Identification and characterisation of novel transcripts involved in the proliferation, differentiation and developmental networks of the mouse cerebral cortex

King Hwa Ling

Bachelor of Science (Hons.) (Biomedical Sciences)

Master of Science (Genetics)

A thesis submitted for the Degree of Doctor of Philosophy

School of Medicine

(Discipline of Medicine)

Faculty of Health Sciences

University of Adelaide, South Australia.

June 2011

CHAPTER 1

Literature review

1.1 Introduction

Complex behavioural tasks, from perception of sensory input and the control of motor output to cognitive functions such as learning and memory are dependant on the meticulous innumerable interconnections of neuronal networks in the brain. These neuronal networks are precisely developed during embryonic growth under specific guidance of imprinted mechanisms, which are triggered spatiotemporally. Development of the central nervous system particularly the cerebral cortex (also known as cerebral corticogenesis) requires both intrinsic and extrinsic factors [reviewed in 1, 2, 3]. Intrinsic factors include cell surface and secreted molecules that control the fate of neighbouring cells as well as transcription factors that regulate the abundance of transcripts at the nucleic acid level. On the other hand, extrinsic factors including nutrients, secreted factors and sensory stimuli also interact with intrinsic factors for proper differentiation and maturation of nerve cells.

Complex regulatory elements as well as epigenetic factors are important intrinsic determinants of brain development. Molecular regulatory networks at chromosomal (chromosomal packaging or remodelling, histone acetylation and deacetylation) [5-7], DNA (transcription machinery and processes, chromatin insulation, DNA methylation) or RNA (mRNA processing and post-transcriptional regulation) [5, 6, 8, 9] sequences and protein (regulation of translation, protein modifications and ubiquitination pathway) [5, 10] levels play a pivotal role in determining when, where and how a multipotent progenitor cell should proliferate, differentiate, migrate and settle at a designated position in the cortex. These molecular regulatory networks are further complicated by various recently discovered alternative and genome-wide processes such as natural antisense transcripts (NATs) [11] and microRNA (miRNA) regulations [12], nuclear-specific paraspeckle proteins and nuclear RNA retention [13, 14]. Questions arise on what is actually triggering these regulatory networks and when would these networks be turned off? As embryonic corticogenesis lays the foundation for dedicated postnatal cognitive, learning and memory functions, any perturbed crucial molecular networks could cause various levels of detrimental effects postnatal such as mental retardation.

Consequently, it is important to investigate any molecular markers at various stages of brain development. The expression profile of these markers is an important determinant of their roles spatiotemporally, thus leading to the quest of hunting for common regulator(s) and the mechanisms involved, where these markers are peculiar to a specific developmental stage. Since the emergence of virtually complete human and mouse genome sequences, extensive molecular investigation has been carried out throughout brain development [15-19]. However, many questions still remain unanswered, especially on what intrinsic factors are involved and how they regulate the timing of cell proliferation, differentiation and guidance of neuronal migration. Information remains scarce on how cells know where and when to stop and aggregate and what specialised neurones they should differentiate into? What are the factors that guide axonal tract growth and how do they know where to stop and with what target they should form synapses? The following sections briefly review some of the fundamental events during the development of the brain with special focus on the cerebral cortex.

1.2 Development and anatomy of the brain

The total mouse gestation period is about 20 days. The cerebral cortex is developed as a part of the central nervous system starting on gestational day (also known as embryonic day, E) 10. Prior to the development of the central nervous system, the developing notochord (Figure 1.1) induces the development of the neuroectoderm, which later develops into the primitive neural tube (E7.5-8) [20, 21]. At this stage, a longitudinal groove on the neural plate deepens as the neural folds from both sides of the embryonic axis become elevated. Subsequently, these neural folds fuse together and transform the groove into a closed neural tube with broad cephalic and long caudal portions that eventually form the future brain and spinal cord, respectively (Figure 1.1).

At E8.5, the cephalic portion of the neural tube undergoes three dilatations giving rise to prosencephalon (primitive forebrain), mesencephalon (primitive midbrain) and rhombencephalon (primitive hindbrain) (Figure 1.2). Between E9-11, the neural tube undergoes further dilatations with the rostral part of the prosencephalon developing into two telencephalic hemispheres (future cerebrum), whereas the caudal part develops into the diencephalon (future thalamus and hypothalamus). The mesencephalon of the neural tube remains, but the rhombencephalon develops into two regions, namely, metencephalon (future cerebellum and pons) and myelencephalon (future medulla oblongata) (Figures 1.2 and 1.3). The dilatation processes transform the three-vesicle stage neural tube into a five-vesicle stage neural tube.

NOTE:
This figure is included on page 4 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.1 **Neural tube formation.**
1: Notochord, 2: Neural plate, 3: Ectoderm, 4: Neural grove, 5: Neural fold, 6: Neural tube, 7: Neurocoele, 8: Neural crest, 9: Anterior neuropore and 10: Posterior neuropore. The figure is modified from Cohen HS, 1999 [21].

NOTE:
This figure is included on page 5 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.2 **Neural tube at three- and five-vesicle stages.**
Transverse sections show dilatation of the cephalic portion of the neural tube at three-vesicle (left) (consisting of prosencephalon, mesencephalon and rhombencephalon) and five-vesicle (right) (consisting of telencephalon, diencephalon, mesencephalon, metencephalon and mylencephalon) stages. The figure is modified from Cohen HS, 1999 [21].

From E11 and onwards, the five-vesicle stage neural tube develops into various important functional regions of an adult brain [20, 21]. At E12.5, the telencephalon starts to develop into the olfactory cortex, hippocampus and cerebral cortex that are involved in olfaction, learning, memory, cognitive functions and motor-sensory responses (Figure 1.3). Between E12 and E15.5, the diencephalon gives rise to the thalamus, hypothalamus, caudate putamen, ganglionic eminence and neurohypophysis (part of pituitary gland), which are involved in various functions ranging from hormonal regulation, sleep-wake cycle or circadian rhythm modulation to refinement of movement. Both telencephalon and diencephalon derivatives form the forebrain whereas mesencephalon develops into the midbrain (a part of the brain stem). The midbrain functions as a motor or sensory nerve relay centre associated with vision, hearing, eye movement and body movement. Together with the midbrain in the brain stem are the pons and medulla oblongata. The pons is a derivative of metencephalon (at E11.5) and

contains nerve tracts conducting signals among the cerebrum, thalamus, cerebellum and medulla oblongata. The medulla oblongata is a derivative of the myelencephalon (at E12), which is the caudal-most structure of the brain located in between the pons and the spinal cord. It houses various control centres for vital functions such as breathing and heartbeat. At E12.5, the metencephalon gives rise to the cerebellum, which is an important centre for body movement coordination. Development of all brain structures and the communication among them are precisely coordinated within such a short timeframe. The underlying mechanisms governing such complex events are influenced by spatiotemporally expressed genes/proteins as well as interactions between various molecular networks, which require further in-depth investigations.

Both human and mouse brains undergo very similar developmental pathways and the corresponding timeline for the human brain in comparison to the mouse brain timeline is presented in Figure 1.3. Despite the similarity between the human and mouse brain, the surface anatomy of a human brain is very different from the mouse brain. The human brain has various gyri (ridges) and sulci (depressions or fissures) that increase the surface area as compared to the smooth surface of the mouse brain (Figure 1.4). The olfactory bulb, is another signatory difference between human and mouse brains. The murine olfactory bulb is located at the rostro-most region of the brain and is relatively large when compared to the cerebrum. In humans, the olfactory bulb is located at the inferior side (bottom) of the brain and is relatively small when compared to the cerebrum. Even the directional/anatomical planes or loci of the mouse brain differ from the human brain (Figure 1.4); therefore they may need to be perceived differently in different contexts.

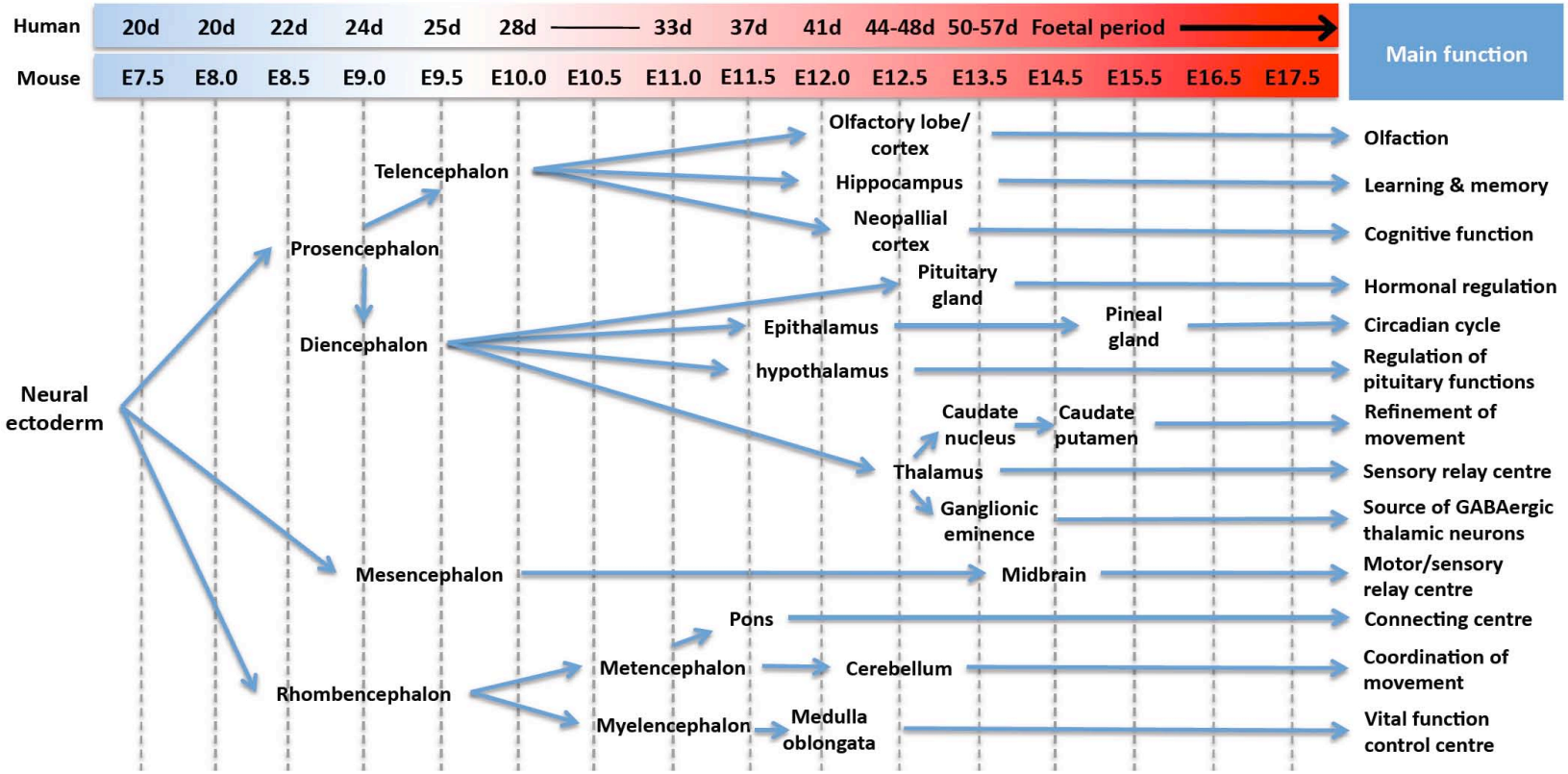


Figure 1.3 **Timeline for the brain development.**
The two panels at the top denote the matched developmental timeline for both the mouse (embryonic day, E) and human (day, d). The blue colour panel denotes the timeline for neural tube dilatation processes followed by development of specific brain regions (red panel). The blue arrows in the lower panel denote the development process and the labels at a specific timepoint denote the beginning of the brain region. The main role of each of the brain regions (but not limited to) is depicted in the right panel. This figure is modified from Kaufman and Bard, 1999 [4].

(A)

NOTE:
This figure is included on page 8 of the print copy of the thesis held in the University of Adelaide Library.

(B)

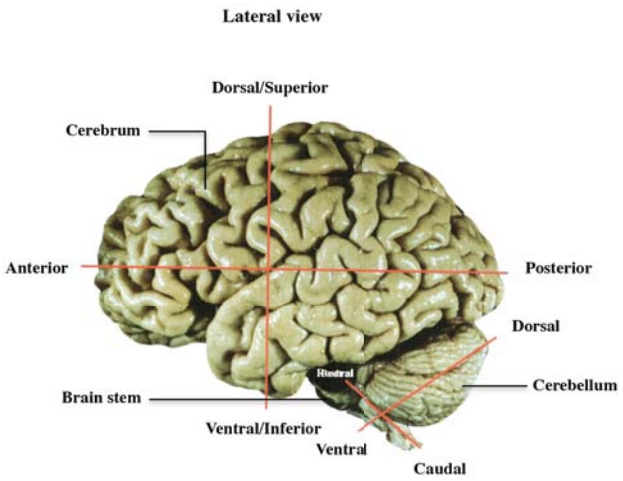


Figure 1.4 **Anatomy of the brain.**
(A) The dorsal (left) and lateral (right) views of the adult mouse brain show various brain regions. The block arrows show the rostral and caudal anatomical loci of the brain. These brain figures were obtained from The Mouse Brain Library website [22]. (B) The lateral view of the adult human brain (modified from Nadeau *et al*, 2004 [23]).

1.3 Cerebral corticogenesis

Cerebral corticogenesis involves a series of events that are spatiotemporally regulated. This includes 1) proliferation of multipotent progenitors; 2) neuronal or glial fate determination; 3) migration of postmitotic cells; 4) cells aggregation and cytodifferentiation or morphogenesis (formation of dendrites and axonal tracts); 5) gliogenesis and synaptogenesis and; 6) reorganisation, elimination and stabilisation of neuronal networks [21]. These events are meticulously implemented under the influence of various molecular factors at different levels of gene expression throughout cerebral corticogenesis.

1.3.1 Proliferation of multipotent progenitors

The cerebral cortex comprises of two main populations of neurones, namely glutamatergic- and excitatory-based projection neurones (pyramidal cells) and GABAergic- and inhibitory-based interneurones. The projection neurones are derived from the progenitor cells located in cortical ventricular zone (VZ) whereas most of the interneurones originate from progenitors located beyond cortical VZ especially the ventral telencephalon [24, 25], which migrated tangentially to reach their final position in the developing cortex. However, some interneurones are generated in the cortical VZ as reported in humans [26]. Other non-neuronal cortical components such as astrocytes are also derived from the cortical VZ progenitor cells, which are multipotent stem cells as demonstrated in retrovirus-based lineage studies and *in vitro* culture experiments [27-29]. On the other hand, non-neuronal oligodendrocytes are derived from the same stem cells located in the ventral telencephalon, which also give rise to the GABAergic interneurones [30, 31]. In contrast, many adult oligodendrocytes originate from cortical progenitors in the postnatal subventricular zone (SVZ) that retained proliferative capability as compared to the non-proliferative adult cortical VZ [32, 33].

Before the neuronal fate determination takes place, the multipotent progenitor cells initially divide symmetrically (non-terminal) to expand the pool of early multipotent stem cells, with the plane of cell division perpendicular to the ventricular surface. In this mode of cell division, two identical daughter cells are

generated [34]. This non-terminal symmetric cell division is important for self-renewal during the early expansion phase of neuronal development. Two other mode of divisions known as terminal symmetric cell division and asymmetric cell division further leads to cell differentiation and migration of neuronal cells out of the VZ. As neural induction endures, the nervous system becomes patterned along the rostro-caudal (RC) and dorso-ventral (DV) axes in response to gradients of signalling molecules from neighbouring tissues. Multipotent stem cells at the earliest stages of development will acquire positional identity first and express gene appropriate for their region of origin. This positional identity influences the type of neurones that arise from precursors in different parts of the nervous system; for example, the RC [35, 36] and DV positioning in the spinal cord [37] and the basal forebrain stem cells of the cerebral cortex [30].

Different modes of multipotent stem cells division are important to generate appropriate numbers of neurones in the correct spatial and temporal patterns. Therefore, it is critical to regulate the number of progenitor cells and this can be achieved by regulating the onset of differentiation, survival and/or proliferation of progenitors. Suppression of proneural factors that promote expression of neuronal differentiation factors, basic helix-loop-helix (bHLH) genes, can delay cell cycle exit, thus differentiation. Negative regulators that constrain bHLH factor expression through the inhibition of proneural gene expressions as demonstrated in lateral inhibition of Notch signalling pathway [38], prevent neighbouring cells from differentiating at the same time, thus preventing the depletion of progenitors for later-born cell types. A second class of negative regulators consists of inhibitors of DNA binding (Id) proteins that encode for antagonistic basic domain lacking HLH factors that form non-functional dimers with E proteins and prevent DNA binding. Both *Id1* and *Id3* mutants demonstrated premature neuronal differentiation of progenitors and cell cycle exit [39].

1.3.2 Differentiation of multipotent progenitors

Both terminal symmetric and asymmetric cell divisions give rise to postmitotic cells that are ready for differentiation. Postmitotic cells are capable of specialising into neurone or glial cells. bHLH genes, coupled with other cyclin-dependent kinase (*Cdk*) inhibitors promote cell cycle exit [40] followed by cell

differentiation [41]. Proneural proteins control the expression of these bHLH genes during neuronal fate determination.

Three proneural genes have been discovered in the mouse cerebral cortex, namely *Neurogenin1* (*Ngn1*), *Neurogenin2* (*Ngn2*) and *Mash1*. *Ngn2* appears to be the most important proneural gene among the three as only *Ngn2* mutants present a distinct corticogenesis defect [42, 43], and it is also involved in the regulation of *Ngn1* and *Mash1* in multipotent progenitors. *Ngn2* and *Mash1* double-mutant mice showed reduced corticogenesis due to the reduction of the size of the cortical plate as well as *NeuroD* expression, a marker for neuronal precursor cells. Other features include a reduced number of neurogenic progenitors due to affected *Hes5* in the Notch signalling pathway, resulting in compensatory increments of astrocytic progenitors or progenitors that remain bipotent. The findings from *Ngn2* and *Mash1* double mutant mice studies demonstrate the crucial role of proneural proteins in neural fate determination as well as the astrocytic fate through the pertubation of Notch signalling pathway in multipotent progenitors [43].

In contrast to *Ngn2, Ngn1* promotes neurogenesis and inhibit astrogenesis through distinct mechanisms [44]. *Ngn1* promotes neurogenesis through conventional intact DNA binding activation of the neuronal differentiation gene, *NeuroD*, whereas promoting antiastrogliogenic activity by complex intracellular interactions with components of several signalling pathways, which require no DNA binding activities. *Ngn1* has been shown to block the activity of *Smad1* downstream of BMP in the BMP-Smad signalling pathway, thus interfering with the activation of STAT1 and STAT3 by LIF/CNTF in the LIF-CNTF-JAK-STAT signalling pathway [44]. Both the BMP-Smad1 and LIF-CNTF-JAK-STAT signalling pathways are vital for astrocytes differentiation.

The generic programme of neuronal differentiation (the programme regulating the acquisition of features common to all neurones; axonal and dendrite growth) as regulated by bHLH genes or other neurone differentiation genes is understudied. The best candidate neuronal differentiation gene, which also belongs to the bHLH class, is *NeuroD1* [45]. *NeuroD1* is sequentially expressed as neurones differentiate [46]. Other bHLH genes, that are sequentially expressed in

differentiating cortical neurones and are probably involved in their differentiation, include *NeuroD2*, *Math2*, *Math3* and *Nscl1* [47]. Other non-bHLH genes such as T-box genes, *Tbr1* and *Tbr2* [48, 49], also have the similar expression profiles and depend on *Ngn1* and *Ngn2* for their normal cortical expression [47]. However, all the bHLH and *Tbr* differentiation genes regulated by both *Ngn1* and *Ngn2* in the cortex are confined to cortical progenitors and neurones in the embryonic telencephalon and are not expressed by subcortical progenitors or neurones [47]. The definitive roles of these candidate genes remain unknown and evidence of molecular interactions or networks upstream and downstream of these genes are scarce and therefore warrant further investigations using global approaches such as large scale gene expression profiling.

1.3.3 Neuronal migration

After neuronal fate determination, differentiated neuronal cells migrate out of VZ towards the pial surface, the outermost layer of the cerebral cortex. The neuronal cells in the cerebral cortex adopt two types of migration, radial and tangential migration. There are two modes of radial migration; the glial independent nuclear/somal translocation and glial-dependent locomotion [50]. In a developing cerebral cortex of mouse, the first wave of postmitotic cells will hook their leading processes onto the pial surface at E11 and migrate by nuclear/somal translocation radially towards the dorsal region of the cortex. This migration mode causes the shortening of leading processes as the cells migrate towards the pial surface (Figure 1.5). The migration of the first postmitotic cells form the preplate, which is then split into a marginal zone (MZ) and subplate (SP) by the second postmitotic wave of cell migration (E13) via locomotion mode. The layer in between MZ and SP is known as cortical plate (CP). As the cerebral cortex widens dorsally, the later derived postmitotic cells migrate in an inside-out manner within the CP with assistance of glial cells (E14-E18) [51, 52]. These cells migrate past their predecessors to settle underneath the MZ. When these cells reach near to the pial surface, they may switch their mode of migration into nuclear/somal translocation. The migration processes peak between E15 and E16 and give rise to a six-layered cerebral cortex. The SP layer in the brain degenerates after birth.

NOTE:
This figure is included on page 13 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.5 **The development and organisation of the embryonic and adult neocortices into distinct neuronal layers.**
CP = cortical plates; E = embryonic day development; IZ = intermediate zone; MZ = marginal zone; PP = preplate; PS = pial surface; SP = subplate; vertical bars = glial cells; VZ = ventricular zone. The numbers in the adult stage cortex panel denote the distinct layers of neuronal organisation. This figure is modified from Gupta *et al*, 2002 [50].

There are also neuronal cells, which are tangentially migrated from outside the cerebral cortex during cortical development. These neuronal cells are mainly derived from the lateral ganglionic eminence (LGE) and medial ganglionic eminence (MGE) of the ventral forebrain (Figure 1.6). These tangentially migrated cells mainly form the interneurones of the developing cortex at the intermediate zone (IZ) and MZ. The tangential migration can be influenced by chemoattractants in the VZ as well as repellent signals expressed in the CP [51] before acquiring positional information that results in further migration into appropriate layers of the developing cortex.

Molecular guidance of neuronal migration is also understudied. Mechanisms intrinsic to the migratory process related to the adhesion properties may dictate the cessation of migration and, therefore, establish the definitive location of the different neuroblasts [53]. To date, only a few genes have been extensively explored and shown to play a pivotal role in cytoskeleton organisation during neuronal migration. These well-described genes are *Lis1* [54, 55], *Dcx* [56, 57] and *Dab1* [58, 59]. The former two are involved in microtubule polymerisation while the latter is associated with the Reelin signalling pathway.

NOTE:
This figure is included on page 14 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.6 **Tangential migration of interneurones.**
Interneurones arise in the ganglionic eminence (GE) and migrate into the neocortex through the intermediate zone (blue cells) and the marginal zone (pink cells). CP = cortical plate. VZ = ventricular zone. This figure is modified from Nadarajah & Parnavelas, 2002 [51].

1.3.4 Cell aggregation, differentiation, axonogenesis and synaptogenesis

Before neuronal cells undergo morphogenesis, they have to know when to stop migrating and with what other cells they have to aggregate together. Little is known on how neuronal cells are aggregated into nuclear or laminar arrangements. It has been postulated that cells featuring similar functions will aggregate together. This selective aggregation has been demonstrated in *in vitro* experiments using cells from different regions of the brain that are dissociated together and when mixed, preferentially re-aggregate with cells from the same origin [60]. The common recognition mechanism between cells as well as migration termination signals is yet to be determined. Once neuronal cells settle at their final destination, they start to differentiate morphologically, which is highly

dependent on the neuronal polarity (direction of impulse transmission). Both dendrites (postsynaptic end) and axons (presynaptic end) differ at ultrastructural level with ribosomes and Golgi elements preferentially present in dendrites but not the axon. Others biochemical differences between the two regions also further distinguish one from another [61]. The intrinsic factors that direct compartmentalisation of organelles are not known and microtubule organisation in neuronal processes has been speculated to account for the attributes.

Both dendrites and axons navigate to their targets through the growth cone (enlarged tips of growing nerve extensions). Various factors mediated growth cone guidance. This includes cell adhesion molecules (N-CAM, L1 and Fasciclins), extracellular matrix molecules and related receptors (collagens, laminins or integrins), tyrosine kinase receptors and related ligands (neurotrophins, Eph-receptors and ephrins) [62], netrins, semaphorins and related receptors [63-65]. The growth cone navigation and guidance is a prime event prior to reaching any target neurones or tissues before the commencement of synaptogenesis.

Three main mechanisms that complete the synapse formation are the induction, assembly and maturation of synapse. An important aspect of synaptogenesis induction is target recognition. This includes the ability of axons from different parts of the brain to grow into their respective target fields and synapse with the correct cell type where functional presynaptic boutons are formed. Secreted proteins or molecules demonstrated to have synaptogenic activity include neuronal activity-regulated pentraxin (Narp) that not only localises to synapses, but also binds to the extracellular domains of subunits of the α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-type glutamate receptor [66]. Another molecule with synaptogenic activity is EphrinB, which also plays an important role during axonal growth cone guidance. This molecule promotes the clustering of subunits of N-methyl-D-aspartate (NMDA) type of glutamate receptor [67]. The EphrinB-mediated aggregation of EphB receptors leads to coaggregation of NMDA receptors during dendritic spine development [68] and maturation [69]. In contrast to the limited activities of Narp and Ephrin, other secreted proteins such as Wnts [70, 71], Fgfs [72], SynCAM [73, 74] and neuroligin (cell-adhesion proteins) [75] induce presynaptic differentiation and

result in more complete functional differentiation of the presynaptic active zone including regional axon and dendrite arborisation.

Subsequent to induction is the molecular assembly of the synaptic junction for delivery of pre- and postsynaptic components. In presynaptic assembly, the vesicular delivery of proteins plays a critical role. These include small, clear-centred vesicles, tubulovesicular structures and 80-nm dense core vesicles. The exact compositions within these morphologically distinct vesicle types are not known but they appear to be somewhat specified with different delivery priori to the plasma membrane [76-78]. Contrarily, assembly in postsynaptic appears to occur primarily by gradual accumulation of molecules [78, 79] such as calcium calmodulin-dependent protein kinase II (CaMKII) and scaffolding proteins of the PSD-95 family followed by recruitment of NMDA-type and AMPA-type glutamate receptors which are independently regulated by interacting proteins [63].

Formed synapses will expand in size (e.g. the bouton volume, number of synaptic vesicles, docked vesicles, active zone area, postsynaptic density area and spine head volume) and mature. These synaptic connections are subjected to rearrangement through either decrease of convergence onto target, segregation of inputs across targets or complete loss of projection to a target. These processes are believed to improve the overall effectiveness of synaptic networks. Intracellular signalling pathways are sensitive to the activity states of synapses, including ubiquitin-mediated degradation, not only to regulate the turnover of synaptic components but also to promote synapse elimination. However, synapse elimination is not a universal feature of neural development as it has not been observed in all systems and regions investigated [60].

1.4 Central dogma of molecular biology

Gene-to-protein expression involves different inter-connected hierarchical levels of regulation that govern the course of central dogma of molecular biology; from DNA to RNA to protein (Figure 1.7). The genomic sequence, chromatin structure and packaging, and RNA transcripts affect the way a gene is expressed. Regulation at the sequence level includes the recruitment and binding of transcriptional machineries to the promoter region and the initiation of transcription process whereas the packaging and remodelling of chromatin can affect the accessibility of these transcription factors to the promoter region. After transcription, mRNAs are processed into mature transcripts before being transported into the cytoplasm for translation. The translocation of transcribed mRNAs from the nucleus to the cytoplasm as well as the half-life of these mRNAs can affect the level of translation and eventually the level of functional protein in the cell.

Since early 2000s, the information flow within the central dogma of molecular biology is professed as a more complex event than what was first perceived. Due to the growing importance of a new class of RNA known as noncoding RNA (ncRNA), it has been instituted as one of the most important elements within the central dogma. NcRNAs interact with the DNA, mRNA and protein in both reversible as well as irreversible manners to ultimately affect the level of targeted mRNAs or proteins in the cell (Figure 1.7). From the moment transcriptional machineries assemble themselves at promoter regions until the translation of mRNAs into proteins, all the involved processes are meticulously orchestrated partly by ncRNAs to ensure appropriate cellular spatial and temporal expression of functional proteins and RNAs.

NOTE:
This figure is included on page 18 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.7 **Central dogma of molecular biology.**
(1) Transcription process: DNA information is copied into RNA. (2) RNA processing: splicing of introns, 5' capping and 3' polyadenylation of a pre-mRNA into a mature form mRNA (intronless). (3) mRNA translation: translating mRNA information to synthesise polypeptides that eventually form proteins. (4) DNA replication: DNA information is copied to ensure inheritance of genetic information. Information flow from (1) to (3) requires inter-connected regulation of various transcription and translation factors, which (5) ncRNAs play an important regulatory role (at DNA, RNA and protein levels) throughout the course of central dogma of molecular biology. This figure is modified from http://www.emc.maricopa.edu/ [80].

1.4.1 Transcription factors

Transcription factors are proteins with the capability of recognising and binding to specific DNA sequences within the genome. Upon binding to the specific DNA sequences generally known as promoter or enhancer regions, these factors either promote or block the recruitment of RNA polymerase to a TATA box or GC-rich region located upstream of a gene, thus activating or suppressing the transcription of DNA into mRNA or non-coding RNA. All transcription factors have one or more DNA binding domain(s) and usually require cofactors during regulation of gene expression.

Depending on the spatiotemporal development or functional state of cells, the transcription process is further fine-tuned by recruiting specific transcription activators to distal control element(s) known as enhancer regions (Figure 1.8). Many regulatory regions also employ a DNA-bending protein to facilitate DNA looping and hence interactions between transcription activators and the basal transcription machinery at the promoter region. The basal transcription machinery consists of general transcription factors and RNA polymerase II. The interactions between activators and the basal transcription machinery subsequently form the preinitiation complexes at the promoter region. The basal transcription machinery contains TATA-binding proteins, which recognise and bind the TATA box, thus positioning RNA polymerase II at the transcription start site and 'kick-off' the transcription process to synthesise RNA. It should be noted that the majority of promoters do not contain TATA boxes, but rather GC-rich regions, but still use the TATA-binding protein to aid in transcription without it binding to a TATA box.

Examples of transcription factors that are expressed in the brain and regulate metazoan brain development and functions are discussed in Section 1.3. Some transcription factors are expressed in restricted areas within the neocortex during development such as Emx2, Pax6, Coup-tf1, Fgfs and Bmps [81, reviewed in 82, reviewed in 83]. These transcription factors play a crucial role in brain arealisation during development leading to different functional specialisation within the postnatal or adult brains. On the other hand, Sox4 and Sox11 that are involved in pan-neuronal gene expression regulation and glial cell differentiation have a more generalised expression during central nervous system development [84-86]. These transcription factors are found only in stipulated developmental stages of the brain and their onset or termination of expression are controlled by other regulatory networks consisting mainly of transcription factors [87].

NOTE:
This figure is included on page 20 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.8 **Initiation of the transcription.**
Initiation of the transcription process starts with the assembly of preinitiation complexes. The complexes consist of transcription activators that recognise an enhancer region (A), general or basal transcription factors that recognise the TATA box (or GC-rich regions), and the RNA polymerase II (B). The interactions between transcription activators, basal transcription factors and RNA polymerase II are mediated by a group of mediator proteins. When the DNA loops out to facilitate the assembly of the basal transcription machinery, basal transcription factors recognise and bind to the TATA box (or GC-rich regions) and position RNA polymerase II at the transcription start site to initiate RNA synthesis (C). This figure is modified from http://www.smc.edu/ [88].

1.4.2 Epigenetics

The term epigenetic refers to the study of phenotypic changes of cells or tissues due to stable alteration of gene expression profiles without affecting the DNA sequence within the genome [89]. The changes can be passed on indefinitely in subsequent cell divisions or only for a limited number of generations. Developmental processes, environmental chemicals, ageing and diet can affect epigenetic mechanisms. Epigenetic changes affect the expression of specific genes and enable a group of homogenous totipotent stem cells to differentiate into

pluripotent stem cells and subsequently into a heterogenous and complex organism with different cell types and functions [90]. These processes require epigenetics reprogramming, which is observed throughout embryogenesis [7, 91]. Epigenetics changes play a critical role during embryonic development of the brain as well as being responsible for the dynamics of brain function and plasticity after birth [19, reviewed in 92].

The mechanism of epigenetics reprogramming is usually mediated by DNA methylation or histone modifications. DNA methylation involves the transfer of a methyl group (an epigenetic factor found in dietary sources) to DNA at CpG dinucleotide sites (or CpG islands) to convert cytosine to 5-methylcytosine [reviewed in 93]. DNA methylation of cytosine within CpG islands near to promoters of genes may lead to inaccessibility of transcription complexes and therefore negatively regulate gene expression. It has been proposed that this mechanism plays a critical role in silencing unwanted elements especially transposable elements found throughout the genome thus preventing insertional inactivation or ectopic activation of genes due to transposition of mobile elements [94, 95]. In addition, epigenetic mechanisms are also mediated by histone modifications. DNA wraps around histone proteins to form a more compacted structure known as chromatin. These histone proteins contain flexible and charged tails, which are subjected to posttranslational modifications via methylation, acetylation, phosphorylation, ubiquitination or ADP-ribosylation [reviewed in 5]. Postranslational modifications to histone proteins not only alter the extent to which DNA is wrapped, but can also cause a region of chromatin to undergo nuclear compartmentalisation, hence affecting the accessibility of transcription factors and other regulators to the DNA. Collectively, both DNA methylation and histone modifications cause chromatin remodeling and provide another level of gene expression regulation within the nuclear compartment.

1.4.3 Noncoding RNAs

The human genome comprises of ~3 billion bases with less than 2% of them are genes encode for ~20,000 proteins. Until the last decade or so, the non-gene sequences were considered ‘junk’ or bearing no significant information pertaining to the development, biological function or complexity of human beings. A few years ago, the ENCyclopedia Of DNA Elements (ENCODE) project jointly carried out by hundreds of researchers worldwide has revealed that as high as 90% of the re-analysed 30 million bases (1%) of the human genome sequences were actually transcribed. The landmark finding of the study suggested the previously termed ‘junk DNA’ has much more to offer, predominantly in its non-protein coding RNA (also known as noncoding RNAs/ncRNAs). These noncoding RNAs have very different genomic and transcript features as compared to protein-coding RNAs or mRNAs. These noncoding RNAs are generally categorised into two large groups according to their size. Noncoding RNAs with 200nt or less are known as ‘small RNAs’ whereas any noncoding RNAs greater than 200nt are known as ‘long noncoding RNAs’ (lncRNAs). Table 1.1 shows the type of small RNAs and lncRNAs and their general role in the cell.

NcRNAs are abundantly expressed in the nucleus and cytoplasm of prokaryotic and eukaryotic cells. Their functional role, however, remains under-characterised. Only a few classes of well-studied ncRNAs such as miRNAs, siRNAs and piRNAs were reported to have spatiotemporal expressions in the mammalian system [96-99]. In this study, only miRNAs, siRNAs and lncRNAs are discussed.

Table 1.1: Type of noncoding RNAs.

