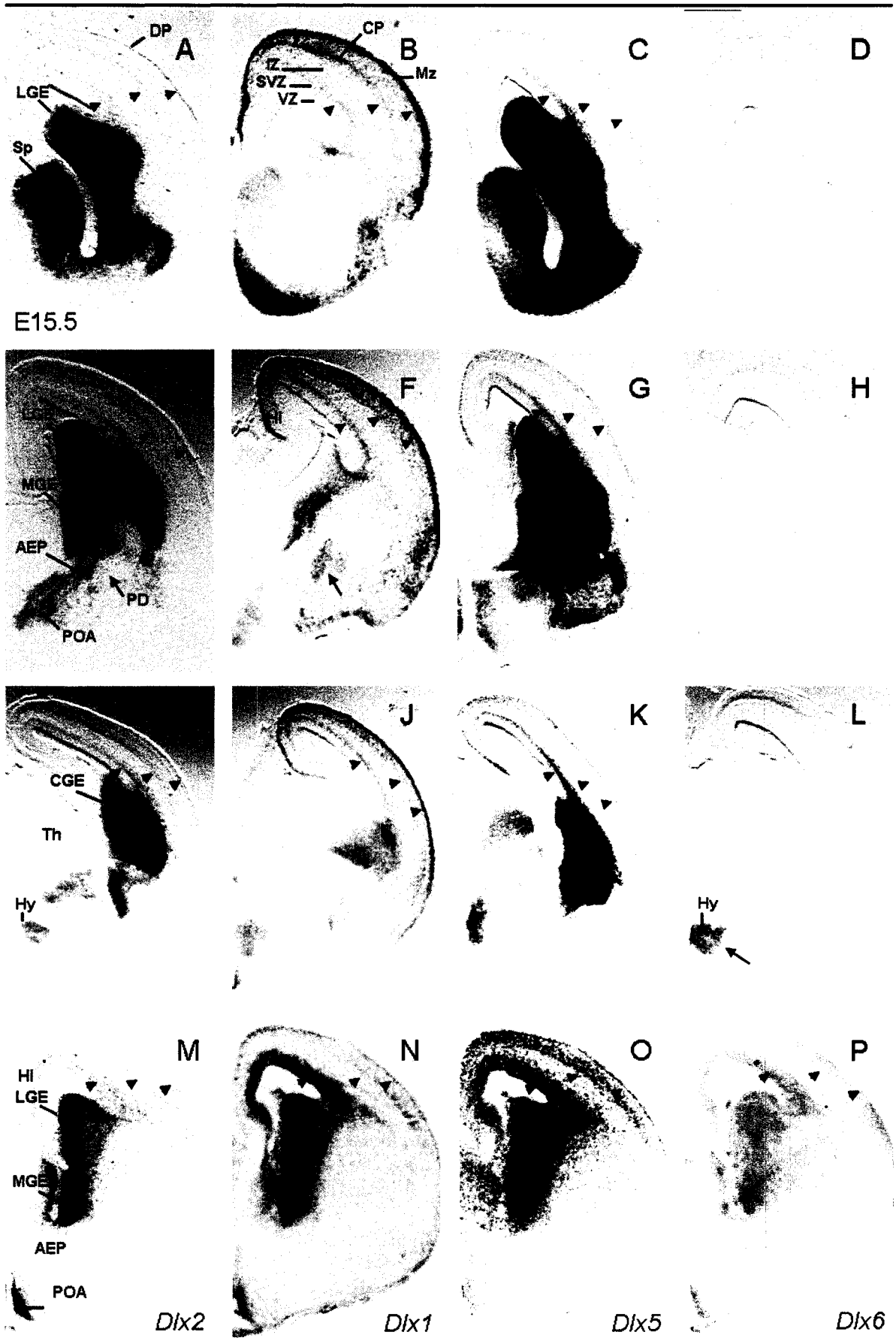

URE2
I12b
I56i
I56ii
A
B
C
D
DP
LGE
Sp
CP
IZ
SVZ
VZ
Mz
E15.5
F
G
H
MGE
AEP
PD
POA
J
K
L
CGE
Th
Hy
M
N
O
P
HI
LGE
MGE
AEP
POA
Dlx2
Dlx1
Dlx5
Dlx6

Figure 4.7. Comparison of the activities of four *Dlx* CREs in the subpallial telencephalon of transgenic mice at birth (P0)

(A-L) Coronal hemisections showing *lacZ* expression when under the control of URE2 (A, E, I), I12b (B, F, J), I56i (C, G, K) and I56ii (D, H, L) enhancers at rostral (A-D), medial (E-H) and caudal (I-L) levels. At this age, *I12b-lacZ* expression is very weak in the septum (Sp) and caudate putamen (CP) (B, F, J), however, *URE2-lacZ* (A, E, I) and *I56i-lacZ* (C, G, K) lines maintain strong *lacZ* expression in these domains. In contrast, *lacZ* staining is stronger in the DP (neocortex) of *I12b-lacZ lines* (B, F, J) compared with *URE2-lacZ* (A, E, I) and *I56i-lacZ* lines (C, G, K). Only a small group of hypothalamic cells is *lacZ*-positive in the *I56ii-lacZ* lines at this age (L; arrow). Legend as in Figure 4.4.

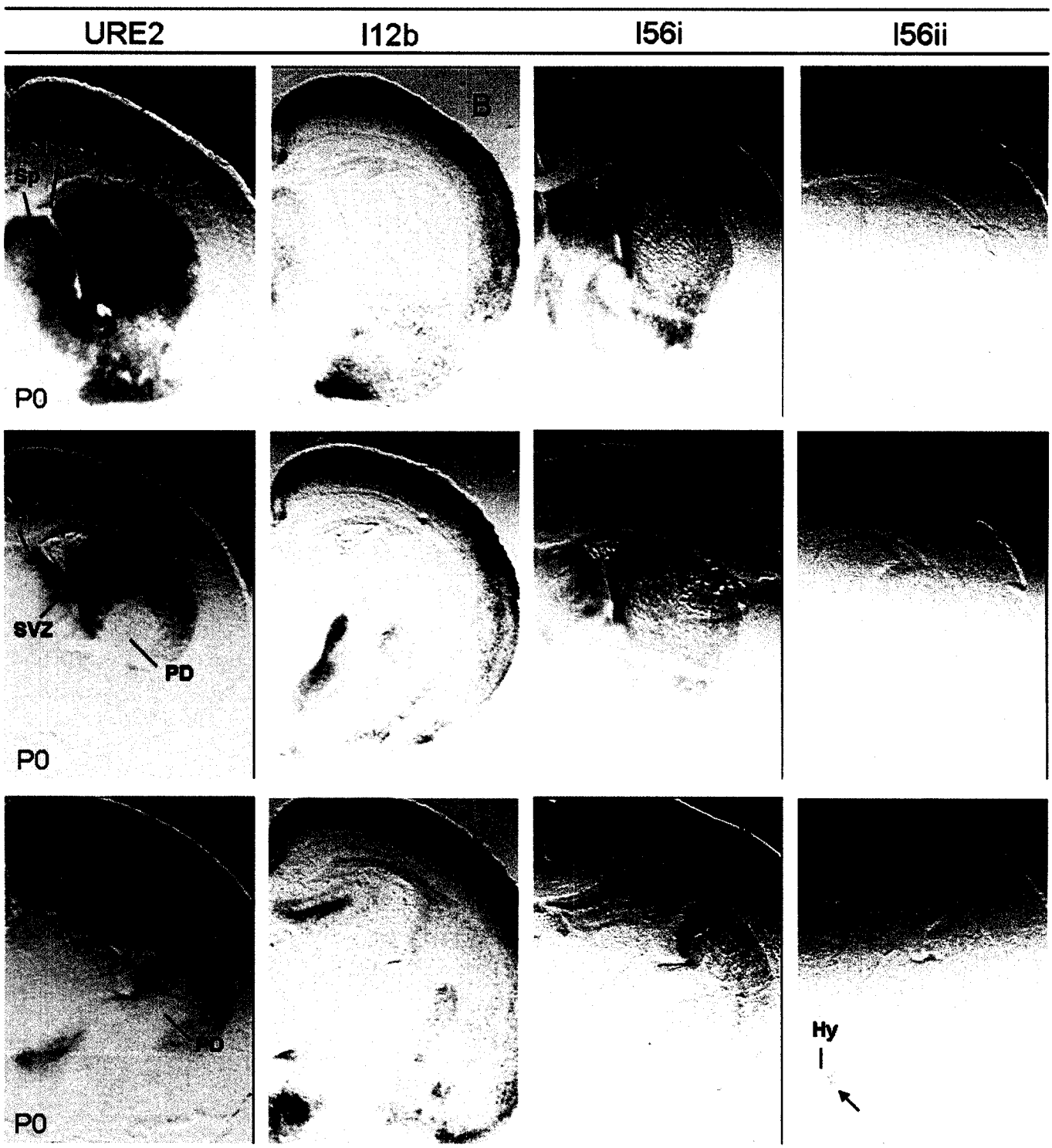
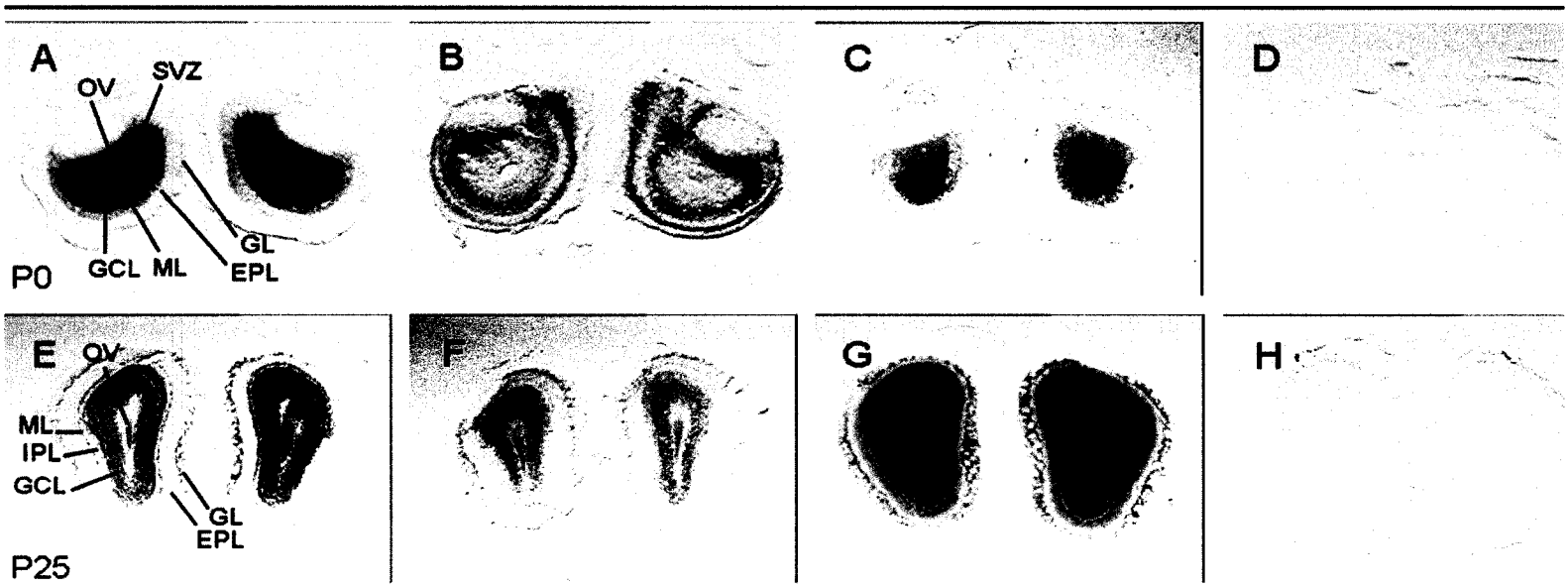

URE2
I12b
I56i
I56ii
Sp
B
P0
SVZ
PD
P0
P0
Hy

Figure 4.8. Comparison of the activities of four *Dlx* CREs in the olfactory bulb of transgenic mice at P35

(A-H) Coronal sections showing *lacZ* expression when under the control of URE2 (A, E), I12b (B, F), I56i (C, G) and I56ii (D, H) enhancers at P0 (A-D) and P25 (E-H). Robust expression from the *URE2-lacZ* (A, E), *I56i-lacZ* (C, G) and *I12b-lacZ* (B, F) transgenes is observed in all the layers of the olfactory bulb containing GABAergic neurons. In contrast, there is no *lacZ* expression in the olfactory bulb in *I56ii-lacZ* lines (D, H). Abbreviations: EPL, external plexiform layer; GCL, granule cell layer; GL, glomerular layer; IPL, internal plexiform layer; ML, mitral layer; OV, olfactory ventricle; SVZ, subventricular zone.


URE2
I12b
I56i
I56ii
A
OV
SVZ
GCL
ML
EPL
GL
P0
B
C
D
E
OV
ML
IPL
GCL
GL
EPL
P25
F
G
H

I56i activity remains strong in the striatum from E13.5-P0 (Figures 4.4., 4.6., 4.7.; panels C and G, Table 4.1. B and data not shown), and persists at P25 as a complex high rostroventral-low caudodorsal patchy pattern (Table 4.1. B and data not shown). Like I12b, I56i expression appears in the pallidum at E15.5 (compare Figures 4.4. G and 4.6. G), and persists at P0 and P25, albeit at lower levels at the later ages (Figure 4.7. G; Table 4.1. C and data not shown).

I56ii activity in the extended pallidum rapidly declines by E15.5 and was not detectable in the telencephalon at later ages (Figures 4.6., 4.7.; panesl D, H and L, Table 4.1. and data not shown). Diencephalic expression remains strong in a few areas (Figures 4.6. L, 4.7. L; arrows, and data not shown).

No major regional differences between the endogenous expression patterns of *Dlx* genes were detected at E13.5 and E15.5 compared with earlier ages.

4.4. Discussion

We have performed here a detailed spatio-temporal analysis of *lacZ* expression under the control of four *Dlx* CREs, URE2, I12b, I56i and I56ii, in the telencephalon of transgenic mice. Comparison of the activities of these enhancers revealed highly overlapping patterns as well as several regional and laminar differences in *lacZ* expression throughout development, thus, reflecting: 1) similarities between the activities of the enhancers and the endogenous *Dlx* gene expression; 2) both similar and distinct role(s) for each enhancer; 3) dynamic and complex *Dlx* gene regulation during embryonic development.

4.4.1. Different activities of the *Dlx* CREs between E10.5 and E12.5 suggest distinct roles for *Dlx* paralogs

We demonstrated that E10 marks the onset of *lacZ* expression in all four CRE lines, which is restricted to a diagonal cluster of cells in the primordia of the ventral

thalamus (Figure 4.1. A-D; arrowheads). This matches the onset of endogenous *Dlx* expression in space but occurs slightly later than E9, when the mRNA of *Dlx1*, which is the first *Dlx* gene to be expressed, is detectable (Price et al., 1991). Although the β-galactosidase protein may be translated at that time, this difference could be due to the fact that *lacZ* staining is only detected after the accumulation of the protein. Alternatively, the CREs described here could be active slightly later in development and other unidentified CRE(s) is/are active earlier. As of E10.5, all lines revealed highly overlapping *lacZ* staining in the subpial layer of the neuroepithelium along the primordia of the LGE and MGE but with spatial differences (Figure 4.1. I-K). First, a few positive cells are scattered in the subpial region of the LGE but not the MGE in *I56ii-lacZ* lines, suggesting that this enhancer may be selectively expressed in a subset of cells at this time point (Figure 4.1. L; arrows). Second, I56i lines display a unique *lacZ* pattern of expression in the VZ of the neuroepithelium where many β-galactosidase-positive cells are arranged in radial columns suggesting a clonal identity (Figure 4.1. K; arrowheads). While such arrangement is rarely observed with the URE2 or I12b lines at this age (Figure 4.1. I-J; arrowheads), we speculate that a *Dlx5/Dlx6* lineage may derive directly from progenitor cell(s) in the VZ and produce the early born striosomal neurons in a *Dlx1/Dlx2* independent manner, as suggested by Anderson et al., 1997a. In contrast, later on, at E12.5, when late-born neuronal progenitors begin to proliferate and differentiate (second wave of neurogenesis), the expression of *Dlx5* and *Dlx6* becomes highly dependent on the expression of *Dlx1* and *Dlx2*. Hence, a radial cell arrangement is described in the VZ at this age but with cells expressing DLX1 and DLX2 proteins. At this time, the DLX5-positive cells are a subset of the DLX1-positive cells, which, in turn, are a subset of the DLX2 positive cells (Bulfone et al., 1993b; Porteus et al., 1994; Eisenstat et al., 1999). Our data supports these observations considering that the number of *lacZ*-positive cells in the VZ of both eminences increases greatly at E11.5 and E12.5 in the URE2 lines compared with E10.5 (compare Figure 4.1. I, and, 4.2. E and 4.3. E), while expression from the *Dlx5/Dlx6* enhancer lines is restricted to the SVZ and MZ of the basal ganglia (Figures 4.2. and 4.3.; G and H). Previous studies demonstrated that

expression of *Dlx5* and *Dlx6* is highly dependent on *Dlx1* and *Dlx2* expression at E12.5. Thus, in the *Dlx1/Dlx2* double mutants, the expression of *Dlx5* and *Dlx6* is completely lost in the SVZ and partially lost in the MZ (Anderson et al., 1997b; Zerucha et al., 2000). We, and others, have shown that this regulation is mediated, at least in part, via the interaction of DLX1 and/or DLX2, either directly or indirectly, with I56i (Zerucha et al., 2000; Zhou et al., 2004).

Taken together, the above results suggest that, depending on the spatio-temporal context, the same regulatory element(s) (e.g. I56i) could be required for the regulation of *Dlx* expression in a differential manner during neurogenesis. *Dlx* paralogs may thus be involved in different function(s) at various developmental stages. Still, it remains unknown how this specific control is established at the molecular level and whether different trans-acting factors are implicated.

4.4.2. Similarities between the laminar activities of the *Dlx* CREs and the endogenous expression of *Dlx* genes at E12.5

We describe that, starting at E11.5, the transgene expression patterns conferred by all four CRE lines, although similar with respect to the A/P and D/V boundaries, differed along the axis of the lateral ventricle. Indeed, the order of expression of *lacZ* with respect to the distance from the lateral ventricle at E11.5 and E12.5 is as follows: *URE2-lacZ* / *I12b-lacZ* / *I56i-lacZ* / *I56ii-lacZ* (Figures 4.2. and 4.3. A-L). Interestingly, the four *Dlx* genes expressed in the telencephalon also follow an endogenous order of expression with respect to the ventriculo-pial axis. This order is derived from a hypothesized genetic and biochemical pathway that proposes the sequential role of *Dlx2, Dlx1, Dlx5* and *Dlx6* at different stages of neuronal differentiation in the basal ganglia and the developing cortex (Figure 4.3. M-P) (reviewed in Panganiban and Rubenstein, 2000). However, direct mechanistic evidence is still lacking at this point in order to specifically link the regulation conferred by one enhancer to one or more *Dlx* genes. Nevertheless, we speculate that, starting at E11.5, late-born *Dlx*-positive cells may use different combination(s) of enhancer(s) depending on their stage of development. Hence, URE2 is

the first enhancer to be active in cells that express *Dlx* gene(s) as suggested by its activity in a subset(s) of proliferating cells (progenitor cells) lining the VZ of the LGE, MGE and CGE (Figures 4.2. and 4.3.; A, E and I). I12b becomes active in *Dlx*-positive cells at a slightly later time point, that is, when these cells reach the second proliferative zone or SVZ of the ganglionic eminences (Figures 4.2. and 4.3.; B, F and J). Indeed, we have shown elsewhere that URE2 and I12b mark identical as well as distinct *Dlx*-subpopulations of cells between E12.5 and E13.5 in the LGE, MGE and CGE (Figures 4.5. L and chapter 3 of this thesis). Yet, it is not clear whether each enhancer regulates *Dlx1* and/or *Dlx2* expression in the spatio-temporal context described above. There is indirect evidence that URE2 may specifically regulate *Dlx1* expression at E12.5. In fact, in the *Dlx1/Dlx2* double mutants where the homeodomain of both genes and the intergenic region were removed (including I12b), *Dlx2* but not *Dlx1* expression was abolished (Anderson et al., 1997b; Zerucha et al., 2000). Therefore, URE2 is not sufficient by itself to maintain *Dlx2* expression at this age but may do so in combination with other enhancer(s) such as I12b. The regulation of *Dlx2* expression may also require the presence of the normal form of *Dlx2* and/or *Dlx1* protein(s) in case auto-regulation is needed. Furthermore, we have shown that I56i and I12b mark primarily identical subpopulation(s) of *Dlx*-expressing cells in the SVZ and the MZ of the basal ganglia (Figure 4.5. I and chapter 3 of this thesis). Most of these I12b+/I56i+ cells have probably initiated their differentiation program. However, I12b may act before I56i in the regulation cascade controlling *Dlx* expression between E11.5 and E13.5. This is evidenced by the presence of cross-regulation among *Dlx* orthologs. In fact, *Dlx1* and *Dlx2* are thought to control directly the expression of *Dlx5* and *Dlx6* both *in vitro* and *in vivo*, through binding to I56i, thus, acting as upstream regulators (Liu et al., 1997; Zhang et al., 1997; Zerucha et al., 2000; Zhou et al., 2004). Moreover, the *Dlx1/Dlx2* double mutant lack *Dlx5* and *Dlx6* expression in the SVZ of telencephalon except in a narrow stripe of early born cells in the deep mantle of the rostral telencepahlon (Anderson et al., 1997b; Zerucha et al., 2000).

4.4.3. I56ii labels distinct group(s) of cells in the MZ at E12.5 and E13.5

Unlike the other CREs that are uniformaly expressed in the mantle of the neuroepithelium between E10 and E11.5, *I56ii-lacZ* expression is restricted to a subset of cells located in the mantle of the vLGE and dMGE at these ages (Figures 4.1. L and 4.2. H). In addition, I56ii labels a group(s) of cells located in the MZ of the vLGE and MGE at E12.5 and E13.5 that do not express *I12b-AP* (Figure 4.5. A-F). We have shown elsewhere that I12b and I56i are active in primarily overlapping sets of cells in the ganglionic eminences at E12.5 and E13.5 (Figure 4.5. G-I and chapter 3 of this thesis). Therefore, it is likely that I56i and I56ii may also mark distinct group(s) of cells and regulate separate functions of *Dlx5/Dlx6* genes. In fact, all DLX proteins are able to bind I56i but not I56ii and activate transcription in reporter assays performed *in vitro* (Zerucha et al., 2000). Furthermore, DLX1 and DLX2 were shown to bind I56i but not I56ii *in vivo* using chromatin immunoprecipitation and nuclear extracts derived from the ganglionic eminences at E13.5 (Zhou et al., 2004). In addition, the two *Dlx* putative binding sites found in I56i and I12b were only partially conserved in I56ii (Ghanem et al., 2003). Like I12b and I56i, URE2 activity is weak and absent in the mantle region where I56ii is active between E12.5 and E13.5, and thus, may not co-label the I56ii+ group(s) of cells (Figures 4.3., 4.4.; E and H). We consider that the MZ of the MGE roughly corresponds to the pallidum, which includes a rostro-caudally extended series of nuclei; the nucleus accumbens (part of it), bed nucleus stria terminalis and globus pallidus. These pallidal nuclei, and components of nuclei, have in common that they autonomously require *Nkx2.1* function (Sussel et al., 1999; Marin et al., 2000). As a result, I56ii may be exclusively required for the differentiation of subset(s) of cells that are born in the primordia of the pallidum and belonging to one or more of the above nuclei. Moreover, the regulation conferred by I56ii is time-dependent since this enhancer is not active in the telencephalon after E14.5 and is only active in a group of hypothalamic cells (Figure 4.6., 4.7.; H and L, Table 4.1. C and data not shown). The role played by I56ii+ cells is under investigation (Garel S, Ghanem N and Ekker M, unpublished observations).

4.4.4. Dynamic regulation of *Dlx* expression in space and time between E13.5 and P25

The laminar differences in *lacZ* activity detected between E13.5 and P25 among the CRE lines in the subpallium are underlined by a change in the order of enhancer activity with respect to the lateral ventricle from *URE2-lacZ / I12b-lacZ / I56i-lacZ / I56ii-lacZ* at E12.5 to *URE2-lacZ / I56i-lacZ / I12b-lacZ / I56ii-lacZ* starting at E13.5. This is due to the fact that I12b but not I56i activity regresses considerably in the proliferative zones of the ganglionic eminences after E13.5 (Figures 4.4. F, G; 4.5. G, H; 4.6. F, G, J and K; Table 4.1. and data not shown). Morevoer, while tangential migrations are well underway, *lacZ* expression from the I12b and I56i transgenes is widely distributed in the MGE at E12.5 (Figure 4.3. F and G), whereas from E13.5 until E15.5, MGE expression, especially from the *I12b-lacZ* is reduced, being largely restricted to the SVZ (Figures 4.4. F, G and 4.6. F, G and data not shown). The reduction in MGE expression in *I12b-lacZ* lines suggests that many of the transgene-expressing neurons that migrate tangentially have already been produced and migrated away from this region.

Several regional differences were also highlighted between the activities of the CREs after E15.5 such as the absence of URE2 but not I56i nor I12b activity in the pallidum (Figures 4.6. E-G, arrows; 4.7. E-G and Table 4.1. C) and the weak (residual) activity of I12b but not of I56i in the striatum (Figures 4.6. F, G; 4.7. F, G and Table 4.1. B).

Since no change in the laminar or regional distributions of endogenous *Dlx* expression is observed between E13.5 and E15.5 compared to earlier time points (Figures 4.3., 4.4. and 4.6.; M-P), our results suggest that: 1) *Dlx* regulation, as controlled by the four *Dlx* CREs undergoes dynamic spatio-temporal changes during development, yet, it is still unclear if each CRE regulates specifically the expression of one or more *Dlx* genes; 2) the changes in enhancer activity are likely to affect distinct subpopulation(s) of *Dlx*-positive cells that may express one or more *Dlx* gene and/or *Dlx* enhancer; 3) the *Dlx* CREs may possess distinct(s) regulatory roles at different time points. For instance, our data suggest that I56i plays a differential role in *Dlx5/Dlx6* regulation at E10.5 and E12.5

as discussed earlier. It will therefore be interesting to determine whether the shift in the activities of I12b and I56i observed after E13.5 is correlated with functional modification(s) associated with the DLX proteins such as a possible regulation of *Dlx1/Dlx2* by *Dlx5/Dlx6* in the basal ganglia.