Noncoding RNAs	Size	Function/Descriptions	References
microRNA (miRNA)	20-25nt	Causes translational repression or mRNA degradation via RNA interference pathway.	[12, 100]
endogenous small interference RNA (endo-siRNA)	20-25nt	Commonly found in plants and involved in transcriptional gene silencing and causes mRNA degradation.	[101, 102]
PIWI-interacting RNA (piRNA)	26-31nt	Transcriptional gene silencing of retrotransposons in germ line cells.	[96, 103, 104]
transcription initiation RNA (tiRNA)	~18nt	Unknown function but targets the promoter region or transcription start site vicinity.	[105]
microRNA-offset RNA (moRNA)	19-20nt	Derived from either end of pre-miRNA with unknown function.	[106, 107]
splice-site RNA (spliRNA)	17-18nt	3' termini of spliRNAs are mapped precisely to splice donor sites and are believed to promote RNA polymerase II pausing and backtracking.	[106]
telomere specific small RNA (tel-sRNA)	~24nt	Dicer-independent, and 2′-*O*-methylated at the 3′ terminus sRNA with specificity toward telomere G-rich strand.	[108]
trans-acting small interfering RNA (tasiRNA)	21-24nt	Represses gene expression via post-transcriptional gene silencing in plants.	[109]
repeat associated small interfering RNA (rasiRNA)	24-29nt	Involved in the establishment and maintenance of heterochromatin structure and silencing of transposon and retrotransposon elements.	[110, 111]
cis-acting small interfering RNA (casiRNA)	24nt	Involved in chromatin modifications in plants.	[112, 113]
small-scan RNA (scnRNA)	~28nt	Involved in histone methylation leading to DNA elimination during cell conjugation in *Tetrahymena thermophila*.	[114-116]
small modulatory RNA (smRNA)	~20nt	Modulates transcription possibly via direct interaction with transcription factors.	[117, 118]
tiny non-coding RNA (tncRNA)	~21nt	Very similar to tasiRNA with unknown biogenesis and function.	[118, 119]
centromere repeat associated short interacting RNAs (crasiRNA)	34-42nt	Associated with the formation and maintenance of centromeric heterochromatin in vertebrates.	[120]
long noncoding RNA (lncRNA)	>200nt	Has a very diverse role in regulating gene expression at genomic, transcription, post-transcription and translation levels during cellular development and disease progression.	[121-123]
cis natural antisense transcript (*cis*-NAT)	>200nt	Overlapping noncoding RNA, which is transcribed from the same genomic locus together with the sense transcript.	[121-123]
trans natural antisense transcript (*trans*-NAT)	>200nt	Overlapping noncoding RNA, which is transcribed from a genomic locus different from the genomic locus of the sense transcript.	[121-123]

1.4.3.1 MicroRNAs, biogenesis and mechanism of action

MicroRNA (miRNA) is a small RNA with ~22nt in length and play an important role in repressing cellular protein translation and transcription processes [reviewed in 12, reviewed in 118]. The first miRNA was discovered in 1993 known as *lin*-4 RNA, which complements the 3' UTRs of *lin*-14 and *lin*-28 transcripts and negatively regulate the level of LIN-14 and LIN-28 proteins, thus affecting the developmental timing in *Caenorhabditis elegans* [124, 125]. The second miRNA, *let-7*, was discovered seven years later [126] involved in the regulation of *lin-41* and *hbl-1* genes, which are important for the transitional development of *C. elegans* larvae [127, 128]. The significance of miRNAs in organismal development is gaining more interest from scientists of various backgrounds and subsequently share the 'fame' of another small RNA class known as small interfering RNAs (siRNAs), which involve in RNA interfering (RNAi) pathway or related mechanisms in animals and plants [129, 130]. Since then, a huge amount of miRNAs have been discovered using high-throughput sequencing and bioinformatics approaches in arthropods, nematodes, animals and plants [reviewed in 12, 131, 132].

The biogenesis of miRNAs begins with the transcription of miRNA genes by RNA polymerase II into primary form miRNA (pri-miRNA) (Figure 1.9) [reviewed in 133]. Any secondary structures within the pri-miRNA that resemble a hairpin is first recognised by Pasha (also known as DGCR8) and then cropped by an RNase type III enzyme, Drosha, into a free form precursor miRNA (pre-miRNA) [reviewed in 12, reviewed in 133, 134]. Pre-miRNA is transported into the cytoplasm via exportin-5 (Exp5) nuclear export receptor. In the cytoplasm, the pre-miRNA is diced by another type III RNase, Dicer, into ~22bp miRNA duplex containing the mature form miRNA in one strand and a passenger strand of no clear function in the other. The mature miRNA will serve as a guide for Argonaute 2 catalytic protein within the RNA-induced silencing complex (RISC) to target mRNA and initiate the RNA interfering mechanism whereas the passenger strand is commonly rapidly degraded. The degree of sequence complementary between the miRNA and the mRNA target site determine the type of molecular action performed by the RISC machinery; either repress translation or degrade the targeted mRNA (Figure 1.10).

NOTE:
This figure is included on page 25 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.9 **Biogenesis of miRNAs.**
The production of a miRNA involves the processing of primary-miRNA (pri-miRNA) into the precursor (pre-miRNA) and mature forms miRNA. Artificial siRNA can enter the system in the form of siRNA duplex (equivalent to miRNA duplex) or as artificial pri-miRNAs (shRNA-mir). Only the main factors involved in the pathway are shown. Pol II = RNA polymerase II, Exp5 = Exportin-5, Ago2 = Argonaute 2 and RISC = RNA-induced silencing complex. This figure is adapted from Cullen BR, 2005 [135].

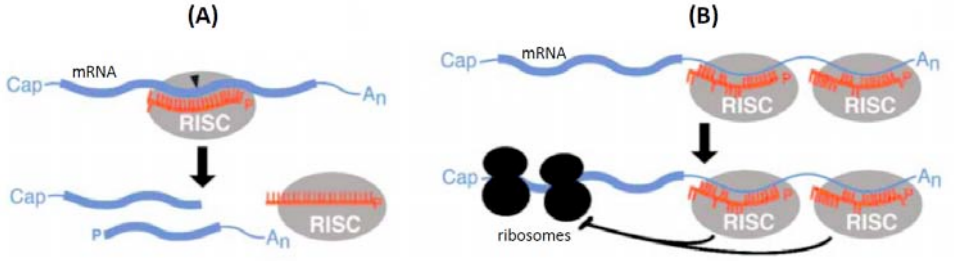


Figure 1.10 **The action of miRNAs.**
(A) Extensive complementary between a miRNA and its target site within the coding or untranslated regions of an mRNA will normally lead to mRNA degradation. (B) Partial or short complementary segments between miRNAs and their target sites in the 3' UTR causes translational repression. This figure is modified from Bartel DP, 2009 [100].

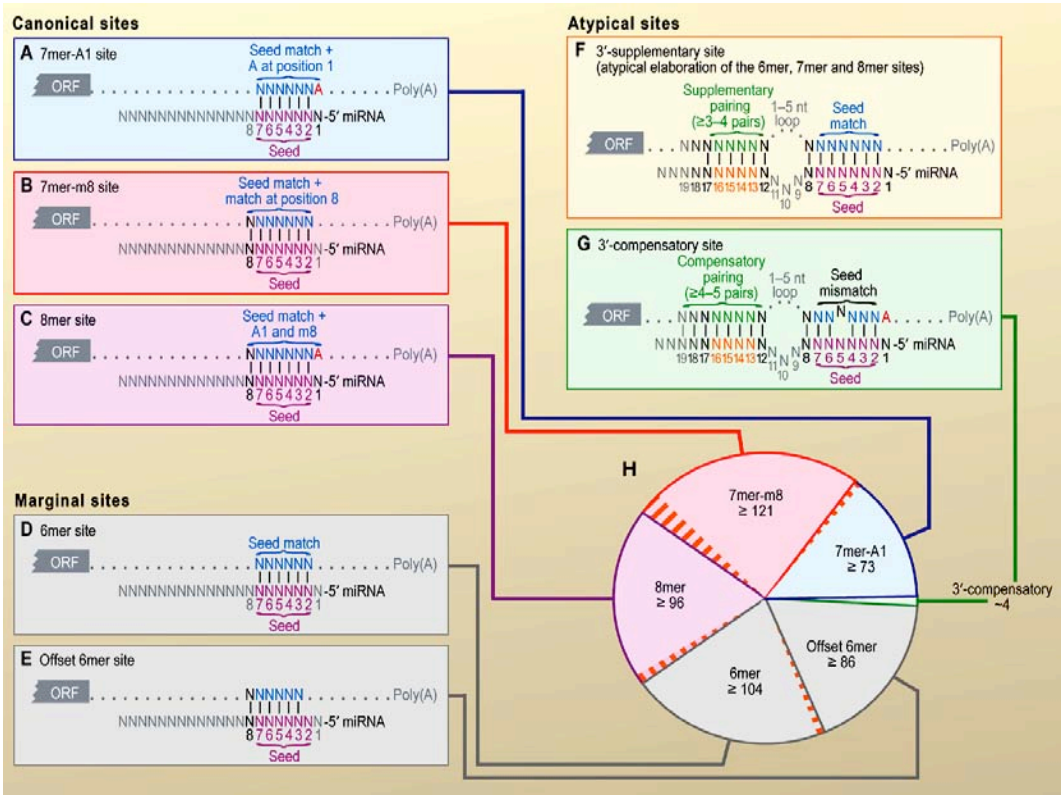


Figure 1.11 **The different types of miRNA binding site.**
(A-C) Canonical binding sites with 7-8mer of Watson-Crick pairing at the miRNA seed region. (D-E) Marginal binding sites with 6mer of Watson-Crick pairing at the miRNA seed region. (F-G) Atypical binding sites for miRNA involving seed regions as well as supplementary or compensatory binding of the sequences at the 3' end. (H) Number of preferentially conserved mammalian sites matching a highly conserved miRNA [136]. The orange-hatched subsectors indicate the fraction of conserved sites with preferential conserved 3'-supplementary pairing. This figure is adapted from Bartel DP, 2009 [100].

The first 8nt of the 5' end of the miRNA is defined as the seed region. This region plays a crucial role in establishing the binding site within the targeted mRNAs. The rest of the miRNA sequences toward the 3' end will provide supplementary or compensatory binding and determine the course of actions carried out by the RISC machinery. The binding site is generally categorised into canonical, marginal and atypical sites featuring the contiguous Watson-Crick pairing of 6-8nt (also known as 6-8mer pairing) of the seed region (Figures 1.11 A-E) [reviewed in 100]. Both pairings at the seed as well as at the supplementary regions will influence the efficacy of the pairing thus creating a functional site for initiating the RNA interference mechanism. At the seed region, a 7-8mer pairing site generally has a higher efficacy as compared to 6mer pairing. A supplementary

pairing of 3 or more pairs at the 3' end will usually increase the efficacy of the canonical or marginal binding at the seed region whereas the efficacy of a 6mer pairing with a mismatch at the seed region can be improved by contiguous compensatory pairing of 4 or more pairs at the 3' end [reviewed in 100, 137, 138]. When an extensive pairing occurs between a miRNA and its target mRNA, RISC machinery generally degrade the mRNA leading to no translation at all (Figure 1.10). On the other hand, partial pairing between them leads to translational repression. Mechanisms underlying these choices of RISC activities, however, remain unknown and could depend on the physical characteristics of the miRNA-target site pairing lies within the catalytic site of Argonaute 2 [reviewed in 100, 139].

MiRNA expression has been implicated in various organisms and is commonly regulated in a spatiotemporal format. MiRNAs play an important role in the development of the central nervous system and they are localised at specific regions or cells within the brain. For example, *Mir134* is localised to the synaptodendritic compartment of rat hippocampal neurones and has been associated with synaptic development, maturation or plasticity [140], *Mir9* regulates the patterning and neurogenesis at the midbrain-hindbrain boundary in zebrafish [141] and *Mir124* triggers alternative pre-mRNA splicing in neuronal cells leading to neurodifferentiation in the mouse [142]. In addition, some miRNAs are associated with neurological disorders such as schizophrenia [143], Hungtington's disease [144], glioblastoma [145] and Fragile X mental retardation [146]. Identification of various miRNAs in the central nervous system and their association to neurological disorders suggest that this class of small RNAs are crucial in the regulation of cellular development and function in the brain.

1.4.3.2 Endogenous small interfering RNAs

Both miRNAs and endogenous small interfering RNAs (endo-siRNAs) share a similar central biogenesis mechanism and are not distinguishable in terms of their chemical compositions as well as their mechanism of action. However, the most distinct difference between the two lies within their very similar biogenesis. Endo-siRNAs are produced when Dicer performs multiple cleavages on long double stranded RNAs (dsRNAs) to give rise to 20-25nt endo-siRNA duplexes in

the cytoplasm (Figure 1.12). Comparatively to miRNA, the endo-siRNA biogenesis does not require the formation of a hairpin/precursor structure as a result of Drosha and Pasha activities. Other distinguishable properties between miRNAs and endo-siRNAs include their origin within the genome, the number of small RNAs produced from each primary transcript/dsRNA pair, sequence conservation in related organisms and their mode of transcriptional/translational silencing (Table 1.2) [reviewed in 12].

Endo-siRNAs are commonly found in plants [147, reviewed in 148] and viruses [reviewed in 149] due to the present of dsRNAs. Both plants and viruses can synthesise dsRNAs via the activity of RNA-dependent RNA polymerase (RdRP) enzyme. The enzyme is, however, not found in the mammalian system. Therefore, the origins of mammalian endo-siRNAs have always been originated from the exogenous dsRNAs or transcripts with naturally occurring dsRNA regions within the cell such as mRNAs [102], sense-antisense transcript pairs, transposons and repetitive elements with inverted repeat sequences [reviewed in 12]. To date, mammalian endo-siRNAs have been discovered mainly in the germ cells [102, 150] and embryonic stem cells [97]. Their role in the mammalian cells remains undercharacterised and is believed to play a part in the silencing of retrotransposons. However, their role is better defined in plants, fungi and ciliates where endo-siRNAs are not only involved in translational repression within the cytoplasm but also associated with DNA methylation [151], heterochromatin formation [152, 153] and DNA rearrangement [154], respectively. Endo-siRNAs within the mammalian cells may have similar mechanism of actions that are yet to be confirmed.

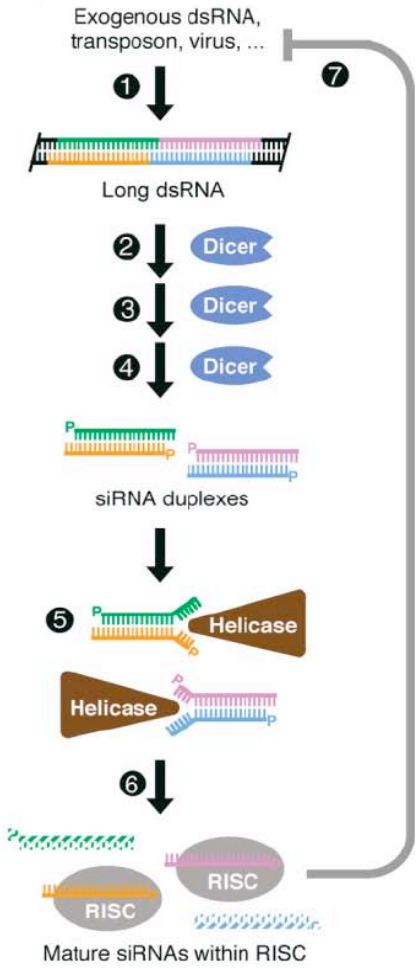


Figure 1.12 **The biogenesis of endo-siRNAs.**
(1) Endo-siRNAs often derive from mRNAs, exogenous dsRNA, transposons and viruses and heterochromatic DNA. (2)-(4) Primary transcripts for endo-siRNAs, normally in long dsRNA pairs, are subjected to multiple cleavages by Dicer in the cytoplasm to form multiple siRNA duplexes. (5) Helicase enzyme unwound the endo-siRNA duplex into single-stranded endo-siRNA, known as the mature form endo-siRNA. (6) Incorporation of mature endo-siRNAs into ribonucleoprotein complex known as RNA-induced silencing complex (RISC), thus initiate the RNA interference pathway. This figure is modified from Bartel DP, 2004 [12].

Table 1.2: Differences between miRNAs and endo-siRNAs [reviewed in 12].

Properties	miRNAs	endo-siRNAs
Origin in the genome	Derived from genomic loci, which is different from the protein-coding gene such as intergenic regions.	Often derived from mRNAs, exogenous dsRNA, transposons and viruses and heterochromatic DNA.
Biogenesis	Involved an intermediary precursor structures resemble a hairpin. The precursor structure is produced from the pri-miRNA transcript as a result of Drosha and Pasha activities within the nucleus.	Do not require intermediary structures. Require dsRNA pairs as templates and Dicer to cleave them into endo-siRNA duplexes in the cytoplasm.
Number of mature sRNA produced	Each pre-miRNA generally produces one mature miRNA.	One long dsRNA pair template produces many mature endo-siRNAs.
Sequence conservation	Generally conserved in related organisms.	Rarely conserved in related organisms.
Mode of interference	Initiate 'hetero-silencing' mode. A miRNA silences the gene/mRNA that is very different from the gene/transcript of its own origin. Often leads to translational repression.	Initiate 'auto-silencing' mode. An endo-siRNA silences the opposite strand of a dsRNA pairs from which it is derived due to high sequence complementary. Often leads to transcript degradation.

1.4.3.3 Long noncoding RNAs

Long noncoding RNAs (lncRNAs) were first reported in mice based on the analysis of full-length cDNA libraries [155]. LncRNAs are generally greater than 200nt and apparently have little or no protein-coding capacity. Majority of lncRNAs are 5' capped, 3' polyadenylated and spliced in a manner that is similar to protein-coding transcripts [155-157]. The noncoding transcripts have low sequence and genomic organisation conservation patterns as compared to protein-coding genes [156, 157] indicating that lncRNAs are under greater selective pressure to undergo rapid evolution [158]. In addition, lncRNAs are localised in both the nucleus and cytoplasm, and are spatiotemporally expressed in a very low level in developing cells, tissues and organs, suggesting that they have functions affecting a wide spectrum of processes [159, 160].

A large proportion of lncRNAs complement other transcripts and form dsRNAs. They are known as natural antisense transcripts (NATs). There are two types of NATs, namely *cis* and *trans*, which are categorised based on their origins within the genome (Figure 1.13). *Cis*-NATs complement their sense transcripts (usually mRNAs), which are transcribed from the same genomic locus (often in the opposite strand) for the NATs. On the other hand, *trans*-NATs and their sense transcript counterparts are transcribed from two different loci within the genome.

NOTE:
This figure is included on page 31 of the print copy of the thesis held in the University of Adelaide Library.

Figure 1.13 Natural antisense transcripts (NATs).
Gene structures for sense and antisense transcripts, and overlapping between sense and (A) *cis*-NATs or (B) *trans*-NATs. This figure is modified from Okamura & Lai, 2008 [161].

Approximately 8-20% of well-defined human protein-coding genes have at least one overlapping NAT [121, 162]. These overlapping NATs especially the *cis*-NATs form a part of the regulatory system for protein-coding gene expression. Both sense and overlapping NATs are regulated by independent promoters and are often simultaneously expressed [123, 163]. NATs are normally expressed in between <1 to 10 copies per cell and their origins vary from the intergenic regions to intronic or exonic regions of known genes (Figure 1.13) [164]. Due to their low expression level and unclear sequence or genomic organisation conservation, the biological function of NATs remains controversial. Differential regulation of

NAT expression has been associated with various human disorders ranging from breast [165], colon [166] and renal [165] cancers to follicular B-cell lymphoma [167], α-thalassemia [168], Beckwith-Wiedermann syndrome [169], Alzheimer's disease [170] and Fragile X-associated tremor and ataxia syndrome [171]. Although these NATs are related to the up- or downregulation as well as DNA methylation of their sense transcript/genomic locus counterparts, the underlying mechanisms that lead to these observations remain unknown.

To date, only a limited number of NATs were characterised and have been shown to interact with the nuclear genome, proteins and mRNAs in the cell. For examples, embryonic ventral forebrain-2 (*Evf-2*), regulates gene expression of distal-less homeobox 2 (*Dlx2*) through transcriptional activation [172] whereas hox transcript antisense RNA **(***HOTAIR*) from the homeobox C (*HoxC*) locus represses transcription, in *trans*, across 40 kilobases at the *HoxD* locus [173]. At chromatin level, potassium voltage-gated channel, subfamily Q, member 1 overlapping transcript 1 (*Kcnq1ot1*) and antisense *Igf2r* RNA (*Air*) can both elicit chromatin modifications during development [174, 175]. The most classical example of functional NATs are the inactive X specific transcripts (*Xist*) and XIST antisense RNA (*Tsix*) ncRNAs, which are involved in X chromosome inactivation [176, 177]. Besides the interactions with the chromatin and transcriptional machineries, there is another NAT known as non-protein coding RNA, repressor of NFAT (*NRON*) that interacts with nuclear importin receptors and may regulate the nuclear trafficking of nuclear factor of activated T-cells (NFAT) proteins [178]. Although abundance of NATs was found in the cytoplasm, most of the previously characterised NATs have functional roles mainly in the nuclear compartment of the cells. The role of NATs within the cytoplasm remains understudied. It has been suggested that NATs can form partial or full overlapping dsRNA pairs with the sense transcripts in the cytoplasm and exert either 1) catalytic degradation of dsRNA pairs into smaller RNAs by Dicer enzyme, 2) the masking of miRNA target sites within the 3' UTR of the sense transcripts, or 3) the mechanism to repress translation process. It has been shown that beta-site APP cleaving enzyme 1 (*BACE1*) NAT prevents the induction of miRNA-mediated translational repression or degradation of *BACE1* in the cytoplasm *in vitro* and this finding has been proposed as the underlying mechanism, which improves the stability of the highly expressed *BACE1* mRNA

in Alzheimer's patients [170]. With their diverse mechanism of actions, lncRNAs or NATs may contribute to the wide spectrum of gene expression dosages and subsequently influence the phenotypic outcomes at cellular or organismal level. Therefore, characterisation of lncRNAs, particularly the NATs involved in cellular processes or disease development may provide resourceful insight regarding causative/potential therapeutic targets for the treatment of complex neurological disorders.

1.4.4 Other factors

The ability to export mRNAs into cytoplasm and maintain their integrity from degradation is the primary step to a successful translation in the cytoplasm. Sequence or structural variations such as different 3' UTR lengths and alternative sites for polyadenylation (poly(A)) among mRNAs that code for the same protein could affect the process of protein translation. Different 3' UTR lengths of mRNAs may harbour different regulatory motifs such as AU-rich elements and miRNA target sites leading to cell-specific regulation, different nuclear or cytoplasmic mRNA stability and translation rates [179, 180]. An mRNA with a shorter 3' UTR length is potentially harbouring less target sites for miRNA as compared to any mRNA variants with a greater 3' UTR length, therefore is less likely to undergo miRNA-mediated translational repression or mRNA degradation. In addition, AU-rich elements within the 3' UTR play a crucial role in mRNA stability through activation of deadenylation, decapping or 3'→5' decay processes [181]. The appropriate length of the poly(A) tail determines the identity of the mRNA and plays a critical role in nucleus-to-cytoplasm mRNA translocation [182]. Poly(A)-binding proteins target poly(A) tails and mediate the exportation of mRNA into cytoplasm, recruit translation factors and prevent mRNA degradation [183, 184]. In a rare case, dominant cataract 2 transcribed nuclear RNA (CTN-RNAs) transcribed from the mouse cationic amino acid transporter 2 (mCAT2) gene can be retained at the paraspeckles within the nuclear compartment due to the adenosine-to-inosine editing within the 3' UTR of CTN-RNA, resulting in the regulation of its protein-coding partner expression [13, 14]. Collectively, these features provide an additional level of post-transcriptional regulation leading to the expression of the ultimate product of gene expression.

1.5 Global transcriptome profiling

1.5.1 Gene discovery and expression studies

Transcriptome profiling involves the quantification of global transcript expression (both known and novel transcripts) in samples of normal or peculiar physiological states. The underlying mechanisms leading to the different physiological states could be dissected molecularly by identifying differentially expressed transcripts (DETs) between these samples. DETs in one sample or another will shed light on the causative or disrupted multilayer gene regulation networks essentially for the distinct physiological conditions. Various methods can be used to compare the gene expression profile between two samples including the subtractive cDNA library construction and screening, northern- or southern-blotting approaches, differential display and PCR-based analysis. These methods allow only qualitative comparison and some of them rely on the gene or transcript information for probe preparation, hence limit themselves as low throughput methods for targeted approach or as validation tools.

To truly perform the global transcriptome profiling analysis in samples of interest, each transcript molecule present in the cell (or a group of cells) should be sampled, quantitated and compared. It is almost possible to achieve this aim especially by employing high-throughput technologies to generate information near to global representation of the original pool of transcripts. The discovery of the automated sequencing technologies and the latest next-generation sequencing platforms have enabled high-throughput expression profiling approaches such as large scale Expressed Sequenced Tags (ESTs) or full-length cDNA library sequencing, Serial Analysis of Gene Expression (SAGE) and direct sequencing of the transcriptome. These methods do not require any prior knowledge of the transcripts and are very useful in discovering novel genes. The completion of various genome sequencing projects such as the human, mouse, *Drosophila*, *Caenorhabditis elegans* and other, provides an increasing amount of data for both functional and structural gene annotations, comparative genomics and transcriptomics, and evolutionary studies.

The application of automated next-generation sequencing technologies for global transcriptome profiling provides a powerful approach to discover novel as well as characterise known transcripts in samples of interest. This approach allows a global and unbiased analysis of a transcriptome and is best performed on SAGE cDNA libraries to provide extensive coverage or true representation of all the transcripts in the sample of interest [185-188]. However, the sequencing approach often requires more starting RNA materials, higher operational costs and more complex statistical/bioinformatics algorithms for downstream sequence analysis. In contrast to the sequencing approach, the utilisation of array-based technologies such as cDNA and oligonucleotide microarrays have gained their own reputation. The availability of gene information and related full-length cDNA clones has led to the employment of the hybridisation-based approach to profile the expression level of thousands of genes in parallel by measuring the intensity of signal emitted from each of the thousands spots fixed on a glass slide. Each spot represents a signature belonging to a known transcripts or gene. Although this method is limited to known transcripts/genes, the method can be customised to suit the needs for cost- and labour-effective routine screening of a set of transcripts/genes of interest across a large number of samples. Depending on the availability of suitable resources and desired outcomes for each gene expression profiling/discovery experiment, the right approach must be carefully evaluated and tailored to suit the requirements of each study.

1.5.2 Serial Analysis of Gene Expression (SAGE)

Serial Analysis of Gene Expression (SAGE) technique was first described by Velculescu et al in 1995 [187]. SAGE is a global expression profiling technique that quantitatively measures the level of both known as well as novel transcripts without the need of any prior knowledge about their sequences. The technique utilises 10-14bp unique sequences (known as tags) isolated from unique positions of transcripts, for example at the 3’ most *Nla*III restriction enzyme site, to determine their level of expression. These tags can be concatenated to form long serial molecules of unique tags that can be cloned and sequenced. Quantification of the number of times a unique tag being sequenced or observed represents the level of expression for the corresponding transcript of which the tag was isolated.

Global expression profiling using SAGE technique involves multiple steps in library preparation (Figure 1.14). The technique has greater throughput and feasibility in data analysis compared to other conventional methods such as reverse-transcription polymerase chain reaction (RT-PCR), subtractive hybridisation analysis, EST sequencing and hybridisation-blotting method. Besides, SAGE generates absolute expression data in digital format, which can be reanalysed together with other SAGE databases produced in different timepoints or by different laboratories from all around the world. Since the technique depends heavily on the number of sequences generated, the sensitivity and specificity of SAGE correlate positively with the amount of tags generated. It remains difficult to justify the amount of tags that need to be sequenced in various types of tissues in order to sample all the species of expressed transcripts especially lowly expressed and short-lived transcripts. Although it has been shown that the number of unique transcripts is approaching zero when the total sequenced tags near to ~600,000 [186], the sequencing process remains costly and time consuming especially when a few biological replicates have to be examined together.

In addition to the cost and time involved, there are other challenges need to be considered when the SAGE technique is employed for transcriptomic profiling such as data preprocessing and analysis. Data preprocessing includes the extraction of tag sequences from sequencing files, exclusion of low quality tags due to sequencing errors, removal of duplicated ditags and anchoring enzymes or linker sequences. There are various SAGE analysis softwares available publicly for this purpose such as those provided at the SAGE Genie [189], SAGEmap [190], MD Anderson Cancer Center [191] and Melbourne Brain Genome Project sites [192]. Extracted tag sequences are usually clustered to generate unique tags with corresponding count numbers, which represent the expression level. Due to the fact that different sizes of SAGE libraries are often generated from different experimental replicates, there is a need for library size normalisation prior to statistical comparisons. Some of the limitations of the SAGE technique remain difficult to resolve such as tags generated from repetitive elements or from the same region of different transcript variants. These tags are usually difficult to annotate and wrongly clustered as the same unique tag.

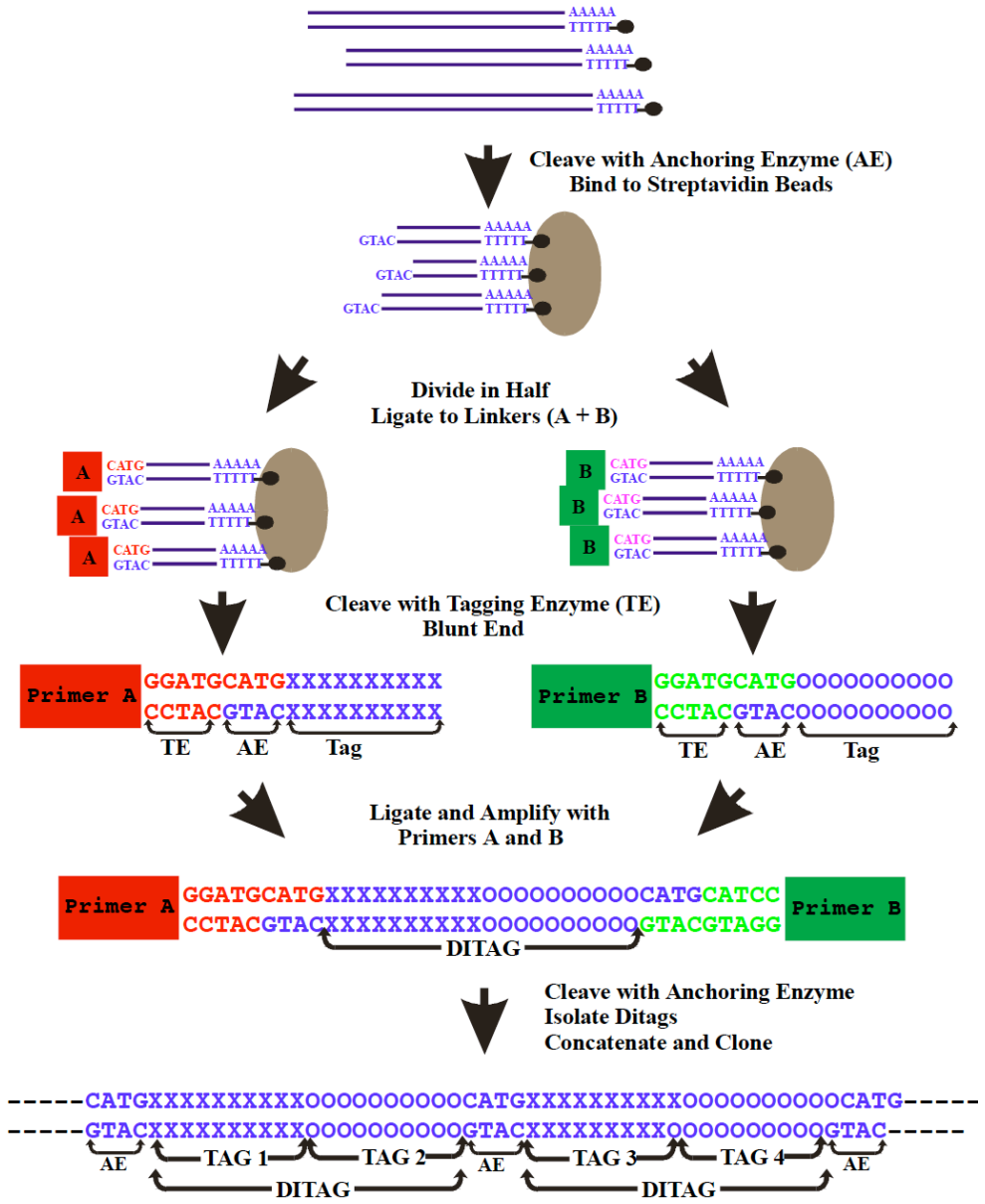


Figure 1.14 **An outline of SAGE library construction.**
Preparation of SAGE library for sequencing involves multiple steps; (1) mRNA isolation from samples of interest, (2) cDNA synthesis using streptavidin-bound oligo-d[T]$_{20}$, (3) digestion of cDNAs using anchoring enzyme (AE), (4) ligation of linker adaptors A and B to cDNA, (5) releasing the bound cDNA by enzymatic digestion using tagging enzyme (TE), (6) ligation of tags into ditags, (7) amplification of ditags using universal primers for adaptors A and B, (8) release ditags from linker adaptors A and B using anchoring enzyme (AE), (8) concatenation of isolated ditags, (9) cloning of ditags, (10) sequencing of serially concatenated ditags, and (11) computational analysis of the absolute count for each unique tags. This figure is adapted from Velculescu *et al*, 2000 [193].

Since its initial introduction in 1995, various adaptations have been made to the technique leading to new approaches such as microSAGE [194], SAGE-Lite [195], 5' and 3' LongSAGE [196], Cap Analysis of Gene Expression (CAGE) [197], Polony Multiplex Analysis of Gene Expression (PMAGE) [185, 198], DeepSAGE [199] and rapid analysis of 5' transcript ends (5'-RATE) [200] in transcriptome profiling. Each of these modified SAGE techniques utilises different preparation enzymes and can be implemented to answer more specialised research questions. The conventional SAGE and newly adapted approaches have been widely used to profile the transcriptome from different tissues such as different brain regions [201-205], pancreas [187], human embryonic stem cells [206], heart [198] and breast tumour cells [207]. Transcriptome profiling based on the SAGE principles has led to the discovery of novel transcripts or disease markers including differentially expressed, rare, alternatively spliced or fusion transcripts [208-211]. With the emergence of new ultra high-throughput next generation sequencing technologies, SAGE technique has been enhanced with improved sequencing coverage and reduced implementation cost or time, hence the method of choice for global transcriptome profiling and gene discovery.

1.5.3 Massively Parallel Sequencing (MPS) and next-generation sequencing platforms

Since the introduction of chain termination sequencing method by Sanger's group in mid 1970s [212, 213], it remains the mostly used DNA sequencing method to date and was the core for sequencing technologies employed in the human genome project [214]. It took about 30 years before the new sequencing methods were introduced to the market broadly known as next-generation sequencing technologies, which offer tremendous sequencing throughput, cost- and time-effectiveness at the expense of read lengths. To date, there are many commercially available next-generation sequencing platforms but only four are discussed in this section; 1) 454 GS FLX pyrosequencing based instruments by Roche Applied Science, 2) the Illumina HiSeq2000 analyser by Illumina Inc., 3) 5500xl SOLiD™system offered by Applied Biosystems, and 4) the HeliScope™platform from Helicos Biosciences Corporation. More details regarding these latest sequencing platforms are listed in Table 1.3.

Table 1.3: Available sequencing platforms.
(Information was extracted from individual manufacturer's website)

Platform	Sequencing approach	Read length	Throughput	Manufacturer
ABI 3730xl	Chain termination using dideoxy dye chemistry	Up to 900bp	~96kb per run	Applied Biosystems [215]
454 GS FLX	Massively parallel pyrosequencing of emulsion-based clonally amplified single stranded DNA/cDNA library	400bp	400Mb per run	Roche Applied Science [216]
HiSeq2000	Massively parallel sequencing by synthesis using reversible terminator-based method	35-100bp	25Gb per day	Illumina Inc. [217]
5500xl SOLiD™	Massively parallel sequencing by ligation principles	35-75bp	20-30Gb per day	Applied Biosystems [215]
Heliscope™	Massively parallel direct sequencing of single molecule by synthesis	25-55bp	21-35Gb per run	Helicos Biosciences Corporation [218]

Massively parallel sequencing technologies evolved rapidly since its first introduction in year 2000 [219]. The original concept involves simultaneous sequencing carried out on a million microbeads, each liaise with a single DNA/cDNA template to generate 16-20 bases of signature sequences after repeated cycles of enzymatic cleavage with a type IIs restriction endonuclease, adaptor ligation, and sequence interrogation by encoded hybridisation probes. These signature sequences can be quantitated and mapped back to the genome to annotate their origins for gene expression studies. This approach is directly or indirectly adapted in the latest next-generation sequencing platforms. With modifications to the sequencing chemistry, cloning approach and bioinformatics analysis, the next-generation sequencing platforms become very efficient in generating ultra high-throughput sequences for various applications such as *de novo* sequencing of new genomes, metagenomic characterisation of complex organisms, whole genome resequencing, gene expression analysis and discovery of novel or rare transcripts. Unfortunately, the read length is the major setback for these platforms. Short sequences prevent effective *de novo* genome assembly at the repetitive regions or during gene annotations. However, modification of the sample preparation and cloning strategy itself to generate paired ends or mate pairs sequences has been shown to improve the mapping process tremendously [220, 221]. Additionally, read length setback is not affecting sequencing projects related to small noncoding RNAs such as siRNAs, miRNAs and piRNAs (between 18- to 30-bases) in plants and mammals [103, 144, 222].