4.4.5. Is/Are there additional/unknown *Dlx* enhancer(s) active in the forebrain?

We have identified and characterized four *Dlx* forebrain-specific CREs regulating the expression of four *Dlx* genes. However, in the *Dlx5/Dlx6* null mutant mice where *lacZ* was inserted in the *Dlx6* locus, and, the homeodomains of both genes plus the intergenic region were deleted, there is residual β-galactosidase staining in the forebrain (Dye C and Rubenstein J, unpublished observations). This suggests that additional *Dlx* forebrain-specific CRE(s) may exist elsewhere in the *Dlx5/Dlx6* locus. In fact, preliminary results showed that at least one such element may be located in the 5' flanking region of *Dlx6* (Jarinova O, Ekker M, unpublished observations). This finding is interesting considering that URE2 is located in the 5' flanking region of *Dlx1* and that both genes belong to the same *Dlx* clade: *Dlx1/Dlx4/Dlx6* (Stock et al., 1996). It is probable that this region in both genes may have derived from the duplication event that took place before the divergence of modern vertebrates and which gave rise to the *Dlx* bigene clusters. Thus, it is interesting to determine if any regulatory element(s) was/were also duplicated and whether they have diverged in sequence and/or function.

Acknowledgments

We thank Adrianna Gambarotta for generating the transgenic mice and Elaine Coderre for technical assistance with sectioning. N.G. was supported in part by a scholarship from the Lebanese University, Beyrouth. This work is supported by grants from the Canadian Institutes of Health Research (MOP14460). M.E. was supported by an Investigator award of the CIHR during part of this work.

5. Conclusions

The *Dlx* genes are homeobox transcription factors required for normal embryonic development during mid to late gestation. In the forebrain, *Dlx* expression and function are primarily associated with the development of virtually all neurons that use GABA as their neurotransmitter. Inhibitory GABAergic interneurons are indispensable for proper neocortical function since they are the regulators of principal neuron activity (excitatory pyramidal cells). In the present research project, I have studied the regulatory elements that govern expression of the four *Dlx* genes, *Dlx1, Dlx2, Dlx5* and *Dlx6*, in the subpallial telencephalon of vertebrates. As a result, I have contributed to our understanding of the role played by four *Dlx* regulatory elements in the forebrain, URE2, I12b, I56i and I56ii, two of which are identified and characterized for the first time: URE2 and I12b. Thus, I have shown that *Dlx* expression is under the control of a complex and dynamic regulation that changes in a spatial and temporal manner during embryogenesis. More importantly, I have demonstrated that this regulation is tightly linked to *Dlx* function in regional specification of distinct progenitors born in the ventral cortex. These progenitors will subsequently give rise to most neurons of GABAergic phenotype including nearly all interneurons in the telencephalon.

Although some aspects of *Dlx* regulation and function were addressed in this study, several other aspects remain to be uncovered. In the light of the results shown here, I will highlight some of the interesting questions that remain unanswered and may be the goal of future studies.

Additional conserved elements in the *Dlx* bigene clusters

As mentioned above, I and others have identified thus far four *Dlx*-forebrain-specific enhancers in the *Dlx1/Dlx2* and *Dlx5/Dlx6* loci of vertebrates. However, these enhancers may not be the only elements to control *Dlx* expression in the forebrain. There is evidence that additional conserved element(s), at least in the case of *Dlx5/Dlx6*, may

exist (personal communication for C. Dye and J. Rubenstein). One such element is found in the 5' flanking region of *Dlx6* (Jarinova O, Ekker M. unpublished observations). Furthermore, element(s) with less conserved motifs and/or shorter sequence (<100bp) could have escaped the resolution of our phylogenetic screening. In fact, we have performed in the first study a pairwise sequence alignment using relatively simple programs such as PIPMAKER or Best fit and Mapplot programs. Recent advances in bioinformatics led to the development of more sophisticated algorithms and sequence comparison tools. Hence, more powerful programs are now available that provide a better resolution to evaluate the sensitivity and specificity of the constraint based-alignment such as phastCons and multi-species conserved sequences -MCSs- scores and the pattern-matching regulatory potential (RP) score (Elnitski et al., 2003; Margulies et al., 2003; Siepel et al., 2005). The increased capacity of these programs to detect *cis*-regulatory modules was compared and evaluated in a recent study conducted by King et al. 2005 (King et al., 2005). In addition, in order to improve the search resolution, other groups have successfully combined '*phylogenetic footprinting*' with other detection methods such as *cis*-regulatory module (CRM) detection (Krivan and Wasserman, 2001) and *de novo* motif discovery (Blanchette and Tompa, 2002), resulting in a significant increase in the specificity of predictions. Therefore, using these tools, it will be easier to refine the search for *Dlx* regulatory element(s) in the future and, check whether these elements have diverged in sequence and/or function.

Conserved function(s) of the *Dlx* CREs in the forebrain of vertebrates

I have shown, in collaboration with Olga Jarinova, that zI56i and zI56ii (zebrafish orthologs), exhibit conserved function(s) and consequently, are active in the forebrain of zebrafish including the telencephalon. This finding is intriguing since mammals and fish species have diverged over 450 millions years ago. However, we did not check if conserved activity of the enhancers is also true for I12b and URE2. Besides that, it will be interesting to study further the role played by these enhancers in zebrafish. It is not known if they are needed to achieve similar and/or other telencephalic functions in fish or

other vertebrate species such as the development of cortical interneurons. The high degree of sequence conservation of these enhancers between fish and mammals suggest that they may do so. However, one caveat is that brain organization and neuronal migration still need to be studied in zebrafish. Alternatively, the *Dlx* enhancers may serve also different functions. It has been shown that equivalent regulatory regions from the same gene in different species may have evolved different components to potentiate their activity in combination with a selection of core components. Indeed, conservation as well as diversity in sequence and/or function between distantly related species was described in *cis*-regulatory networks of the *Hoxa2* gene (Tumpel et al., 2002) and the *HoxD* cluster (Spitz et al., 2003). This may explain also how the *Dlx* paralogous and/or orthologous enhancers could have diverged in sequences and functions, yet, retained highly overlapping activities. From an evolutionary perspective, studying the role of these enhancers in distantly related species such as zebrafish will hopefully reveal how *Dlx* function in the telencephalon has evolved since the divergence of modern vertebrates.

Correlation between the activity of the *Dlx* CREs and *Dlx* expression

I have shown that the four *Dlx* enhancers follow generally a defined order of activity in the proliferative zones of the subpallium with respect to the lateral ventricle during development: URE2/ I12b or I56i/ I56ii. Thus, URE2 is active in scattered cells in the VZ (undifferentiated cells), a substantial number of cells in the SVZ and in scattered cells in the MZ (differentiated cells). I12b and I56i are primarily active in differentiated cells in the SVZ and MZ whereas I56ii activity is restricted to a subset of cells in the deep MZ. Interestingly, *Dlx* genes follow also a specific order of expression *in vivo* with respect to the lateral ventricle as shown here and elsewhere: *Dlx2/ Dlx1/ Dlx5/ Dlx6*. Like URE2, *Dlx2* is mainly expressed in undifferentiated cells; its expression is found in scattered cells in the VZ, in most cells in the SVZ and few cells in the MZ. Like I12b and I56i, *Dlx1* and *Dlx5* are mainly expressed in the SVZ whereas *Dlx6* is primarily expressed in differentiated cells in the MZ. However, it is not clear whether a correlation exists between the combination of *Dlx* enhancer(s) that is/are active in one cell type and

Dlx genes expressed by these cells. It is likely that these enhancers are *cis*-acting regulatory elements, and, for instance, URE2 or I12b could be active in group(s) of cells where both *Dlx1* and *Dlx2* are expressed. The same applies to I56i and I56ii with respect to *Dlx5* and *Dlx6*. Still, this remains to be proven at the molecular level. Alternatively, it is important to check whether each enhancer could regulate the expression of one flanking *Dlx* paralog but not of the other, depending on the context. There is evidence that, for instance, URE2 and *Dlx1* label differentially two out of three major subtypes of adult cortical interneurons. Cobos et al. 2005 showed that, in one month old mice, *Dlx1* labels the majority of CR- and SOM- but not PV-expressing interneurons. In contrast, I found that URE2 labels CR- and PV- but not SOM-expressing interneurons whereas I12b labels all three subtypes. Thus, URE2/*Dlx1* and I12b/*Dlx1* may define the identity of distinct subtype(s) of cortical interneurons. It remains to be determined if this is true also with *Dlx2*, I12b and URE2 as well as the *Dlx* enhancers and genes from the *Dlx5/Dlx6* cluster. Future experiments such as co-labeling of the various combinations of *Dlx* genes and enhancers in various *Dlx* lineages, in both the embryonic and adult brains, will be required to understand the development of these lineages.

Spatio-temporal regulation of *Dlx* expression and regionalization in the subpallium

I have highlighted major regional and laminar differences in the activity of the four *Dlx* CREs at all ages examined (E10.5-P25). In fact, labeling of distinct groups or subgroups of *Dlx*-progenitor cells by each enhancer in the ventral cortex was the most striking difference. Differential enhancer activity was also detected in tangentially migrating cells that are derived from these progenitors. Moreover, I showed that this differential control occurs in a time- and space-dependent manner, and led to the identification of new subdivisions within the MGE and CGE (dMGE, vMGE, dCGE, mCGE and vCGE) where distinct *Dlx*-progenitors seem to be born. My data suggests that *Dlx* expression marks several subpopulations of progenitors that may give rise to distinct adult cortical interneurons. One or more *Dlx* enhancer(s) and gene(s) can be active in

each subpopulation. Identification of regional markers such as transcription factors that are specifically expressed in each subdivision of the MGE and CGE will enforce directly the model proposed in this study. Altenatively, the enhancer activity could be unrelated to the generation of distinct lineage(s) of adult interneurons and merely changes in a spatio-temporal manner depending on the context. Thus, the existence of several *Dlx*-progenitors and the involvement of each enhancer in distinct aspect(s) of *Dlx* regulation can be directly demonstrated by the deletion of one and/or more enhancer(s) in mice ('Knock-out' experiments). Such experiments are ongoing in the laboratory and may uncover the consequence(s) of loss of one enhancer or combination of enhancers on *Dlx* expression and function as well as the development of various groups of GABAergic neurons.

Uncovering of the *Dlx* regulatory mechanisms

Despite high overlapping activities between three out of four enhancers, these elements displayed little sequence similarities. These are restricted to two putative *Dlx* binding sites (A/C/G)TAATT(G/A)(C/G) and several potential homeodomain binding motifs (TAAT/ATTA) in each sequence. We have hypothesized that regulatory elements may have been included in the duplication events that gave rise to multiple *Dlx* bigene cluster. None of the enhancers described in this study seem to be found in invertebrates such as *Drosophila* and *C. elegans* suggesting that, either they were not duplicated from ancestral element(s), or, more likely, they were duplicated but have rapidly diverged in sequence early in chordate/vertebrate evolution. Furthermore, this divergence was probably followed by fixation of *cis*-acting regulatory elements as suggested by high degree of sequence conservation between orthologous enhancers (77-99%). Still, it is not clear whether the same and/or distinct regulatory mechanisms mediate the role of paralogous as well as orthologous enhancers. Furthermore, the *trans*-acting factors that are the key players in these mechanisms are still unknown. Ongoing studies in the laboratory are aimed at identifying the binding motifs in each enhancer and thereafter, the upstream regulators of *Dlx* paralogs. DNA-footprinting experiments coupled to

electrophoretic mobility shift assays (EMSA) followed by mutagenesis experiments in mice proved to be successful in achieving some of the above goals (L. Poitras, N. Shipley, M.Ekker, unpublished observations). As a result, several potential binding sites were identified in the sequence of I12b including a binding motif for the pro-neural determinant *Mash*-1 [(Poitras, 2006), submitted]. This is considered the first molecular evidence that this gene is regulating directly the expression of *Dlx1/Dlx2* through binding to I12b, and lends support to the existence of a *Mash-1/ Dlx/ GAD* cascade. As progress is made in such experiments, our understanding of the molecular mechanisms and upstream targets of *Dlx* genes will increase.

Finally, it will be interesting to apply the findings made in this study as well as future studies to learn more about *Dlx* function with regard to the development of GABAergic neurons in primates and, ultimately, humans. In fact, it was shown a few years ago that about half of cortical interneurons in the embryonic human brain are derived from progenitors located in the subpallial telencephalon (Letinic et al., 2002). Therefore, it would be useful to investigate whether DLX proteins achieve similar function(s) in the human brain considering that all four enhancers studied here are highly conserved in our species. A better understanding of the development of GABAergic neurons will advance future research that is aimed at finding therapeutic cures for brain dysfunction associated with absence/reduction of inhibition such as in cases of epilepsy, schizophrenia and, maybe, autism.

6. References

Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., and Brulet, P. (1995). Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. Development *121*, 3279-3290.

Acampora, D., Merlo, G. R., Paleari, L., Zerega, B., Postiglione, M. P., Mantero, S., Bober, E., Barbieri, O., Simeone, A., and Levi, G. (1999). Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5. Development *126*, 3795-3809.

Ahituv, N., Prabhakar, S., Poulin, F., Rubin, E. M., and Couronne, O. (2005). Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. Hum Mol Genet *14*, 3057-3063.

Akimenko, M. A., Ekker, M., Wegner, J., Lin, W., and Westerfield, M. (1994). Combinatorial expression of three zebrafish genes related to distal-less: part of a homeobox gene code for the head. J Neurosci *14*, 3475-3486.