The rapid development of next-generation sequencing platforms promises a broad range of applications in functional genomics and comparative transcriptomics. Unfortunately, the downstream data analysis pipelines do not develop in accordance with the throughput of these technologies. More studies are needed to address the data storage issue, increased consumption of computation power and the robustness of statistical or bioinformatics algorithms to deal with assembly, mapping and integration of various datasets generated using next-generation sequencing platforms. As soon as these issues are taken care of, the next-generation sequencing platform has enormous potential to reform the existing landscape of genomics or transcriptomics researches.

1.5.4 Profiling of the brain transcriptome

Various attempts have been carried out to characterise the transcriptome of the mammalian brain. Since a decade ago, the analysis of brain transcriptomes has evolved from the low throughput detection using semi-quantitative RT-PCR, *in situ* hybridisation and northern analysis techniques [223-225] to high-throughput approach involving the sequencing and annotations of ESTs, clones from subtraction hybridisation libraries or SAGE tags [204, 226-228], utilisation of microarray for quantification of transcript abundances [229, reviewed in 230, 231] and sequencing of millions of transcript signatures using the massively parallel and next-generation sequencing technologies [232, 233].

The most comprehensive expression profiling of gene in the mammalian brain is described in the online Allen Brain Atlas [234], a project initiated by the Allen Institute of Brain Science, which was established in 2003 [235]. The project has characterised the expression profile of more than 21,000 genes in the adult (P56) mouse brain using the high-throughput and semi-automated colorimetric *in situ* hybridisation method described previously [236]. Recently, the institute has embarked in the expression profiling of several hundred genes in developing mouse embryos. In parallel with the Allen Brain Atlas project are the Gene Expression Nervous System Atlas (GENSAT) initiated by the National Institutes of Health (NIH) [237], and GenePaint atlas led by Gregor Eichele at the Max-Plank-Institute of Biophysical Chemistry in Germany [236, 238]. Both GENSAT and GenePaint atlases were not set to comprehensively catalogue the expression of the number of genes as championed by the Allen Brain Atlas. GENSAT project aims to profile the gene expression using radiometric *in situ* hybridisation on brains obtained from four different stages of mouse development. The group also used Bacterial Artificial Chromosome (BAC) to clone up to 200kb of upstream sequences of genes of interest and use enhanced Green Fluorescent Protein (eGFP) reporter protein to study gene expression in the brain of transgenic mice. Alternatively, GenePaint atlas utilises both methods used in Allen Brain and GENSAT atlases on whole E14.5 mouse embryos with additional information for selected genes in other stages of development. Despite the comprehensive cataloguing of gene expression profiles, all atlases provide minimal information at cellular level. Interpretation of the expression profile for the same gene using all

the three atlases can be tricky and requires proper streamlining in terms of the age, anatomical section and subregions of the brain section being referred to. Another problem faced by all three atlases is the lack of common and agreeable neuroanatomical and functional annotations as well as nomenclature leading to confusion among non-specialists.

The profiling of gene expression in the mammalian brain has also been investigated by various studies utilising SAGE and microarray technologies [201-205]. Melbourne Brain Genome Project was initiated in 2002 [192] and set to generate millions of tags from various normal and diseased mouse brains using both short and longSAGE techniques [188, 196]. Under the Cancer Genome Anatomy Project, large-scale SAGE datasets generated from various human and mouse organs including the brain are deposited in the SAGE genie database, which also provides various visualisation and mapping tools [189]. With the emergence of various next-generation sequencing platforms and the introduction of a massively parallel sequencing approach pioneered by Brenner and colleagues [219], the brain transcriptome can now be assessed by generating millions of transcript signatures (e.g. SAGE tags, paired ends sequences, direct sequences etc.) by performing deep sequencing. This approach has been broadly practised to profile both known and novel mRNA as well as small noncoding RNA such as miRNA and endo-siRNA [144, 232, 239]. To further enhance the transcriptome information, Harlt and colleagues (2008) analysed both the transcriptome and proteome of the brain obtained from E9.5, E11.5 and E13.5 mouse embryos to correlate the gene expression with protein expression during brain development [240]. They found metabolism and cell cycle related gene products were down-regulated as the brain developed from E9.5 to E13.5 where multipotent progenitor cells committed to the neuronal lineage and underwent differentiation.

Despite various reports on brain transcriptome profiling, the outcomes of the studies are lacking coherence in terms of many aspects such as the type of organisms studied, the brain regions selected for the analysis, choices of different developmental stages of the brain, disease models, types of RNA transcript of interest and platforms used for transcriptome profiling. Dissimilarities in experimental approaches lead to difficulty in direct comparison of datasets generated from different studies.

1.6 Problem statement and the aim of the study

Neurological, psychiatric, and developmental disorders such as epilepsy, stroke, schizophrenia, unipolar depression, bipolar disorder, Alzheimer's disease, Parkinson's disease, mental retardation, cerebral palsy and autism cause a profound economic and social burden worldwide. In early 2000s, the brain disorder itself was estimated to affect about 1.5 billion people globally and responsible for at least 27% of total years lived with disability in developing countries [241]. In almost all cases, no effective treatments or clear preventive guidelines can be prescribed due to the nature of the disorders, which have a broad spectrum clinical observations and affect people from as young as newborns to ageing population. In addition, the very limited and ill-defined disease progression and pathogenesis as well as lack of understanding on the underlying genetic mechanisms responsible for the onset of these disorders also contribute to the inadequacy of any effective intervention measurements.

Unlike single-gene disorders, genetic factors that lead to the development of these disorders are of low penetrance. The appreciable symptoms in these disorders are usually the outcomes of complex interactions between molecular networks and various environmental factors throughout lifetime. In the quest of finding the causative mechanism, it is crucial that we understand the normal mechanism, which underlies the development of the brain leading to optimal physiological functions. To achieve this aim, global transcriptome analysis can be performed to characterise the type of genes expressed during different developmental stages of the brain with greater emphasis on the cerebral cortex, the centre of cognitive and intellectual abilities. Both human and mouse share about 85% of gene contents and therefore the mouse serves as a perfect model to comprehend the molecular factors involved in brain development, which allow the findings to be extrapolated to human in the future. With special focus on transcription factors that are involved in the regulation of brain development and function, we would be able to molecularly characterise the genetic networks that govern normal brain development processes. Therefore, this study aims to identify genes involved in the proliferation, differentiation and developmental networks of the mouse brain.

1.7 Thesis outline

The thesis contains four result chapters (Chapters 3-6). Chapter 3 describes the global transcriptome analysis on four developmental stages of the cerebral cortex using the short and long SAGE techniques, which were carried out in the early phase of Melbourne Brain Genome Project in 2002 [192]. The study revealed many important differentially expressed transcripts including four genomic clusters (at *Nrgn*, *Camk2n1*, *Sox4* and *Sox11* gene loci) that are differentially transcribed during cerebral corticogenesis. These genomics loci feature multiple overlapping sense and antisense transcripts with different 3' untranslated regions (UTR) lengths and expression levels. These features were validated based on the observations for *Sox4* and *Sox11* clusters of transcripts. Chapter 4 is the continuation of the previous result chapter, which further characterised the expression profiles for *Nrgn* and *Camk2n1* clusters of transcripts during cerebral corticogenesis. The most significant finding of the study is the ability of *Nrgn* and *Camk2n1* sense and antisense transcripts to form dsRNA aggregates in the cytoplasm of various brain cells. In Chapter 5, the results show similar findings for *Sox4* sense and antisense transcripts with the ability to form dsRNA aggregates in the cytoplasmic region of various types of brain cells. The study also described the role of these dsRNA in the biogenesis of a novel endo-siRNA with a specific expression profile within the embryo especially in the brain throughout development. Finally, Chapter 6 reports the expression profiling of small RNAs isolated from a developing brain at age E15.5. This study employed next-generation sequencing technology involving the generation of approximately 3.7 million small RNA sequences. The study revealed four candidate novel miRNAs from the mouse brain and one of them, M1181, was validated as a novel miRNA and may play a critical role during embryogenesis especially in cerebral corticogenesis.

1.8 References

1. Dehay C, Kennedy H: **Cell-cycle control and cortical development**. *Nat Rev Neurosci* 2007, **8**(6):438-450.
2. Rakic P: **Specification of cerebral cortical areas**. *Science* 1988, **241**(4862):170-176.
3. O'Leary DD: **Do cortical areas emerge from a protocortex?** *Trends Neurosci* 1989, **12**(10):400-406.
4. Kaufman MH, Bard JBL: **The Anatomical Basis of Mouse Development**, 1 edn. San Diego: Elsevier Science; 1999.
5. Bartova E, Krejci J, Harnicarova A, Galiova G, Kozubek S: **Histone modifications and nuclear architecture: a review**. *J Histochem Cytochem* 2008, **56**(8):711-721.
6. van Driel R, Fransz PF, Verschure PJ: **The eukaryotic genome: a system regulated at different hierarchical levels**. *J Cell Sci* 2003, **116**(Pt 20):4067-4075.
7. Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development**. *Science* 2001, **293**(5532):1089-1093.
8. **Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs**. *Nature* 2009, **457**(7232):1028-1032.
9. Amaral PP, Dinger ME, Mercer TR, Mattick JS: **The eukaryotic genome as an RNA machine**. *Science* 2008, **319**(5871):1787-1789.
10. Conaway RC, Brower CS, Conaway JW: **Emerging roles of ubiquitin in transcription regulation**. *Science* 2002, **296**(5571):1254-1258.
11. Lapidot M, Pilpel Y: **Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms**. *EMBO Rep* 2006, **7**(12):1216-1222.
12. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.
13. Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI: **Paraspeckles: a novel nuclear domain**. *Curr Biol* 2002, **12**(1):13-25.
14. Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL: **Regulating gene expression through RNA nuclear retention**. *Cell* 2005, **123**(2):249-263.
15. D'Souza CA, Chopra V, Varhol R, Xie YY, Bohacec S, Zhao Y, Lee LL, Bilenky M, Portales-Casamar E, He A *et al*: **Identification of a set of genes showing regionally enriched expression in the mouse brain**. *BMC Neurosci* 2008, **9**:66.
16. Hill RS, Walsh CA: **Molecular insights into human brain evolution**. *Nature* 2005, **437**(7055):64-67.
17. Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D *et al*: **Regional and cellular gene expression changes in human Huntington's disease brain**. *Hum Mol Genet* 2006, **15**(6):965-977.
18. Koester SE, Insel TR: **Mouse maps of gene expression in the brain**. *Genome Biol* 2007, **8**(5):212.
19. Ladd-Acosta C, Pevsner J, Sabunciyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP: **DNA methylation signatures within the human brain**. *Am J Hum Genet* 2007, **81**(6):1304-1315.

20. Bard JBL, Kaufmatz MH: **The Mouse**. In: *EMBRYO: Colour Atlas of Development.* Edited by Bard JBL, 2nd edn. London: Wolfe Publishing Ltd.; 1994: 183-206.
21. Cohen HS: **Neuroscience for Rehabilitation**, 2 edn. New York: Lippincott Williams & Wilkins; 1999.
22. **The Mouse Brain Library** [http://www.mbl.org/]
23. Nadeau SE, Ferguson TS, Valenstein E, Vierck CJ, Petruska JC, Streit WJ, Ritz LA: **Medical Neuroscience**, 1st edn. Philadelphia: Saunders; 2004.
24. Marin O, Rubenstein JL: **A long, remarkable journey: tangential migration in the telencephalon**. *Nat Rev Neurosci* 2001, **2**(11):780-790.
25. Gorski JA, Talley T, Qiu M, Puelles L, Rubenstein JL, Jones KR: **Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage**. *J Neurosci* 2002, **22**(15):6309-6314.
26. Letinic K, Zoncu R, Rakic P: **Origin of GABAergic neurons in the human neocortex**. *Nature* 2002, **417**(6889):645-649.
27. Davis AA, Temple S: **A self-renewing multipotential stem cell in embryonic rat cerebral cortex**. *Nature* 1994, **372**(6503):263-266.
28. Williams BP, Price J: **Evidence for multiple precursor cell types in the embryonic rat cerebral cortex**. *Neuron* 1995, **14**(6):1181-1188.
29. Temple S: **The development of neural stem cells**. *Nature* 2001, **414**(6859):112-117.
30. He W, Ingraham C, Rising L, Goderie S, Temple S: **Multipotent stem cells from the mouse basal forebrain contribute GABAergic neurons and oligodendrocytes to the cerebral cortex during embryogenesis**. *J Neurosci* 2001, **21**(22):8854-8862.
31. Ross SE, Greenberg ME, Stiles CD: **Basic helix-loop-helix factors in cortical development**. *Neuron* 2003, **39**(1):13-25.
32. Ivanova A, Nakahira E, Kagawa T, Oba A, Wada T, Takebayashi H, Spassky N, Levine J, Zalc B, Ikenaka K: **Evidence for a second wave of oligodendrogenesis in the postnatal cerebral cortex of the mouse**. *J Neurosci Res* 2003, **73**(5):581-592.
33. Marshall CA, Suzuki SO, Goldman JE: **Gliogenic and neurogenic progenitors of the subventricular zone: who are they, where did they come from, and where are they going?** *Glia* 2003, **43**(1):52-61.
34. Chenn A, McConnell SK: **Cleavage orientation and the asymmetric inheritance of Notch1 immunoreactivity in mammalian neurogenesis**. *Cell* 1995, **82**(4):631-641.
35. Kalyani A, Hobson K, Rao MS: **Neuroepithelial stem cells from the embryonic spinal cord: isolation, characterization, and clonal analysis**. *Dev Biol* 1997, **186**(2):202-223.
36. Kalyani AJ, Piper D, Mujtaba T, Lucero MT, Rao MS: **Spinal cord neuronal precursors generate multiple neuronal phenotypes in culture**. *J Neurosci* 1998, **18**(19):7856-7868.
37. Lee SK, Pfaff SL: **Transcriptional networks regulating neuronal identity in the developing spinal cord**. *Nat Neurosci* 2001, **4 Suppl**:1183-1191.
38. Dorsky RI, Chang WS, Rapaport DH, Harris WA: **Regulation of neuronal diversity in the Xenopus retina by Delta signalling**. *Nature* 1997, **385**(6611):67-70.
39. Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, Bader BL, Hynes RO, Zhuang Y, Manova K *et al*: **Id1 and Id3 are required for**

neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* 1999, **401**(6754):670-677.
40. Halevy O, Novitch BG, Spicer DB, Skapek SX, Rhee J, Hannon GJ, Beach D, Lassar AB: **Correlation of terminal cell cycle arrest of skeletal muscle with induction of p21 by MyoD**. *Science* 1995, **267**(5200):1018-1021.
41. Mutoh H, Naya FJ, Tsai MJ, Leiter AB: **The basic helix-loop-helix protein BETA2 interacts with p300 to coordinate differentiation of secretin-expressing enteroendocrine cells**. *Genes Dev* 1998, **12**(6):820-830.
42. Fode C, Ma Q, Casarosa S, Ang SL, Anderson DJ, Guillemot F: **A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons**. *Genes Dev* 2000, **14**(1):67-80.
43. Nieto M, Schuurmans C, Britz O, Guillemot F: **Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors**. *Neuron* 2001, **29**(2):401-413.
44. Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME: **Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms**. *Cell* 2001, **104**(3):365-376.
45. Lee JE: **Basic helix-loop-helix genes in neural development**. *Curr Opin Neurobiol* 1997, **7**(1):13-20.
46. Bertrand N, Castro DS, Guillemot F: **Proneural genes and the specification of neural cell types**. *Nat Rev Neurosci* 2002, **3**(7):517-530.
47. Schuurmans C, Armant O, Nieto M, Stenman JM, Britz O, Klenin N, Brown C, Langevin LM, Seibt J, Tang H *et al*: **Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways**. *Embo J* 2004, **23**(14):2892-2902.
48. Hevner RF, Shi L, Justice N, Hsueh Y, Sheng M, Smiga S, Bulfone A, Goffinet AM, Campagnoni AT, Rubenstein JL: **Tbr1 regulates differentiation of the preplate and layer 6**. *Neuron* 2001, **29**(2):353-366.
49. Englund C, Fink A, Lau C, Pham D, Daza RA, Bulfone A, Kowalczyk T, Hevner RF: **Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex**. *J Neurosci* 2005, **25**(1):247-251.
50. Gupta A, Tsai LH, Wynshaw-Boris A: **Life is a journey: a genetic look at neocortical development**. *Nat Rev Genet* 2002, **3**(5):342-355.
51. Nadarajah B, Parnavelas JG: **Modes of neuronal migration in the developing cerebral cortex**. *Nat Rev Neurosci* 2002, **3**(6):423-432.
52. Wichterle H, Turnbull DH, Nery S, Fishell G, Alvarez-Buylla A: **In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain**. *Development* 2001, **128**(19):3759-3771.
53. Hatten ME: **Riding the glial monorail: a common mechanism for glial-guided neuronal migration in different regions of the developing mammalian brain**. *Trends Neurosci* 1990, **13**(5):179-184.
54. Hirotsune S, Fleck MW, Gambello MJ, Bix GJ, Chen A, Clark GD, Ledbetter DH, McBain CJ, Wynshaw-Boris A: **Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality**. *Nat Genet* 1998, **19**(4):333-339.
55. Fleck MW, Hirotsune S, Gambello MJ, Phillips-Tansey E, Suares G, Mervis RF, Wynshaw-Boris A, McBain CJ: **Hippocampal abnormalities**

and enhanced excitability in a murine model of human lissencephaly. *J Neurosci* 2000, **20**(7):2439-2450.
56. Gleeson JG, Lin PT, Flanagan LA, Walsh CA: **Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons**. *Neuron* 1999, **23**(2):257-271.
57. Francis F, Koulakoff A, Boucher D, Chafey P, Schaar B, Vinet MC, Friocourt G, McDonnell N, Reiner O, Kahn A *et al*: **Doublecortin is a developmentally regulated, microtubule-associated protein expressed in migrating and differentiating neurons**. *Neuron* 1999, **23**(2):247-256.
58. Sheldon M, Rice DS, D'Arcangelo G, Yoneshima H, Nakajima K, Mikoshiba K, Howell BW, Cooper JA, Goldowitz D, Curran T: **Scrambler and yotari disrupt the disabled gene and produce a reeler-like phenotype in mice**. *Nature* 1997, **389**(6652):730-733.
59. Howell BW, Hawkes R, Soriano P, Cooper JA: **Neuronal position in the developing brain is regulated by mouse disabled-1**. *Nature* 1997, **389**(6652):733-737.
60. Cabana T: **Development of the nervous system**. In: *Neuroscience for rehabilitation.* Edited by Cohen H, 2 edn. Philadelphia: Lippincott Williams & Wilkins; 1998: 369-399.
61. Black MM, Baas PW: **The basis of polarity in neurons**. *Trends Neurosci* 1989, **12**(6):211-214.
62. Tessier-Lavigne M: **Eph receptor tyrosine kinases, axon repulsion, and the development of topographic maps**. *Cell* 1995, **82**(3):345-348.
63. Waites CL, Craig AM, Garner CC: **Mechanisms of vertebrate synaptogenesis**. *Annu Rev Neurosci* 2005, **28**:251-274.
64. Bagri A, Tessier-Lavigne M: **Neuropilins as Semaphorin receptors: in vivo functions in neuronal cell migration and axon guidance**. *Adv Exp Med Biol* 2002, **515**:13-31.
65. Pascual M, Pozas E, Barallobre MJ, Tessier-Lavigne M, Soriano E: **Coordinated functions of Netrin-1 and Class 3 secreted Semaphorins in the guidance of reciprocal septohippocampal connections**. *Mol Cell Neurosci* 2004, **26**(1):24-33.
66. O'Brien RJ, Xu D, Petralia RS, Steward O, Huganir RL, Worley P: **Synaptic clustering of AMPA receptors by the extracellular immediate-early gene product Narp**. *Neuron* 1999, **23**(2):309-323.
67. Dalva MB, Takasu MA, Lin MZ, Shamah SM, Hu L, Gale NW, Greenberg ME: **EphB receptors interact with NMDA receptors and regulate excitatory synapse formation**. *Cell* 2000, **103**(6):945-956.
68. Murai KK, Nguyen LN, Irie F, Yamaguchi Y, Pasquale EB: **Control of hippocampal dendritic spine morphology through ephrin-A3/EphA4 signaling**. *Nat Neurosci* 2003, **6**(2):153-160.
69. Penzes P, Beeser A, Chernoff J, Schiller MR, Eipper BA, Mains RE, Huganir RL: **Rapid induction of dendritic spine morphogenesis by trans-synaptic ephrinB-EphB receptor activation of the Rho-GEF kalirin**. *Neuron* 2003, **37**(2):263-274.
70. Krylova O, Herreros J, Cleverley KE, Ehler E, Henriquez JP, Hughes SM, Salinas PC: **WNT-3, expressed by motoneurons, regulates terminal arborization of neurotrophin-3-responsive spinal sensory neurons**. *Neuron* 2002, **35**(6):1043-1056.
71. Hall AC, Lucas FR, Salinas PC: **Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling**. *Cell* 2000, **100**(5):525-535.

72. Umemori H, Linhoff MW, Ornitz DM, Sanes JR: **FGF22 and its close relatives are presynaptic organizing molecules in the mammalian brain**. *Cell* 2004, **118**(2):257-270.
73. Biederer T, Sara Y, Mozhayeva M, Atasoy D, Liu X, Kavalali ET, Sudhof TC: **SynCAM, a synaptic adhesion molecule that drives synapse assembly**. *Science* 2002, **297**(5586):1525-1531.
74. Scheiffele P: **Cell-cell signaling during synapse formation in the CNS**. *Annu Rev Neurosci* 2003, **26**:485-508.
75. Scheiffele P, Fan J, Choih J, Fetter R, Serafini T: **Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons**. *Cell* 2000, **101**(6):657-669.
76. Hannah MJ, Schmidt AA, Huttner WB: **Synaptic vesicle biogenesis**. *Annu Rev Cell Dev Biol* 1999, **15**:733-798.
77. Huttner WB, Ohashi M, Kehlenbach RH, Barr FA, Bauerfeind R, Braunling O, Corbeil D, Hannah M, Pasolli HA, Schmidt A *et al*: **Biogenesis of neurosecretory vesicles**. *Cold Spring Harb Symp Quant Biol* 1995, **60**:315-327.
78. Ziv NE, Garner CC: **Cellular and molecular mechanisms of presynaptic assembly**. *Nat Rev Neurosci* 2004, **5**(5):385-399.
79. Bresler T, Shapira M, Boeckers T, Dresbach T, Futter M, Garner CC, Rosenblum K, Gundelfinger ED, Ziv NE: **Postsynaptic density assembly is fundamentally different from presynaptic active zone assembly**. *J Neurosci* 2004, **24**(6):1507-1520.
80. **Department of Biology, University of Miami, Florida, USA** [http://www.bio.miami.edu/~cmallery/255/255hist/mcb4.1.dogma.jpg]
81. Bishop KM, Goudreau G, O'Leary DD: **Regulation of area identity in the mammalian neocortex by Emx2 and Pax6**. *Science* 2000, **288**(5464):344-349.
82. Mallamaci A, Stoykova A: **Gene networks controlling early cerebral cortex arealization**. *Eur J Neurosci* 2006, **23**(4):847-856.
83. O'Leary DD, Chou SJ, Sahara S: **Area patterning of the mammalian cortex**. *Neuron* 2007, **56**(2):252-269.
84. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J: **The establishment of neuronal properties is controlled by Sox4 and Sox11**. *Genes Dev* 2006, **20**(24):3475-3486.
85. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ: **Roles of Sox4 in central nervous system development**. *Brain Res Mol Brain Res* 2000, **79**(1-2):180-191.
86. Potzner MR, Griffel C, Lutjen-Drecoll E, Bosl MR, Wegner M, Sock E: **Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system**. *Mol Cell Biol* 2007, **27**(15):5316-5326.
87. Wegner M, Claus Stolt C: **Sox transcription factors in neural development**. In: *Transcription Factors in the Nervous System.* Edited by Thiel G, 1st edn. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA; 2006: 181-203.
88. **Santa Monica College, Caifornia, USA** [http://homepage.smc.edu/hodson_kent/central_dogma/Transcription/TranFactors.htm]
89. Bird A: **Perceptions of epigenetics**. *Nature* 2007, **447**(7143):396-398.
90. Reik W: **Stability and flexibility of epigenetic gene regulation in mammalian development**. *Nature* 2007, **447**(7143):425-432.

91. Lucifero D, Chaillet JR, Trasler JM: **Potential significance of genomic imprinting defects for reproduction and assisted reproductive technology**. *Hum Reprod Update* 2004, **10**(1):3-18.
92. Keverne EB, Curley JP: **Epigenetics, brain evolution and behaviour**. *Front Neuroendocrinol* 2008, **29**(3):398-412.
93. Scarano MI, Strazzullo M, Matarazzo MR, D'Esposito M: **DNA methylation 40 years later: Its role in human health and disease**. *J Cell Physiol* 2005, **204**(1):21-35.
94. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: **Origin of a substantial fraction of human regulatory sequences from transposable elements**. *Trends Genet* 2003, **19**(2):68-72.
95. Ukai H, Ishii-Oba H, Ukai-Tadenuma M, Ogiu T, Tsuji H: **Formation of an active form of the interleukin-2/15 receptor beta-chain by insertion of the intracisternal A particle in a radiation-induced mouse thymic lymphoma and its role in tumorigenesis**. *Mol Carcinog* 2003, **37**(2):110-119.
96. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T *et al*: **A novel class of small RNAs bind to MILI protein in mouse testes**. *Nature* 2006, **442**(7099):203-207.
97. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: **Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs**. *Genes Dev* 2008, **22**(20):2773-2785.
98. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T: **Identification of tissue-specific microRNAs from mouse**. *Curr Biol* 2002, **12**(9):735-739.
99. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M *et al*: **A mammalian microRNA expression atlas based on small RNA library sequencing**. *Cell* 2007, **129**(7):1401-1414.
100. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.
101. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R: **Control of translation and mRNA degradation by miRNAs and siRNAs**. *Genes Dev* 2006, **20**(5):515-524.
102. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T *et al*: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes**. *Nature* 2008.
103. Girard A, Sachidanandam R, Hannon GJ, Carmell MA: **A germline-specific class of small RNAs binds mammalian Piwi proteins**. *Nature* 2006, **442**(7099):199-202.
104. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ: **Developmentally regulated piRNA clusters implicate MILI in transposon control**. *Science* 2007, **316**(5825):744-747.
105. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K *et al*: **Tiny RNAs associated with transcription start sites in animals**. *Nat Genet* 2009, **41**(5):572-578.
106. Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJ, Rasko JE *et al*: **Nuclear-localized tiny RNAs are**

associated with transcription initiation and splice sites in metazoans. *Nat Struct Mol Biol* 2010, **17**(8):1030-1034.
107. Shi W, Hendrix D, Levine M, Haley B: **A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate**. *Nat Struct Mol Biol* 2009, **16**(2):183-189.
108. Cao F, Li X, Hiew S, Brady H, Liu Y, Dou Y: **Dicer independent small RNAs associate with telomeric heterochromatin**. *RNA* 2009, **15**(7):1274-1281.
109. Yoshikawa M, Peragine A, Park MY, Poethig RS: **A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis**. *Genes Dev* 2005, **19**(18):2164-2175.
110. Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE: **Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response**. *Dev Cell* 2007, **12**(1):45-55.
111. Aravin A, Tuschl T: **Identification and characterization of small RNAs involved in RNA silencing**. *FEBS Lett* 2005, **579**(26):5830-5840.
112. Ghildiyal M, Zamore PD: **Small silencing RNAs: an expanding universe**. *Nat Rev Genet* 2009, **10**(2):94-108.
113. Llave C, Kasschau KD, Rector MA, Carrington JC: **Endogenous and silencing-associated small RNAs in plants**. *Plant Cell* 2002, **14**(7):1605-1619.
114. Chalker DL, Yao MC: **Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila**. *Genes Dev* 2001, **15**(10):1287-1298.
115. Tomaru Y, Hayashizaki Y: **Cancer research with non-coding RNA**. *Cancer Sci* 2006, **97**(12):1285-1290.
116. Choudhuri S: **Lesser known relatives of miRNA**. *Biochem Biophys Res Commun* 2009, **388**(2):177-180.
117. Kuwabara T, Hsieh J, Nakashima K, Taira K, Gage FH: **A small modulatory dsRNA specifies the fate of adult neural stem cells**. *Cell* 2004, **116**(6):779-793.
118. Kim VN: **MicroRNA biogenesis: coordinated cropping and dicing**. *Nat Rev Mol Cell Biol* 2005, **6**(5):376-385.
119. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D: **MicroRNAs and other tiny endogenous RNAs in C. elegans**. *Curr Biol* 2003, **13**(10):807-818.
120. Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ *et al*: **A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres**. *Chromosoma* 2009, **118**(1):113-125.
121. Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD: **Over 20% of human transcripts might form sense-antisense pairs**. *Nucleic Acids Res* 2004, **32**(16):4812-4820.
122. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: **The antisense transcriptomes of human cells**. *Science* 2008, **322**(5909):1855-1857.
123. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J *et al*: **Antisense transcription in the mammalian transcriptome**. *Science* 2005, **309**(5740):1564-1566.

124. Moss EG, Lee RC, Ambros V: **The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA**. *Cell* 1997, **88**(5):637-646.
125. Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14**. *Cell* 1993, **75**(5):843-854.
126. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G: **The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans**. *Nature* 2000, **403**(6772):901-906.
127. Abrahante JE, Daul AL, Li M, Volk ML, Tennessen JM, Miller EA, Rougvie AE: **The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs**. *Dev Cell* 2003, **4**(5):625-637.
128. Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G: **The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor**. *Mol Cell* 2000, **5**(4):659-669.
129. Hamilton AJ, Baulcombe DC: **A species of small antisense RNA in posttranscriptional gene silencing in plants**. *Science* 1999, **286**(5441):950-952.
130. Zamore PD, Tuschl T, Sharp PA, Bartel DP: **RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals**. *Cell* 2000, **101**(1):25-33.
131. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans**. *Science* 2001, **294**(5543):858-862.
132. Lee RC, Ambros V: **An extensive class of small RNAs in Caenorhabditis elegans**. *Science* 2001, **294**(5543):862-864.
133. Cullen BR: **Transcription and processing of human microRNA precursors**. *Mol Cell* 2004, **16**(6):861-865.
134. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S *et al*: **The nuclear RNase III Drosha initiates microRNA processing**. *Nature* 2003, **425**(6956):415-419.
135. Cullen BR: **RNAi the natural way**. *Nat Genet* 2005, **37**(11):1163-1165.
136. Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs**. *Genome Res* 2009, **19**(1):92-105.
137. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP: **MicroRNA targeting specificity in mammals: determinants beyond seed pairing**. *Mol Cell* 2007, **27**(1):91-105.
138. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output**. *Nature* 2008, **455**(7209):64-71.
139. Diederichs S, Haber DA: **Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression**. *Cell* 2007, **131**(6):1097-1108.
140. Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME: **A brain-specific microRNA regulates dendritic spine development**. *Nature* 2006, **439**(7074):283-289.
141. Leucht C, Stigloher C, Wizenmann A, Klafke R, Folchert A, Bally-Cuif L: **MicroRNA-9 directs late organizer activity of the midbrain-hindbrain boundary**. *Nat Neurosci* 2008, **11**(6):641-648.

142. Makeyev EV, Zhang J, Carrasco MA, Maniatis T: **The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing**. *Mol Cell* 2007, **27**(3):435-448.
143. Hansen T, Olsen L, Lindow M, Jakobsen KD, Ullum H, Jonsson E, Andreassen OA, Djurovic S, Melle I, Agartz I *et al*: **Brain expressed microRNAs implicated in schizophrenia etiology**. *PLoS ONE* 2007, **2**(9):e873.
144. Marti E, Pantano L, Banez-Coronel M, Llorens F, Minones-Moyano E, Porta S, Sumoy L, Ferrer I, Estivill X: **A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing**. *Nucleic Acids Res* 2010.
145. Ciafre SA: **MiRNAs in glioblastoma**. In: *MicroRNAs: From Basic Science to Disease Biology*. Edited by Appasani K, 1st edn. Cambridge: Cambridge University Press; 2008: 350-362.
146. Szulwach K, Jin P: **Role of microRNA pathway in Fragile X mental retardation**. In: *MicroRNAs: From Basic Science to Disease Biology*. Edited by Appasani K, 1st edn. Cambridge: Cambridge University Press; 2008: 363-371.
147. Held MA, Penning B, Brandt AS, Kessans SA, Yong W, Scofield SR, Carpita NC: **Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley**. *Proc Natl Acad Sci U S A* 2008, **105**(51):20534-20539.
148. Vaucheret H: **Post-transcriptional small RNA pathways in plants: mechanisms and regulations**. *Genes Dev* 2006, **20**(7):759-771.
149. Ahlquist P: **RNA-dependent RNA polymerases, viruses, and RNA silencing**. *Science* 2002, **296**(5571):1270-1273.
150. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H: **Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes**. *Genes Dev* 2006, **20**(13):1732-1743.
151. Zilberman D, Cao X, Jacobsen SE: **ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation**. *Science* 2003, **299**(5607):716-719.
152. Reinhart BJ, Bartel DP: **Small RNAs correspond to centromere heterochromatic repeats**. *Science* 2002, **297**(5588):1831.
153. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA: **Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi**. *Science* 2002, **297**(5588):1833-1837.
154. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA: **Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena**. *Cell* 2002, **110**(6):689-699.
155. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H *et al*: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs**. *Nature* 2002, **420**(6915):563-573.
156. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al*: **The transcriptional landscape of the mammalian genome**. *Science* 2005, **309**(5740):1559-1563.

157. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR: **Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays**. *Genome Res* 2005, **15**(7):987-997.
158. Pang KC, Frith MC, Mattick JS: **Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function**. *Trends Genet* 2006, **22**(1):1-5.
159. Lee S, Bao J, Zhou G, Shapiro J, Xu J, Shi RZ, Lu X, Clark T, Johnson D, Kim YC *et al*: **Detecting novel low-abundant transcripts in Drosophila**. *Rna* 2005, **11**(6):939-946.
160. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP *et al*: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals**. *Nature* 2009, **458**(7235):223-227.
161. Okamura K, Lai EC: **Endogenous small interfering RNAs in animals**. *Nat Rev Mol Cell Biol* 2008, **9**(9):673-678.
162. Dahary D, Elroy-Stein O, Sorek R: **Naturally occurring antisense: transcriptional leakage or real overlap?** *Genome Res* 2005, **15**(3):364-368.
163. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ *et al*: **Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs**. *Cell* 2004, **116**(4):499-509.
164. Gingeras TR: **Origin of phenotypes: genes and transcripts**. *Genome Res* 2007, **17**(6):682-690.
165. Thrash-Bingham CA, Tartof KD: **aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia**. *J Natl Cancer Inst* 1999, **91**(2):143-151.
166. Yamamoto T, Manome Y, Nakamura M, Tanigawa N: **Downregulation of survivin expression by induction of the effector cell protease receptor-1 reduces tumor growth potential and results in an increased sensitivity to anticancer agents in human colon cancer**. *Eur J Cancer* 2002, **38**(17):2316-2324.
167. Capaccioli S, Quattrone A, Schiavone N, Calastretti A, Copreni E, Bevilacqua A, Canti G, Gong L, Morelli S, Nicolin A: **A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t(14;18) cell lines**. *Oncogene* 1996, **13**(1):105-115.
168. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR: **Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease**. *Nat Genet* 2003, **34**(2):157-165.
169. Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, Smallwood AC, Joyce JA, Schofield PN, Reik W *et al*: **A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome**. *Proc Natl Acad Sci U S A* 1999, **96**(14):8064-8069.
170. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, Cookson MR, St-Laurent G, 3rd, Wahlestedt C: **Evidence for natural antisense transcript-mediated inhibition of microRNA function**. *Genome Biol* 2010, **11**(5):R56.

171. Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS, Hagerman RJ, Tassone F, Tapscott SJ, Filippova GN: **An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals**. *Hum Mol Genet* 2007, **16**(24):3174-3187.
172. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD: **The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator**. *Genes Dev* 2006, **20**(11):1470-1484.
173. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al*: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.
174. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes**. *Nature* 2002, **415**(6873):810-813.
175. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: **Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation**. *Mol Cell* 2008, **32**(2):232-246.
176. Mlynarczyk SK, Panning B: **X inactivation: Tsix and Xist as yin and yang**. *Curr Biol* 2000, **10**(24):R899-903.
177. Panning B, Dausman J, Jaenisch R: **X chromosome inactivation is mediated by Xist RNA stabilization**. *Cell* 1997, **90**(5):907-916.
178. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG: **A strategy for probing the function of noncoding RNAs finds a repressor of NFAT**. *Science* 2005, **309**(5740):1570-1573.
179. Winter J, Kunath M, Roepcke S, Krause S, Schneider R, Schweiger S: **Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1**. *BMC Mol Biol* 2007, **8**:105.
180. Abdel Wahab N, Gibbs J, Mason RM: **Regulation of gene expression by alternative polyadenylation and mRNA instability in hyperglycaemic mesangial cells**. *Biochem J* 1998, **336 (Pt 2)**:405-411.
181. Wilusz CJ, Wilusz J: **Bringing the role of mRNA decay in the control of gene expression into focus**. *Trends Genet* 2004, **20**(10):491-497.
182. Fuke H, Ohno M: **Role of poly (A) tail as an identity element for mRNA nuclear export**. *Nucleic Acids Res* 2008, **36**(3):1037-1049.
183. Coller JM, Gray NK, Wickens MP: **mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation**. *Genes Dev* 1998, **12**(20):3226-3235.
184. Siddiqui N, Mangus DA, Chang TC, Palermino JM, Shyu AB, Gehring K: **Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein**. *J Biol Chem* 2007, **282**(34):25067-25075.
185. Velculescu VE, Kinzler KW: **Gene expression analysis goes digital**. *Nat Biotechnol* 2007, **25**(8):878-880.
186. Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM *et al*: **Analysis of human transcriptomes**. *Nat Genet* 1999, **23**(4):387-388.

187. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression**. *Science* 1995, **270**(5235):484-487.
188. Scott HS, Chrast R: **Global transcript expression profiling by Serial Analysis of Gene Expression (SAGE)**. *Genet Eng (N Y)* 2001, **23**:201-219.
189. **The Cancer Genome Anatomy Project: SAGE genie** [http://cgap.nci.nih.gov/SAGE]
190. **National Center for Biotechnology Information: SAGEmap** [http://www.ncbi.nlm.nih.gov/projects/SAGE/]
191. **MD Anderson Cancer Center** [spi.mdanderson.org]
192. **Melbourne Brain Genome Project** [http://www.mbgproject.org/]
193. Velculescu VE, Zhou W, Traverso G, Croix BS, Vogelstein B, Kinzler KW: **Serial Analysis of Gene Expression: Detailed Protocol**. Baltimore, USA; 2000.
194. Datson NA, van der Perk-de Jong J, van den Berg MP, de Kloet ER, Vreugdenhil E: **MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue**. *Nucleic Acids Res* 1999, **27**(5):1300-1307.
195. Peters DG, Kassam AB, Yonas H, O'Hare EH, Ferrell RE, Brufsky AM: **Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite**. *Nucleic Acids Res* 1999, **27**(24):e39.
196. Wei CL, Ng P, Chiu KP, Wong CH, Ang CC, Lipovich L, Liu ET, Ruan Y: **5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation**. *Proc Natl Acad Sci U S A* 2004, **101**(32):11701-11706.
197. Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M *et al*: **CAGE: cap analysis of gene expression**. *Nat Methods* 2006, **3**(3):211-222.
198. Kim JB, Porreca GJ, Song L, Greenway SC, Gorham JM, Church GM, Seidman CE, Seidman JG: **Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy**. *Science* 2007, **316**(5830):1481-1484.
199. Nielsen KL, Hogh AL, Emmersen J: **DeepSAGE--digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples**. *Nucleic Acids Res* 2006, **34**(19):e133.
200. Gowda M, Li H, Alessi J, Chen F, Pratt R, Wang GL: **Robust analysis of 5'-transcript ends (5'-RATE): a novel technique for transcriptome analysis and genome annotation**. *Nucleic Acids Res* 2006, **34**(19):e126.
201. Sun T, Patoine C, Abu-Khalil A, Visvader J, Sum E, Cherry TJ, Orkin SH, Geschwind DH, Walsh CA: **Early asymmetry of gene transcription in embryonic human left and right cerebral cortex**. *Science* 2005, **308**(5729):1794-1798.
202. Tan SS, Gunnersen J, Job C: **Global gene expression analysis of developing neocortex using SAGE**. *Int J Dev Biol* 2002, **46**:653-660.
203. Peters BA, St Croix B, Sjoblom T, Cummins JM, Silliman N, Ptak J, Saha S, Kinzler KW, Hatzis C, Velculescu VE: **Large-scale identification of novel transcripts in the human genome**. *Genome Res* 2007, **17**(3):287-292.
204. Gunnersen JM, Augustine C, Spirkoska V, Kim M, Brown M, Tan SS: **Global analysis of gene expression patterns in developing mouse neocortex using serial analysis of gene expression**. *Mol Cell Neurosci* 2002, **19**(4):560-573.

205. Boon WM, Beissbarth T, Hyde L, Smyth G, Gunnersen J, Denton DA, Scott H, Tan SS: **A comparative analysis of transcribed genes in the mouse hypothalamus and neocortex reveals chromosomal clustering**. *Proc Natl Acad Sci U S A* 2004, **101**(41):14972-14977.
206. Hirst M, Delaney A, Rogers SA, Schnerch A, Persaud DR, O'Connor MD, Zeng T, Moksa M, Fichter K, Mah D *et al*: **LongSAGE profiling of nine human embryonic stem cell lines**. *Genome Biol* 2007, **8**(6):R113.
207. Bosma AJ, Weigelt B, Lambrechts AC, Verhagen OJ, Pruntel R, Hart AA, Rodenhuis S, van 't Veer LJ: **Detection of circulating breast tumor cells by differential expression of marker genes**. *Clin Cancer Res* 2002, **8**(6):1871-1877.
208. Voth H, Oberthuer A, Simon T, Kahlert Y, Berthold F, Fischer M: **Identification of DEIN, a novel gene with high expression levels in stage IVS neuroblastoma**. *Mol Cancer Res* 2007, **5**(12):1276-1284.
209. Cho M, Ishida K, Chen J, Ohkawa J, Chen W, Namiki S, Kotaki A, Arai N, Arai K, Kamogawa-Schifter Y: **SAGE library screening reveals ILT7 as a specific plasmacytoid dendritic cell marker that regulates type I IFN production**. *Int Immunol* 2008, **20**(1):155-164.
210. Ruzanov P, Jones SJ, Riddle DL: **Discovery of novel alternatively spliced C. elegans transcripts by computational analysis of SAGE data**. *BMC Genomics* 2007, **8**:447.
211. Oberthuer A, Skowron M, Spitz R, Kahlert Y, Westermann F, Mehler K, Berthold F, Fischer M: **Characterization of a complex genomic alteration on chromosome 2p that leads to four alternatively spliced fusion transcripts in the neuroblastoma cell lines IMR-5, IMR-5/75 and IMR-32**. *Gene* 2005, **363**:41-50.
212. Sanger F, Coulson AR: **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. *J Mol Biol* 1975, **94**(3):441-448.
213. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proc Natl Acad Sci U S A* 1977, **74**(12):5463-5467.
214. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**(6822):860-921.
215. **Applied Biosystems** [http://www.appliedbyosystems.com]
216. **Roche Applied Science** [http://www.rocheapplied-science.com]
217. **Illumina Inc.** [http://www.illumina.com/]
218. **Helicos Biosciences Corporation** [http://www.helicosbio.com/]
219. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M *et al*: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays**. *Nat Biotechnol* 2000, **18**(6):630-634.
220. Mardis ER: **Anticipating the 1,000 dollar genome**. *Genome Biol* 2006, **7**(7):112.
221. Ng P, Tan JJ, Ooi HS, Lee YL, Chiu KP, Fullwood MJ, Srinivasan KG, Perbost C, Du L, Sung WK *et al*: **Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes**. *Nucleic Acids Res* 2006, **34**(12):e84.
222. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ: **Elucidation of the small RNA component of the transcriptome**. *Science* 2005, **309**(5740):1567-1569.

223. Kappler J, Stichel CC, Gleichmann M, Gillen C, Junghans U, Kresse H, Muller HW: **Developmental regulation of decorin expression in postnatal rat brain**. *Brain Res* 1998, **793**(1-2):328-332.
224. Ford-Perriss M, Guimond SE, Greferath U, Kita M, Grobe K, Habuchi H, Kimata K, Esko JD, Murphy M, Turnbull JE: **Variant heparan sulfates synthesized in developing mouse brain differentially regulate FGF signaling**. *Glycobiology* 2002, **12**(11):721-727.
225. Richard D, Huang Q, Sanchis D, Ricquier D: **Brain distribution of UCP2 mRNA: in situ hybridization histochemistry studies**. *Int J Obes Relat Metab Disord* 1999, **23 Suppl 6**:S53-55.
226. Qiu P, Benbow L, Liu S, Greene JR, Wang L: **Analysis of a human brain transcriptome map**. *BMC Genomics* 2002, **3**:10.
227. Nobis W, Ren X, Suchyta SP, Suchyta TR, Zanella AJ, Coussens PM: **Development of a porcine brain cDNA library, EST database, and microarray resource**. *Physiol Genomics* 2003, **16**(1):153-159.
228. Majda BT, Meloni BP, Rixon N, Knuckey NW: **Suppression subtraction hybridization and northern analysis reveal upregulation of heat shock, trkB, and sodium calcium exchanger genes following global cerebral ischemia in the rat**. *Brain Res Mol Brain Res* 2001, **93**(2):173-179.
229. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH: **Functional organization of the transcriptome in human brain**. *Nat Neurosci* 2008, **11**(11):1271-1282.
230. Mirnics K, Levitt P, Lewis DA: **DNA microarray analysis of postmortem brain tissue**. *Int Rev Neurobiol* 2004, **60**:153-181.
231. Miller JA, Horvath S, Geschwind DH: **Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways**. *Proc Natl Acad Sci U S A* 2010, **107**(28):12698-12703.
232. Twine NA, Janitz K, Wilkins MR, Janitz M: **Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease**. *PLoS One* 2011, **6**(1):e16266.
233. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE *et al*: **Mammalian microRNAs: experimental evaluation of novel and previously annotated genes**. *Genes Dev* 2010, **24**(10):992-1009.
234. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ *et al*: **Genome-wide atlas of gene expression in the adult mouse brain**. *Nature* 2007, **445**(7124):168-176.
235. **Allen Institute for Brain Science** [http://mouse.brain-map.org/]
236. Visel A, Thaller C, Eichele G: **GenePaint.org: an atlas of gene expression patterns in the mouse embryo**. *Nucleic Acids Res* 2004, **32**(Database issue):D552-556.
237. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME *et al*: **A gene expression atlas of the central nervous system based on bacterial artificial chromosomes**. *Nature* 2003, **425**(6961):917-925.
238. **GenePaint.org** [http://www.genepaint.org/]
239. Shao NY, Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Li N, Chen W, Khaitovich P: **Comprehensive survey of human brain microRNA by deep sequencing**. *BMC Genomics* 2010, **11**:409.

240. Hartl D, Irmler M, Romer I, Mader MT, Mao L, Zabel C, de Angelis MH, Beckers J, Klose J: **Transcriptome and proteome analysis of early embryonic mouse brain development**. *Proteomics* 2008, **8**(6):1257-1265.
241. Committee on Nervous System Disorders in Developing Countries BoGH, Institute of Medicine: **Neurological, Psychiatric, and Developmental Disorders: Meeting the Challenge in the Developing World**, 1st edn. Washington, D. C.: The National Academy Press; 2001.

CHAPTER 2

Materials and methods

2.1 Introduction

This chapter presents general protocols used in routine experimental setups. Unless otherwise specified in Chapters 3-6, all materials and methods described here were utilised throughout all the studies presented in the thesis. Manufacturers of kits, reagents, chemicals, instruments or service providers cited in this chapter are listed below:

1. Abcam – Sapphire Bioscience Pty Ltd, Waterloo, NSW, Australia.
2. Ajax Finechem – Ajax Finechem Pty Ltd, Taren Point, NSW, Australia.
3. Ambion® – Applied Biosystems, Scoresby, VIC, Australia.
4. Bio-Rad – Bio-Rad Laboratories (Pacific) Pty Ltd, Gladesville, NSW, Australia.
5. Carl Zeiss – Carl Zeiss Inc., Oberkochen, Germany.
6. Clonetech – Scientifix Pty Ltd, VIC, Australia.
7. Dako – Dako Australia Pty Ltd, Campbellfield, VIC, Australia.
8. Exiqon – GeneWorks, Hindmarsh, SA, Australia.
9. GE Healthcare – GE Healthcare Bio-Sciences Pty Ltd, Rydalmere, NSW, Australia.
10. GeneWorks – GeneWorks, Hindmarsh, SA, Australia.
11. Invitrogen – Invitrogen Australia Pty Ltd, Mulgrave, VIC, Australia.
12. Kodak – Kodak, Sydney, NSW, Australia.
13. Kreatech – Diagnostic Technology Pty Ltd, Belrose, NSW, Australia.
14. Menzel-Glaser – Thermo Fisher Scientific, Scoresby, VIC, Australia.
15. Olympus – Olympus Australia Pty Ltd, Richmond, SA, Australia.
16. Pierce – Thermo Fisher Scientific, Scoresby, VIC, Australia.
17. Promega – Promega Corporation, Alexandria, NSW, Australia.
18. ProSciTech – ProSciTech, Thuringowa, QLD, Australia.
19. QIAGEN – Q AGEN Pty Ltd, Doncaster, VIC, Australia.
20. Roche Diagnostics – Roche Diagnostics Australia Pty Ltd, Castle Hill, NSW, Australia.
21. Santa Cruz – Quantum Scientific, Murarrie, QLD, Australia.
22. Savant Instrument – Aurora BioScience Pty Ltd, Baulkham Hills, NSW, Australia.

23. Sigma Aldrich – Sigma-Aldrich Pty Ltd, Sydney, NSW, Australia.
24. Stratagene – Integrated Sciences, Melbourne, VIC, Australia.
25. Thermo Scientific – Thermo Fisher Scientific, Scoresby, VIC, Australia.
26. Vector Laboratories – Abacus ALS, East Brisbane, QLD, Australia.

2.2 Animals handling and tissue processing

2.2.1 Animals

All experiments were performed according to protocols approved by the Melbourne Health Animal Ethics Committee (Project numbers 2001.045 and 2004.041) and the University of Adelaide Animal Ethics Committee (S-086-2005). All mice used in the study were C57BL/6 unless otherwise specified. Mice were kept under a 12 hour light/12 hour dark cycle with unlimited access to food and water and were culled by either CO_2 asphyxiation or by cervical dislocation. Female mice were considered pregnant when the appearance of vaginal plugs were detected, at which time the embryos were designated embryonic day E0.5.

2.2.2 Procurement of mouse tissues

Tissues procurement was carried out in a sterile environment. Procurement of the mouse cerebral cortex is described in Chapter 3. Procurement of other mouse organs was carried out according to the standard mouse necropsy protocol accessible at National Institute of Allergy and Infectious Diseases (NIAID) website [1].

2.2.3 Tissue processing

Tissues for the preparation of paraffin sections were fixed in 4% (w/v) paraformaldehyde (PFA, pH7.4) in phosphate buffered saline (PBS) solution. Embryonic cerebral cortices, whole brains or whole embryos were fixed for 2 days at 4^{o}C whereas adult brains were initially perfused with 4% (w/v) PFA followed by 2 days fixation in the same solution at 4^{o}C. For total RNA or protein lysate preparation, tissues were snap frozen on dry ice with minimal amount of PBS solution and stored at -80^{o}C until subsequent extraction step was carried out.

For whole mount *In situ* RNA Hybridisation, mouse embryos were dissected in cold PBS and fixed in 4% (w/v) PFA in PBS followed by washes in PBS (3x for 5 minutes each). Embryos were then dehydrated in a series of methanol starting with 25% (v/v) methanol and finishing with washes in 100% (v/v) methanol prepared in PBS with 0.1% (v/v) Tween-20 (PBT). Embryos were stored at -20^{o}C until used.

2.3 Preparation and manipulation of nucleic acids

2.3.1 DNA isolation

All genomic DNA (gDNA) extraction from tissues or cell lines was carried out using QIAGEN DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's protocol. Plasmid DNA isolation was performed using QIAGEN Plasmid DNA Mini or Midi Kits (QIAGEN) according to the manufacturer's protocol.

2.3.2 Total RNA isolation

Small RNA free total RNAs were extracted from the mouse cerebral cortex, whole brain or whole embryo using RNeasy Lipid Tissue Mini or Midi Kits (QIAGEN) according to the manufacturer's protocol. Small RNA enriched total RNAs were extracted from similar tissues using TRIzol® Reagent (Invitrogen) according to the manufacturer's protocol. For extraction of small RNA enriched total RNA from small quantities of tissues, 200μg of glycogen was added into the isopropyl alcohol precipitation step.

2.3.3 DNA or RNA purification

DNA purification was performed using either QIAquick PCR Purification or Gel Extraction Kits (QIAGEN) according to the manufacturer's protocol. In other circumstances, manual purification or concentration of DNA was carried out using isopropyl alcohol method. Briefly, three volumes of isopropyl alcohol was added into the DNA, mixed and incubated for 20 minutes at room temperature.

The mixture was centrifuged for 1 hour at 4°C followed by a 75% (v/v) ethanol wash before the DNA pellet was dissolved in nuclease-free water.

RNA purification or concentration steps were performed using RNeasy MinElute Cleanup Kit (QIAGEN) according to the manufacturer's protocol. In other circumstances, concentration of RNA was also carried out in Savant SVC 100H Speed Vacuum Concentrator (Savant Instrument).

2.3.4 Reverse-transcription of total RNA and small RNA

The concentration and purity of isolated total RNA was measured using ND2000 NanoDrop Spectrophotometer (Thermo Scientific) and only RNA with A260/A280 between 1.8-2.1 was used for complementary DNA (cDNA) synthesis. cDNA was synthesised from 50ng-3µg of high quality total RNA using either oligo-d$[T]_{15}$ or random hexamers, and the SuperScript® III Reverse Transcriptase Kit (Invitrogen) according to the manufacturer's protocol.

Reverse transcription of small RNA was performed based on modified methods [2, 3] first published in Brown *et al*, 2010 [4]. CDNA was synthesised from 150ng-1µg of small RNA enriched total RNA using 0.05µM of an in-house designed stem loop primer (5'-GTTGGCTCT GGTAGGATG CCGCTCTCA GGGCATCCT ACCAGAGCCA ACNNNNNN-3', GeneWorks) and the Superscript® III Reverse Transcriptase Kit (Invitrogen) according to the manufacturer's protocol. The last 6bp at the 3' end complements with the last 6bp at the 3' end of a specific small RNA. The stem loop RT primer contains a target site for a universal reverse primer (5'-GTAGGATGCC GCTCTCAGG-3', GeneWorks), which was used in subsequent cDNA amplication step together with a specific forward primer. List of stem loop primers, their corresponding small RNA targets and the specific forward primer are listed in Table 2.1. Briefly, cDNA synthesis was performed at 16°C for 30 minutes followed by 60 cycles of 20°C for 30 seconds, 42°C for 30 seconds and 50°C for 1 second. Final incubation at 75°C for 15 minutes was performed to inactivate the reverse transcriptase enzyme.

Table 2.1: List of stem loop RT primers, targeted small RNAs and specific forward primers.

Specific forward primer (5'→3')	Stem loop primer (only the modified unique sequences are shown)	Targeted small RNA
AGGCGGAGAG TAGACGGG	5′ – G…… CCCGTC – 3′	*Sox4_sir1*
CCACTGGGGT TGTACGAA	5′ – …… TTCGTA – 3′	*Sox4_sir2*
TCAAGGACAG CGACAAGATT C	5′ – …… ACGGAATC – 3′	*Sox4_sir3*
TCAGGGAAAG GGGTGGGGGA	5′ – …… TCCCCCAC – 3′	*Sox4_sir4*
AGACGATGTC GCTTTCCTGA	5′ – …… TCAGGA – 3′	*Sox4_sir5*
CGGTAGGACT TAGGCGCTAGAG	5′ – G…… CTCTAG – 3′	*Sox4_sir6*
AGGCGCTAGA GACGATGT	5′ – …… ACATCG – 3′	*Sox4_sir7*
CGCGTAGGCT AGAGAGAGGT	5′ – …… TCCCCA – 3′	*M1181*

2.3.5 Restriction enzyme digestion

List of restriction enzymes used, their manufacturers, sources and restriction conditions are summarised in Chapters 3-6 and Appendices A-D. All restriction enzyme reactions were prepared according to their respective manufacturers' protocols and were carried out at 37^{o}C on a heating block followed by heat inactivation at 75^{o}C for 20 minutes.

2.3.6 DNA cloning

Double-stranded DNA with 3'-T/A overhang ends were ligated into pGEM-T Easy vector using the pGEM-T Easy Vector System Kit (Promega) according to the manufacturer's protocol. For overexpression of *Sox4* antisense transcripts studies in Chapter 5, blunt-end amplicons from long range PCR reactions were first treated with T4 Polynucleotide Kinase (Promega) according to the manufacturer's protocol before cloned into pcDNA3 vector (Invitrogen) using T4 DNA Ligase and buffer system provided in pGEM-T Easy Vector System Kit (Promega).

Irrespective of vector used, sticky end ligation was carried out at room temperature for 15 minutes whereas blunt-end ligation was carried out at 16^{o}C for 16 hours. After the ligation reaction, ligated product was transformed into JM109 chemically competent cells (Promega) using the heat-shock approach. Briefly, ligated product was gently mixed with JM109 cells and incubated for 20 minutes

on ice followed by 45 seconds at 42°C in water bath and back on ice for additional 2 minutes. Transformed cells were pre-cultured in Super Optimal Broth with Catabolite repression (SOC) medium (Invitrogen) at 37°C for 2 hours followed by plating an appropriate amount on Luria Bertani (LB) agar plate with 100µg/ml of ampicillin (Sigma Aldrich), 0.1M isopropyl thiogalactoside (IPTG, Sigma Aldrich) and 20mg/ml of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal, Sigma Aldrich). LB agar plates containing transformed cells were incubated at 37°C for 16-18 hours. Recombinant clones were chosen based on the blue-white selection approach. White colonies were picked and dipped into 100µl LB broth supplied with 100 µg/ml of ampicillin (Sigma Aldrich) and a 20µl PCR reaction mix for screening for the clone with correct insert size and orientation. Details of colony PCR are described in Section 2.5.1.

2.5 Polymerase Chain Reaction (PCR)

2.5.1 Routine PCR

PCR reactions for routine genetic screening were performed in 10-50µl volume containing 0.4-2.0U of FastStart Taq DNA Polymerase, 1X concentration PCR reaction buffer with 2mM $MgCl_2$, 200µM of each dNTP (Roche Diagnostics) and 0.25µM of each forward and reverse primers. PCR was performed with initial denaturation at 95°C for 5 minutes followed by 10 cycles at 95°C for 10 seconds, 65°C (-1°C per cycle) for 30 seconds and 72°C for 30-60 seconds (depending on the size of amplicon), additional 20-30 cycles of similar conditions but with an annealing step at 55°C and a final elongation step at 72°C for 7 minutes.

For colony PCR screening of specific recombinant clones, T7 (5'-GTAATACGAC TCACTATAGG GC -3', GeneWorks) or SP6 (5'-ATTTAGGTGA CACTATAGAA TACTC-3', GeneWorks) universal primers were used in combination of a gene specific primer for orientation specific screening. Colony PCR reaction consists of 1X LightCycler 480 (LC480) Probe Master mix (Roche Diagnostics) and 0.25µM of each forward and reverse primers. Similar conditions for routine PCR reactions were adopted with 10 cycles touch down PCR steps were carried out from 60°C to 50°C (-1°C per cycle) followed by additional 30 cycles of PCR at annealing temperature of 50°C.

2.5.2 Real-Time quantitative PCR (RT-qPCR)

Real-Time quantitative PCR (RT-qPCR) was carried out according to the approach described elsewhere [5]. A 10µl reaction volume consisting of 1X LightCycler 480 (LC480) Probe Master mix (Roche Diagnostics), 0.1µM of a relevant Universal ProbeLibrary probe (Roche Diagnostics), 0.25µM of each forward and reverse primers and 1µl of 0.1X of synthesised cDNA was prepared. Reactions were prepared in 384-well plate and RT-qPCR was performed using LightCycler® 480 Real Time PCR System instrument (Roche Diagnostics). RT-qPCR was performed with an initial denaturation at 95^{o}C for 10 minutes followed by 45 cycles at 95^{o}C for 10 seconds, 60^{o}C for 30 seconds and 72^{o}C for 10 seconds, and a final step at 40^{o}C for 1 second.

Real-Time amplification signals were acquired during the elongation step and recorded live using LightCycler® 480 Software version 1.5 (Roche Diagnostics). The cycle threshold or crossing point (Cp) from each signal was calculated based on the Second Derivative Maximum method [6]. A 4-data point standard curve was constructed using a serially diluted pooled cDNAs for each primer sets used in RT-qPCR in each run. The standard curve was used to determine the PCR efficiency and reproducibility of each PCR system. *Pgk1*, *Hmbs*, *Psmb2* and *Hprt1* genes were routinely used for data normalisation. Primer sequences, cutoff criteria, detailed data normalisation, linear modeling and empirical Bayesian moderated T-test used in the analysis were described in Chapter 3 and Appendix A [7]. Cp comparative method was used in RT-qPCR analysis in Chapter 3 whereas relative quantification method was adopted for other chapters.

2.5.3 Long-ranged PCR

Amplification of long amplicons was carried out in 25-50µl reaction volume using Expand Long Template PCR System Kit (Roche Diagnostics) according to the manufacturer's protocol. Briefly, PCR reactions were prepared in the presence of 3.75U of Expand Long Template enzyme mix, 350µM of each dNTPs, 300nM of each forward and reverse primers, 1X PCR buffer with $MgCl_2$ system I, II or III

and appropriate amount of cDNA (1µl of 0.1-1X of synthesised cDNA) or gDNA (50-500ng). Thermal cycling was carried out with an initial denaturation step at 94ºC for 2 minutes, 10 cycles of 94ºC for 10 seconds, 55ºC-65ºC (depending on primers used) for 30 seconds, 68ºC for 1-8 minutes (depending on the amplicon's size, ~1-6kb), followed by additional 25 cycles of similar PCR cycle conditions but increasing elongation time at 68ºC (additional 20 seconds per cycle). The thermal cycling was ended with a final elongation step at 68ºC for 10 minutes.

2.6 Rapid Amplification of cDNA Ends (RACE)

For 3' Rapid Amplification of cDNA Ends (RACE) analysis in Chapter 3, a custom made oligo-d$[T]_{15}$ with adaptor (5'-TACGACGTCT GCTAGGACTG-3', GeneWorks) was used during total RNA reverse transcription step. Protocols for reverse transcription and amplification steps are provided in Section 2.3.4, Chapter 3 and Appendix A [7]. RACE analysis in other chapters were carried out using SMART™ or SMARTer™ RACE Amplification Kit (Clonetech) according to the manufacturer's protocol. After reverse transcription step, subsequent amplification reactions were carried out using the Advantage® 2 PCR Enzyme System Kit (Clonetech) according to manufacturer's protocol.

2.7 *In situ* hybridisation (ISH)

2.7.1 Preparation of probes

Relevant clones from the National Institute on Aging (NIA) 15K Mouse cDNA Clone Set (May 16, 2000) were purchased from Australian Genome Research Facilities (AGRF), Australia, for *in vitro* transcription. Other genes that were not available from the clone set were constructed by cloning the specific PCR amplicon. All clones used are described in Chapters 3-6 and Appendices A-D [7]. For RNA Fluorescent *In situ* Hybridisation (FISH), probes for *Sox4*, *Nrgn*, *Camk2n1* and *Hmbs* were prepared from DNA amplicons amplified from either mouse gDNAs or cDNAs. Primers used for RNA FISH probes preparation are listed in Table 2.2. All clones were linearised using restriction enzymes described in Section 2.3.5.

Table 2.2: Primers used to amplify *Sox4*, *Nrgn*, *Camk2n1* and *Hmbs* amplicons for RNA FISH probe preparation.

Probe sequence (5'→3')	Description	Amplicon size
GGTTCGAAGT TAAAATCCAG GTC	*Sox4* forward primer 1	761bp
GGAGAAGGGC GACAAGGT	*Sox4* reverse primer 1	
TGATGTTGGT GGTGGCTAAA	*Sox4* forward primer 2	614bp
TTTGCACAGA CCCCAGGCGG AG	*Sox4* reverse primer 2	
AACACCGGCA ATGGACTG	*Nrgn* forward primer	863bp
CAGGCAGGTA TGGGATAGGA	*Nrgn* reverse primer	
TGAACCACTT CCCGAAAGC	*Camk2n1* forward primer	729bp
ATTCCTCACG CAAGAGGCAT T	*Camk2n1* reverse primer	
GCCTGTTTAC CAAGGAGCTA GA	*Hmbs* forward primer	1016bp
AAGGTGAGGC ATATCTTCCA AA	*Hmbs* reverse primer	

Non-radioactive double-end digoxigenin (DIG)-labeled miRCURY™Locked Nucleic Acid (LNA) probes for detection of small RNAs were custom made and supplied by Exiqon (GeneWorks). These LNA probes are listed in Table 2.3.

Table 2.3: Custom made double-end DIG-labeled LNA probe from Exiqon.

Probe sequence (5'→3')	Small RNA target	Exiqon Catalog Number	Melting temperature
TCCCCAACCT CTCTCTAGCC TA	*M1181*	EQ-70532	75°C
ACGGAATCTT GTCGCTGTCC TTGA	*Sox4-sir3*	EQ-70537	81°C
GTGTAACACG TCTATACGCC CA	*Scramble-miR*	EQ-67685	78°C

Radioactive-labeled single stranded complementary RNA (cRNA) probes were *in vitro* transcribed using Riboprobe® Combination System - SP6/T7 RNA Polymerase (Promega) and [^{35}S]-dUTP (GE Healthcare) according to the manufacturer's protocol. After *in vitro* transcription, the probe was precipitated (using 0.9M of sodium perchlorate, 0.18μg/ul of *Escherichia coli* (*E. coli*) tRNA and 43% (v/v) isopropanol) followed by a wash in 70% (v/v) ethanol. The pelleted probe was hydrolysed (using 40mM of sodium bicarbonate, 60mM of sodium carbonate and 10mM of β-mercapto-ethanol at pH10) at 60°C for 30-40 minutes before the reaction was stopped (using 3.75M of sodium perchlorate and

2% (v/v) glacial acetic acid). The hydrolysed probe was precipitated and resuspended in 10mM dithiothreitol (DTT). The probe was stored at -20^{o}C until used (within the same week).

RNA probes for FISH were fragmented by sonication and labelled using Platinum*Bright*™Nucleic Acid Labeling Kit (Kreatech) according to the manufacturer's protocol. Two different Universal Linkage System (ULS) haptens/dyes were used for labeling of RNA probes: Dynomics547-ULS (absorbance at 547nm and emission at 565nm) for the sense RNA probe and Dynomics415-ULS (absorbance at 415nm and emission at 472nm) for the antisense RNA probe.

2.7.2 Section radioactive RNA ISH

Paraffin embedded sections (8μm) were used for radioactive RNA ISH. All steps were performed in an RNAse-free environment and at room temperature unless otherwise specified. Sections were de-paraffinised with washes in histolene (2x for 5 minutes each) followed by 100% (v/v) ethanol (2x for 2 minutes each), 96% (v/v) ethanol (2 minutes), 90% (v/v) ethanol (2 minutes), 70% (v/v) ethanol (2 minutes) and 0.86% (w/v) NaCl (5 minutes). Sections were then digested in Proteinase K buffer (20mM of Tris HCl pH7.5, 1mM of ethylenediaminetetraacetic acid (EDTA) and 20μg/ml of Proteinase K) for 10 minutes followed by dehydration steps involving washes in 0.86% (w/v) of NaCl, 70% (v/v) ethanol (2 minutes), 90% (v/v) ethanol (2 minutes) and 100% (v/v) ethanol (2 minutes). Sections were air dried and ready for the hybridisation step.

Hybridisation was performed in hybridisation solution (0.02% (w/v) of Ficoll, 0.02% (w/v) of polyvinylpyrrolidone, 0.02% (w/v) of BSA fraction V, 0.138% (w/v) of monosodium phosphate monohydrate, 5mM of EDTA, 0.3M of NaCl, 0.01M of Tris HCl pH7.5, 10% (w/v) dextran sulfate and 50% (v/v) formamide) containing [^{35}S]-UTP (GE Healthcare) labeled RNA probe with activity of 1x10^{7} counts per minute (CPM). The hybridisation solution was denatured at 80^{o}C for 2 minutes before added onto sections and covered with coverslip. Hybridisation was performed in a humidified chamber (50% (v/v) formamide) in Hybaid hybridisation oven (Thermo Scientific) at 55^{o}C for 16 hours.

After hybridisation, sections were washed in SSCF solution (50% (v/v) formamide, 2X sodium chloride/sodium citrate (SSC) and 0.14% (v/v) β-mercaptol-ethanol; 2x for 30 minutes each at 60°C) followed by NTE solution (0.5M of NaCl, 10mM of Tris HCl pH7.5 and 5mM of EDTA; 10 minutes at 37°C), NTE solution with added 20μg/ml of RNAse A (30 minutes at 37°C), SSCF (30 minutes at 60°C), 2X SSC (30 minutes at 60°C) and 0.1X SSC (30 minutes at 60°C). Sections were then dehydrated in 70% (v/v) ethanol (2 minutes), 90% (v/v) ethanol (2 minutes), 96% (v/v) ethanol (2 minutes) and finally in 100% (v/v) ethanol (2 minutes).

In a darkroom, slides containing probed sections were dipped into 50% (v/v) Kodak NTB-2 emulsion (Kodak) with 1% (v/v) glycerol prepared in water (at 42°C) and were allowed to dry in dark for about 2 hours before placed in a dark slide box with desiccants. Liquid film was exposed for 14-21 days in refrigerator (4°C). After appropriate exposure, liquid film coated slides were developed in 16% (w/v) of Kodak D-19 (Kodak) developer (3 minutes at 42°C) followed by washes in 1% (v/v) acetic acid (1 minute), 30% (v/v) sodium thiosulfate (5 minutes) and finally in tap water (10 minutes). Slides were counterstained with haematoxylin for 45 seconds, then washed in running tap water and dehydrated through a series of ascending ethanol concentrations (70%, 90%, 96% and 100% (v/v) ethanol; 2 minutes each). Slides were cleared in histolene and coverslips were mounted using DEPEX mounting media (ProSciTech). Slides were stored in room temperature until microscopic analysis.