Altman, B. S. a. (1991). 'Neocortical Development' (New York).

Altmann, C. R., and Brivanlou, A. H. (2001). Neural patterning in the vertebrate embryo. Int Rev Cytol *203*, 447-482.

Amemiya, C. T., Amores, A., Ota, T., Mueller, G., Garrity, D., Postlethwait, J. H., and Litman, G. W. (2001). Generation of a P1 artificial chromosome library of the Soutehrn pufferfish. Gene *272*, 283-289.

Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y.-L., *et al.* (1998). Zebrafish *hox* clusters and vertebrate genome evolution. Science *282*, 1711-1714.

Anderson, S. A., Eisenstat, D. D., Shi, L., and Rubenstein, J. L. (1997a). Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. Science *278*, 474-476.

Anderson, S. A., Kaznowski, C. E., Horn, C., Rubenstein, J. L., and McConnell, S. K. (2002). Distinct origins of neocortical projection neurons and interneurons in vivo. Cereb Cortex *12*, 702-709.

Anderson, S. A., Marin, O., Horn, C., Jennings, K., and Rubenstein, J. L. (2001). Distinct cortical migrations from the medial and lateral ganglionic eminences. Development *128*, 353-363.

Anderson, S. A., Qiu, M., Bulfone, A., Eisenstat, D. D., Meneses, J., Pedersen, R., and Rubenstein, J. L. (1997b). Mutations of the homeobox genes Dlx-1 and Dlx-2 disrupt the striatal subventricular zone and differentiation of late born striatal neurons. Neuron *19*, 27-37.

Ang, E. S., Jr., Haydar, T. F., Gluncic, V., and Rakic, P. (2003). Four-dimensional migratory coordinates of GABAergic interneurons in the developing mouse cortex. J Neurosci *23*, 5805-5815.

Ang, S. L., Jin, O., Rhinn, M., Daigle, N., Stevenson, L., and Rossant, J. (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. Development *122*, 243-252.

Anthony, T. E., Klein, C., Fishell, G., and Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. Neuron *41*, 881-890.

Arnone, M. I., and Davidson, E. H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851-1864.

Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G., and Pfeifer, D. (2001). Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. Genomics *78*, 73-82.

Barton, L. M., Gottgens, B., Gering, M., Gilbert, J. G., Grafham, D., Rogers, J., Bentley, D., Patient, R., and Green, A. R. (2001). Regulation of the stem cell leukemia (SCL) gene: a tale of two fishes. Proc Natl Acad Sci U S A *98*, 6747-6752.

Basler, K., Edlund, T., Jessell, T. M., and Yamada, T. (1993). Control of cell pattern in the neural tube: regulation of cell differentiation by dorsalin-1, a novel TGF beta family member. Cell *73*, 687-702.

Beanan, M. J., and Sargent, T. D. (2000). Regulation and function of Dlx3 in vertebrate development. Dev Dyn *218*, 545-553.

Bei, M., and Maas, R. (1998). FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. Development *125*, 4325-4333.

Ben-Ari, Y. (2002). Excitatory actions of gaba during development: the nature of the nurture. Nat Rev Neurosci *3*, 728-739.

Benardo, L. S. (1994). Separate activation of fast and slow inhibitory postsynaptic potentials in rat neocortex in vitro. J Physiol *476*, 203-215.

Bendall, A. J., and Abate-Shen, C. (2000). Roles for Msx and Dlx homeoproteins in vertebrate development. Gene *247*, 17-31.

Benson, M. D., Bargeon, J. L., Xiao, G., Thomas, P. E., Kim, A., Cui, Y., and Franceschi, R. T. (2000). Identification of a homeodomain binding element in the bone sialoprotein gene promoter that is required for its osteoblast-selective expression. J Biol Chem *275*, 13907-13917.

Bielle, F., Griveau, A., Narboux-Neme, N., Vigneau, S., Sigrist, M., Arber, S., Wassef, M., and Pierani, A. (2005). Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. Nat Neurosci *8*, 1002-1012.

Bishop, K. M., Goudreau, G., and O'Leary, D. D. (2000). Regulation of area identity in the mammalian neocortex by Emx2 and Pax6. Science *288*, 344-349.

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

Blumberg, B., Bolado, J., Jr., Moreno, T. A., Kintner, C., Evans, R. M., and Papalopulu, N. (1997). An essential role for retinoid signaling in anteroposterior neural patterning. Development *124*, 373-379.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L., and Rubin, E. M. (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science *299*, 1391-1394.

Borg-Graham, L. J., Monier, C., and Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. Nature *393*, 369-373.

Bouwmeester, T., Kim, S., Sasai, Y., Lu, B., and De Robertis, E. M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. Nature *382*, 595-601.

Brasset, E., and Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. Heredity *94*, 571-576.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., and Aparicio, S. (1993). Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. Nature *366*, 265-268.

Bulfone, A., Kim, H. J., Puelles, L., Porteus, M. H., Grippo, J. F., and Rubenstein, J. L. (1993a). The mouse Dlx-2 (Tes-1) gene is expressed in spatially restricted domains of the forebrain, face and limbs in midgestation mouse embryos. Mech Dev *40*, 129-140.

Bulfone, A., Puelles, L., Porteus, M. H., Frohman, M. A., Martin, G. R., and Rubenstein, J. L. (1993b). Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. J Neurosci *13*, 3155-3172.

Bulfone, A., Wang, F., Hevner, R., Anderson, S., Cutforth, T., Chen, S., Meneses, J., Pedersen, R., Axel, R., and Rubenstein, J. L. (1998). An olfactory sensory map develops in the absence of normal projection neurons or GABAergic interneurons. Neuron *21*, 1273-1282.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A *99 Suppl 4*, 16433-16437.

Butt, S. J., Fuccillo, M., Nery, S., Noctor, S., Kriegstein, A., Corbin, J. G., and Fishell, G. (2005). The temporal and spatial origins of cortical interneurons predict their physiological subtype. Neuron *48*, 591-604.

Cajal, R. (1911). Histology of the nervous system, vol. 2).

Campbell, K. (2003). Dorsal-ventral patterning in the mammalian telencephalon. Curr Opin Neurobiol *13*, 50-56.

Caracciolo, A., Di Gregorio, A., Aniello, F., Di Lauro, R., and Branno, M. (2000). Identification and developmental expression of three Distal-less homeobox containing genes in the ascidian Ciona intestinalis. Mech Dev *99*, 173-176.

Casarosa, S., Fode, C., and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. Development *126*, 525-534.

Cauli, B., Audinat, E., Lambolez, B., Angulo, M. C., Ropert, N., Tsuzuki, K., Hestrin, S., and Rossier, J. (1997). Molecular and physiological diversity of cortical nonpyramidal cells. J Neurosci *17*, 3894-3906.

Chase, M. B., Fu, S., Haga, S. B., Davenport, G., Stevenson, H., Do, K., Morgan, D., Mah, A. L., and Berg, P. E. (2002). BP1, a homeodomain-containing isoform of DLX4, represses the beta-globin gene. Mol Cell Biol *22*, 2505-2514.

Chen, C. Y., and Schwartz, R. J. (1995). Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. J Biol Chem *270*, 15628-15633.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature *383*, 407-413.

Cobos I, Long JE, Thwin MT, and JL., R. (2006). Cellular patterns of transcription factor expression in developing cortical interneurons. Cereb Cortex *suppl 1*, i82-88.

Cobos, I., Calcagnotto, M. E., Vilaythong, A. J., Thwin, M. T., Noebels, J. L., Baraban, S. C., and Rubenstein, J. L. (2005). Mice lacking Dlx1 show subtype-specific loss of interneurons, reduced inhibition and epilepsy. Nat Neurosci *8*, 1059-1068.

Cobos, I., Puelles, L., and Martinez, S. (2001). The avian telencephalic subpallium originates inhibitory neurons that invade tangentially the pallium (dorsal ventricular ridge and cortical areas). Dev Biol *239*, 30-45.

Cobos, I., Shimamura, K., Rubenstein, J. L., Martinez, S., and Puelles, L. (2001). Fate map of the avian anterior forebrain at the four-somite stage, based on the analysis of quail-chick chimeras. Dev Biol *239*, 46-67.

Cohen, S. M., Bronner, G., Kuttner, F., Jurgens, G., and Jackle, H. (1989). Distal-less encodes a homoeodomain protein required for limb development in Drosophila. Nature *338*, 432-434.

Cohen, S. M., and Jurgens, G. (1989). Proximal-distal pattern formation in Drosophila: cell autonomous requirement for Distal-less gene activity in limb development. Embo J *8*, 2045-2055.

Connors, B. W., Malenka, R. C., and Silva, L. R. (1988). Two inhibitory postsynaptic potentials, and GABAA and GABAB receptor-mediated responses in neocortex of rat and cat. J Physiol *406*, 443-468.

Corbin, J. G., Gaiano, N., Machold, R. P., Langston, A., and Fishell, G. (2000). The Gsh2 homeodomain gene controls multiple aspects of telencephalic development. Development *127*, 5007-5020.

Corbin, J. G., Nery, S., and Fishell, G. (2001). Telencephalic cells take a tangent: non-radial migration in the mammalian forebrain. Nat Neurosci *4 Suppl*, 1177-1182.

Couly, G. F., and Le Douarin, N. M. (1987). Mapping of the early neural primordium in quail-chick chimeras. II. The prosencephalic neural plate and neural folds: implications for the genesis of cephalic human congenital abnormalities. Dev Biol *120*, 198-214.

Cox, W. G., and Hemmati-Brivanlou, A. (1995). Caudalization of neural fate by tissue recombination and bFGF. Development *121*, 4349-4358.

Crackower, M. A., Scherer, S. W., Rommens, J. M., Hui, C. C., Poorkaj, P., Soder, S., Cobben, J. M., Hudgins, L., Evans, J. P., and Tsui, L. C. (1996). Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. Hum Mol Genet *5*, 571-579.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292-301.

Davis, A. A., and Temple, S. (1994). A self-renewing multipotential stem cell in embryonic rat cerebral cortex. Nature *372*, 263-266.

de Carlos, J. A., Lopez-Mascaraque, L., and Valverde, F. (1996). Dynamics of cell migration from the lateral ganglionic eminence in the rat. J Neurosci *16*, 6146-6156.

de Melo, J., Du, G., Fonseca, M., Gillespie, L. A., Turk, W. J., Rubenstein, J. L., and Eisenstat, D. D. (2005). Dlx1 and Dlx2 function is necessary for terminal differentiation and survival of late-born retinal ganglion cells in the developing mouse retina. Development *132*, 311-322.

de Melo, J., Qiu, X., Du, G., Cristante, L., and Eisenstat, D. D. (2003). Dlx1, Dlx2, Pax6, Brn3b, and Chx10 homeobox gene expression defines the retinal ganglion and inner nuclear layers of the developing and adult mouse retina. J Comp Neurol *461*, 187-204.

DeFelipe, J. (1999). Chandelier cells and epilepsy. Brain *122 (Pt 10)*, 1807-1822.

DeFelipe, J., and Farinas, I. (1992). The pyramidal neuron of the cerebral cortex: morphological and chemical characteristics of the synaptic inputs. Prog Neurobiol *39*, 563-607.

Dellovade, T. L., Pfaff, D. W., and Schwanzel-Fukuda, M. (1998). Olfactory bulb development is altered in small-eye (Sey) mice. J Comp Neurol *402*, 402-418.

Depew, M. J. (2002b). In Mouse Development.).

Depew, M. J., Liu, J. K., Long, J. E., Presley, R., Meneses, J. J., Pedersen, R. A., and Rubenstein, J. L. (1999). Dlx5 regulates regional development of the branchial arches and sensory capsules. Development *126*, 3831-3846.

Depew, M. J., Lufkin, T., and Rubenstein, J. L. R. (2002a). Specification of jaw subdivisions by *Dlx* genes. Science *298*, 381-385.

Di Gregorio, A., Spagnuolo, A., Ristoratore, F., Pischetola, M., Aniello, F., Branno, M., Cariello, L., and Di Lauro, R. (1995). Cloning of ascidian homeobox genes provides evidence for a primordial chordate cluster. Gene *156*, 253-257.

Dickinson, M. E., Selleck, M. A., McMahon, A. P., and Bronner-Fraser, M. (1995). Dorsalization of the neural tube by the non-neural ectoderm. Development *121*, 2099-2106.

Dickmeis, T., and Muller, F. (2005). The identification and functional characterisation of conserved regulatory elements in developmental genes. Brief Funct Genomic Proteomic *3*, 332-350.

Dodig, M., Kronenberg, M. S., Bedalov, A., Kream, B. E., Gronowicz, G., Clark, S. H., Mack, K., Liu, Y. H., Maxon, R., Pan, Z. Z., *et al.* (1996). Identification of a TAAT-containing motif required for high level expression of the COL1A1 promoter in differentiated osteoblasts of transgenic mice. J Biol Chem *271*, 16422-16429.

Dodou, E., Xu, S. M., and Black, B. L. (2003). mef2c is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. Mech Dev *120*, 1021-1032.

Dolle, P., Price, M., and Duboule, D. (1992). Expression of the murine Dlx-1 homeobox gene during facial, ocular and limb development. Differentiation *49*, 93-99.

Dong, P. D., Chu, J., and Panganiban, G. (2000). Coexpression of the homeobox genes Distal-less and homothorax determines Drosophila antennal identity. Development *127*, 209-216.

Doniach, T. (1993). Planar and vertical induction of anteroposterior pattern during the development of the amphibian central nervous system. J Neurobiol *24*, 1256-1275.

Dou, C. L., Li, S., and Lai, E. (1999). Dual role of brain factor-1 in regulating growth and patterning of the cerebral hemispheres. Cereb Cortex *9*, 543-550.

Eagleson, G. W., and Harris, W. A. (1990). Mapping of the presumptive brain regions in the neural plate of Xenopus laevis. J Neurobiol *21*, 427-440.

Eddy, S. R. (2002). Computational genomics of noncoding RNA genes. Cell *109*, 137-140.

Edmondson, J. C., and Hatten, M. E. (1987). Glial-guided granule neuron migration in vitro: a high-resolution time-lapse video microscopic study. J Neurosci *7*, 1928-1934.

Eisenstat, D. D., Liu, J. K., Mione, M., Zhong, W., Yu, G., Anderson, S. A., Ghattas, I., Puelles, L., and Rubenstein, J. L. (1999). DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation. J Comp Neurol *414*, 217-237.

Ekker, M., Akimenko, M. A., Bremiller, R., and Westerfield, M. (1992). Regional expression of three homeobox transcripts in the inner ear of zebrafish embryos. Neuron *9*, 27-35.

Elgar, G. (2004). Identification and analysis of cis-regulatory elements in development using comparative genomics with the pufferfish, Fugu rubripes. Semin Cell Dev Biol *15*, 715-719.

Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B., and Brenner, S. (1996). Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). Trends Genet *12*, 145-150.

Ellies, D. L., Langille, R. M., Martin, C. C., Akimenko, M. A., and Ekker, M. (1997a). Specific craniofacial cartilage dysmorphogenesis coincides with a loss of dlx gene expression in retinoic acid-treated zebrafish embryos. Mech Dev *61*, 23-36.

Ellies, D. L., Stock, D. W., Hatch, G., Giroux, G., Weiss, K. M., and Ekker, M. (1997b). Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish dlx genes. Genomics *45*, 580-590.

Elnitski, L., Hardison, R. C., Li, J., Yang, S., Kolbe, D., Eswara, P., O'Connor, M. J., Schwartz, S., Miller, W., and Chiaromonte, F. (2003). Distinguishing regulatory DNA from neutral sites. Genome Res *13*, 64-72.

Fairen, A. (1984). Nonpyramidal neurons: general account. In: Cerebral Cortex, vol. 1, Cellular components of the cerebral cortex).

Fairen, A., Cobas, A., and Fonseca, M. (1986). Times of generation of glutamic acid decarboxylase immunoreactive neurons in mouse somatosensory cortex. J Comp Neurol *251*, 67-83.

Fan, G., Martinowich, K., Chin, M. H., He, F., Fouse, S. D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., *et al.* (2005). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. Development *132*, 3345-3356.

Feldman, M. (1984). Morphology of the neocortical pyramidal neuron. In: Cerebral Cortex).

Feledy, J. A., Beanan, M. J., Sandoval, J. J., Goodrich, J. S., Lim, J. H., Matsuo-Takasaki, M., Sato, S. M., and Sargent, T. D. (1999a). Inhibitory patterning of the anterior neural plate in Xenopus by homeodomain factors Dlx3 and Msx1. Dev Biol *212*, 455-464.

Feledy, J. A., Morasso, M. I., Jang, S. I., and Sargent, T. D. (1999b). Transcriptional activation by the homeodomain protein distal-less 3. Nucleic Acids Res *27*, 764-770.

Fernandez, A. S., Pieau, C., Reperant, J., Boncinelli, E., and Wassef, M. (1998). Expression of the Emx-1 and Dlx-1 homeobox genes define three molecularly distinct domains in the telencephalon of mouse, chick, turtle and frog embryos: implications for the evolution of telencephalic subdivisions in amniotes. Development *125*, 2099-2111.

Ferrari, D., Harrington, A., Dealy, C. N., and Kosher, R. A. (1999). Dlx-5 in limb initiation in the chick embryo. Dev Dyn *216*, 10-15.

Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R. J., Hardison, R., Miller, W., Philipsen, S., K.C., T.-U., McMorrow, T., *et al.* (2001). Comparative genome analysis delimits a chromosomal domain and idnetifies key regualtory elements in the alpha globin cluster. Hum Mol Genet *10*, 371-382.

Fode, C., Ma, Q., Casarosa, S., Ang, S. L., Anderson, D. J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes Dev *14*, 67-80.

Freund, T. F., and Buzsaki, G. (1996). Interneurons of the hippocampus. Hippocampus *6*, 347-470.

Fu, S., Stevenson, H., Strovel, J. W., Haga, S. B., Stamberg, J., Do, K., and Berg, P. E. (2001). Distinct functions of two isoforms of a homeobox gene, BP1 and DLX7, in the regulation of the beta-globin gene. Gene *278*, 131-139.

Fukuchi-Shimogori, T., and Grove, E. A. (2001). Neocortex patterning by the secreted signaling molecule FGF8. Science *294*, 1071-1074.

Furuta, Y., Piston, D. W., and Hogan, B. L. (1997). Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. Development *124*, 2203-2212.

Gaiano, N., Kohtz, J. D., Turnbull, D. H., and Fishell, G. (1999). A method for rapid gain-of-function studies in the mouse embryonic nervous system. Nat Neurosci *2*, 812-819.

Galceran, J., Miyashita-Lin, E. M., Devaney, E., Rubenstein, J. L., and Grosschedl, R. (2000). Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. Development *127*, 469-482.

Gall, C. M., Hendry, S. H., Seroogy, K. B., Jones, E. G., and Haycock, J. W. (1987). Evidence for coexistence of GABA and dopamine in neurons of the rat olfactory bulb. J Comp Neurol *266*, 307-318.

Gallera, J. (1971). Primary induction in birds. Adv Morphog *9*, 149-180.

Gerfen, C. a. W., CJ (1996). The basal ganglia. In: Handbook of chemical neuroanatomy Vol. 12, Integrates systems in the CNS).

Ghanem, N., Jarinova, O., Amores, A., Hatch, G., Park, B. K., Rubenstein, J. L. R., and Ekker, M. (2003). Regulatory roles of conserved intergenic domains in vertebrate *Dlx* bigene clusters. Genome Res *13*, 533-543.

Gilbert, S. F. (2000). Developmental Biology, Sixth Edition edn, Sinauer Associates, Inc).

Glazko, G. V., Koonin, E. V., Rogozin, I. B., and Shabalina, S. A. (2003). A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. Trends Genet *19*, 119-124.

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

Golden, J. A., Bracilovic, A., McFadden, K. A., Beesley, J. S., Rubenstein, J. L., and Grinspan, J. B. (1999). Ectopic bone morphogenetic proteins 5 and 4 in the chicken forebrain lead to cyclopia and holoprosencephaly. Proc Natl Acad Sci U S A *96*, 2439-2444.

Gomez-Skarmeta, J. L., Lenhard, B., and Becker, T. S. (2006). New technologies, new findings, and new concepts in the study of vertebrate cis-regulatory sequences. Dev Dyn *235*, 870-885.

Gonchar, Y., and Burkhalter, A. (1997). Three distinct families of GABAergic neurons in rat visual cortex. Cereb Cortex *7*, 347-358.

Gorfinkiel, N., Morata, G., and Guerrero, I. (1997). The homeobox gene Distal-less induces ventral appendage development in Drosophila. Genes Dev *11*, 2259-2271.

Gorfinkiel, N., Sanchez, L., and Guerrero, I. (1999). Drosophila terminalia as an appendage-like structure. Mech Dev *86*, 113-123.

Gorski, J. A., Talley, T., Qiu, M., Puelles, L., Rubenstein, J. L., and Jones, K. R. (2002). Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. J Neurosci *22*, 6309-6314.

Gotz, M., and Barde, Y. A. (2005). Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons. Neuron *46*, 369-372.

Harafuji, N., Keys, D. N., and Levine, M. (2002). Genome-wide identification of tissue-specific enhancers in the *Ciona* tadpole. Proc Natl Acad Sci USA *99*, 6802-6805.

Hardison, R. C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet *16*, 369-372.

Harland, R. (2000). Neural induction. Curr Opin Genet Dev *10*, 357-362.

Haubensak, W., Attardo, A., Denk, W., and Huttner, W. B. (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. Proc Natl Acad Sci U S A *101*, 3196-3201.

He, F., Ge, W., Martinowich, K., Becker-Catania, S., Coskun, V., Zhu, W., Wu, H., Castro, D., Guillemot, F., Fan, G., *et al.* (2005). A positive autoregulatory loop of Jak-STAT signaling controls the onset of astrogliogenesis. Nat Neurosci *8*, 616-625.

He, W., Ingraham, C., Rising, L., Goderie, S., and Temple, S. (2001). Multipotent stem cells from the mouse basal forebrain contribute GABAergic neurons and oligodendrocytes to the cerebral cortex during embryogenesis. J Neurosci *21*, 8854-8862.

Heimer, L. (1995). The basal ganglia. In: The Rat Nervous System.).

Hendry, S. H., Schwark, H. D., Jones, E. G., and Yan, J. (1987). Numbers and proportions of GABA-immunoreactive neurons in different areas of monkey cerebral cortex. J Neurosci *7*, 1503-1519.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. Nat Rev Neurosci *6*, 877-888.

Hevner, R. F., Daza, R. A., Englund, C., Kohtz, J., and Fink, A. (2004). Postnatal shifts of interneuron position in the neocortex of normal and reeler mice: evidence for inward radial migration. Neuroscience *124*, 605-618.

Holland, N. D., Panganiban, G., Henyey, E. L., and Holland, L. Z. (1996). Sequence and developmental expression of AmphiDll, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. Development *122*, 2911-2920.

Horike, S., Cai, S., Miyano, M., Cheng, J. F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet *37*, 31-40.

Iler, N., Rowitch, D. H., Echelard, Y., McMahon, A. P., and Abate-Shen, C. (1995). A single homeodomain binding site restricts spatial expression of Wnt-1 in the developing brain. Mech Dev *53*, 87-96.

Inoue, T., Nakamura, S., and Osumi, N. (2000). Fate mapping of the mouse prosencephalic neural plate. Dev Biol *219*, 373-383.

Ivanova, A., Nakahira, E., Kagawa, T., Oba, A., Wada, T., Takebayashi, H., Spassky, N., Levine, J., Zalc, B., and Ikenaka, K. (2003). Evidence for a second wave of oligodendrogenesis in the postnatal cerebral cortex of the mouse. J Neurosci Res *73*, 581-592.

Jimenez, D., Lopez-Mascaraque, L. M., Valverde, F., and De Carlos, J. A. (2002). Tangential migration in neocortical development. Dev Biol *244*, 155-169.

Källen, B. (1951a). Embryological studies on the nuclei and their homologization in the vertebrate forebrain. Kgl Fysiogr Sällsk Lund Handl NF *62*, 3-34.

Källen, B. (1951b). The nuclear development in the mammalian forebrain with special regard to the subpallium. Kgl Fysiogr Sällsk Lund Handl NF *61*, 1-43.

Källen, B. (1951c). On the ontogeny of the reptilian forebrain. Nuclear structures and ventricular sulci. J Comp Neurol *95*, 307-347.

Kang, Y., Kaneko, T., Ohishi, H., Endo, K., and Araki, T. (1994). Spatiotemporally differential inhibition of pyramidal cells in the cat motor cortex. J Neurophysiol *71*, 280-293.

Kaphingst, K., and Kunes, S. (1994). Pattern formation in the visual centers of the Drosophila brain: wingless acts via decapentaplegic to specify the dorsoventral axis. Cell *78*, 437-448.

Kawaguchi, Y. (1997a). Neostriatal cell subtypes and their functional roles. Neurosci Res *27*, 1-8.

Kawaguchi, Y., and Kubota, Y. (1997b). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cereb Cortex *7*, 476-486.

Kawaguchi, Y., Wilson, C. J., Augood, S. J., and Emson, P. C. (1995). Striatal interneurones: chemical, physiological and morphological characterization. Trends Neurosci *18*, 527-535.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

Kessaris, N., Fogarty, M., Iannarelli, P., Grist, M., Wegner, M., and Richardson, W. D. (2006). Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. Nat Neurosci *9*, 173-179.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution.).

Kimura, S., Hara, Y., Pineau, T., Fernandez-Salguero, P., Fox, C. H., Ward, J. M., and Gonzalez, F. J. (1996). The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev *10*, 60-69.

King, D. C., Taylor, J., Elnitski, L., Chiaromonte, F., Miller, W., and Hardison, R. C. (2005). Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res *15*, 1051-1060.

Kisvarday, Z. F., and Eysel, U. T. (1993). Functional and structural topography of horizontal inhibitory connections in cat visual cortex. Eur J Neurosci *5*, 1558-1572.

Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., *et al.* (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat Genet *32*, 359-369.

Kohtz, J. D., Baker, D. P., Corte, G., and Fishell, G. (1998). Regionalization within the mammalian telencephalon is mediated by changes in responsiveness to Sonic Hedgehog. Development *125*, 5079-5089.

Kornack, D. R., and Rakic, P. (2001). The generation, migration, and differentiation of olfactory neurons in the adult primate brain. Proc Natl Acad Sci U S A *98*, 4752-4757.

Krivan, W., and Wasserman, W. W. (2001). A predictive model for regulatory sequences directing liver-specific transcription. Genome Res *11*, 1559-1566.