2.7.3 Section Locked Nucleic Acid (LNA) ISH

Paraffin embedded sections (8μm) were used for LNA ISH. All steps were performed in an RNAse-free environment and at room temperature unless otherwise specified. Sections were de-paraffinised with washes in xylene (3x for 5 minutes each), 100% (v/v) ethanol (2x for 5 minutes each), 70% (v/v) ethanol (5 minutes), 50% (v/v) ethanol (5 minutes), 25% (v/v) ethanol (5 minutes) and diethylpyrocarbonate (DEPC)-treated water (1 minute). Subsequently, sections were re-fixed in 4% (w/v) PFA (pH7.0) in PBS (10 minutes) followed by washes in PBS (3x for 5 minutes each). Proteinase K digestion was carried out in

Proteinase K buffer (6.7µg/ml of Proteinase K, 50mM of Tris HCl pH7.5, 5mM of EDTA) for 30 minutes. After digestion, sections were re-fixed in 4% (w/v) PFA in PBS for 5 minutes followed by washes in PBS (3x for 5 minutes each). Acetylation step was carried out in acetylation buffer (0.1M of triethanolamine, 0.178% (v/v) of concentrated HCl and 0.25% (v/v) of acetic anhydride) followed by additional washes in PBS (3x for 5 minutes).

Pre-hybridisation step was carried out in a humidified chamber (50% (v/v) formamide, 5X sodium chloride/sodium citrate, SSC) inside a Hybaid hybridisation oven (Thermo Scientific) at 21-22^{o}C lower than the calculated melting temperature for LNA probes. Amersham Rapid-hyb™ Buffer (GE Healthcare) was used for pre-hybridisation with additional *E. coli* tRNA (Sigma Aldrich) and Herring Sperm DNA (Promega) to a final concentration of 0.1µg/µl each. After 1-2 hours of pre-hybridisation, LNA probes were added to the buffer to a final concentration of 0.020pmol/µl. Hybridisation was carried out in the oven for 16-20 hours.

After hybridisation step, sections were washed in 5X SSC (20 minutes at hybridisation temperature) followed by 0.2X SSC (3 hours at hybridisation temperature). Sections were then rinsed in fresh 0.2X SSC for 5 minutes and in pre-blocking buffer (0.1M of Tris HCl pH7.5, 0.15M of NaCl and 240µg/ml of levamisole) for another 5 minutes. In a humidified chamber, sections were blocked in 20% (v/v) foetal calf serum (Sigma Aldrich) and 2% (w/v) blocking powder (Roche Diagnostics) in maleate buffer for 1 hour. After blocking, sections were incubate with 0.0002X (0.00015U) anti-DIG antibody with alkaline phosphatase, Fab fragments (Roche Diagnostics) in blocking buffer for 16 hours in dark. Consequently, sections were washed in NTMT buffer (3x for 10 minutes each; 0.1M Tris HCl pH9.5, 0.1M NaCl, 0.05M $MgCl_2$, 1% (v/v) Tween-20 and 240µg/ml levamisole) followed by nitro blue tetrazolium chloride (NBT)/ 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP) colour reaction (0.375mg/ml of NBT and 0.188mg/ml of BCIP in prepared in NTMT buffer) for 3 hours to 5 days. After colour reaction step, sections were washed with Tris EDTA buffer pH8.0 (0.01M of Tris HCl pH7.5 and 0.001M EDTA pH8.0) for 10 minutes and were mounted in Entellan® media (ProSciTech). Sections were stored in dark until microscopic examination.

2.7.4 Whole-mount LNA ISH

All steps were performed in an RNAse-free environment and at room temperature unless otherwise specified. Embryos were rehydrated in a series of methanol concentrations prepared in PBT starting from 75% (v/v) followed by 50% (v/v) and 25% (v/v) methanol (5 minutes wash each). Embryos were then bleached in 6% (v/v) hydrogen peroxide (Ajax Finechem) (1 hour) and digested with 10µg/ml proteinase K in PBT (15 minutes). Embryos were subjected to washes in fresh 2mg/ml glycine (5 minutes) followed by fixation in 0.2% (v/v) glutaraldehyde in 4% (v/v) PFA prepared in PBT (20 minutes) and washes in PBT (2x for 5 minutes each).

Pre-hybridisation step was carried out in Hybaid hybridisation oven (Thermo Scientific) in pre-hybridisation buffer (50% (v/v) formamide, 5X SSC, 0.1% Tween-20, 1µg/ml porcine heparin) for 1 hour at 70^{o}C. For hybridisation, LNA probes were added to the buffer to a final concentration of 0.020pmol/µl. Other blocking agents such as 0.1µg/µl of each *E. coli* tRNA (Sigma) and Herring Sperm DNA (Promega) were also added into the hybridisation buffer. Hybridisation was carried out in the oven for 16-20 hours at a temperature of 21-22^{o}C lower than the calculated melting temperature for LNA probe used.

An initial washing step was carried out using hybridisation buffer (without LNA probes and blocking agents, thereafter known as HB) at hybridisation temperature (3 minutes) followed by 15 minutes wash at hybridisation temperature in each 75% (v/v) HB in 2X SSC, 50% (v/v) HB in 2X SSC, 25% (v/v) HB in 2X SSC and 2X SSC buffers. Additional washes were carried out at hybridisation temperature for 30 minutes in 0.2X SSC followed by washes at room temperature for 10 minutes in each 25% (v/v) PBST in 0.2X SSC, 50% (v/v) PBST in 0.2X SSC, 75% PBST in 0.2X SSC and PBST buffers. Embryos were then pre-blocked with 10% (v/v) heat-inactivated sheep serum (HISS) prepared in 1X TBST (0.1M of Tris HCl pH7.5, 0.5M of NaCl and 0.1% (v/v) Tween-20) for 2 hours followed by a 16-20 hours 4^{o}C incubation with 0.0005X (0.000375U) anti-DIG antibody with alkaline phosphatase, Fab fragments (Roche Diagnostics) in 1% (v/v) HISS prepared in 1X TBST. Consequently, embryos were washed in 1X TBST (8x for

20 minutes each) and left in 1X TBST washing buffer at 4°C for 16-20 hours followed by washes in NTMT buffer (3x for 10 minutes each) and colour reaction in NBT/BCIP solution (0.45mg/ml of NBT and 0.175mg/ml of BCIP) prepared in NTMT for 3 hours to 3 days. After the completion of colour reaction, embryos were washed with NTMT (2x for 10 minutes each) in dark followed by 16-20 hours washing in PBT solution before fixed in 0.2% (v/v) glutaraldehyde in 4% (v/v) PFA prepared in PBT (1 hour). Embryos were washed with PBT solution (2x for 5 minutes each) to get rid of fixatives and stored at 4°C in PBS until microscopic examination.

2.7.5 Tri-colour DNA/RNA Fluorescent ISH (FISH)

Dissected brain tissues were digested in 0.025% (w/v) of trypsin prepared in PBS for 5-10 minutes at 37°C with occasional pippetting to homogenise the samples. Immediately after digestion, cells were pelleted and washed with PBS followed by 30 minutes fixation in 4% (w/v) PFA in PBS at 4°C. Then, cells were smeared directly onto SuperFrost® Plus microscope glass slide (Menzel-Glaser) and dried on 37°C heating block. Cells were dehydrated through a series of ethanol concentrations starting from 70% (v/v) ethanol followed by 80% (v/v) ethanol, 95% (v/v) ethanol and 100% (v/v) ethanol at room temperature for 2 minutes each step. Slides were dried on 37°C heating block.

Approximately 200µl of hybridisation buffer (2X SSC, 2mg/ml of bovine serum albumin, 10% (v/v) dextran sulfate and 50% (v/v) formamide) containing 100µg/ml of Herring sperm DNA (Promega) and all the labeled DNA, sense and antisense RNA probes to a final concentration of 30ng/ml each. Coverslip was placed over the buffer and the edges were sealed with rubber cement. Slides with hybridisation buffer were denatured at 85°C for 10 minutes followed by 16-20 hours incubation in a humidified chamber (50% (v/v) formamide, 5X SSC) at 55°C. After hybridisation, coverslips were carefully removed and slides were washed in 2X SSC with 50% (v/v) formamide at 55°C (3x for 5 minutes each) followed by 2X SSC at 55°C (3x for 5 minutes each), 1X SSC at room temperature for 10 minutes. Slides were dried on 37°C heating block and coverslips were mounted with VECTASHIELD mounting medium with 1.5µg/ml of 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories). The edges of the

coverslip were sealed with fast drying nail polish. Slides were stored at 4°C in dark until microscopic analysis.

2.8 Southern analysis

2.8.1 Preparation of probes

Oligonucleotide or PCR amplicon probes were used for southern blotting analysis. Probes used for southern analysis in Chapters 3-5 are described in Appendices A-D [7]. For oligonucleotide probes, 50pmol of single stranded 5' dephosphorylated oligonucleotides were labeled using 20U of T4 Polynucleotide Kinase (Promega) in 1X kinase buffer (Promega) and 50pmol of [γ-^{32}P]-dATP (GE Healthcare) with 3000Ci/mmol. The mixture was incubated at 37°C for 10 minutes followed by an inactivation step (by adding 2µl of 0.5M EDTA). Labeled probes were purified using illustra MicroSpin™ G-25 Columns (GE Healthcare) according to manufacturer's protocol. For labeling of probes larger than 200bp such as PCR amplicons, the DECAprime™ II Random Priming DNA Labeling Kit (Ambion®) was used. Approximately 25ng of linearised amplicons were labeled with [α-^{32}P]-dCTP according the manufacturer's protocol. Then, probes were purified using illustra MicroSpin™ G-50 Columns (GE Healthcare) according to manufacturer's protocol. All probes were stored in lead container at -20°C until used (within 3 days).

2.8.2 Membrane blotting, hybridisation, washing, visualisation and storage

PCR amplicons were electrophoresed on 1% (w/v) agarose gel prepared in Tris-borate-EDTA at 50V for 1 hour followed by 30V for approximately 16 hours. After electrophoresis, the gel was depurinated in 1.1% (v/v) HCl for 10 minutes followed by washes in denaturation buffer (1.5M NaCl and 0.5M NaOH for 30 minutes) and neutralisation buffer (1.5M NaCl and 0.5M Tris HCl pH7.5 for 30 minutes). A neutral transfer method using 10X SSC was adopted and setup according to standard protocol [8]. DNA was transferred onto Hybond-N+ nylon membrane (GE Healthcare) through capillary action for approximately 18 hours. After the transfer, membrane was briefly washed in 2X SSC and DNA was UV-crosslinked (Stratagene) onto the membrane. Membrane was then sandwiched in between filter papers and dried for 2 hours.

Pre-warmed Amersham Rapid-hyb™ Buffer (GE Healthcare) with 100µg/ml of herring sperm DNA (Promega) was added (at 125µl/cm^2) onto a glass tube containing the membrane with the DNA side facing inward. Pre-hybridisation step was carried out at 42°C for oligonucleotide (20 minutes) or at 65°C for large probes (1 hour) in a Hybaid oven (Thermo Scientific). After pre-hybridisation, $2x10^6$ dpm/ml labeled probe was added into the buffer and was let to hybridise onto the membrane for 2 hours (oligonucleotide probe) or 18 hours (large probe). For oligonucleotide probe hybridisation, the membrane was removed and washed in 5X SSC with 0.1% (w/v) sodium dodecyl sulfate (SDS) (20 minutes at 37°C) followed by 1X SSC with 0.1% (w/v) SDS (2x for 15 minutes each at 65°C). For large probe hybridisation, the membrane was washed in 2X SSC with 0.1% (w/v) SDS (30 minutes at 65°C) followed by 1X SSC with 0.1% (w/v) SDS (30 minutes at 65°C) and 0.1X SSC with 0.1% (w/v) SDS (multiple times for 30 minutes each time at 65°C) until the recorded radioactive activity (with a Geiger counter) dropped below 10 CPM.

After the washing step, the membrane was rinsed in 2X SSC, wrapped in cling-wrap and exposed to Amersham Hyperfilm MP (GE Healthcare) in an intensifying cassette at -80°C. Exposed films were developed using automated

film developer machine. Alternatively, the membrane was exposed to a storage phosphor screen in a cassette at room temperature before it was scanned by Typhoon 9200 (GE Healthcare). To reuse the membrane for another detection, the hybridised probe was stripped in hot 0.1% (w/v) SDS solution (multiple times for 20 minutes each at 85^{o}C) until radioactive activity was no longer detectable (using Geiger counter). Stripped membrane was rinse in 2X SSC before wrapped in cling-wrap and stored in an intensifying cassette in -80^{o}C.

2.9 Northern analysis

2.9.1 Preparation of probes

Most of the probes used for northern analysis were the same probe used for ISH. Double stranded DNA (dsDNA) probes were also used in some circumstances and were generated by PCR. For small RNA northern analysis, single stranded oligonucleotides were used for hybridisation. Probes used in the northern analysis in Chapters 3, 5 and 6 are described in Appendices A, C and D [7].

Radioactive labeled single stranded cRNA probes were *in vitro* transcribed using Riboprobe® Combination System - SP6/T7 RNA Polymerase (Promega) and [α-^{32}P]-dATP (GE Healthcare) according to the manufacturer's protocol. For dsDNA probe, 25ng of linearised PCR amplicons were denatured and labeled with the DECAprime™ II Random Priming DNA Labeling Kit (Ambion®) and [α-^{32}P]-dCTP (GE Healthcare) according the manufacturers' protocol. Both cRNA and dsDNA probes were purified using illustra MicroSpin™ G-50 Columns (GE Healthcare) according to manufacturer's protocol. All probes were stored in lead container in -20^{o}C until used (within 3 days). Method for labelling single stranded oligonucleotide probes is described in Section 2.8.1.

2.9.2 Northern analysis for total RNA

All steps were performed in an RNAse-free environment and at room temperature unless otherwise specified. For total RNA northern analysis, MOPS gel electrophoresis was performed. 1X MOPS (pH7.0) consists of 40mM 3-(N-morpholino)propanesulfonic acid, 5mM of sodium acetate and 1mM of EDTA.

Denaturation gel was used (1% (w/v) agarose dissolved in 1X MOPS and 0.6M formaldehyde) and electrophoresis was performed in 1X MOPS running buffer with 0.2M of formaldehyde. Approximately 15-30μg total RNA from each sample was denatured in 1X Ambion Gel Loading Buffer II (Ambion®) at 85°C for 3 minutes before subjected to electrophoresis. Electrophoresis was performed at 70V for 3-4 hours or at 30-40V for 8-10 hours.

After electrophoresis, the gel was soaked in DEPC treated water (2x for 10 minutes each) followed by ethidium bromide staining and gel imaging. Next, capillary transfer method using 20X SSC was setup based on the standard protocol [8]. Capillary transfer was performed for 16 hours and electrophoresed RNAs were blotted onto Hybond-N+ nylon membrane (GE Healthcare). After the transfer step, membrane was briefly rinsed with 2X SSC and RNAs on the membrane were immobilised by UV-crosslinking (Stratagene). Alternatively, the membrane can be sandwiched in between two filter papers and placed in an oven at 55°C for 2 hours.

Pre-hybridisation, hybridisation, washing and visualisation steps were performed according to the large probe method outlined for southern blotting analysis section. A different method for northern analysis of *Sox4* and *Sox11* transcripts was described in Chapter 3 and Appendix A [7].

2.9.3 Northern analysis for small RNA

All steps were performed in an RNAse-free environment and at room temperature unless otherwise specified. Approximately 30μg of total RNA from each sample was denatured in 1X Ambion Gel Loading Buffer II (Ambion®) at 85°C for 3 minutes. RNAs were electrophoresed in 15% acrylamide/urea gel (48% (w/v) urea, 15% (v/v) acrylamide, 0.05% (w/v) ammonium persulfate and 0.1% (v/v) tetramethylethylenediamine (TEMED) prepared in 1X TBE) in 1X TBE buffer at 300V for 90 minutes followed by ethidium bromide staining and gel imaging. After electrophoresis, RNAs greater than 200nt was trapped in or near the well. Separated small RNAs in the gel were then transferred onto Hybond-N+ nylon membrane (GE Healthcare) using Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) at a constant 0.4V for 45 minutes. After the transfer,

RNA was immobilised on the membrane using UV-crosslinking (Stratagene) followed by 2 hours incubation at 55^{o}C. Pre-hybridisation, hybridisation, washing and visualisation steps were performed according to the oligonucleotide probe method outlined for southern blotting analysis section.

2.10 Western analysis

2.10.1 Purification of total protein and Bradford's assay

Prior to cell lysate preparation, cells were rinsed with cold PBS for 2 times. About 250µl of cold lysis solution was added directly on the cells. Cells were dislodged using cells scraper and lysates were transferred to microfuge tubes, vortexed and incubated on ice for at least 10 minutes. Cell lysates were spun at 13,000rpm for 15 minutes at 4^{o}C and supernatants were collected into new microfuge tubes. Cell lysates were then resuspended in 1X sample buffer, boiled and total protein concentrations were assayed using Bradford Reagent (BioRad) according to the manufacturer's protocol followed by colorimetric measurement using EL808 Ultra Microplate reader (Bio-Tek Instruments, Inc).

2.10.2 Membrane blotting, hybridisation, washing and visualisation

Equal amounts of protein (~20µg) were loaded onto 10% acrylamide gels, separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Electrophoresed proteins were transferred onto Amersham Hybond™-P hydrophobic polyvinylidene difluoride (PVDF) membrane followed by blocking with 5% (w/v) skim milk powder with 0.1% (v/v) Triton-X100 prepared in 1X PBS. The membrane was probed with a primary polyclonal rabbit antibody directed against Sox4 (Cat. No.: AB80261, Abcam) or polyclonal goat antibody directed against actin (Cat. No.: SC-1616, Santa Cruz) at 4^{o}C overnight followed by ImmunoPure® Goat Anti-Rabbit IgG, (H+L) Peroxidase Conjugate (Cat. No.: 31460, Pierce) or Polyclonal Rabbit Anti-Goat Immunoglobulins/HRP (Cat. No.: P0449, Dako) secondary antibody. Reactive bands were detected using Amersham ECL Plus™ Western Blotting Detection Reagents (GE Healthcare) according to

the manufacturer's protocol. Membranes were stripped using 0.5M of sodium hydroxide before reusing it for another detection.

2.11 Cell culture

2.11.1 P19 teratocarcinoma cells

P19 teratocarcinoma cells were propagated and neurodifferentiated according to method described previously [9]. Briefly, P19 cells were grown in complete medium [Minimum Essential Medium Alpha (alpha-MEM; Cat number: 12571-071; Invitrogen) supplemented with 10% (v/v) heat-inactivated foetal calf serum (FCS; Cat number: 10438-026; Invitrogen), 100 units/ml penicillin, 100 µg/ml streptomycin and 2 mM L-glutamine]. For induction of neuronal differentiation, $1x10^6$ P19 cells were cultured in suspension form using bacteriological Petri dishes. The complete medium with additional supplementation of $5x10^{-7}$M all-trans retinoic acid (ATRA; Cat number: R-2625; Sigma) was used for the induction. After 4 days, P19 cells formed embryoid body stages (EBs). EBs were collected from suspension cultures and replated in adherent culture flasks in the complete medium with only 5% (v/v) foetal bovine serum (FBS) and without ATRA supplementation. The cells were allowed to differentiate for 5 days.

2.11.2 Neural progenitor/stem cells (NPSCs)

Mouse used for generation of neurospheres had a mixed genetic background including MF1, 129SvEv, C57BL/6 and CBA. Cerebral cortices from E14.5 embryos were dissected out into calcium-magnesium free PBS. The tissue was mechanically dissociated and centrifuged. The cells were plated in complete neuroculture medium [1X Neurobasal™ Medium (Cat number: 21103-049, Invitrogen) containing 2% (v/v) 50X B-27 serum-free supplement (Cat number: 17504-044, Invitrogen), 1% (v/v) 200mM L-glutamine, 1% (v/v) 200mM Glutamax (Cat number: 35050-061, Invitrogen), 100 units/ml penicillin, 100 µg/ml streptomycin, 20ng/ml EGF (BD Biosciences) and 20ng/ml bFGF (R & D Systems)] for 4 days followed by induction of neuronal differentiation. These cells were then plated on poly-D-lysine (Cat number: P6407, Sigma) and laminin (Cat number: 23017-015, Invitrogen) coated culture dishes in neuroculture

medium with the presence of 2% (v/v) FBS without epidermal growth factor and basic fibroblast growth factor. The differentiation was allowed to proceed for 5 days.

2.11.3 NIH/3T3 mouse embryo fibroblast

NIH/3T3 cells were obtained from American Type Culture Collection (www.atcc.org) and maintained in Dulbecco's Modified Eagle's Medium (Sigma Aldrich) supplemented with 10% (v/v) heat-inactivated FCS (Invitrogen), 100 units/ml penicillin, 100 μg/ml streptomycin and 2mM L-glutamine. Cells were subcultured into new dishes when they reached 80% or less confluence using approximately 3-5x10^3 cells/cm^2 inoculums.

2.11.4 Others

2.11.4.1 Mouse embryonic stem cells with Dicer1 conditional allele

In Chapters 5-6, the mouse embryonic stem (mES) cells with DICER1 activity were obtained from a line heterozygous for a conditionally mutant *Dicer1* allele (*Dicer1*^{c}) and a null *Dicer1* allele (*Dicer1*$^{-}$). These genetic modifications were described previously [10]. mES cells without DICER1 activity were produced by transient transfection of this *Dicer1*$^{c/-}$ line with Cre recombinase to produce *Dicer1*$^{-/-}$ subclones (JRM and DMM, unpublished data). The mES cells were propagated as previously described [11]. These procedures were performed by Drs Jeff R Mann and Deidre M Mattiske from the Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, Victoria, Australia.

2.11.4.2 Harvesting of E3.5 blastocysts

Female mice aged 3-4 weeks were superovulated using 5IU of Folligon (PMSG) followed by 5IU of Chorulon (HCG) 47.5 hours later and mated with B6D2F1 entire stud males. Microdrop culture dishes were set up to equilibrate in 37^{o}C, 5% CO_2 incubator 4 hours prior to culture. EmbryoMax® KSOM (Millipore) media was used in 20µl droplets in a 35mm dish, overlaid with Embryo Tested Mineral Oil (Sigma). Superovulated female mice were sacrificed after 2.5 days of superovulation induction and mating, and oviducts were collected into M2 handling media (Millipore). Oviducts were flushed using M2 media, a blunt 30G needle and a 1ml syringe. Morulae were collected and cultured in pre-equilibrated KSOM. Blastocysts were collected from culture a day later under a dissecting microscope. These were considered E3.5 blastocysts. These procedures were performed by Ms Sandra Piltz from the School of Molecular and Biomedical Science, Faculty of Sciences, University of Adelaide, Adelaide, South Australia.

2.12 Lipofectamine-mediated transfection

All transfections of plasmid DNA vectors were carried out using the Lipofectamine™ 2000 Transfection Reagent (Invitrogen) strictly according to manufacturer's protocols. All transfections in Chapter 5 were performed on 3T3 mouse embryo fibroblast cells grown in the 6-well culture vessel at a density of 8-10 x10^4 cells/cm^2.

2.13 Microscopic analysis

General microscopic examinations of living cells were carried out using Olympus CKX41 inverted and Olympus CX41 optical microscopes (Olympus). For bright-field and fluorescent microscopic examination and imaging, Zeiss Axioplan 2 fluorescent microscope with Z-stacking ability (Carl Zeiss) was used and operated according to manufacturer's manuals. The micrographs were processed using Axio Vison 4.7 software (Carl Zeiss).

2.14 DNA Sequencing

DNA sequencing has been performed on both PCR products and plasmid DNA using specific primers and T7/SP6 universal primers, respectively. Cycle sequencing of samples was carried out using 1/8 dilution of BigDye® terminator v3.1 (Applied Biosystems) and the relevant primers according to manufacturer's recommended cycle conditions for 50 cycles. All sample purification steps after cycle sequencing were carried out using ExoSAP-IT kit (GE Healthcare) according to manufacturer's protocol. Cleaned samples were submitted to DNA sequencing services using the ABI3730 genetic analyser platform at DNA Sequencing Unit, Institute of Medical and Veterinary Science, Adelaide, South Australia.

2.15 Other services

GeneWorks performed the construction of small RNAs library and provide the next-generation sequencing service. Small RNAs with size ranges from 16-30nt were isolated from 10μg total RNA using polyacrylamide gel electrophoresis. The complementary small RNA library was constructed using the Small RNA Sample Prep Kit version 1.0 (Illumina) according to the manufacturer's protocol with the 5'-GTTCAGAGTT CTACAGTCCG ACGATC-3' and 5'- TCGTATGCCG TCTTCTGCTT GT-3' adapters at the 5' and 3' ends, respectively. Sequencing was carried out using a Genome Analyzer II (Illumina).

2.16 Statistical analysis

For SAGE tags comparison – Fisher's exact test or a Bayesian model was used [12]. Empirical Bayesian moderated t statistic and linear modeling analysis [13] was used to compared log2 normalised expression of genes in various groups in Chapter 3 and 4. T-tests were used to compare log2-normalised expression of gene or small RNAs between two groups in Chapter 5 whereas analysis of variance (ANOVA) was used to compare these values in different groups or samples in Chapters 6. Multiple testing corrections were also carried out in all comparisons unless specified otherwise.

2.17 Bioinformatics analysis and public databases

List of bioinformatics tools and publicly available databases used in Chapters 3-6 are presented in Table 2.4.

Table 2.4: List of bioinformatics tools and publicly available databases used.

Bioinformatics resource	Description	Reference
Allen Institute for Brain Science	An interactive, genome-wide image database of gene expression.	[14]
Basic Local Alignment Search Tool (BLAST)	For finding regions of local similarity between sequences.	[15]
Brain Gene Expression Map (BGEM)	A database that contains gene expression patterns assembled from mouse nervous tissues.	[16]
DIANA microT v3.0	For microRNA downstream targets prediction.	[17]
Ensembl Genome Browser	For visualisation, mining and customising data in relation to reference genomes.	[18]
Gene Expression Nervous System Atlas (GENSAT)	A database for gene expression atlas of the developing and adult central nervous system in the mouse.	[19]
Gene Expression Omnibus (GEO)	A repository for gene expression databases of various platforms.	[20]
GenePaint	A digital atlas of gene expression patterns in the mouse.	[21]
In silico PCR	Searches a sequence database with a pair of PCR primers, using an indexing strategy for fast performance.	[22]
Ingenuity Pathway Analysis (IPA)	Pathway analysis of differentially expressed molecules.	[23]
miRBase	A microRNA repository.	[24]
Online Mendelian Inheritance in Man (OMIM)	A database of human genes and genetic disorders.	[25]
Paired-End Ditags (PETs) from the FANTOM Consortium	Paired-end ditags of polyA+ RNA information obtained from FANTOM Consortium.	[26]
Primer3	For designing primers/probes.	[27]
RNA22	For microRNA target detection & precursor prediction.	[28]
RNAfold web server	For the prediction of secondary structures of single stranded RNA or DNA sequences.	[29]
Tagmapping program	For mapping SAGE tags to genes using EST and genomic information.	[30]
The R Project for Statistical Computing	A programming environment for statistical computing and graphics.	[31]
UCSC genome browser	For visualisation, mining and customising data in relation to reference genomes.	[22]
UniversalProbe Library (UPL) Assay Design Centre	For designing primers for UPL assays.	[32]

2.18 Media and solutions

Values in parentheses are final concentrations.

LB broth (1000ml)

Tryptone, Sigma	10g	(1% w/v)
Yeast extract, Sigma	5g	(0.5% w/v)
Sodium chloride	10g	(1% w/v)
Ampicillin (100 mg/ml), Sigma	1ml	(100µg/ml)
pH to 7.4 using NaOH.		
Filtered water	Top up to 1000ml	

LB Agar (1000ml)

Agar	15g	(1.5% w/v)
LB broth	Top up to 999ml	
Boil to dissolve agar.		
Ampicillin (100 mg/ml)	1ml	(100µg/ml)

Proteinase K buffer (50ml)

0.5M EDTA	5ml	(0.05M)
1M Tris.HCl pH7.5	5ml	(0.1M)
20mg/ml Proteinase K	3µl	(1.2ug/ul)
DEPC-water	Top up to 50ml	

Acetylation buffer (50ml)

7.53M triethanolamine, Sigma	664µl	(1.3% v/v)
Concentrated HCl	89µl	(0.178% v/v)
Acetic anhydride, Sigma	127µl	(0.25% v/v)
DEPC-water	Top up to 50ml	

SSCF solution (200ml)

Low grade formamide, Roche	100ml	(50% v/v)
20X SSC	50ml	(5X SSC)
DEPC-water	50ml	

Pre-hybridisation buffer (50ml)

Deoinised formamide, Sigma	25ml	(50% v/v)
20X SSC	7.5ml	(3X SSC)
50X Denhardt's solution	1ml	(1X)
0.02M phosphate buffer	5ml	(0.002M)
Dextran sulphate, Sigma	5g	(10% w/v)
Yeast total RNA, Roche	50mg	(1mg/ml)
10mg/ml Herring sperm DNA, Promega	5ml	(1mg/ml)
DEPC-water	Top up to 50ml	

50X Denhardt's solution (100ml)

BSA	1g	(1% w/v)
Ficoll	1g	(1% w/v)
Polyvinylpyrrolidone	1g	(1% w/v)
DEPC-water	Top up to 100ml	

Filter sterilise with 0.2µm cartridge.

Blocking buffer I (200ml)

1M Tris.HCl pH7.5	20ml	(0.1M)
5M NaCl	6ml	(0.15M)
1000X levamisole stock	200µl	(1X)
DEPC-water	Top up to 200ml	

1000X levamisole stock (1ml)

Levamisole, Sigma	240mg	(240mg/ml)
DEPC-water	Top up to 1ml	

Blocking buffer II (1ml)

2.5% w/v blocking powder (Roche)		
in maleate buffer	800µl	(2% w/v)
Foetal calf serum	200µl	(20% v/v)

Maleic acid buffer (1000ml)

Maleic acid	11.607g	(100mM)
NaCl	8.76g	(150mM)
DEPC-water	Top up to 1000ml	

pH to 7.5 with NaOH.

Blocking buffer III (10ml)

1M Tris.HCl pH9.5	1ml	(0.1M)
5M NaCl	200ul	(0.1M)
1M $MgCl_2$	500ul	(0.05M)
Tween-20	100ul	(1% v/v)
1000X levamisole stock	10ul	(1X)
DEPC-water	Top up to 10ml	

Colour reaction buffer (1ml)

Blocking buffer III	1ml
NBT stock, Roche	3.75ul
BCIP stock, Roche	3.76ul

Mix well and spin down. NBT/BCIP are very toxic solutions.

1X TE buffer pH8.0 (1000ml)

1M Tris.HCl pH7.5	10ml	(0.01M)
0.5M EDTA pH8.0	2ml	(0.001M)
DEPC-water	Top up to 1000ml	

1M Tris.HCl pH7.5 (crystallised free base) (500ml)

Tris(hydroxymethyl) aminomethane	60.57g	(1M)
DEPC-water	Top up to 500ml	

pH to 7.5 using HCl.

Sterilised by autoclaving.

0.5M EDTA pH8.0 (100ml)

Diaminoethane tetraacetic acid	18.6g	(0.5M)
DEPC-water	Top up to 100ml	

pH to 8.0 using NaOH.

Sterilised by autoclaving.

30% acrylamide (300ml)

Acrylamide	87.6g	(29.2% w/v)
N'N'-bis-methylene acrylamide	2.4g	(0.8% w/v)
Filtered water	Top up to 300ml	

Pass through 0.8µm filter (use 60ml syringe).

Cover with foil and store at 4°C.

1.5M Tris.HCl pH8.8 (150ml)

Tris base	27.23g	(1.5M)
Filtered water	80ml	

Adjust to pH8.8 with 1M HCl.

Filtered water	Top up to 150ml

0.5M Tris.HCl pH6.8 (100ml)

Tris base	6g	(0.5M)
Filtered water	60ml	

Adjust to pH6.8 with 1M HCl.

Filtered water	Top up to100ml

10% SDS (400ml)

Sodium dodecyl sulphate (SDS)	40g	(10% w/v)
Filtered water	Top up to 400ml	

Filter-sterilised with 0.2µm cartridge

10% APS (6ml)

Ammonium persulfate (APS)	0.6g	(10% w/v)
Filtered water	Top up to 6ml	

Aliquots of 30µl and 50µl. Store at -20°C.

Lysis buffer for protein lysate preparation

1M Tris/HCL pH7.5	10ml	(50mM)
NP40	2ml	(1% v/v)
5M NaCl	6ml	(150mM)
200mM EGTA	2ml	(1mM)
500mM NaF	40ml	(100mM)
100mM Na Pyrophosphate	20ml	(10mM)
100mM $NaVO_4$*	2ml	(1/100 dilution)
Proteinase inhibitor*	4ml	(1/50 dilution)
Filtered water	114ml	
Na Azide	400µl	(1:500 dilution)

** add fresh each time before use*

10x running buffer for SDS-PAGE (1000ml)

Tris base	30.3g	(3% w/v)
Glycine	144g	(14.4% w/v)
SDS	10g	(1% w/v)
Filtered water	800ml	
pH to 8.3 with 1M HCl/1M NaOH		
Filtered water	Top up to 1000ml	

1X transfer buffer (4000ml)

Tris base	12.12g	(0.3% w/v)
Glycine	57.6g	(1.44% w/v)
Filtered water	3000ml	
Stir until dissolve.		
Methanol	800ml	(20% v/v)
Filtered water	Top up to 4000 ml	
Store at 4^{o}C.		

PBS-T (1000ml)

Phosphate buffered saline (PBS)	1000ml	
Tween-20	1ml	(0.1% v/v)

Blocking buffer for western blotting (1000ml)

Non-fat dried milk powder	50g	(5% w/v)
PBS-T	Top up to 1000ml	

5X sample buffer (SDS reducing) (8ml)

Filtered water	4ml	
500 mM Tris.HCl pH6.8	1ml	(62.5mM)
100% glycerol	0.8ml	(10% v/v)
10% w/v SDS	1.6ml	(2% w/v)
0.05% w/v bromophenol blue	0.2ml	(0.00125% w/v)
2-β-mercaptoethanol	0.4ml	(5% v/v)

2.19 References

1. **National institute of Allergy and Infectious Disease** [http://www3.niaid.nih.gov/labs/aboutlabs/cmb/InfectiousDiseasePathogenesisSection/mouseNecropsy/]
2. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR *et al*: **Real-time quantification of microRNAs by stem-loop RT-PCR**. *Nucleic Acids Res* 2005, **33**(20):e179.
3. Tang F, Hajkova P, Barton SC, Lao K, Surani MA: **MicroRNA expression profiling of single whole embryonic stem cells**. *Nucleic Acids Res* 2006, **34**(2):e9.
4. Brown CY, Sadlon T, Gargett T, Melville E, Zhang R, Drabsch Y, Ling M, Strathdee CA, Gonda TJ, Barry SC: **Robust, reversible gene knockdown using a single lentiviral short hairpin RNA vector**. *Hum Gene Ther* 2010, **21**(8):1005-1017.
5. Ling KH, Hewitt CA, Kinkel SA, Smyth GK, Scott HS: **High-throughput and complex gene expression validation using the Universal ProbeLibrary and the LightCycler® 480 system**. *Biochemica* 2008, **2**:23-26.
6. Luu-The V, Paquet N, Calvo E, Cumps J: **Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction**. *Biotechniques* 2005, **38**(2):287-293.
7. Ling KH, Hewitt CA, Beissbarth T, Hyde L, Banerjee K, Cheah PS, Cannon PZ, Hahn CN, Thomas PQ, Smyth GK *et al*: **Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling**. *Genome Biol* 2009, **10**(10):R104.
8. Sambrook J, Russell DW: **Molecular cloning: a laboratory manual**, 3rd edn. New York: Cold Spring Harbor Laboratory; 2001.
9. Rudnicki MA, McBurney MW: **Cell culture methods and induction of differentiation of embryonal carcinoma cell lines**. In: *Teratocarcinomas and embryonic stem cells A practical approach.* Edited by Robertson EJ. Oxford: IRL Press; 1987: 19-49.
10. Mattiske DM, Han L, Mann JR: **Meiotic maturation failure induced by DICER1 deficiency is derived from primary oocyte ooplasm**. *Reproduction* 2009, **137**(4):625-632.
11. Mann JR: **Deriving and propagating mouse embryonic stem cell lines for studying genomic imprinting**. *Methods Mol Biol* 2001, **181**:21-39.
12. Vencio RZ, Brentani H, Patrao DF, Pereira CA: **Bayesian model accounting for within-class biological variability in Serial Analysis of Gene Expression (SAGE)**. *BMC Bioinformatics* 2004, **5**:119.
13. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments**. *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
14. **Allen Institute for Brain Science** [http://mouse.brain-map.org/]
15. **BLAST: Basic Local Alignment Seach Tool** [http://blast.ncbi.nlm.nih.gov/Blast.cgi]
16. **Brain Gene Expression Map** [http://www.stjudebgem.org/]

17. Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K *et al*: **DIANA-microT web server: elucidating microRNA functions through target prediction**. *Nucleic Acids Res* 2009, **37**(Web Server issue):W273-276.
18. **Ensembl Genome Browser** [http://www.ensembl.org/]
19. **Gene Expression Nervous System Atlas** [http://www.gensat.org/]
20. **National Centre for Biotechnology Information (NCBI): Gene Expression Omnibus** [http://www.ncbi.nlm.nih.gov/geo/]
21. **GenePaint.org** [http://www.genepaint.org/]
22. **UCSC Genome Bioinformatics** [http://genome.ucsc.edu/]
23. **Ingenuity Pathway Analysis** [http://www.ingenuity.com]
24. **miRBase** [http://microrna.sanger.ac.uk/]
25. **OMIM: Online Mendelian Inheritance in Man** [http://www.ncbi.nlm.nih.gov/omim]
26. **PETs from the FANTOM Consortium** [http://www.ensembl.org/Mus_musculus/ditags/FANTOM_GSC_PET.html]
27. **Primer3** [http://frodo.wi.mit.edu/primer3/]
28. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I: **A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes**. *Cell* 2006, **126**(6):1203-1217.
29. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W70-74.
30. **Mapping SAGE tags to genes using EST and genomic information** [http://www.mbgproject.org/tagmapping/]
31. **The R Project for Statistical Computing.** [http://www.r-project.org/]
32. **Assay Design Center for Universal ProbeLibrary Assay** [https://www.roche-applied-science.com]

CHAPTER 3

Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling

King-Hwa Ling, Chelsee A Hewitt, Tim Beissbarth, Lavinia Hyde, Kakoli Banerjee, Pike-See Cheah, Ping Z Cannon, Christopher N Hahn, Paul Q Thomas, Gordon K Smyth, Seong-Seng Tan, Tim Thomas and Hamish S Scott.

Published in: *Genome Biology* 2009, 10(10): R104.

3.1 Summary

The development of the cerebral cortex is meticulously regulated by many genes expressed spatiotemporally. Understanding of mechanisms that govern the molecular and cellular processes underlie cerebral corticogenesis will lead us to a better comprehension of the onset and progression of various neurological disorders. Therefore, this study aims to dissect the molecular properties of the mouse cerebral cortex during development by adopting the global transcriptome analysis approach.

Both short and long 3' Serial Analysis of Gene Expression (SAGE) technologies were carried out to identify differentially expressed regulatory elements in cerebral cortices obtained from 4 selected developmental stages: E15.5, E17.5, P1.5 and 4-6 months old. Temporally co-regulated gene clusters, novel molecular networks and associated pathways, novel candidates in regionalized development and genomic clustering of SAGE tags were also reported in this study. The landmark finding of this study includes,

a. the discovery and validation of 70 differentially expressed transcripts (DETs) involved in 6 interconnected molecular networks associated with known human neurological disorders.
b. the identification and validation of genes with regionalized expression profiles in the embryonic cerebral cortex.
c. the identification and validation of 2 embryonic stage specific genomic clusters at SRY (sex determining region Y)-box 4 (*Sox4*) and SRY (sex determining region Y)-box 11 (*Sox11*) gene loci featuring spatiotemporally expressed multiple overlapping sense and antisense transcripts with numerous novel transcription start sites and polyadenylation sites.
d. The sense and antisense transcripts were differentially regulated between the different brain regions, adult mouse organs, proliferating and differentiating neural stem/progenitor cells (NSPCs) and P19 (embryonal carcinoma) cells.

The study has successfully characterised the molecular landscape of the developing mouse cerebral cortex and has led to the discovery of *Sox4* and *Sox11*

novel sense and antisense transcripts with peculiar gene organization structure, regulation and expression profile. It was based on this study that further characterisation of selected transcripts were performed and reported in the subsequent chapters within the thesis. For instance, adult specific genomic clusters at neurogranin (*Nrgn*) and calcium/calmodulin-dependent protein kinase II inhibitor 1 (*Camk2n1*) gene loci are described in Chapter 4 whereas functional characterisation of *Sox4* genomic cluster is reported in Chapter 5.

3.2 Notes

The authors' declaration and 8 additional data files have been included as Appendix A according to the following order:

1. Authors' declaration.
2. Additional data file 1: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s1.doc).
3. Additional data file 2: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s2.xls).
4. Additional data file 3: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s3.xls).
5. Additional data file 4: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s4.xls).
6. Additional data file 5: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s5.xls).
7. Additional data file 6: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s6.xls).

8. Additional data file 7: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s7.xls).
9. Additional data file 8: (Original file is accessible at http://genomebiology.com/content/supplementary/gb-2009-10-10-r104-s8.xls).

3.3 Permission to Reuse Published Materials

The published manuscript is an open access article distributed under the terms of the Creative Common Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited (http://creativecommons.org/license/by/2.0/).

3.4 The published article

Please refer to the next page for the following published article:

Ling KH, Hewitt CA, Beissbarth T, Hyde L, Banerjee K, Cheah PS, Cannon PZ, Hahn CN, Thomas PQ, Smyth GK, Tan SS, Thomas T, Scott HS: **Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling.** *Genome Biol* 2009, 10:R104.

Research

Open Access

Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of *Sox4* and *Sox11* novel antisense transcripts revealed by transcriptome profiling

King-Hwa Ling[*†‡#], Chelsee A Hewitt[*††], Tim Beissbarth[§‡‡], Lavinia Hyde[§§§], Kakoli Banerjee[¶], Pike-See Cheah[¶¥], Ping Z Cannon[*], Christopher N Hahn[#], Paul Q Thomas[¶], Gordon K Smyth[§], Seong-Seng Tan[**], Tim Thomas[*] and Hamish S Scott[*†#]

Addresses: [*]Molecular Medicine Division, The Walter and Eliza Hall Institute of Medical Research, Royal Parade, Parkville, Victoria 3052, Australia. [†]The School of Medicine, The University of Adelaide, SA, 5005, Australia. [‡]Department of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia. [§]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Royal Parade, Parkville, Victoria 3052, Australia. [¶]School of Molecular and Biomedical Science, Faculty of Sciences, University of Adelaide, Adelaide, SA 5005, Australia. [¥]Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia. [#]Department of Molecular Pathology, The Institute of Medical and Veterinary Science and The Hanson Institute, Adelaide, SA 5000, Australia. [**]Howard Florey Institute, The University of Melbourne, Parkville, Victoria 3010, Australia. [††]Current address: Pathology Department, The Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. [‡‡]Current address: Department of Medical Statistics (Biostatistics), University of Göttingen, Humboldtalle 32, 37073 Göttingen, Germany. [§§]Current address: The Bioinformatics Unit, Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Victoria 3052, Australia.

Correspondence: Hamish S Scott. Email: hamish.scott@imvs.sa.gov.au

Published: 2 October 2009

*Genome **Biology*** 2009, **10:**R104 (doi:10.1186/gb-2009-10-10-r104)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2009/10/10/R104

Received: 6 June 2009
Revised: 20 July 2009
Accepted: 2 October 2009

© 2009 Ling *et al.*; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Development of the cerebral cortex requires highly specific spatio-temporal regulation of gene expression. It is proposed that transcriptome profiling of the cerebral cortex at various developmental time points or regions will reveal candidate genes and associated molecular pathways involved in cerebral corticogenesis.

Results: Serial analysis of gene expression (SAGE) libraries were constructed from C57BL/6 mouse cerebral cortices of age embryonic day (E) 15.5, E17.5, postnatal day (P) 1.5 and 4 to 6 months. Hierarchical clustering analysis of 561 differentially expressed transcripts showed regionalized, stage-specific and co-regulated expression profiles. SAGE expression profiles of 70 differentially expressed transcripts were validated using quantitative RT-PCR assays. Ingenuity pathway analyses of validated differentially expressed transcripts demonstrated that these transcripts possess distinctive functional properties related to various stages of cerebral corticogenesis and human neurological disorders. Genomic clustering analysis of the differentially expressed transcripts identified two highly transcribed genomic loci, *Sox4* and *Sox11*, during embryonic cerebral corticogenesis. These loci feature unusual overlapping sense and antisense transcripts with alternative polyadenylation sites and differential expression. The *Sox4* and *Sox11* antisense transcripts were highly expressed in the brain compared to other mouse organs and are

differentially expressed in both the proliferating and differentiating neural stem/progenitor cells and P19 (embryonal carcinoma) cells.

Conclusions: We report validated gene expression profiles that have implications for understanding the associations between differentially expressed transcripts, novel targets and related disorders pertaining to cerebral corticogenesis. The study reports, for the first time, spatio-temporally regulated *Sox4* and *Sox11* antisense transcripts in the brain, neural stem/progenitor cells and P19 cells, suggesting they have an important role in cerebral corticogenesis and neuronal/glial cell differentiation.

Background

Complex behavioral tasks, from perception of sensory input and the control of motor output to cognitive functions such as learning and memory, are dependent on the precise development of innumerable interconnections of neuronal networks in the cerebral cortex. The development of the cerebral cortex (also known as cerebral corticogenesis) involves the specific influence of intrinsic and extrinsic mechanisms, which are triggered spatio-temporally [1-4]. Between embryonic day 11 (E11) and 18 (E18), the mouse cerebral cortex develops from a relatively homogenous band of mitotic multipotent progenitor cells into a complex laminated structure containing various classes of neuronal cells [2,4-6]. Cerebral corticogenesis involves: proliferation of multipotent progenitors (E11 to E16.5); migration of postmitotic cells (E11 to E17); cell morphogenesis (E13 to E18); gliogenesis and synaptogenesis (E16 until early postnatal period); and reorganization, elimination and stabilization of neuronal networks (up to adulthood).

The mouse cerebral cortex develops in the latero-medial and rostro-caudal axes [7,8]. At E11, the primordial plexiform layer begins to form in the most lateral part of the neural wall. Its growth continues in the latero-medial axis to the medial part of the telencephalon by E13. The primordial plexiform layer is also expanded in the rostro-caudal axis. The growth in this axis is always less than the growth in the latero-medial axis. The first wave of migratory neuronal cells form the cortical plate 2 days later after the development of the primordial plexiform layer. These events are followed by the development of the cortical plate into an organization of six distinct layers that forms the adult cerebral cortex. Generally, the rostral-most regions of the adult cerebral cortex consist of areas involved in executive functions and motor coordination, whilst the caudal-most regions consist of areas involved in sensory perception such as visual function. Although distinct functional arealization of the cerebral cortex do not fully apply to rodents, mounting evidence suggesting that regulated arealization exists has been shown in mice involving transcription factors such as empty spiracles homolog 2 (*Emx2*; *Drosophila*) [9,10], paired box gene 6 (*Pax6*) [9], COUP transcription factor 1 (*Coup-tf1*) [11], Sp8 transcription factor (*Sp8*) [12,13], distal-less homeobox 1/2 (*Dlx1/2*) and gastrulation brain homeobox 2 (*Gbx2*) [14]. The extensive cyto-architectural and anatomical changes occurring in a spatio-temporal manner during the peak (E15) and at the end (E17) of embryonic cerebral neurogenesis as well as during early postnatal (P1) corticogenesis through to adulthood involves complex underlying molecular regulatory networks.

Complex molecular regulatory elements are important determinants in both spatial and temporal cerebral corticogenesis. These elements regulate gene expression at the chromatin, DNA or RNA, and protein levels through chromatin packaging or remodeling, histone acetylation and deacetylation, chromatin insulation, DNA methylation, post-transcriptional regulation and post-translational protein modification or degradation signaling pathways [15-17]. Other processes involved in such regulation include pre-mRNA processing and nuclear mRNA retention by nuclear-specific paraspeckle complexes [18,19], microRNAs (miRNAs) that interfere with mRNA translation [20-22], and a new class of under-characterized non-coding RNA transcripts known as natural antisense transcripts (NATs) [23,24]. These regulatory networks play a pivotal role in establishing when, where and how a multipotent progenitor cell should proliferate, migrate and settle at a designated position in the cortex. The information regarding regulatory networks during cerebral corticogenesis, however, remains incomplete and does not provide a comprehensive view of the underlying regulatory elements throughout this complex event.

In this study, we employed both short and long 3' serial analysis of gene expression (SAGE) technologies [25,26] to identify differentially expressed regulatory elements by comparing transcriptomes of cerebral cortices generated from four selected developmental stages: E15.5, E17.5, P1.5 and 4 to 6 months old. We also compared rostral to caudal regions of E15.5 and left to right regions of adult cerebral cortices. We report temporally co-regulated gene clusters, novel molecular networks and associated pathways, novel candidates in regionalized development and genomic clustering of SRY-box containing gene 4 (*Sox4*) and SRY-box containing gene 11 (*Sox11*) sense and antisense transcripts. The genomic clustering analysis led us to the discovery of spatio-temporal regulation of novel *Sox4* and *Sox11* antisense transcripts as well as differential regulation of these transcripts in proliferating and differentiating neural stem/progenitor cells (NSPCs) and P19 (embryonal carcinoma) cells.

Results

Generation and analysis of SAGE tags

We constructed 10 SAGE libraries from the cerebral cortex of E15.5, E17.5 and 4- to 6-month-old adult C57BL/6 mice (N = 10; Table 1). The data from two additional SAGE libraries generated from E15.5 and P1.5 cerebral cortices from Gunnersen *et al.* [27] were also incorporated into our analysis. These SAGE libraries represent key stages of cerebral corticogenesis and are accessible from the Gene Expression Omnibus (GEO) website [GEO: GSE15031] [28]. The libraries contain a total of 531,266 SAGE tags (approximately 44,000 tags per SAGE library), 223,471 of which are unique (non-redundant) after screening for artifacts and mapping of short SAGE tags to long SAGE tags (Table 1). The number of unique tags in each library ranges from approximately 7,200 to 32,000 due to the variation in library size (approximately 13,500 to 70,000). The distribution of tag abundance, however, is similar in all libraries (Figures S1 and S2 in Additional data file 1), in which the majority of tags were detected only once (58 to 76% or approximately 5,500 to approximately 24,000 tags), representing a trend comparable with previously reported SAGE analyses of mouse neocortices [27,29]. Of all unique tags, only 5,199 (approximately 2.4%) are common to all developmental stages. The low number of common unique tags is most probably due to the high abundance of tags that occur only once.

Analysis of differentially expressed transcripts/tags

To identify differentially expressed tags/transcripts (DETs), we considered only those 25,165 unique tags with a count >2 across all libraries. Under stringent analyses (Table S1 in Additional data file 1), we identified a total of 561 DETs in various comparisons between developmental stages (Table 2; Figure S3 in Additional data file 1). A full list of DETs with assigned IDs is provided in Additional data file 2. Greater numbers of DETs are observed when the interval of two comparative developmental stages becomes wider. We find the largest number of DETs (approximately 59% or 328 DETs) in the embryonic versus adult stages (E versus Ad) followed by P1.5 versus Ad (approximately 34% or 192 DETs), E15.5 versus P1.5 (approximately 6% or 36 DETs) and E15.5 versus E17.5 (approximately 7% or 38 DETs) comparisons. These indicate distinctive transcript signatures during cerebral cortex development. Comparisons between rostral and caudal E15.5 (R versus C), and left and right adult cerebral cortices (L versus Ri) are described in a different section below.

Approximately 69% of DETs have been assigned to known genes and 6% to expressed sequence tags (ESTs), while 10% of DETs have multiple matches (tag matching multiple gene identifiers) and 8% having ambiguous matches (tag matching the same gene identifier at multiple chromosome loci). Approximately 8% of DETs have no matches and it is most

Table 1

SAGE library information

SAGE library*	Sex	Age	Tissue	Generated sequences	Generated SAGE tag (library)	Unique tags	Unique tags (after scaling to 100,000 tags/library)	GEO accession number
E15_1†	U	E15.5	Rostral cerebral cortex	2,044	43,327	15,664	36,153	GSM375449
E15_2†	U	E15.5	Caudal cerebral cortex	2,925	39,314	19,929	50,692	GSM375450
E15_3†‡	U	E15.5	Cerebral cortex	1,920	44,332	15,787	35,611	GSM375451
E17_1†	U	E17.5	Cerebral cortex	384	13,573	7,214	53,150	GSM375638
E17_2†	U	E17.5	Cerebral cortex	1,053	47,733	19,508	40,869	GSM375639
P1_1†‡	U	P1.5	Cerebral cortex	4,194	42,869	20,465	47,738	GSM375640
Ad_1	M	4 months	Left cerebral cortex	2,016	50,760	19,032	37,494	GSM375458
Ad_2	M	4 months	Left cerebral cortex	1,536	52,476	19,157	36,506	GSM375459
Ad_3	F	5-6 months	Left cerebral cortex	2,688	30,914	15,998	51,750	GSM375460
Ad_4	F	5 months	Left cerebral cortex	1,728	43,592	17,262	39,599	GSM375461
Ad_5	F	5-6 months	Right cerebral cortex	2,684	53,292	21,693	40,706	GSM375462
Ad_6	F	6 months	Right cerebral cortex	3,264	69,084	31,762	45,976	GSM375463

*Each library was constructed by using an independent mouse cerebral cortex. †Short tags were generated from these libraries. ‡These libraries were obtained from [27]. F: female; M: male; U: undetermined sex.

Table 2

Summary of tag classification into various categories and comparisons

		DETs in various comparisons*						
Category	**Number of DETs†**	**E15.5 R versus C**	**Adult L versus Ri**	**E15.5 versus E17.5**	**E15.5 versus P1.5**	**P1.5 versus Ad**	**E versus Ad**	**Total**
Gene	386 (68.8)	25 (56.8)	10 (58.8)	24 (63.2)	16 (44.4)	114 (59.4)	253 (77.1)	442
EST	33 (5.9)	2 (4.5)	2 (11.8)	1 (2.6)	1 (2.8)	9 (4.7)	24 (7.3)	39
Multiple matches	55 (9.8)	9 (20.5)	3 (17.6)	5 (13.2)	4 (11.1)	21(10.9)	23 (7.0)	65
Ambiguous	44 (7.8)	4 (9.1)	2 (11.8)	6 (15.8)	4 (11.1)	14 (7.3)	20 (6.1)	50
No match	43 (7.7)	4 (9.1)	0 (0)	2 (5.3)	11 (30.6)	34 (17.7)	8 (2.4)	59
Total	561	44	17	38	36	192	328	655

Tag classification into various categories and comparisons was based on the mouse genome assembly released in July 2007. Values in parentheses represent the 'percentage' across a column. *Total number of tags for various comparisons is 655 (rather than 561) due to the same tag being statistically significant in more than one comparison. †Differentially expressed tags. Ad: adult stage; C: caudal region; E: embryonic day/stage; L: adult left hemisphere; P: postnatal days; R: rostral region; Ri: adult right hemisphere.

likely that these DETs belong to transcripts from less defined regions, such as centromeric and telomeric areas of a chromosome or assembly gaps of the mouse genome.

Hierarchical clustering of DETs

To identify co-regulated genes, all 561 DETs were hierarchically grouped into 12 clusters based on the calculation of the Euclidean distance of logged normalized counts (Figure 1). Clusters 1, 5 and 6 consist of embryonic-specific DETs that exhibit the highest expression during embryonic development of the cerebral cortex. DETs in cluster 1 are expressed throughout all stages of development but exhibit the lowest expression in the adult cortex. Expression of DETs in cluster 5 ceases prior to birth, whereas DETs in cluster 6 are expressed up to early postnatal stage. On the other hand, clusters 4, 8 and 10 consist of adult specific DETs, showing very similar temporal expression profiles, but with different magnitudes (for example, highest expression in cluster 10). Clusters 2 and 7 are termed 'gene-switching' clusters as they show interesting expression-switching profiles. Cluster 2 shows an expression switch between P1.5 and adult stages whilst cluster 7 shows an expression switch between embryonic and adult stages. Clusters 3 and 9 consist of DETs showing region- (caudal region of E15.5 cerebral cortex) and stage-specific (P1.5 only) expression. Clusters 11 and 12 were excluded from subsequent analyses because they contained very few annotated tags. DETs within the same cluster may be co-regulated and/or involved in similar biological functions during cerebral corticogenesis.

We performed a systematic gene ontology functional clustering using the standardized Gene Ontology term analysis tool DAVID (Database for Annotation, Visualization and Integrated Discovery) [30] (Additional data file 3). Functional analysis of these gene clusters reveals that they have distinctive roles during cerebral corticogenesis. Embryonic-specific gene clusters (1, 5 and 6) are dominated by genes associated with cellular protein and macromolecule metabolic processes or biosynthesis, and nervous system and neuron development. These results match with the expected functional ontologies during embryonic cerebral cortex development in which neuronal migration, differentiation and axonogenesis events are at their peak. In contrast, adult-specific gene clusters (4, 8 and 10) consist of genes related to biological processes in the mature cerebral cortex, such as ion homeostasis, synaptic transmission and regulation of neurotransmitter level. In addition, these gene clusters are also enriched for ribonucleotide/nucleotide binding activity and components of cytoplasmic membrane-bound vesicles. These functional ontologies are in accordance with adult cerebral cortex function, which features synaptogenesis and nerve impulse transmission at synapses. Gene-switching clusters 2 and 7 are enriched with gene ontologies that are similar to both the embryonic- and adult-specific gene clusters. In addition, these gene clusters are also enriched for microtubule cytoskeleton organization and biogenesis, nucleotide biosynthesis and regulation of mRNA translation processes.

Quantitative RT-PCR validation of DETs and gene clusters

To ascertain the robustness of the SAGE datasets, we selected 136 candidate DETs and two additional genes of interests (ATPase, Cu++ transporting, alpha polypeptide, *Atp7a*, and cullin-associated and neddylation-dissociated 2, *Cand2*) for validation after considering both stage-to-stage and hierarchical based analyses (Table S2 in Additional data file 1). The selected DETs are transcription regulators, chromatin modifiers or post-translational regulators, such as ubiquitination pathway related molecules. Seventeen DETs are ESTs, which have been identified in brain-related mouse cDNA libraries or transcriptomes. Independent quantitative RT-PCR (RT-qPCR) validation was carried out using three biological repli-

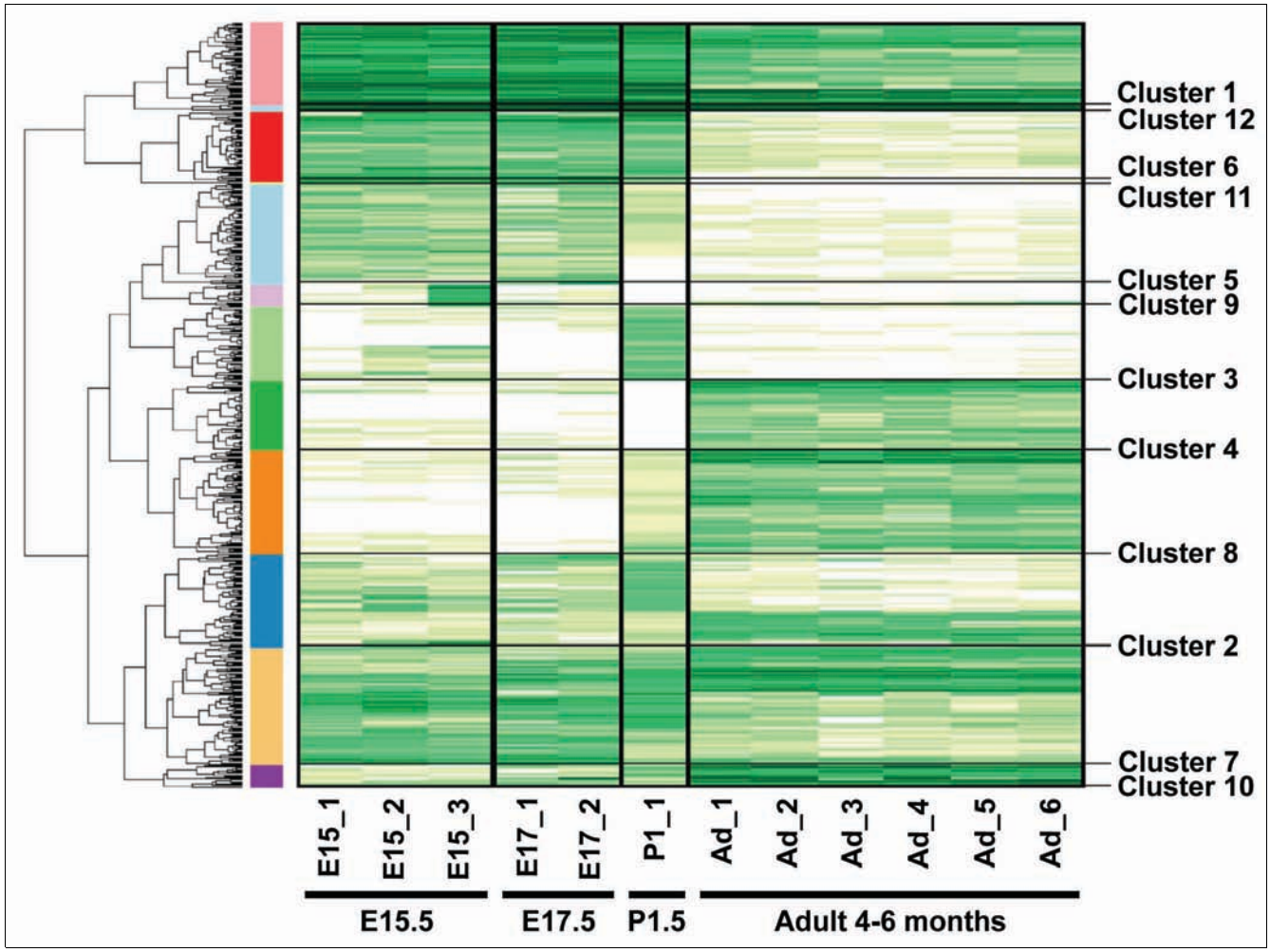


Figure 1
Hierarchical clustering of 561 differentially expressed transcripts/tags. Log2 of normalized counts of DETs from cerebral cortices of various developmental stages were clustered. Dark green clusters denote high levels of expression whereas light green to white clusters denote low levels of expression. The x-axis represents the SAGE libraries whereas the y-axis represents the SAGE tags. The panel on the right shows the 12 different clusters.

cates of unpooled cerebral cortex total RNA for each developmental stage. We validated 70 DETs (including 10 ESTs) from SAGE profiles of comparisons between two developmental stages (Additional data file 3) after considering various stringent criteria and cutoffs. The RT-qPCR results for all the 70 validated DETs, *Atp7a* and *Cand2* are presented in Tables 3, 4, 5 and 6 (Additional data file 4). To validate the expression profiles of gene clusters from the hierarchical clustering analysis, we performed additional RT-qPCR analyses on 65 validated DETs (based on the analysis of two developmental stages), *Atp7a* and *Cand2* by including other developmental stages (Additional data file 5). The analysis validated 62 out of 65 DETs (Figure 2). Of these, 22 are from embryonic gene clusters, 26 from adult gene clusters, and 14 from gene-switching clusters. We assigned *Atp7a* and *Cand2* to clusters 2 and 5, respectively, based on their RT-qPCR expression profiles across all the developmental stages. See Tables 3, 4, 5 and 6 for complete list of DETs and their full gene names.

According to the comparison between two developmental stages, the most abundantly expressed and validated DETs in the E versus Ad analysis are *Camk2a*, *Egr1* and *Plp1* (Table 3). In the adult cerebral cortex, the expression of these DETs is more than 100-fold greater than that at E15.5. Other validated adult-specific DETs with expression levels of approximately tenfold or greater than those in the embryonic brain are *Camk2n1*, *Cryab*, *Nrgn*, BQ176089, CD802535, *Sncb* and *Nptxr*. Conversely, *Sox11*, an embryonic specific DET, is expressed in the E15.5 and E17.5 cerebral cortices with an expression level of at least 100-fold greater than that in the adult. Other validated embryonic specific DETs are *Dcx*, *Zfp57*, *Ezh2*, *Sfrp1* and *Cdk4*, which have expression levels approximately tenfold or greater than those in the adult.

In the P1.5 versus Ad analysis, *Dcx* is expressed at a level approximately 80-fold greater in P1.5 compared to the adult cerebral cortex (Table 4). Other P1.5 enriched and validated

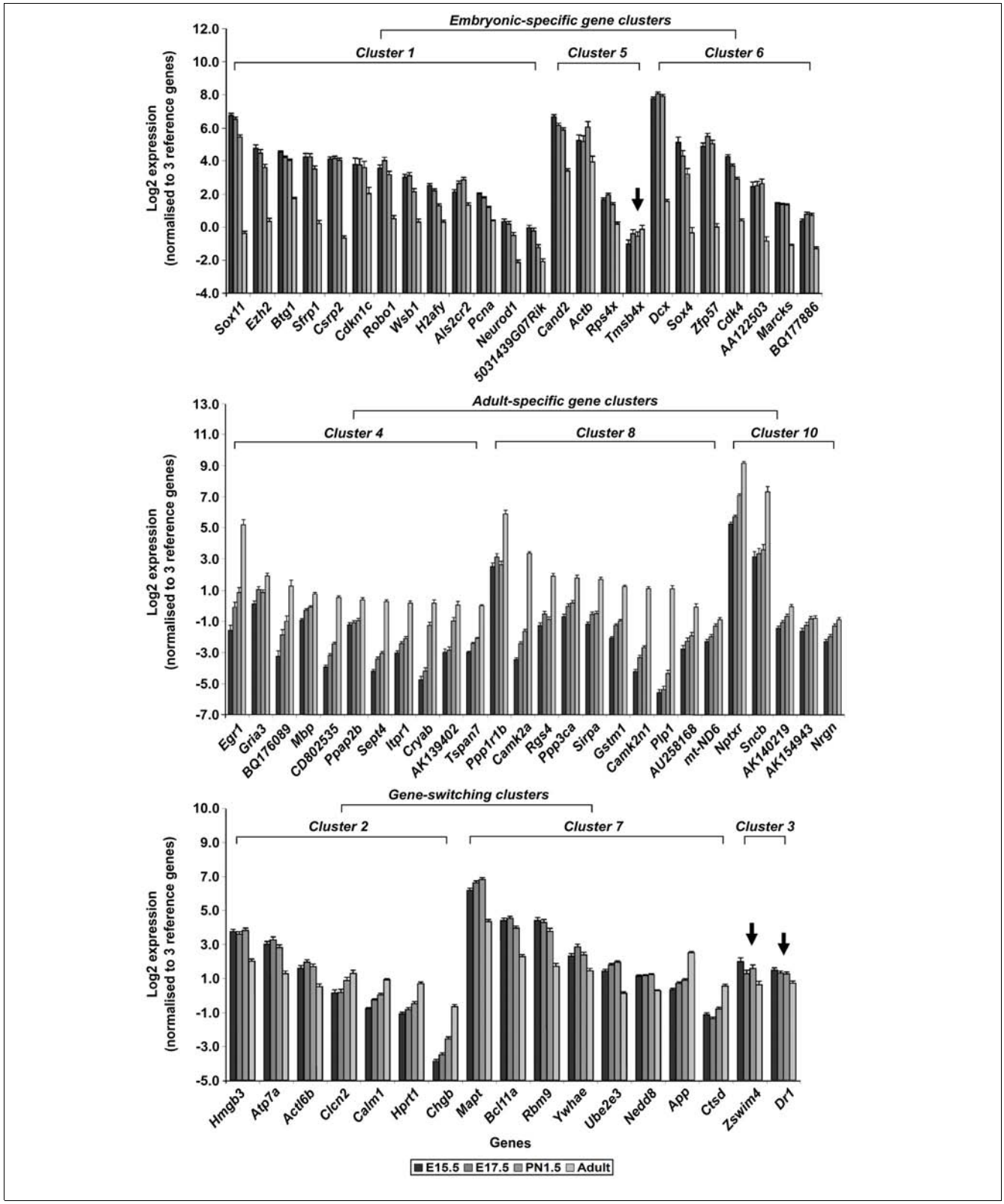


Figure 2
High-throughput RT-qPCR validation of gene clusters. All validations were based on DETs for canonical mRNA. Failed validation of DETs according to hierarchical clustering expression profiles is indicated by arrows. N = 3 and data are presented as mean ± standard error of the mean.