Kubota, Y., Hattori, R., and Yui, Y. (1994). Three distinct subpopulations of GABAergic neurons in rat frontal agranular cortex. Brain Res *649*, 159-173.

Kumar, S., and Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917-920.

Ladher, R. K., Anakwe, K. U., Gurney, A. L., Schoenwolf, G. C., and Francis-West, P. H. (2000). Identification of synergistic signals initiating inner ear development. Science *290*, 1965-1967.

Lamb, T. M., and Harland, R. M. (1995). Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior-posterior neural pattern. Development *121*, 3627-3636.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lavdas, A. A., Grigoriou, M., Pachnis, V., and Parnavelas, J. G. (1999). The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. J Neurosci *19*, 7881-7888.

Lee, D. U., Avni, O., Chen, L., and Rao, A. (2004). A distal enhancer in the interferon-gamma (IFN-gamma) locus revealed by genome sequence comparison. J Biol Chem *279*, 4802-4810.

Lee, S. M., Tole, S., Grove, E., and McMahon, A. P. (2000). A local Wnt-3a signal is required for development of the mammalian hippocampus. Development *127*, 457-467.

Letinic, K., Zoncu, R., and Rakic, P. (2002). Origin of GABAergic neurons in the human neocortex. Nature *417*, 645-649.

Lewin, B. (1997). Genes VI, Oxford University Press).

Lewis, D. A. (2000). GABAergic local circuit neurons and prefrontal cortical dysfunction in schizophrenia. Brain Res Brain Res Rev *31*, 270-276.

Liem, K. F., Jr., Jessell, T. M., and Briscoe, J. (2000). Regulation of the neural patterning activity of sonic hedgehog by secreted BMP inhibitors expressed by notochord and somites. Development *127*, 4855-4866.

Liem, K. F., Jr., Tremml, G., and Jessell, T. M. (1997). A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. Cell *91*, 127-138.

Liem, K. F., Jr., Tremml, G., Roelink, H., and Jessell, T. M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell *82*, 969-979.

Lien, C. L., McAnally, J., Richardson, J. A., and Olson, E. N. (2002). Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. Dev Biol *244*, 257-266.

Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B., and Bartel, D. P. (2003). Vertebrate microRNA genes. Science *299*, 1540.

Liu, J. K., Ghattas, I., Liu, S., Chen, S., and Rubenstein, J. L. (1997). Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation. Dev Dyn *210*, 498-512.

Lois, C., and Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. Science *264*, 1145-1148.

Long, J. E., Garel, S., Depew, M. J., Tobet, S., and Rubenstein, J. L. R. (2003). DLX5 regulates development of peripheral and central components of the olfactory system. J Neurosci *23*, 568-578.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science *288*, 136-140.

Luskin, M. B., Pearlman, A. L., and Sanes, J. R. (1988). Cell lineage in the cerebral cortex of the mouse studied in vivo and in vitro with a recombinant retrovirus. Neuron *1*, 635-647.

Ma, Q., Sommer, L., Cserjesi, P., and Anderson, D. J. (1997). Mash1 and neurogenin1 expression patterns define complementary domains of neuroepithelium in the developing CNS and are correlated with regions expressing notch ligands. J Neurosci *17*, 3644-3652.

Malatesta, P., Hack, M. A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., and Gotz, M. (2003). Neuronal or glial progeny: regional differences in radial glia fate. Neuron *37*, 751-764.

Mallamaci, A., Muzio, L., Chan, C. H., Parnavelas, J., and Boncinelli, E. (2000). Area identity shifts in the early cerebral cortex of Emx2-/- mutant mice. Nat Neurosci *3*, 679-686.

Margulies, E. H., Blanchette, M., Haussler, D., and Green, E. D. (2003). Identification and characterization of multi-species conserved sequences. Genome Res *13*, 2507-2518.

Marin, O., Anderson, S. A., and Rubenstein, J. L. (2000). Origin and molecular specification of striatal interneurons. J Neurosci *20*, 6063-6076.

Marin, O., Plump, A. S., Flames, N., Sanchez-Camacho, C., Tessier-Lavigne, M., and Rubenstein, J. L. (2003b). Directional guidance of interneuron migration to the cerebral cortex relies on subcortical Slit1/2-independent repulsion and cortical attraction. Development *130*, 1889-1901.

Marin, O., and Rubenstein, J. L. (2001b). A long, remarkable journey: tangential migration in the telencephalon. Nat Rev Neurosci *2*, 780-790.

Marin, O., and Rubenstein, J. L. (2003a). Cell migration in the forebrain. Annu Rev Neurosci *26*, 441-483.

Marin, O., Yaron, A., Bagri, A., Tessier-Lavigne, M., and Rubenstein, J. L. (2001a). Sorting of striatal and cortical interneurons regulated by semaphorin-neuropilin interactions. Science *293*, 872-875.

Marin, O. a. R., J. (2002). Patterning, Regionalization, and Cell Differentiation in the Forebrain. In Mouse Development (Rossant J and Tam P)).

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nat Rev Neurosci *5*, 793-807.

Marshall, C. A., Suzuki, S. O., and Goldman, J. E. (2003). Gliogenic and neurogenic progenitors of the subventricular zone: who are they, where did they come from, and where are they going? Glia *43*, 52-61.

Masuda, Y., Sasaki, A., Shibuya, H., Ueno, N., Ikeda, K., and Watanabe, K. (2001). Dlxin-1, a novel protein that binds Dlx5 and regulates its transcriptional function. J Biol Chem *276*, 5331-5338.

Matsuo, I., Kuratani, S., Kimura, C., Takeda, N., and Aizawa, S. (1995). Mouse Otx2 functions in the formation and patterning of rostral head. Genes Dev *9*, 2646-2658.

McBain, C. J., and Fisahn, A. (2001). Interneurons unbound. Nat Rev Neurosci *2*, 11-23.

McCormick, D. A. (1989). GABA as an inhibitory neurotransmitter in human cerebral cortex. J Neurophysiol *62*, 1018-1027.

McGuinness, T., Porteus, M. H., Smiga, S., Bulfone, A., Kingsley, C., Qiu, M., Liu, J. K., Long, J. E., Xu, D., and Rubenstein, J. L. (1996). Sequence, organization, and transcription of the Dlx-1 and Dlx-2 locus. Genomics *35*, 473-485.

Meinecke, D. L., and Peters, A. (1987). GABA immunoreactive neurons in rat visual cortex. J Comp Neurol *261*, 388-404.

Merlo, G. R., Zerega, B., Paleari, L., Trombino, S., Mantero, S., and Levi, G. (2000). Multiple functions of Dlx genes. Int J Dev Biol *44*, 619-626.

Metin, C., Baudoin, J. P., Rakic, S., and Parnavelas, J. G. (2006). Cell and molecular mechanisms involved in the migration of cortical interneurons. Eur J Neurosci *23*, 894-900.

Meyer, G., Schaaps, J. P., Moreau, L., and Goffinet, A. M. (2000). Embryonic and early fetal development of the human neocortex. J Neurosci *20*, 1858-1868.

Meyers, E. N., Lewandoski, M., and Martin, G. R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat Genet *18*, 136-141.

Miller, M. W. (1985). Cogeneration of retrogradely labeled corticocortical projection and GABA-immunoreactive local circuit neurons in cerebral cortex. Brain Res *355*, 187-192.

Miyama, K., Yamada, G., Yamamoto, T. S., Takagi, C., Miyado, K., Sakai, M., Ueno, N., and Shibuya, H. (1999). A BMP-inducible gene, dlx5, regulates osteoblast differentiation and mesoderm induction. Dev Biol *208*, 123-133.

Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T., and Ogawa, M. (2004). Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. Development *131*, 3133-3145.

Molnar, Z., Metin, C., Stoykova, A., Tarabykin, V., Price, D. J., Francis, F., Meyer, G., Dehay, C., and Kennedy, H. (2006). Comparative aspects of cerebral cortical development. Eur J Neurosci *23*, 921-934.

Monier, C., Chavane, F., Baudot, P., Graham, L. J., and Fregnac, Y. (2003). Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. Neuron *37*, 663-680.

Monuki, E. S., Porter, F. D., and Walsh, C. A. (2001). Patterning of the dorsal telencephalon and cerebral cortex by a roof plate-Lhx2 pathway. Neuron *32*, 591-604.

Moore, K. L. a. T. V. N. P. (1993). *Before we are born: Essentials of Embryology and Birth defects*, W.B. Saunders, Philadelphia).

Morasso, M. I., Grinberg, A., Robinson, G., Sargent, T. D., and Mahon, K. A. (1999). Placental failure in mice lacking the homeobox gene Dlx3. Proc Natl Acad Sci U S A *96*, 162-167.

Morasso, M. I., Jamrich, M., and Sargent, T. D. (1994). The homeodomain gene Xenopus Distal-less-like-2 (Xdll-2) is regulated by a conserved mechanism in amphibian and mammalian epidermis. Dev Biol *162*, 267-276.

Morasso, M. I., Mahon, K. A., and Sargent, T. D. (1995). A Xenopus distal-less gene in transgenic mice: conserved regulation in distal limb epidermis and other sites of epithelial-mesenchymal interaction. Proc Natl Acad Sci U S A *92*, 3968-3972.

Morasso, M. I., Markova, N. G., and Sargent, T. D. (1996). Regulation of epidermal differentiation by a Distal-less homeodomain gene. J Cell Biol *135*, 1879-1887.

Morrow, T., Song, M. R., and Ghosh, A. (2001). Sequential specification of neurons and glia by developmentally regulated extracellular factors. Development *128*, 3585-3594.

Muenke, M., and Beachy, P. A. (2000). Genetics of ventral forebrain development and holoprosencephaly. Curr Opin Genet Dev *10*, 262-269.

Mullen, L. M., Bryant, S. V., Torok, M. A., Blumberg, B., and Gardiner, D. M. (1996). Nerve dependency of regeneration: the role of Distal-less and FGF signaling in amphibian limb regeneration. Development *122*, 3487-3497.

Muller, F., Blader, P., and Strahle, U. (2002). Search for enhancers: teleost models in comparative genomic and transgenic analysis of *cis* regulatory elements. BioEssays *24*, 564-572.

Murthy, A., and Humphrey, A. L. (1999). Inhibitory contributions to spatiotemporal receptive-field structure and direction selectivity in simple cells of cat area 17. J Neurophysiol *81*, 1212-1224.

Muzio, L., DiBenedetto, B., Stoykova, A., Boncinelli, E., Gruss, P., and Mallamaci, A. (2002). Emx2 and Pax6 control regionalization of the pre-neuronogenic cortical primordium. Cereb Cortex *12*, 129-139.

Myojin, M., Ueki, T., Sugahara, F., Murakami, Y., Shigetani, Y., Aizawa, S., Hirano, S., and Kuratani, S. (2001). Isolation of Dlx and Emx gene cognates in an agnathan species,

Lampetra japonica, and their expression patterns during embryonic and larval development: conserved and diversified regulatory patterns of homeobox genes in vertebrate head evolution. J Exp Zool *291*, 68-84.

Nadarajah, B., Alifragis, P., Wong, R. O., and Parnavelas, J. G. (2002). Ventricle-directed migration in the developing cerebral cortex. Nat Neurosci *5*, 218-224.

Nadarajah, B., and Parnavelas, J. G. (2002). Modes of neuronal migration in the developing cerebral cortex. Nat Rev Neurosci *3*, 423-432.

Naegele, J. R., and Barnstable, C. J. (1989). Molecular determinants of GABAergic local-circuit neurons in the visual cortex. Trends Neurosci *12*, 28-34.

Nakamura, S., Stock, D. W., Wydner, K. L., Bollekens, J. A., Takeshita, K., Nagai, B. M., Chiba, S., Kitamura, T., Freeland, T. M., Zhao, Z., *et al.* (1996). Genomic analysis of a new mammalian distal-less gene: Dlx7. Genomics *38*, 314-324.

Neidert, A. H., Virupannavar, V., Hooker, G. W., and Langeland, J. A. (2001). Lamprey Dlx genes and early vertebrate evolution. Proc Natl Acad Sci U S A *98*, 1665-1670.

Nery, S., Corbin, J. G., and Fishell, G. (2003). Dlx2 progenitor migration in wild type and Nkx2.1 mutant telencephalon. Cereb Cortex *13*, 895-903.

Nery, S., Fishell, G., and Corbin, J. G. (2002). The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. Nat Neurosci *5*, 1279-1287.

Newberry, E. P., Latifi, T., and Towler, D. A. (1998). Reciprocal regulation of osteocalcin transcription by the homeodomain proteins Msx2 and Dlx5. Biochemistry *37*, 16360-16368.

Nieto, M., Schuurmans, C., Britz, O., and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron *29*, 401-413.

Nieuwenhuys R, t. D. H., Nicholson C (1998). The Central Nervous System of Vertebrates., Berlin-Heidelberg-New York, Springer Verlag.).

Nobrega, M. A., Ovcharenko, I., Afzal, V., and Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. Science *302*, 413.

Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S., and Kriegstein, A. R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. Nature *409*, 714-720.

Noctor, S. C., Martinez-Cerdeno, V., Ivic, L., and Kriegstein, A. R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nat Neurosci *7*, 136-144.

Ogbourne, S., and Antalis, T. M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J *331 (Pt 1)*, 1-14.

Olivier, C., Cobos, I., Perez Villegas, E. M., Spassky, N., Zalc, B., Martinez, S., and Thomas, J. L. (2001). Monofocal origin of telencephalic oligodendrocytes in the anterior entopeduncular area of the chick embryo. Development *128*, 1757-1769.