Table 3

RT-qPCR validation of SAGE profile for E versus adult comparison

			Fold change			
SAGE tag	**RefSeq accession**	**Gene ID**	**Ad/E15.5 (SAGE)**	**Ad/E17.5 (SAGE)**	**Ad/E15.5 (RT-qPCR)**	**Ad/E17.5 (RT-qPCR)**
gcttccccaccccctt	NM_177407	calcium/calmodulin-dependent protein kinase II alpha, *Camk2a*	*As*	*As*	111.59	54.74
ggatatgtggtgtgtac	NM_007913	early growth response 1, *Egr1*	*As*	*As*	108.47	38.93
aaattattgggaaatcc	NM_011123	proteolipid protein (myelin) 1, *Plp1*	28.77	*As*	103.15	89.13
gtatttgcaaaaaaaaa	NM_025451	calcium/calmodulin-dependent protein kinase II inhibitor 1, *Camk2n1*	69.59	59.76	40.08	21.46
gcttcatctccagggag	NM_009964	crystallin, alpha B, *Cryab*	*As*	11.59	30.47	20.58
ttaccatactgggttgg	NM_022029	neurogranin, *Nrgn*	24.67	8.47	43.67	17.00
cctcatttcccctgttc	BQ176089	EST from adult C57BL/6 subfornical organ and postrema tissues	*As*	*As*	23.01	8.91
acccggctagtagtgaa	NM_011129	septin 4, *Sept4*	18.97	16.29	22.38	12.96
ctcattataatcaagaa	CD802535	EST from 27-32 days C57BL/6 hippocampus tissue	*As*	9.05	22.03	13.25
aataaagccaatctgac	NM_033610	synuclein, beta, *Sncb*	20.89	7.66	18.20	15.83
gcttttgttaccatctc	NM_030689	neuronal pentraxin receptor, *Nptxr*	17.88	12.24	15.02	10.93
tccctcccttagtatcc	NM_144828	protein phosphatase 1, regulatory (inhibitor) subunit 1B, *Ppp1r1b*	*As*	*As*	10.47	6.88
gccccttcttcattggc	NM_010358	glutathione S-transferase, mu 1, *Gstm1*	*As*	*As*	9.87	5.58
tgactagcgtgacctgt	NM_007694	chromogranin B, *Chgb*	6.13	6.30	9.43	7.22
atttcttttctggatgg	NM_010585	inositol 1,4,5-triphosphate receptor 1, *Itpr1*	14.34	6.16	9.29	6.12
actttgagattgtacct	NM_009062	regulator of G-protein signaling 4, *Rgs4*	12.23	21.01	8.84	5.32
aataattagccttaggt	AK139402	Mus musculus 10 days neonate cortex cDNA	*As*	*As*	8.28	7.55
ctagacagaggcattat	NM_019634	tetraspanin 7, *Tspan7*	13.08	5.61	7.99	5.38
tgtatacacacacgggt	NM_007547	signal-regulatory protein alpha, *Sirpa*	*As*	*As*	7.18	4.63
tgacaagacactgtggc	AU258168	EST from mouse brain	*As*	*As*	6.49	4.58
cttacctcaggtttcct	NM_008913	protein phosphatase 3, catalytic subunit, alpha isoform, *Ppp3ca*	*As*	5.43	5.48	3.45
atagctttctacacact	NM_007471	amyloid beta (A4) precursor protein, *App*	3.98	2.74	4.56	3.49
tttcagcagtgttggct	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1, *Hprt1*	6.90	8.87	3.45	2.85
aggtatgtacaaagttt	NM_016886	glutamate receptor, ionotropic, AMPA3 (alpha 3), *Gria3*	4.97	8.51	3.40	1.81*
tccaacttgtaactata	NM_009790	calmodulin 1, *Calm1*	4.23	3.17	3.24	2.27

Table 3 *(Continued)*

RT-qPCR validation of SAGE profile for E versus adult comparison

cctcagcctggggtaga	NM_009983	cathepsin D, *Ctsd*	3.46	2.04	3.22	3.84
gcttcgtccacacagcg	NM_010777	myelin basic protein, *Mbp*	202.46	173.85	3.19	2.05
tattaaatgtgcttttt	NM_080555	phosphatidic acid phosphatase type 2B, *Ppap2b*	5.71	2.44	3.02	2.76
cttatcctcacctcagc	NC_005089	NADH dehydrogenase 6, mitochondrial, *mt-ND6*	*As*	*As*	2.65	2.13
caaacctccaaaaacca	AK140219	Mus musculus adult male corpora quadrigemina cDNA	29.22	18.62	2.63	2.03
agtggctaattaggtgt	NM_009900	chloride channel 2, *Clcn2*	14.55	4.16	2.23	2.18
accaatgaacaaaaaaa	AK154943	Mus musculus NOD-derived CD11c +ve dendritic cells cDNA	109.40	51.13	1.76*	1.35 (NS)
ccagtacctgaaaaaaa	NM_008453	Kruppel-like factor 3 (basic), *Klf3*	*As*	-52.97	-1.47†‡	-1.27 (NS)
aagaaaacatttaaata	NM_012010	eukaryotic translation initiation factor 2, subunit 3, structural gene X-linked, *Eif2s3x*	-7.70	-10.38	-1.72*	-1.74*
caccctgtgggagctca	NM_172656	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2 (human), *Als2cr2*	-11.03	-12.88	-1.75*	-2.48
cctccatcctttatact	NM_009536	tyrosine 3-monooxygenase/ tryptophan 5-monooxygenase activation protein, epsilon polypeptide, *Ywhae*	-3.19	-2.52	-1.82*	-2.66
tgtgcttccctgtctta	NM_008683	neural precursor cell expressed, developmentally down-regulated gene 8, *Nedd8*	-4.73	-8.41	-1.83	-1.86
ctcctgaaggcatagtt	NM_009454	ubiquitin-conjugating enzyme E2E 3, UBC4/5 homolog (yeast), *Ube2e3*	-4.14	-5.52	-2.51	-3.22
gtgaaactaaaaaaaaa	NM_009094	ribosomal protein S4, X-linked, *Rps4x*	-10.46	-19.19	-2.76	-3.35
aatgtttctgctttaca	NM_011045	proliferating cell nuclear antigen, *Pcna*	-14.21	-8.30	-3.09	-2.62
cgtggatccctctgtca	NM_009876	cyclin-dependent kinase inhibitor 1C (P57), *Cdkn1c*	-13.68	-7.16	-3.39*	-3.30*
cctttgtgacagtggcc	NM_025635	ZW10 interactor, *Zwint*	7.29	8.84	-3.47‡	-4.80‡
gaagccagtgggccatc	NM_001033273	RIKEN cDNA 5031439G07 gene, *5031439G07Rik*	-9.10	-10.00	-4.13	-3.65
gctgtgggtcgctgtgg	NM_010561	interleukin enhancer binding factor 3, *Ilf3*	*As*	*As*	-4.15‡	-2.97‡
acccctgacccttgtt	NM_016707	B-cell CLL/lymphoma 11A (zinc finger protein), *Bcl11a*	-7.88	-5.52	-4.42	-4.78

Table 3 *(Continued)*

RT-qPCR validation of SAGE profile for E versus adult comparison

cggtgtcccccacctcc	NM_012015	H2A histone family, member Y, *H2afy*	-21.48	-27.88	-4.62	-3.70
cagttgcaataaaaata	NM_010894	neurogenic differentiation 1, *Neurod1*	-7.54	-4.96	-5.62	-5.10
aagtttgcaagtctcca	NM_008538	myristoylated alanine rich protein kinase C substrate, *Marcks*	-24.87	-11.76	-5.72	-5.56
ttgctggcttttataaa	NM_053104	RNA binding motif protein 9, *Rbm9*	-6.10	-7.82	-6.50	-5.97
ggttttgtttgtttgac	NM_019653	WD repeat and SOCS box-containing 1, *Wsb1*	-7.88	-11.96	-6.54	-7.00
tatattgattgtggcaa	NM_007569	B-cell translocation gene 1, anti-proliferative, *Btg1*	-15.28	-16.48	-6.93	-5.58
taagaaacct	NM_019413	roundabout homolog 1 (Drosophila), *Robo1*	-28.64	-16.73	-8.23	-11.35
gctttgactgttctctt	AA122503	EST from M2 cells of skin melanoma	-17.17	-18.68	-9.88	-10.15
tggagcgttggctgtat	NM_009870	cyclin-dependent kinase 4, *Cdk4*	-114.55	-80.88	-14.77	-10.00
ctttccctgccaatgta	NM_013834	secreted frizzled-related protein 1, *Sfrp1*	*Es*	*Es*	-16.48	-16.37
tgcagctttctgttcaa	NM_007971	enhancer of zeste homolog 2 (Drosophila), *Ezh2*	-10.63	-5.52	-21.36	-17.59
cacgacacccccaccc	NM_009559	zinc finger protein 57, *Zfp57*	-30.72	-42.33	-29.13	-44.46
tgtgtgaggtgttgtga	NM_010025	doublecortin, *Dcx*	-44.41	-58.81	-73.76	-91.10
cagagtgtagtgtgttg	NM_009234	SRY-box containing gene 11, *Sox11*	*Es*	*Es*	-140.05	-120.35

All RT-qPCR data are statistically significant at $P < 0.001$ unless specified: * $P < 0.01$; †$P < 0.05$. ‡A disagreement between RT-qPCR and SAGE data. *As*, adult-specific expression; *Es*, embryonic-specific expression; NS, no statistically significant difference between two developmental stages. Fold change values of <1.0 are presented in a negative fold change format.

Table 4

RT-qPCR validation of SAGE profile for P1.5 versus adult stages

			Fold change	
SAGE tag	**RefSeq accession**	**Gene ID**	**Adult/P1.5 (SAGE)**	**Adult/P1.5 (RT-qPCR)**
aaattattgggaaatcc	NM_011123	proteolipid protein (myelin) 1, *Plp1*	38.9	43.33
tttcagcagtgttggct	NM_013556	hypoxanthine guanine phosphoribosyl transferase 1, *Hprt1*	6.99	2.26
actcggagccaccagac	NM_009790	calmodulin 1, *Calm1*	4.57	1.83
gcttcgtccacacagcg	NM_010777	myelin basic protein, *Mbp*	*As*	1.79
tccccgtcat	NM_026106	down-regulator of transcription 1, *Dr1*	*Ps*	-1.46
gggaaactaagggagag	NM_172503	zinc finger, SWIM domain containing 4, *Zswim4*	-42.42	-1.96*
gaacgcaagttcagccc	NM_031404	actin-like 6B, *Actl6b*	-5.51	-2.26
gtgaaactaaaaaaaaa	NM_009094	ribosomal protein S4, X-linked, *Rps4x*	-4.72	-2.27
Nil	NM_009726	ATPase, Cu++ transporting, alpha polypeptide, *Atp7a*	*Nil*	-2.91
agaagtgtttggagttt	NM_008253	high mobility group box 3, *Hmgb3*	-20.99	-3.53
cctttgtgacagtggcc	NM_025635	ZW10 interactor, *Zwint*	4.65	-4.04†
acagtctatgttggagg	BQ177886/NM_010487	C57BL/6 whole brain E15.5 (or known as embryonic lethal, abnormal vision, Drosophila-like 3 (Hu antigen C), *Elavl3*)	-41.99	-4.06
gatacttggaatgacta	NM_007393	actin, beta, cytoplasmic, *Actb*	-18.49	-4.35
ctggcttctt	NM_008538	myristoylated alanine rich protein kinase C substrate, *Marcks*	*Ps*	-4.76
Nil	NM_025958	cullin-associated and neddylation-dissociated 2 (putative), *Cand2*	*Nil*	-5.62
gctttgactgttctctt	NM_009870	cyclin-dependent kinase 4, *Cdk4*	-56.55	-5.77
ctcagtaatg	AA122503	EST from M2 cells of skin melanoma	-12.28	-11.12
tggagcgttggctgtat	NM_009238	SRY-box containing gene 4, *Sox4*	-31.51	-11.74
cacgacacccccaccc	NM_007792	cysteine and glycine-rich protein 2, *Csrp2*	*Ps*	-26.10
tgtgtgaggtgttgtga	NM_009559	zinc finger protein 57, *Zfp57*	-16.33	-32.74
gggacctcgtggaagcc	NM_010025	doublecortin, *Dcx*	-24.56	-82.11

All RT-qPCR data are statistically significant at $P < 0.001$ unless specified: * $P < 0.01$. †A disagreement between RT-qPCR and SAGE data. *As:* adult-specific expression; *Nil*: SAGE data not available; *Ps*: P1.5-specific expression. Fold change values of <1.0 are presented in a negative fold change format.

Table 5

RT-qPCR validation of SAGE profile for E15.5 versus P1.5 stages

			Fold change	
SAGE tag	**RefSeq accession**	**Gene ID**	**P1.5/E15.5 (SAGE)**	**P1.5/E15.5 (RT-qPCR)**
Atttctttggtgatttt	NM_010838	microtubule-associated protein tau, *Mapt*	5.51	1.53*
Gcactgttaacaagtgt	NM_009234	SRY-box containing gene 11, *Sox11*	-2.17	-3.32

All RT-qPCR data are statistically significant at $P < 0.001$ unless specified: * $P < 0.05$. Fold change values of <1.0 are presented in a negative fold change format.

Table 6

RT-qPCR validation of SAGE profile for rostral E15.5 and caudal E15.5 regions

			Fold change	
SAGE tag	**RefSeq accession**	**Gene ID**	**Caudal/rostral (SAGE)**	**Caudal/rostral (RT-qPCR)**
gttgttcttccagtcgg	NM_016916	bladder cancer associated protein homolog (human), *Blcap*	2.68	1.43
gtcatagctgttctgtg	BC025816	EST sequence BC025816	*Cs*	1.31 (NS)
aagcttgacatttggaa	NM_026187	ankyrin repeat and zinc finger domain containing 1, *Ankzf1*	*Rs*	1.17† (NS)
gatacttggaatgacta	NM_007393	actin, beta, cytoplasmic, *Actb*	-2.63	-1.32
ttggtgaaggaaaaaac	NM_021278	thymosin, beta 4, X chromosome, *Tmsb4x*	-2.33	-2.13 (NS)

All RT-qPCR data are statistically significant at $P < 0.05$ unless specified. †A disagreement between RT-qPCR and SAGE data. *Cs:* caudal region-specific expression; NS: no statistically significant difference between two developmental stages; *Rs:* rostral region-specific expression. Fold change values of <1.0 are presented in a negative fold change format.

DETs are (in descending order of enrichment) *Zfp57*, *Csrp2*, AA122503, *Cdk4*, *Sox4*, *Marcks*, *Actb*, BQ177889, *Hmgb3*, *Rps4x*, *Actl6b*, *Zswim4* and *Dr1*, whose expression ranges from 33- to 1.4-fold greater than in the adult. On the other hand, the *Plp1* is expressed at a level 40 times greater in the adult cerebral cortex compared to P1.5. Other validated genes that are enriched in the adult cerebral cortex include (in descending order) *Hprt1*, *Calm1* and *Mbp*, with a 2.3- to 1.8-fold greater expression than the P1.5 cerebral cortex. Comparison between E15.5 and P1.5 shows that *Mapt* has a 1.5-fold greater expression level in the P1.5 cerebral cortex while *Sox11* expression is 3.3-fold lower (Table 5).

We were unable to validate all 17 DETs from L versus Ri regions, suggesting the left and right hemispheres of the adult mouse cerebral cortex are highly similar and indistinguishable at the molecular level. SAGE and RT-qPCR analyses for R versus C regions of E15.5 are discussed in a separate section below.

Functional analysis of validated gene clusters using Ingenuity Pathway Analysis

The validated DETs of embryonic, adult and gene-switching clusters were functionally characterized using proprietary software, Ingenuity Pathway Analysis (IPA) from Ingenuity Systems®, to identify enriched molecular networks and canonical pathways. Given a list of input genes (known as focus genes), IPA mapped these genes to a global molecular network developed from information contained in the Ingenuity knowledge base (a manually curated database of experimentally proven molecular interactions from published literature). Networks of these focus genes were then algorithmically generated based on their connectivity. IPA determined the most significantly enriched biological functions and/or related diseases by calculating the *P*-value using Fisher's exact test. Using similar methods, significantly represented canonical pathways in a set of focus genes were also determined using IPA (Section C in Additional data file 1).

From the embryonic-specific gene clusters, we identified two statistically significant molecular networks (made up of 19 focus genes and 47 associated nodes; networks 1 and 2 in Figure 3; Figures S4 and S5 in Additional data file 1). The networks are interconnected through two genes, *Marcks* and *Neurod1*. In general, these networks are associated with cell cycle and DNA replication, recombination and repair processes. Three statistically significant (using $P < 0.05$ as a cut-off) canonical pathways are enriched in these networks (Figure 3); Wnt/β-catenin signaling (*Sox4*, *Sox11* and *Sfrp1*), P53 signaling (*Cdk4* and *Pcna*) and tight junction signaling (*Cdk4* and *Actb*) pathways. Validated DETs such as *Btg1*, *Cdk4*, *Cdkn1c*, *Csrp2*, *Ezh2*, *Neurod1*, *Pcna*, and *Rps4x* are associated with cell cycle control whereas *Actb*, *Ezh2*, *Als2cr2*, *Marcks*, *Robo1* and *Dcx* are associated with cellular assembly and organization. These processes are important in the formation of filopodia, membrane blebs and growth cones during neuronal growth, migration and axonogenesis. Known human neurological disorders associated with the networks, particularly network 2, include X-linked lissencephaly (Online Mendelian Inheritance in Man [OMIM:300067]; DCX), juvenile onset dystonia ([OMIM:607371]; ACTB) and Beckwith-Wiedemann syndrome ([OMIM:130600]; CDKN1C). All the DETs implicated in these networks are expressed in the cortical plate with the exception of *Pcna* (Table S5 in Additional data file 1) [31-42].

In adult-specific gene clusters, two molecular networks (18 focus genes and 50 associated nodes; networks 3 and 4 in Figure 3) were identified and interconnected via a single gene, *Mbp* (Figures S6 and S7 in Additional data file 1). These molecular networks enrich for nine statistically significant canonical pathways ($P < 0.05$) such as synaptic long-term potentiation and depression, calcium signaling, B cell receptor signaling, cAMP-mediated signaling, GM-CSF signaling, amyotrophic lateral sclerosis signaling, G-protein-coupled receptor signaling and xenobiotic metabolism signaling pathways. Validated DETs such as *Camk2a*, *Gria3*, *Itpr1*, *Egr1*,

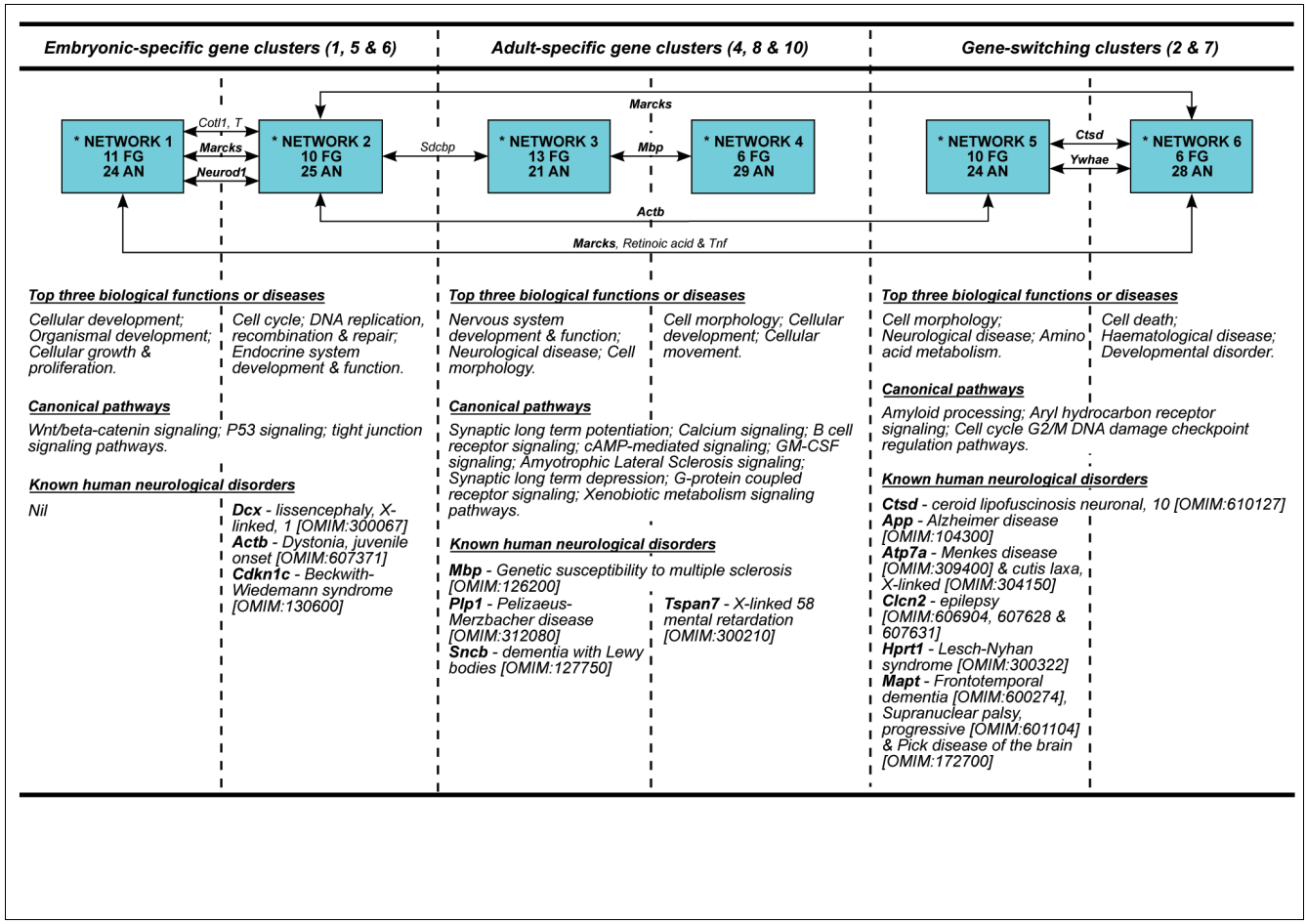


Figure 3

Novel molecular networks involved in cerebral corticogenesis. The figure shows novel molecular networks, related biological functions/diseases, canonical pathways and known human neurological disorders based on Ingenuity Pathway Analysis and OMIM database. Detailed molecular interactions for all networks (indicated by asterisks) are illustrated in Figures S4, S5, S6, S7, S8 and S9 in Additional data file 1. Gene names next to arrow lines refer to common genes shared by two networks. Bold gene name refers to a focus gene. AN: associated nodes; FG: focused genes.

Rgs4, *Gstm1* and *Ppp3ca* are associated with cell-to-cell communication in the adult cerebral cortex. Biological functions such as cell proliferation, movement and death processes are also linked to these networks through the genes *Camk2a*, *Cryab*, *Egr1*, *Gstm1*, *Itpr1*, *Mbp*, *Nrgn*, *Ppp1r1b*, *Ppp3ca*, *Rgs4*, *Sept4*, *Sirpa* and *Sncb*. Known human neurological disorders associated with Network 3 include Pelizaeus-Merzbacher disease ([OMIM:312080]; PLP1) and dementia with Lewy bodies ([OMIM:127750]; SNCB), whereas X-linked mental retardation ([OMIM:300210]; TSPAN7 or TM4SF2) is associated with network 4. MBP, found in both network 3 and 4, is associated with genetic susceptibility to multiple sclerosis [OMIM:126200]. All the DETs implicated in these networks are expressed in layers I to VI of the cerebral cortex, with the exception of *Gstm1* (layers II to V), *Sirpa* and *Ppp3ca* (layers II to VI), and *Nptxr* (no expression data available) (Table S5 in Additional data file 1) [32-34,43-52].

Both networks 5 and 6 (Figure 3) are linked to gene-switching clusters. These two molecular networks (14 focus genes and 52 associated nodes) are associated with cellular morphogenesis, amino acid metabolism, cell death processes and developmental disorders (Figures S8 and S9 in Additional data file 1). These networks are connected by *Ctsd* and *Ywhae* and are implicated in amyloid processing (*App* and *Mapt*), aryl hydrocarbon receptor signaling (*Ctsd* and *Nedd8*) and cell cycle G2/M DNA damage checkpoint regulation pathways (*Ywhae*). *App*, *Atp7a*, *Hprt1*, *Mapt*, *Ctsd* and *Clcn2* are involved in cell morphogenesis, assembly and organization. Among all the networks in the study, network 5 has the greatest number of DETs associated with known human neurological disorders with six out of ten focus genes being associated with known neurological disorders such as neuronal ceroid lipofuscinosis ([OMIM:610127]; CTSD), Alzheimer's disease ([OMIM:104300]; APP), Menkes disease ([OMIM:309400]; ATP7A), X-linked cutis laxa ([OMIM:304150]; ATP7A), epi-

lepsy ([OMIM:606904, 607628 and 607631]; CLCN2), Lesch-Nyhan syndrome ([OMIM:300322]; HPRT1), frontotemporal dementia ([OMIM:600274]; MAPT), progressive supranuclear palsy ([OMIM:601104]; MAPT) and Pick disease of the brain ([OMIM:172700]; MAPT). All the DETs implicated in these networks are expressed in the cortical plate during embryonic development with the exception of *Ube2e3*. In the adult, all DETs are expressed in layers I to VI of the cerebral cortex except for *Bcl11a*, *Clcn2*, *Hprt1* (layers II to VI), *Atp7a* and *Ube2e3* (no expression was detected in the adult cerebral cortex) (Table S5 in Additional data file 1) [32-35,46,53-61].

To refine the functional analysis to a cellular level, we grouped all 40 RT-qPCR validated DETs from networks 3, 4, 5 and 6 (adult specific and gene-switching gene clusters) into three groups according to where they are expressed: only in cortical neurons (N group); only in cortical glia (G group); and in both cortical neurons and glia (B group). None of the DETs from networks 1 and 2 (embryonic-specific gene clusters) were analyzed because most of the cells within the cerebral cortex are committed to the neuronal lineage at E15.5. The methods used to tabulate and group validated DETs are detailed in Section D in Additional data file 1. DETs classified as part of the N group are *Bcl11a*, *Calm1*, *Camk2a*, *Camk2n1*, *Hprt1*, *Itpr1*, *Mapt*, *Nedd8*, *Nrgn*, *Ppp1r1b*, *Ppp3ca*, *Rbm9*, *Rgs4*, *Sncb* and *Ywhae*, whereas those in the G group include *Cryab*, *Mbp*, *Nptxr*, *Plp1*, *Ppap2b*, *Sept4* and *Tspan7*. The B group consists of *App*, *Atp7a*, *Chgb*, *Ctsd*, *Egr1*, *Gria3*, *Gstm1* and *Sirpa*. DETs without any cellular expression information, such as *Actl6b*, *Clcn2*, *Hmgb3*, *Ube2e3*, AK138272, AK139402, AK140219, AK154943, AU258168 and BQ176089, were placed into group B to facilitate downstream analysis. Of the 40 DETs, only 35 were mapped to the IPA knowledgebase and subjected to further analysis. Functional analysis of these 35 DETs show that the neuron-based DETs (N and B groups) are enriched for various human neurological disorders, such as schizophrenia (9 DETs; $P = 4.36E-8$), Huntington's disease (8 DETs; $P = 1.03E-5$), X-linked mental retardation (4 DETs; $P = 3.78E-7$), Parkinson's disease (3 DETs; $P = 6.99E-3$) and Alzheimer's disease (3 DETs; $P = 1.52E-2$) (Figure S10 in Additional data file 1). Thirteen DETs associated with these neurological disorders are also implicated in processes related to cell death. Of these, eight DETs are expressed in neurons only, two in glia only and three in both neurons and glia.

From the above IPA analysis, *Cand2* (embryonic gene clusters), *Camk2n1* (adult gene clusters), *Hmgb3* (gene-switching clusters) and 10 ESTs (BQ176089, CD802535, AK139402, AU258168, AK138272, AK140219, AK154943, AA122503, NM_001033273 and BQ177886) are not currently connected to any networks.

Regionalized expression of DETs in the E15.5 cerebral cortex

Early regionalized development is an important event that could lay the foundations for adult arealization of cerebral functions. To identify genes with regionalized expression profiles, we compared SAGE libraries generated from rostral and caudal regions (equivalent to anterior and posterior regions of the human cerebral cortex) of the E15.5 cerebral cortex. We identified 44 DETs and selected 25 DETs (22 known genes, 1 EST and 2 ambiguous genes; Additional data file 2) and 2 genes of interest (bladder cancer associated protein, *Blcap*, and ankyrin repeat and zinc finger domain containing 1, *Ankzf1*) for RT-qPCR validation and further detailed region-based analysis. This was done using independent biological triplicates of clearly defined regions/quadrants, such as rostro-lateral (RL), rostro-medial (RM), caudo-lateral (CL) and caudo-medial (CM) of the E15.5 cerebral cortex (see Materials and methods). Two positive controls with known regionalised expression were included in the RT-qPCR: RAR-related orphan receptor beta (*Rorb*) and nuclear receptor subfamily 2, group F, member 1 (*Nr2f1*). *Rorb* is highly expressed in the rostral region whereas *Nr2f1* is highly expressed in the caudal region of the cerebral cortex [1,62].

An initial RT-qPCR analysis of combined RL and RM (rostral) as well as CL and CM (caudal) regions shows upregulation of *Rorb* and *Nr2f1* in rostral and caudal regions, respectively (based on fold change direction and magnitude of 1.3 times). The same analysis also confirmed the SAGE data for 3 out of 25 DETs (*Actb*, *Tmsb4x* and BC025816) and *Blcap* (Table 6; Additional data file 6). Both *Actb* and *Tmsb4x* have greater expression in the rostral region whereas *Blcap* and BC025816 are greater in the caudal region. To identify expression profiles in a more refined area and prevent regional compensation due to combined quadrant analysis, we performed a quadrant versus quadrant multiple regions comparison. The largest number of DETs were found in the RL versus CM comparison, as they are the two developmentally most distinct regions within the cerebral cortex at E15.5 compared to others; RL versus CL > RM versus CL > RL versus RM = RM versus CM > CL versus CM. The region-specific expression profiles were plotted for each of the DETs (Figure 4) and grouped into two categories: RL-specific DETs, such as *Actb*, *Tmsb4x* and cytochrome b-245, beta polypeptide (*Cybb*); and CM-specific DETs, including *Blcap*, EST BC025816, *Ankzf1* and cytochrome c oxidase I, mitochondrial (*Cox1*) (Additional data file 6).

To visualize the regionalized expression profiles, we performed *in situ* RNA hybridization (ISH) on all DETs validated by quadrant versus quadrant RT-qPCR analysis, *Blcap*, *Ankzf1* as well as the positive controls *Rorb* and *Nr2f1*. We performed ISH on sagittal and coronal sections (from rostral to caudal regions) of the E15.5 mouse brain (Figure 5). Under dark-field microscopic examination, we confirmed the regionalized expression of *Rorb* (at the rostral cortical plate;

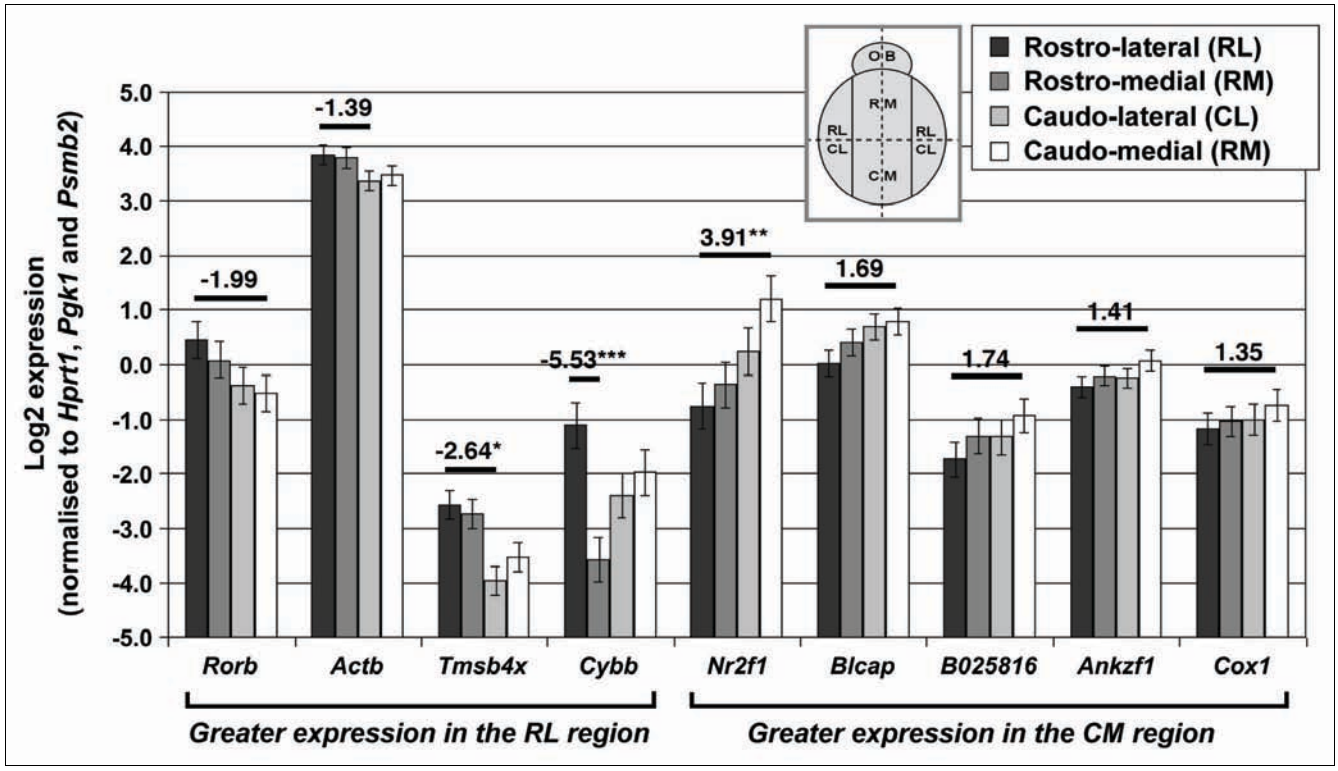


Figure 4
RT-qPCR analysis of all R versus C DETs based on quadrant versus quadrant analysis. RL, RM, CL and CM denote rostro-lateral, rostro-medial, caudo-lateral and caudo-medial regions of the cerebral cortex, respectively. OB denotes olfactory bulb. N = 3 per quadrant and data are presented as mean ± standard error of the mean. Fold change values (normalized to RL) are presented above the comparative bar and any values <1 are presented in the negative fold change format. Only the most significant fold change value is presented for each target gene. * P < 0.05; ** P < 0.01; *** P < 0.001.

Figure 5a) and *Nr2f1* (at the caudal ventricular zone; Figure 5f). From the analysis, *Actb* is highly expressed at the cortical plate and the subplate (Figure 5b). *Tmsb4x* is highly expressed at the cortical plate and the intermediate zone (Figure 5c-e).

ISH analysis showed that both *Blcap* and *Ankzf1* are caudal specific. Serial coronal sections from rostral to caudal regions of the brain (Figure 5g-i) show *Blcap* is weakly expressed in the rostral cerebral cortex, particularly at the intermediate zone, subplate and the cortical plate, but is highly expressed in the hippocampus and thalamus (Figure 5i). *Ankzf1* is expressed specifically in the ventricular zone as well as the cortical plate towards the caudal region of the cerebral cortex (Figure 5j-m). The distinctive expression of *Ankzf1* in both the ventricular zone and cerebral cortex prompted us to extend our ISH analysis to various developmental stages starting from E11.5 to adulthood (Figure 5n-t). *Ankzf1* is expressed in the primordial plexiform and the ventricular zone layers of the telencephalon at E11.5 (Figure 5n). By E13.5, *Ankzf1* is weakly expressed in the ventricular zone, but is highly expressed in the cortical plate (Figure 5o). From E17.5 to P1.5, the expression of *Ankzf1* is maintained in the cerebral cortex (Figure 5p, q). In the adult, *Ankzf1* expression is obvious in the piriform cortex, hippocampus and cerebellum (Figure 5r-t).

Genomic clustering of sense-antisense SAGE tags at the *Sox4* and *Sox11* loci

We performed genomic clustering analysis of the SAGE tags to determine any actively transcribed chromosomal loci throughout cerebral corticogenesis. Probabilities for chance occurrences of two, three, four, and five DETs being clustered within a window of ten adjacent tags present within each chromosomal location, irrespective of genetic distance, were calculated. This analysis was based on the DET lists described above (Additional data file 2). The analysis showed two over-represented chromosome loci at *Sox4* and *Sox11*, which derive from embryonic-specific gene clusters.

At both loci, we observed multiple SAGE tags with both sense and antisense orientations, which signify alternative polyadenylation sites, differential splicing and overlapping antisense transcription. As an initial validation of the antisense messages, we performed strand-specific RT-PCR (Figure 6) using cDNA synthesized from equally pooled total RNAs (three

Figure 5

In situ **RNA hybridization of selected R versus C regions of E15.5 DETs**. ISH was performed on **(a-m)** E15.5 and **(n-t)** E11.5 to P150 brains. (g-m) Coronal sections that are generated from the rostral to caudal axis; (a-f, s-t) sagittal sections. Micrographs of higher magnification are presented directly after any micrographs with an inset box (d-e, m, t). All micrographs are in dark-field except for (e, m), which are bright-field micrographs. C: caudal; CB: cerebellum; CP: cortical plate; DG: dentate gyrus; HPf: hippocampal formation; IZ: intermediate zone; OB: olfactory bulb; R: rostral; VZ: ventricular zone. Arrows show the region with expression or silver grains.

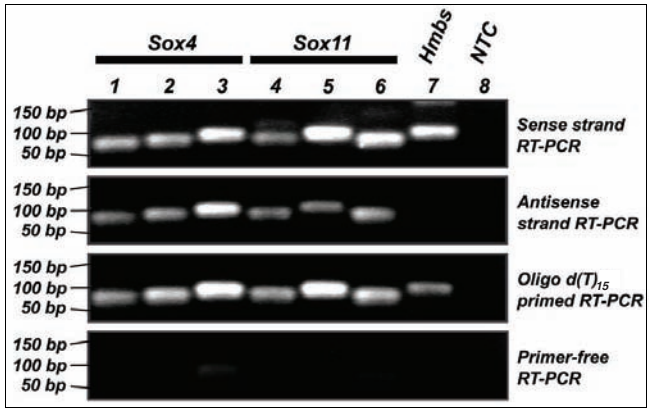


Figure 6

Strand-specific RT-PCR. Lanes 1 to 6 represent the amplification of the *Sox4* and *Sox11* transcripts using more than one probe. Lane 7 shows the amplification of the *Hmbs* housekeeping gene whereas lane 8 represents the amplification of water as non-template control (NTC) using the primer set for *Hmbs*. The first panel consists of amplicons generated from the reverse-transcribed sense strand cDNA whereas the second panel consists of amplicons generated from the reverse-transcribed antisense strand cDNA. The third panel represents amplicons generated from oligo-d $[T]_{15}$ primed reverse-transcribed cDNA, which serves as a positive control. The last panel represents amplicons generated from primer-free reverse-transcription reactions. The numbers on the left indicate the size of the generated bands.