Ovcharenko, I., Stubbs, L., and Loots, G. G. (2004). Interpreting mammalian evolution using Fugu genome comparisons. Genomics *84*, 890-895.

Panganiban, G., and Rubenstein, J. L. (2002). Developmental functions of the Distal-less/Dlx homeobox genes. Development *129*, 4371-4386.

Parent, A., and Hazrati, L. N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res Brain Res Rev *20*, 91-127.

Park, G. T., and Morasso, M. I. (1999). Regulation of the Dlx3 homeobox gene upon differentiation of mouse keratinocytes. J Biol Chem *274*, 26599-26608.

Parnavelas, J. G. (2000). The origin and migration of cortical neurones: new vistas. Trends Neurosci *23*, 126-131.

Parnavelas, J. G., Lieberman, A. R., and Webster, K. E. (1977). Organization of neurons in the visual cortex, area 17, of the rat. J Anat *124*, 305-322.

Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J., and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. Genes Dev *16*, 324-338.

Peduzzi, J. D. (1988). Genesis of GABA-immunoreactive neurons in the ferret visual cortex. J Neurosci *8*, 920-931.

Pellegrini, M., Mansouri, A., Simeone, A., Boncinelli, E., and Gruss, P. (1996). Dentate gyrus formation requires Emx2. Development *122*, 3893-3898.

Pencea, V., Bingaman, K. D., Freedman, L. J., and Luskin, M. B. (2001). Neurogenesis in the subventricular zone and rostral migratory stream of the neonatal and adult primate forebrain. Exp Neurol *172*, 1-16.

Pennacchio, L. A., and Rubin, E. M. (2001). Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet *2*, 100-109.

Pera, E., Stein, S., and Kessel, M. (1999). Ectodermal patterning in the avian embryo: epidermis versus neural plate. Development *126*, 63-73.

Peters A, J. E. (1984). Cellular components of the cerebral cortex).

Peters, A., and Sethares, C. (1991). Organization of pyramidal neurons in area 17 of monkey visual cortex. J Comp Neurol *306*, 1-23.

Peters, A. a. S. M. R. (1984). Smooth and sparsely spinous nonpyramidal cells forming local axonal plexuses. In: Cerebral Cortex, vol. 1).

Pierani, A., Brenner-Morton, S., Chiang, C., and Jessell, T. M. (1999). A sonic hedgehog-independent, retinoid-activated pathway of neurogenesis in the ventral spinal cord. Cell *97*, 903-915.

Pleasure, S. J., Anderson, S., Hevner, R., Bagri, A., Marin, O., Lowenstein, D. H., and Rubenstein, J. L. (2000). Cell migration from the ganglionic eminences is required for the development of hippocampal GABAergic interneurons. Neuron *28*, 727-740.

Poitras, L. G., N Hatch, G and Ekker, M (2006). The pro-neural determinant Mash1 regulates Dlx1/2 forebrain gene expression through the I12b intergenic enhancer. Submitted to Development.

Polleux, F., Whitford, K. L., Dijkhuizen, P. A., Vitalis, T., and Ghosh, A. (2002). Control of cortical interneuron migration by neurotrophins and PI3-kinase signaling. Development *129*, 3147-3160.

Porteus, M. H., Bulfone, A., Liu, J. K., Puelles, L., Lo, L. C., and Rubenstein, J. L. (1994). DLX-2, MASH-1, and MAP-2 expression and bromodeoxyuridine incorporation

define molecularly distinct cell populations in the embryonic mouse forebrain. J Neurosci *14*, 6370-6383.

Powell, E. M., Mars, W. M., and Levitt, P. (2001). Hepatocyte growth factor/scatter factor is a motogen for interneurons migrating from the ventral to dorsal telencephalon. Neuron *30*, 79-89.

Price, J. A., Bowden, D. W., Wright, J. T., Pettenati, M. J., and Hart, T. C. (1998a). Identification of a mutation in DLX3 associated with tricho-dento-osseous (TDO) syndrome. Hum Mol Genet *7*, 563-569.

Price, J. A., Wright, J. T., Kula, K., Bowden, D. W., and Hart, T. C. (1998b). A common DLX3 gene mutation is responsible for tricho-dento-osseous syndrome in Virginia and North Carolina families. J Med Genet *35*, 825-828.

Price, M., Lemaistre, M., Pischetola, M., Di Lauro, R., and Duboule, D. (1991). A mouse gene related to Distal-less shows a restricted expression in the developing forebrain. Nature *351*, 748-751.

Puelles, L., Kuwana, E., Puelles, E., Bulfone, A., Shimamura, K., Keleher, J., Smiga, S., and Rubenstein, J. L. (2000). Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. J Comp Neurol *424*, 409-438.

Puelles, L., Kuwana, E., Puelles, E., and Rubenstein, J. L. (1999). Comparison of the mammalian and avian telencephalon from the perspective of gene expression data. Eur J Morphol *37*, 139-150.

Puelles, L., and Rubenstein, J. L. (2003). Forebrain gene expression domains and the evolving prosomeric model. Trends Neurosci *26*, 469-476.

Qian, X., Shen, Q., Goderie, S. K., He, W., Capela, A., Davis, A. A., and Temple, S. (2000). Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. Neuron *28*, 69-80.

Qiu, M., Bulfone, A., Ghattas, I., Meneses, J. J., Christensen, L., Sharpe, P. T., Presley, R., Pedersen, R. A., and Rubenstein, J. L. (1997). Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: mutations of Dlx-1, Dlx-2, and Dlx-1

and -2 alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches. Dev Biol *185*, 165-184.

Qiu, M., Bulfone, A., Martinez, S., Meneses, J. J., Shimamura, K., Pedersen, R. A., and Rubenstein, J. L. (1995). Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain. Genes Dev *9*, 2523-2538.

Quint, E., Zerucha, T., and Ekker, M. (2000). Differential expression of orthologous Dlx genes in zebrafish and mice: implications for the evolution of the Dlx homeobox gene family. J Exp Zool *288*, 235-241.

Rakic, P. (1971). Guidance of neurons migrating to the fetal monkey neocortex. Brain Res *33*, 471-476.

Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. J Comp Neurol *145*, 61-83.

Rakic, P. (1974). Neurons in rhesus monkey visual cortex: systematic relation between time of origin and eventual disposition. Science *183*, 425-427.

Rakic, S., and Zecevic, N. (2003). Emerging complexity of layer I in human cerebral cortex. Cereb Cortex *13*, 1072-1083.

Rallu, M., Machold, R., Gaiano, N., Corbin, J. G., McMahon, A. P., and Fishell, G. (2002). Dorsoventral patterning is established in the telencephalon of mutants lacking both Gli3 and Hedgehog signaling. Development *129*, 4963-4974.

Rao, S. G., Williams, G. V., and Goldman-Rakic, P. S. (2000). Destruction and creation of spatial tuning by disinhibition: GABA(A) blockade of prefrontal cortical neurons engaged by working memory. J Neurosci *20*, 485-494.

Reid, C. B., Liang, I., and Walsh, C. (1995). Systematic widespread clonal organization in cerebral cortex. Neuron *15*, 299-310.

Roberson, M. S., Meermann, S., Morasso, M. I., Mulvaney-Musa, J. M., and Zhang, T. (2001). A role for the homeobox protein Distal-less 3 in the activation of the glycoprotein hormone alpha subunit gene in choriocarcinoma cells. J Biol Chem *276*, 10016-10024.

Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P. L., Zelus, D., Hughes, S., and Laudet, V. (2001). Euteleost fish genomes are characterized by expansion of gene families. Genome Res *11*, 781-788.

Robinson, G. W., and Mahon, K. A. (1994). Differential and overlapping expression domains of Dlx-2 and Dlx-3 suggest distinct roles for Distal-less homeobox genes in craniofacial development. Mech Dev *48*, 199-215.

Robinson, G. W., Wray, S., and Mahon, K. A. (1991). Spatially restricted expression of a member of a new family of murine Distal-less homeobox genes in the developing forebrain. New Biol *3*, 1183-1194.

Robledo, R. F., Rajan, L., Li, X., and Lufkin, T. (2002). The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev *16*, 1089-1101.

Ross, S. E., Greenberg, M. E., and Stiles, C. D. (2003). Basic helix-loop-helix factors in cortical development. Neuron *39*, 13-25.

Rossant, J. a. T. P. P. L. (2002). Mouse Development: Patterning, Morphogenesis and Organogenesis).

Rubenstein, J. L., and Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav *2*, 255-267.

Rubenstein, J. L., Shimamura, K., Martinez, S., and Puelles, L. (1998). Regionalization of the prosencephalic neural plate. Annu Rev Neurosci *21*, 445-477.

Ryoo, H. M., Hoffmann, H. M., Beumer, T., Frenkel, B., Towler, D. A., Stein, G. S., Stein, J. L., van Wijnen, A. J., and Lian, J. B. (1997). Stage-specific expression of Dlx-5 during osteoblast differentiation: involvement in regulation of osteocalcin gene expression. Mol Endocrinol *11*, 1681-1694.

Santagati, F., Abe, K., Schmidt, V., Schmitt-John, T., Suzuki, M., Yamamura, K., and Imai, K. (2003). Identification of Cis-regulatory elements in the mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved synteny. Genetics *165*, 235-242.

Sasai, Y., and De Robertis, E. M. (1997). Ectodermal patterning in vertebrate embryos. Dev Biol *182*, 5-20.

Sato, H., Katsuyama, N., Tamura, H., Hata, Y., and Tsumoto, T. (1996). Mechanisms underlying orientation selectivity of neurons in the primary visual cortex of the macaque. J Physiol *494 (Pt 3)*, 757-771.

Scherer, S. W., Heng, H. H., Robinson, G. W., Mahon, K. A., Evans, J. P., and Tsui, L. C. (1995). Assignment of the human homolog of mouse Dlx3 to chromosome 17q21.3-q22 by analysis of somatic cell hybrids and fluorescence in situ hybridization. Mamm Genome *6*, 310-311.

Scherer, S. W., Poorkaj, P., Massa, H., Soder, S., Allen, T., Nunes, M., Geshuri, D., Wong, E., Belloni, E., Little, S., and et al. (1994). Physical mapping of the split hand/split foot locus on chromosome 7 and implication in syndromic ectrodactyly. Hum Mol Genet *3*, 1345-1354.

Schier, A. F., Neuhauss, S. C., Harvey, M., Malicki, J., Solnica-Krezel, L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Stemple, D. L., Rangini, Z., *et al.* (1996). Mutations affecting the development of the embryonic zebrafish brain. Development *123*, 165-178.

Schier, A. F., and Shen, M. M. (2000). Nodal signalling in vertebrate development. Nature *403*, 385-389.

Schuurmans, C., and Guillemot, F. (2002). Molecular mechanisms underlying cell fate specification in the developing telencephalon. Curr Opin Neurobiol *12*, 26-34.

Shanmugalingam, S., Houart, C., Picker, A., Reifers, F., Macdonald, R., Barth, A., Griffin, K., Brand, M., and Wilson, S. W. (2000). Ace/Fgf8 is required for forebrain commissure formation and patterning of the telencephalon. Development *127*, 2549-2561.

Shawlot, W., and Behringer, R. R. (1995). Requirement for Lim1 in head-organizer function. Nature *374*, 425-430.

Shigetani, Y., Sugahara, F., Kawakami, Y., Murakami, Y., Hirano, S., and Kuratani, S. (2002). Heterotopic shift of epithelial-mesenchymal interactions in vertebrate jaw evolution. Science *296*, 1316-1319.

Shimamoto, T., Nakamura, S., Bollekens, J., Ruddle, F. H., and Takeshita, K. (1997). Inhibition of DLX-7 homeobox gene causes decreased expression of GATA-1 and c-myc genes and apoptosis. Proc Natl Acad Sci U S A *94*, 3245-3249.

Shimamoto, T., Ohyashiki, K., and Takeshita, K. (2000). Overexpression of the homeobox gene DLX-7 inhibits apoptosis by induced expression of intercellular adhesion molecule-1. Exp Hematol *28*, 433-441.

Shin, J. T., Priest, J. R., Ovcharenko, I., Ronco, A., Moore, R. K., Burns, C. G., and MacRae, C. A. (2005). Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. Nucleic Acids Res *33*, 5437-5445.

Sidow, A. (2002). Sequence first. Ask questions later. Cell *111*, 13-16.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Sillito, A. (1984). Functional considerations of the operation of GABAergic inhibitory processes in the visual cortex. In Cerebral Cortex.).

Simeone, A., Acampora, D., Pannese, M., D'Esposito, M., Stornaiuolo, A., Gulisano, M., Mallamaci, A., Kastury, K., Druck, T., Huebner, K., and et al. (1994). Cloning and characterization of two members of the vertebrate Dlx gene family. Proc Natl Acad Sci U S A *91*, 2250-2254.

Smith, A. D., Sumazin, P., Xuan, Z., and Zhang, M. Q. (2006). DNA motifs in human and mouse proximal promoters predict tissue-specific expression. Proc Natl Acad Sci U S A *103*, 6275-6280.

Smith, J. L., and Schoenwolf, G. C. (1997). Neurulation: coming to closure. Trends Neurosci *20*, 510-517.

Solloway, M. J., and Robertson, E. J. (1999). Early embryonic lethality in Bmp5;Bmp7 double mutant mice suggests functional redundancy within the 60A subgroup. Development *126*, 1753-1768.

Spassky, N., Goujet-Zalc, C., Parmantier, E., Olivier, C., Martinez, S., Ivanova, A., Ikenaka, K., Macklin, W., Cerruti, I., Zalc, B., and Thomas, J. L. (1998). Multiple restricted origin of oligodendrocytes. J Neurosci *18*, 8331-8343.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405-417.

Stenman, J., Toresson, H., and Campbell, K. (2003). Identification of two distinct progenitor populations in the lateral ganglionic eminence: implications for striatal and olfactory bulb neurogenesis. J Neurosci *23*, 167-174.

Stock, D. W., Ellies, D. L., Zhao, Z., Ekker, M., Ruddle, F. H., and Weiss, K. M. (1996). The evolution of the vertebrate Dlx gene family. Proc Natl Acad Sci U S A *93*, 10858-10863.

Storm, E. E., Garel, S., Borello, U., Hebert, J. M., Martinez, S., McConnell, S. K., Martin, G. R., and Rubenstein, J. L. (2006). Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. Development *133*, 1831-1844.

Stoykova, A., Treichel, D., Hallonet, M., and Gruss, P. (2000). Pax6 modulates the dorsoventral patterning of the mammalian telencephalon. J Neurosci *20*, 8042-8050.

Striedter, G. F. (1997). The telencephalon of tetrapods in evolution. Brain Behav Evol *49*, 179-213.

Stuhmer, T., Anderson, S. A., Ekker, M., and Rubenstein, J. L. (2002a). Ectopic expression of the Dlx genes induces glutamic acid decarboxylase and Dlx expression. Development *129*, 245-252.

Stuhmer, T., Puelles, L., Ekker, M., and Rubenstein, J. L. (2002b). Expression from a Dlx gene enhancer marks adult mouse cortical GABAergic neurons. Cereb Cortex *12*, 75-85.

Sumiyama, K., Irvine, S. Q., Stock, D. W., Weiss, K. M., Kawasaki, K., Shimizu, N., Shashikant, C. S., Miller, W., and Ruddle, F. H. (2002). Genomic structure and functional control of the Dlx3-7 bigene cluster. Proc Natl Acad Sci U S A *99*, 780-785.

Sumiyama, K., Kim, C. B., and Ruddle, F. H. (2001). An efficient cis-element discovery method using multiple sequence comparisons based on evolutionary relationships. Genomics *71*, 260-262.

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G., and Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. Cell *104*, 365-376.

Sussel, L., Marin, O., Kimura, S., and Rubenstein, J. L. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. Development *126*, 3359-3370.

Szucsik, J. C., Witte, D. P., Li, H., Pixley, S. K., Small, K. M., and Potter, S. S. (1997). Altered forebrain and hindbrain development in mice mutant for the Gsh-2 homeobox gene. Dev Biol *191*, 230-242.

Tamamaki, N., Fujimori, K. E., and Takauji, R. (1997). Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone. J Neurosci *17*, 8313-8323.

Tamamaki, N., Yanagawa, Y., Tomioka, R., Miyazaki, J., Obata, K., and Kaneko, T. (2003). Green fluorescent protein expression and colocalization with calretinin, parvalbumin, and somatostatin in the GAD67-GFP knock-in mouse. J Comp Neurol *467*, 60-79.

Tamas, G., Lorincz, A., Simon, A., and Szabadics, J. (2003). Identified sources and targets of slow inhibition in the neocortex. Science *299*, 1902-1905.

Tanabe, Y., and Jessell, T. M. (1996). Diversity and pattern in the developing spinal cord. Science *274*, 1115-1123.

Tanaka, D., Nakaya, Y., Yanagawa, Y., Obata, K., and Murakami, F. (2003). Multimodal tangential migration of neocortical GABAergic neurons independent of GPI-anchored proteins. Development *130*, 5803-5813.

Tautz, D. (2000). Evolution of transcriptional regulation. Curr Opin Genet Dev *10*, 575-579.

Tekki-Kessaris, N., Woodruff, R., Hall, A. C., Gaffield, W., Kimura, S., Stiles, C. D., Rowitch, D. H., and Richardson, W. D. (2001). Hedgehog-dependent oligodendrocyte lineage specification in the telencephalon. Development *128*, 2545-2554.

Temple, S. (2001). The development of neural stem cells. Nature *414*, 112-117.

Theil, T., Alvarez-Bolado, G., Walter, A., and Ruther, U. (1999). Gli3 is required for Emx gene expression during dorsal telencephalon development. Development *126*, 3561-3571.

Thomas, B. L., Liu, J. K., Rubenstein, J. L., and Sharpe, P. T. (2000). Independent regulation of Dlx2 expression in the epithelium and mesenchyme of the first branchial arch. Development *127*, 217-224.

Thomas, J. L., Spassky, N., Perez Villegas, E. M., Olivier, C., Cobos, I., Goujet-Zalc, C., Martinez, S., and Zalc, B. (2000). Spatiotemporal development of oligodendrocytes in the embryonic brain. J Neurosci Res *59*, 471-476.

Thomas, L. B., Gates, M. A., and Steindler, D. A. (1996). Young neurons from the adult subependymal zone proliferate and migrate along an astrocyte, extracellular matrix-rich pathway. Glia *17*, 1-14.

Thomsen, G. H. (1997). Antagonism within and around the organizer: BMP inhibitors in vertebrate body patterning. Trends Genet *13*, 209-211.

Thomson, A. M., and Destexhe, A. (1999). Dual intracellular recordings and computational models of slow inhibitory postsynaptic potentials in rat neocortical and hippocampal slices. Neuroscience *92*, 1193-1215.

Tobet, S. A., Chickering, T. W., Hanna, I., Crandall, J. E., and Schwarting, G. A. (1994). Can gonadal steroids influence cell position in the developing brain? Horm Behav *28*, 320-327.

Tole, S., Goudreau, G., Assimacopoulos, S., and Grove, E. A. (2000). Emx2 is required for growth of the hippocampus but not for hippocampal field specification. J Neurosci *20*, 2618-2625.

Tole, S., Ragsdale, C. W., and Grove, E. A. (2000). Dorsoventral patterning of the telencephalon is disrupted in the mouse mutant extra-toes(J). Dev Biol *217*, 254-265.

Toresson, H., Potter, S. S., and Campbell, K. (2000). Genetic control of dorsal-ventral identity in the telencephalon: opposing roles for Pax6 and Gsh2. Development *127*, 4361-4371.

Trinklein, N. D., Aldred, S. J., Saldanha, A. J., and Myers, R. M. (2003). Identification and functional analysis of human transcriptional promoters. Genome Res *13*, 308-312.

Tumpel, S., Maconochie, M., Wiedemann, L. M., and Krumlauf, R. (2002). Conservation and diversity in the cis-regulatory networks that integrate information controlling expression of Hoxa2 in hindbrain and cranial neural crest cells in vertebrates. Dev Biol *246*, 45-56.

Valcanis, H., and Tan, S. S. (2003). Layer specification of transplanted interneurons in developing mouse neocortex. J Neurosci *23*, 5113-5122.

Varga, Z. M., Wegner, J., and Westerfield, M. (1999). Anterior movement of ventral diencephalic precursors separates the primordial eye field in the neural plate and requires cyclops. Development *126*, 5533-5546.

Venkatesh, B., and Yap, W. H. (2005). Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. Bioessays *27*, 100-107.

Warren, N., Caric, D., Pratt, T., Clausen, J. A., Asavaritikrai, P., Mason, J. O., Hill, R. E., and Price, D. J. (1999). The transcription factor, Pax6, is required for cell proliferation and differentiation in the developing cerebral cortex. Cereb Cortex *9*, 627-635.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wehr, M., and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442-446.

Weiss, K. M., Bollekens, J., Ruddle, F. H., and Takashita, K. (1994). Distal-less and other homeobox genes in the development of the dentition. J Exp Zool *270*, 273-284.

Wessely, O., and De Robertis, E. M. (2002). Neural plate patterning by secreted signals. Neuron *33*, 489-491.

White, E. (1989). Synaptic Organization of the Cerebral Cortex).

Whitlock, K. E., and Westerfield, M. (2000). The olfactory placodes of the zebrafish form by convergence of cellular fields at the edge of the neural plate. Development *127*, 3645-3653.

Wichterle, H., Alvarez-Dolado, M., Erskine, L., and Alvarez-Buylla, A. (2003). Permissive corridor and diffusible gradients direct medial ganglionic eminence cell migration to the neocortex. Proc Natl Acad Sci U S A *100*, 727-732.

Wichterle, H., Garcia-Verdugo, J. M., Herrera, D. G., and Alvarez-Buylla, A. (1999). Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. Nat Neurosci *2*, 461-466.

Wichterle, H., Turnbull, D. H., Nery, S., Fishell, G., and Alvarez-Buylla, A. (2001). In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. Development *128*, 3759-3771.

Williams, B. P., and Price, J. (1995). Evidence for multiple precursor cell types in the embryonic rat cerebral cortex. Neuron *14*, 1181-1188.

Wilson, S. W., and Rubenstein, J. L. (2000). Induction and dorsoventral patterning of the telencephalon. Neuron *28*, 641-651.

Wonders, C. P., and Anderson, S. A. (2006). The origin and specification of cortical interneurons. Nat Rev Neurosci.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., *et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol *3*, e7.

Wu, J., and Cohen, S. M. (1999). Proximodistal axis formation in the Drosophila leg: subdivision into proximal and distal domains by Homothorax and Distal-less. Development *126*, 109-117.

Xu, Q., Cobos, I., De La Cruz, E., Rubenstein, J. L., and Anderson, S. A. (2004). Origins of cortical interneuron subtypes. J Neurosci *24*, 2612-2622.

Xu, S. C., Harris, M. A., Rubenstein, J. L., Mundy, G. R., and Harris, S. E. (2001). Bone morphogenetic protein-2 (BMP-2) signaling to the Col2alpha1 gene in chondroblasts requires the homeobox gene Dlx-2. DNA Cell Biol *20*, 359-365.

Yang, L., Zhang, H., Hu, G., Wang, H., Abate-Shen, C., and Shen, M. M. (1998). An early phase of embryonic Dlx5 expression defines the rostral boundary of the neural plate. J Neurosci *18*, 8322-8330.

Yang, Z., Liu, N., and Lin, S. (2001). A zebrafish forebrain-specific zinc finger gene can induce ectopic dlx2 and dlx6 expression. Dev Biol *231*, 138-148.

Yee, S.-P., and Rigby, P. W. J. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. Genes Dev *7*, 1277-1289.

Yoshida, M., Suda, Y., Matsuo, I., Miyamoto, N., Takeda, N., Kuratani, S., and Aizawa, S. (1997). Emx1 and Emx2 functions in development of dorsal telencephalon. Development *124*, 101-111.

Yozu, M., Tabata, H., and Nakajima, K. (2004). Birth-date dependent alignment of GABAergic neurons occurs in a different pattern from that of non-GABAergic neurons in the developing mouse visual cortex. Neurosci Res *49*, 395-403.

Yu, G., Zerucha, T., Ekker, M., and Rubenstein, J. L. (2001). Evidence that GRIP, a PDZ-domain protein which is expressed in the embryonic forebrain, co-activates transcription with DLX homeodomain proteins. Brain Res Dev Brain Res *130*, 217-230.

Yun, K., Fischman, S., Johnson, J., Hrabe de Angelis, M., Weinmaster, G., and Rubenstein, J. L. (2002). Modulation of the notch signaling by Mash1 and Dlx1/2 regulates sequential specification and differentiation of progenitor cell types in the subcortical telencephalon. Development *129*, 5029-5040.

Yun, K., Potter, S., and Rubenstein, J. L. (2001). Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon. Development *128*, 193-205.

Zaki, P. A., Quinn, J. C., and Price, D. J. (2003). Mouse models of telencephalic development. Curr Opin Genet Dev *13*, 423-437.

Zerucha, T., and Ekker, M. (2000). *Distal-less*-related homeobox genes of vertebrates: evolution, function, and regulation. Biochem Cell Biol *78*, 593-601.

Zerucha, T., Muller, J.-P., Chartrand, N., and Ekker, M. (1997). Cross-interactions between two members of the *Dlx* family of homeobox-containing genes during zebrafish development. Biochem Cell Biol *75*, 613-622.

Zerucha, T., Stuhmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., Gambarotta, A., Schultz, J. R., Rubenstein, J. L., and Ekker, M. (2000). A highly conserved enhancer in

the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. J Neurosci *20*, 709-721.

Zhang, H., Hu, G., Wang, H., Sciavolino, P., Iler, N., Shen, M. M., and Abate-Shen, C. (1997). Heterodimerization of Msx and Dlx homeoproteins results in functional antagonism. Mol Cell Biol *17*, 2920-2932.

Zhao, G. Q., Zhao, S., Zhou, X., Eberspaecher, H., Solursh, M., and de Crombrugghe, B. (1994). rDlx, a novel distal-less-like homeoprotein is expressed in developing cartilages and discrete neuronal tissues. Dev Biol *164*, 37-51.

Zhou, Q. P., Le, T. N., Qiu, X., Spencer, V., de Melo, J., Du, G., Plews, M., Fonseca, M., Sun, J. M., Davie, J. R., and Eisenstat, D. D. (2004). Identification of a direct Dlx homeodomain target in the developing mouse forebrain and retina by optimization of chromatin immunoprecipitation. Nucleic Acids Res *32*, 884-892.