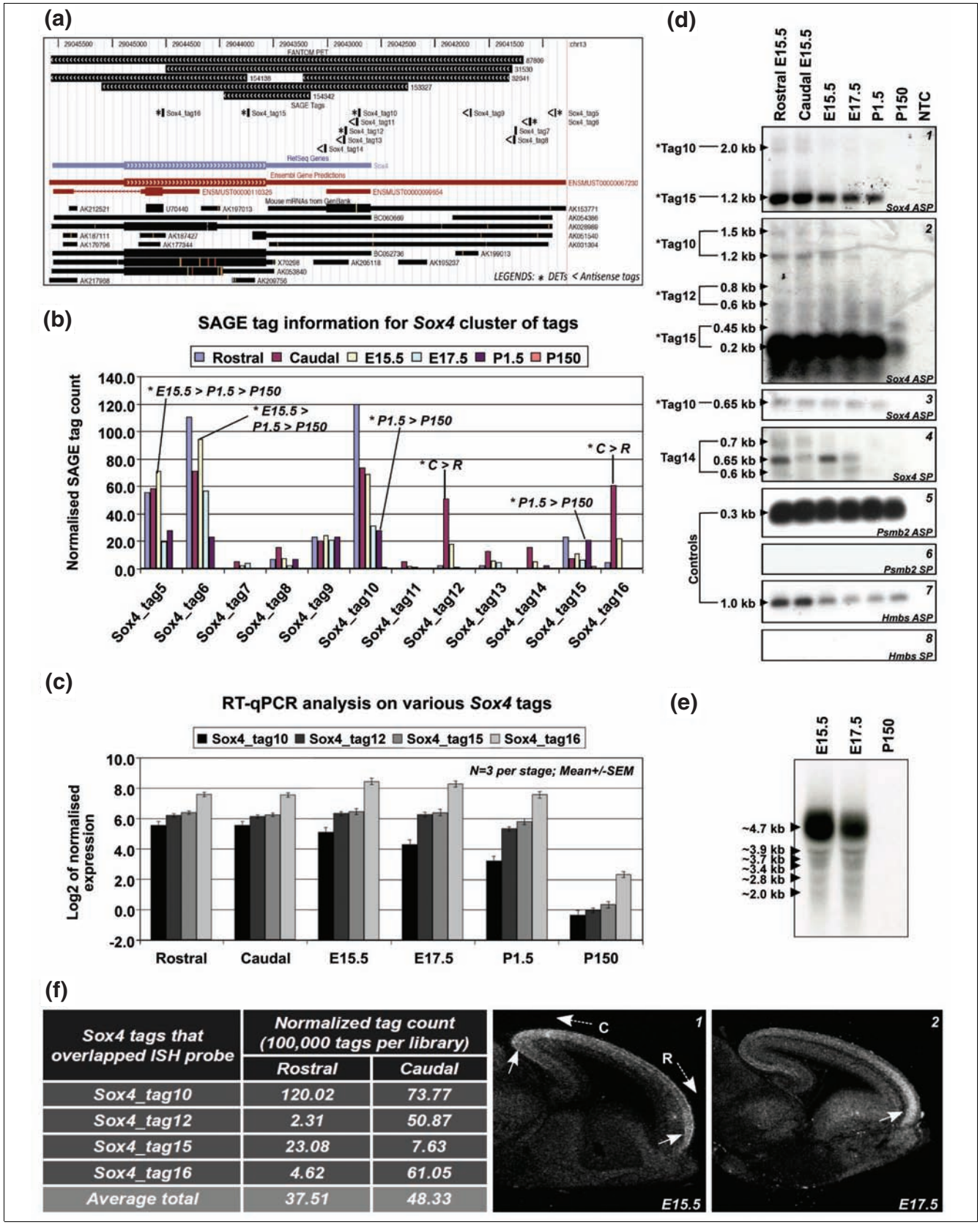
mice from each of E15.5, E17.5, P1.5 and adult stages). Three primer sets were used for each locus (designed at the middle, 5'- and 3'-ends of the canonical transcripts). For both *Sox4* and *Sox11* loci, the analysis showed positive signals from all primer sets used that were complementary to the antisense strand, therefore confirming the presence of one or more antisense transcript(s) that span the canonical transcript. Hydroxymethylbilane synthase gene (*Hmbs*) served as a negative control and there were no positive bands in the antisense strand RT-PCR, confirming the absence of antisense transcripts.

Genomic cluster at the *Sox4* gene locus

We mapped SAGE tags from the genomic cluster to the *Sox4* gene locus using the University of California Santa Cruz (UCSC) genome browser (Figure 7a; Additional data file 7) [63]. Only tags within and around the *Sox4* canonical transcript are shown. Evidence of mapped mouse mRNAs within this locus further justifies the existence of multiple SAGE tags in addition to the canonical *Sox4* transcript. Subsequent validation of the genomic clusters was solely based on the SAGE tags situated within the canonical transcript. Based on the SAGE tag information, 6 out of 12 tags are DETs. These include four DETs within the canonical transcript (Figure 7b), with sox4_tag10 and sox4_tag15 having greater expression in P1.5 compared to P150 (adult stage), whereas sox4_tag12 and sox4_tag16 are both abundantly expressed in the caudal region of the E15.5 cerebral cortex. Based on the RT-qPCR analysis (Figure 7c), sox4_tag10 has greater expression in E15.5 compared to E17.5 (1.76-fold change), P1.5 (3.72-fold change) and Ad (43.67-fold change). For sox4_tag12, sox4_tag15 and sox4_tag16, differences are seen only in P1.5 (-2.00-fold, -1.57-fold and -1.84-fold changes, respectively) and Ad (-82.08-fold, -68.92-fold and -69.41-fold changes, respectively) when compared to E15.5. RT-qPCR analysis on the same tags did not find any differences between rostral and caudal cerebral cortices of E15.5. The differences in fold change between DETs suggest irregular overlapping of various transcript variants at different *Sox4* gene loci.

To further validate the expression profiles of the multiple *Sox4* DETs, we performed 3' rapid amplification of cDNA ends (RACE)-Southern analysis using pooled adaptor-oligo-d $[T]_{15}$ synthesized cDNAs from three mice at each developmental stage. Based on this method, we were able to semi-quantitatively and accurately measure the expression levels of individual SAGE tags at the locus. To show that the amplification was cDNA specific, we performed PCR by using the same primer sets on mouse genomic DNA under the same conditions. In all cases, no amplification was observed (data not shown). This analysis confirmed the presence of four out of seven alternative transcripts for *Sox4* (Figure 7d). Corresponding tags were determined by estimating the amplicon sizes between the strand-specific primers used, the next downstream AAUAAA/AUUAAA polyadenylation signal (if any) and succeeding CATG sequence or SAGE tags. Figure 7d(1)-d(3) confirms the existence of sox4_tag10, sox4_tag12 and sox4_tag15. Of these tags, SAGE expression profiles of sox4_tag10 and sox4_tag15 were validated (embryonic-specific and reduced expression after P1.5) but not sox4_tag12 (E15.5 caudal region-specific). 3' RACE-Southern analysis using a sense probe detected bands in the rostral and caudal regions of E15.5, E15.5 and E17.5 cerebral cortices and, therefore, confirmed the existence of the *Sox4* antisense transcripts (Figure 7d(4)). Even though none of these tags were differentially expressed in between these regions based on the SAGE analysis, our findings show distinctive regionalization for sox4_tag14 expression at the E15.5 rostral cerebral cortex. Proteasome (prosome, macropain) subunit, beta type 2 gene (*Psmb2*) and *Hmbs* were used as controls and no antisense or alternative transcripts were identified at these gene loci (Figure 7d(5)-d(8)).

Since 3' RACE-Southern analysis was dependent on oligo-$[dT]_{15}$ priming, we could not rule out the possibility of amplicons that were generated by false priming on homopolymer-A stretches. Therefore, Northern analyses were performed on equally pooled total RNA extracted from the cerebral cortices of seven mice at E15.5, E17.5 and P150 (negative control). By using a double-stranded DNA probe at the 3' untranslated region (UTR) of *Sox4* (Additional data file 8), we identified six bands ranging from approximately 2 kb to approximately 4.7 kb (Figure 7e). *Sox4* sense transcripts are weakly expressed in



Sox4 tags that overlapped ISH probe	Normalized tag count (100,000 tags per library)	
	Rostral	Caudal
Sox4_tag10	120.02	73.77
Sox4_tag12	2.31	50.87
Sox4_tag15	23.08	7.63
Sox4_tag16	4.62	61.05
Average total	37.51	48.33

Figure 7 *(see legend on next page)*

Figure 7 *(see previous page)*
Genomic cluster at the *Sox4* locus. **(a)** The UCSC genome browser of the over-represented *Sox4* genomic locus. SAGE tags were found in both directions within the *Sox4* reference gene. **(b)** The SAGE expression profiles for each tag in the *Sox4* locus. **(c)** The RT-qPCR validations of selected DETs at various loci within the *Sox4* canonical gene. **(d)** The 3' RACE-Southern blotting analysis. The left panel of the figure shows the sizes of the bands and their corresponding tags. (d1-d3) Amplification of *Sox4* sense transcripts (ASP); (d4) amplification of *Sox4* antisense transcripts (SP). Tags with asterisks are DETs. Both (d5) and (d7) are positive controls exclusively generated from the sense strand of *Psmb2* and *Hmbs* housekeeping genes (ASP), respectively. The corresponding (d6) and (d8) are the antisense expression (negative control) of *Psmb2* and *Hmbs* (SP), respectively. **(e)** Northern analysis of total RNA isolated from pooled mouse cerebral cortices (N = 7). **(f)** The regionalized expression of *Sox4* sense transcripts determined by ISH. The table consists of SAGE information for related *Sox4* sense tags at rostral and caudal regions of the E15.5 cerebral cortex. (f1) and (f2) are sagittal sections obtained from E15.5 and E17.5 mouse brains, respectively. The figures show the expression of the *Sox4* sense transcript variants. Arrows show brain regions with greater *Sox4* sense expression. C: caudal region of the cerebral cortex; R: rostral region of the cerebral cortex.

the adult, but highly expressed in the embryonic cerebral cortex. The number of bands observed is similar and corresponds to overlapping mouse mRNAs as well as publicly available paired-end diTag (PET) sequences downloaded from Ensembl (Figure 7a; Table S6 in Additional data file 1) [64]. Taken together, the analysis confirmed the existence of multiple overlapping variants of *Sox4* sense and antisense transcripts at this gene locus.

To confirm the rostro-caudal expression of *Sox4* sense transcripts, we performed ISH on sagittal sections of mouse brains using a *Sox4* antisense riboprobe that spanned across the sox4_tag10, sox4_tag12, sox4_tag15 and sox4_tag16 SAGE tags. *Sox4* showed regionalized expression at E15.5 and E17.5 (Figure 7f). At E15.5, *Sox4* sense transcripts are expressed more in both the rostral- and caudal-end regions of the cortical plate compared to the intermediate region between them (Figure 7f(1)). By E17.5, expression of *Sox4* sense transcripts is obvious in the rostral cortical plate (Figure 7f(2)). At both stages of development, *Sox4* sense transcripts are uniformly expressed in the intermediate zone of the cerebral cortex. These findings correspond to the SAGE tag counts for E15.5 rostro-caudal regions of the cerebral cortex (Figure 7f). These observations explain the averaged total tag count per 100,000 tags for different *Sox4* sense transcripts, which are predominantly expressed in both rostral and caudal regions of the cerebral cortex. The regionalized expression of *Sox4* in the cortical plate is obvious only at E15.5 and E17.5, but not at other stages of development (Figure S11 in Additional data file 1).

In situ RNA hybridization of *Sox4* sense and antisense transcripts

To further ascertain the antisense expression of *Sox4* in a spatio-temporal manner, we performed ISH on coronal sections obtained from E11.5, E13.5, E15.5, E17.5, P1.5 and P150 mouse brains. Sense and antisense RNA probes were generated from the same clone used in the Northern analysis. At E11.5, *Sox4* sense transcripts are confined to the primordial plexiform layer (Figure 8a). From E13.5 to P1.5, the sense transcripts are expressed throughout the cortical plate (Figure 8b-e). Expression of sense transcripts in the subventricular zones is observed at E17.5 and P1.5 only (Figure 8d, e). There is no observable sense expression in the adult stage (Figure 8f). *Sox4* antisense transcripts are expressed throughout the telencephalon at E11.5 (Figure 8g). From E13.5 to P1.5, *Sox4* antisense expression is confined to the cortical plate only (Figure 8h-k). There is no obvious antisense expression in the cerebral cortex in the adult stage (Figure 8l). A microscopic examination at high magnification showed that *Sox4* antisense transcripts are predominantly localized in the nucleus whereas *Sox4* sense transcripts are found in both the nucleus and cytoplasm (Figures S12, S13 and S14 in Additional data file 1). We used hemoglobin alpha, adult chain 1 (*Hba-a1*) of the corresponding brain region and time-point as a control in the analysis (Figure 8m-r).

Furthermore, *Sox4* antisense expression occurs in the piriform cortex layer II (Figure 9a-c) and dentate gyrus (Figure 9g-i) in the adult brain; however, no sense expression is observed in these regions. At P1.5, we identified complementary expression between *Sox4* sense and antisense transcripts in the olfactory bulb (Figure 9d-f). *Sox4* sense expression was confined to the granular and glomerular layers of the olfactory bulb whereas antisense expression was found only in the outer plexiform layer. We used either *Sox11* or *Hba-a1* of the corresponding brain region and time-point as a control in the analysis (Figure 9c, f, i).

Analysis of the *Sox11* genomic cluster

SAGE tags, which represent multiple overlapping sense and antisense transcript variants at the *Sox11* genomic cluster, were validated using 3' RACE-Southern analysis as described above. See Section F in Additional data file 1 for a full description of the *Sox11* results. ISH analysis did not confirm the expression of antisense transcripts of *Sox11*, but the presence of PETs spanning three out of five antisense tags confirmed the existence of *Sox11* antisense transcripts (Table S8 and Figures S16, S17 and S18 in Additional data file 1). The discrepancy between ISH and RT-qPCR or 3' RACE-Southern analysis suggests that *Sox11* antisense transcripts might be expressed at low levels or at specific locations of the cerebral cortex, and hence can be detectable only by using serial sections or whole mount ISH.

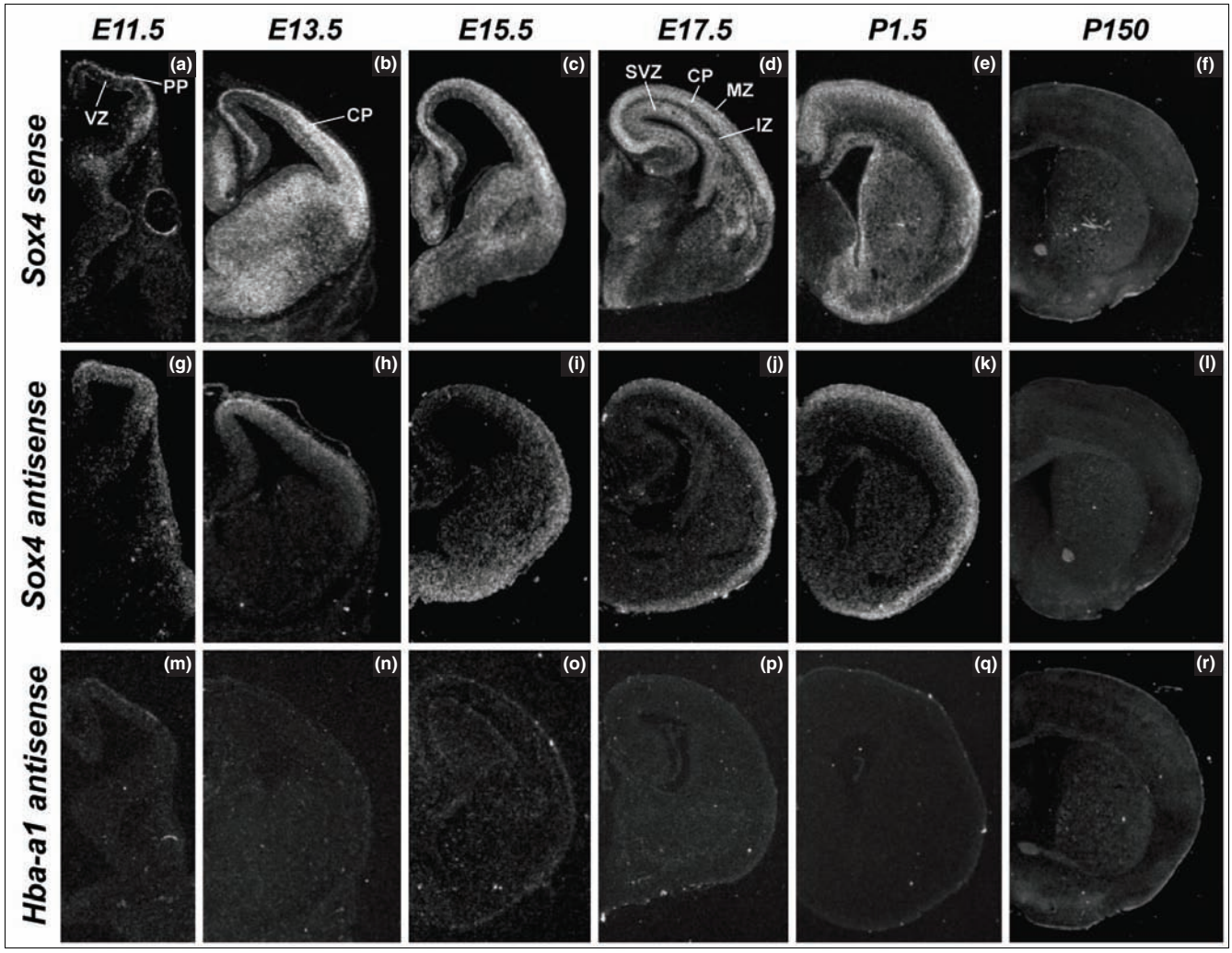


Figure 8
ISH analysis of *Sox4* transcripts in E11.5 to P150 mouse brains. **(a-f)** Expression of the sense transcript for *Sox4*. **(g-l)** The expression of the antisense transcript for *Sox4*. **(m-r)** *Hba-a1* antisense expression (negative control). All micrographs were taken from coronal sections. CP: cortical plate; IZ: intermediate zone; MZ: marginal zone; PP: primordial plexiform layer; SVZ: subventricular zone; VZ: ventricular zone.

Screening of *Sox4* and *Sox11* antisense transcripts in the adult mouse brain, organs, P19 cell line and neurospheres

We screened various adult brain regions (olfactory bulb, cerebellum, medulla, hippocampus, thalamus and cerebral cortex) and selected mouse organs (E15.5 whole brain, heart, kidney, liver, skeletal muscle, skin, spleen, stomach, testis and thymus) for the expression of *Sox4* and *Sox11* antisense transcripts by strand-specific RT-qPCR. Within the adult brain, *Sox4* sense and antisense transcripts are expressed in all regions, with the highest level found in the olfactory bulb, which is approximately four- to nine-fold greater than those in other brain regions (Figure 10a). Expression of *Sox4* antisense transcripts occurs in all mouse organs, with the highest level in the thymus followed by E15.5 whole brain, testis and skin (Figure 10b). *Sox4* sense and antisense expression profiles are similar throughout the entire series of samples screened, with the sense transcripts being consistently expressed at a greater level than the antisense transcripts (approximately 1.7-fold in various brain regions and approximately 2- to 14-fold in various organ comparisons).

Sox11 sense transcripts are expressed at the highest level in the olfactory bulb, approximately two- to seven-fold greater than those in other brain regions (Figure 10a). *Sox11* antisense transcripts, on the other hand, are expressed in all brain regions screened and at a comparable level in the olfactory bulb, hippocampus, thalamus and cerebral cortex. In comparison to other adult mouse organs, *Sox11* sense and antisense transcripts are highly expressed in the E15.5 whole brain, with *Sox11* sense transcript levels at least 100-fold greater than those in other mouse organs (Figure 10b). On the other hand, *Sox11* antisense expression is observed only in the E15.5 whole brain, skin and stomach. Notably, *Sox11* sense

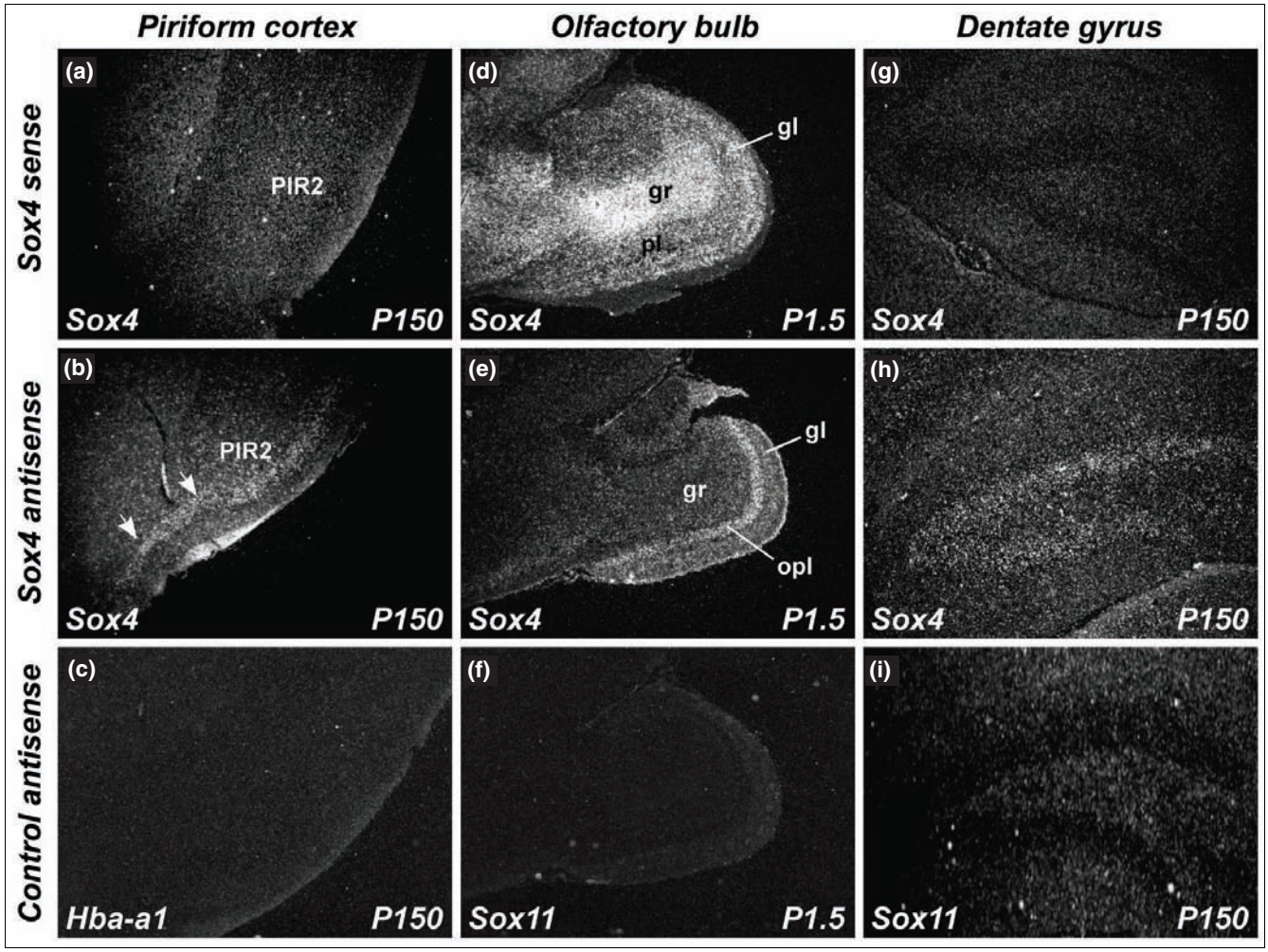


Figure 9

***Sox4* antisense expression in other brain regions**. **(a-i)** *Sox4* antisense expression is also observed in other regions such as the piriform cortex (a-c; arrows), olfactory bulb (d-f) and dentate gyrus (g-i). (a, d, g) *Sox4* sense expression; (b, e, h) *Sox4* antisense expression. (c, f, i) *Hba-a1* or *Sox11* antisense expression (negative controls). All micrographs were taken from sagittal sections except (a-c), which were taken from coronal sections. gl: glomerular layer; gr: granule layer; Opl: outer plexiform layer; PIR2: piriform cortex layer II.

transcripts are expressed more highly than antisense transcripts in the E15.5 whole brain and skin (23- and 4-fold, respectively).

Since both Sox4 and Sox11 are implicated in neuronal differentiation and glial maturation processes [65,66], we examined both *Sox4* and *Sox11* sense and antisense transcript expression in proliferating and differentiating P19 (embryonal carcinoma cells) and in embryonic NSPCs grown as neurospheres. Both *Sox4* sense and antisense transcripts are upregulated during P19 cell differentiation (approximately 5.7- and 1.6-fold upregulation, respectively; Figure 11a) and neurosphere differentiation (approximately 1.9- and 1.8-fold upregulation, respectively; Figure 11b). For *Sox11*, both sense and antisense transcripts are upregulated in the differentiating compared to the proliferating P19 cells by approximately 2.3- and 4.2-fold, respectively (Figure 11c). Both the *Sox11* sense and antisense transcripts are, however, downregulated in the differentiating neurospheres (approximately 2.6- and 1.5-fold, respectively; Figure 11d).

Discussion

In this study, SAGE was used to analyze global gene expression in the normal mouse cerebral cortex at various developmental stages. We report validated spatio-temporal regulation of genes involved in mouse cerebral cortex development from embryo to adulthood. The study highlights four main findings: association of DETs from different gene clusters with known functional processes or signaling pathways and disease-causative genes that are involved in cerebral corticogenesis; *Ankzf1* and *Sox4* sense transcripts are regionally expressed in the E15.5 cerebral cortex; multiple overlapping *Sox4* and *Sox11* sense and antisense transcripts are spatio-

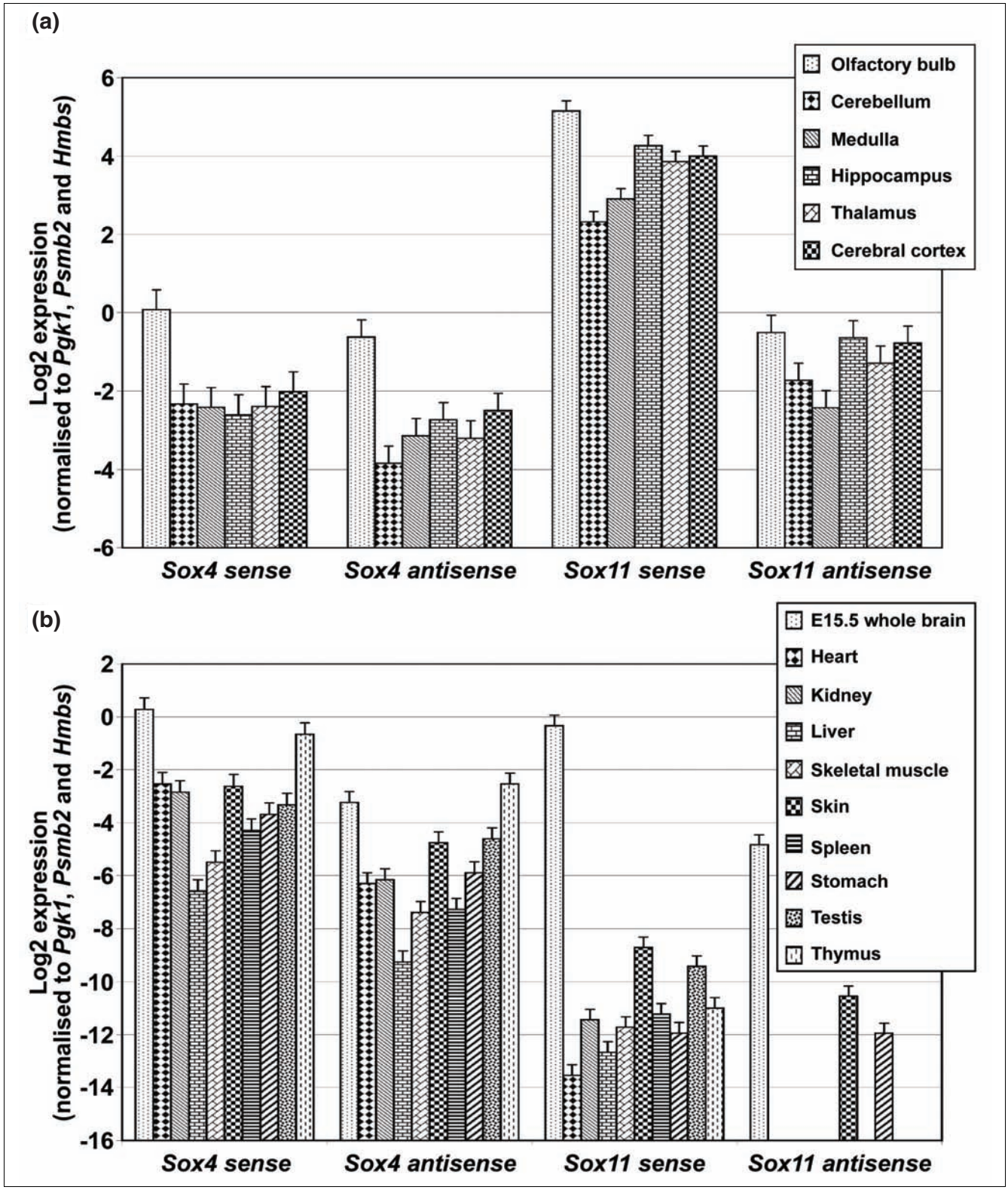


Figure 10

Expression of *Sox4* and *Sox11* transcripts in various mouse organs. **(a)** Strand-specific RT-qPCR screening of *Sox4* and *Sox11* sense and antisense transcript expression in various adult mouse brain regions. N = 2 and data are presented as mean ± standard error of the mean (SEM). **(b)** Strand-specific RT-qPCR screening of *Sox4* and *Sox11* sense and antisense transcripts in various adult mouse organs. N = 3 and data are presented as mean ± SEM.

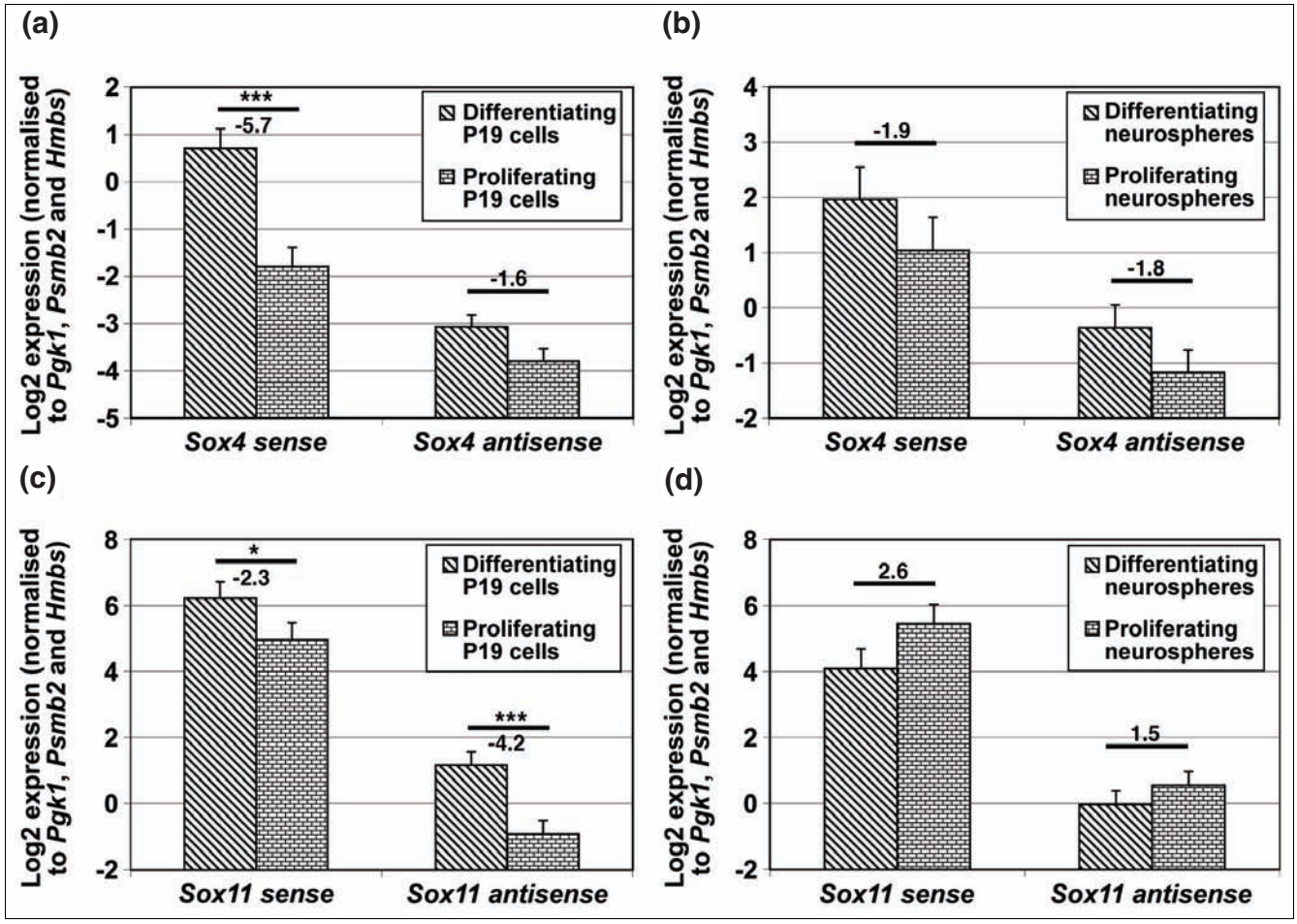


Figure 11

Expression of *Sox4* and *Sox11* transcripts in neuropspheres and P19 cells. (a-d) The figure shows strand-specific RT-qPCR screening of *Sox4* (a, b) and *Sox11* (c, d) sense and antisense transcripts expression in proliferating and differentiating P19 cells (a, c) and neuropspheres (b, d). N = 3 for P19 cells and N = 2 for neuropspheres. All data are presented as mean ± standard error of the mean. Fold change values (normalization to proliferating cells) are presented above the comparative bar and any values <1 are presented in the negative fold change format. * P < 0.05; ** P < 0.01; *** P < 0.001.

temporally regulated during cerebral corticogenesis; and *Sox4* and *Sox11* antisense transcripts are differentially regulated in both proliferating and differentiating embryonic-derived neurospheres and P19 cells.

We have shown that most tags generated in all libraries were singletons. The number of singletons could be reduced by increasing the number of tags sequenced. In mammalian cells, the number of additional unique transcripts identified approached zero when the number of SAGE tags sequenced reached approximately 600,000 [67]. Increasing the number of tags sequenced could improve the sensitivity of the technique to identify weakly expressed or novel transcripts, and the application of massively parallel signature sequencing [68] using a next-generation sequencer would be an ideal solution to accomplish this. In addition, one of the benefits of SAGE is that datasets generated from different groups or in public repositories such as SAGE Genie [69] and GEO [28] are readily comparable and, hence, can increase the tag count and sensitivity of the technique in discovering DETs between SAGE libraries. However, any meta-analyses involving various SAGE datasets are affected by experimental and biological variation; thus, a careful selection of matching libraries is crucial to limit systematic error or biases.

Our SAGE analysis robustly detected DETs with a low false positive rate (for example, <0.001% for comparison between left and right hemispheres of the adult cerebral cortex). Of all the identified DETs, approximately 8% were not mapped to either a single locus in the mouse genome or any unique annotation. This problem could be overcome by generating additional information from the 5' end of the transcript through alternative techniques such as PET sequencing [70], cap analysis gene expression (CAGE) [71] and 5' LongSAGE [72].

We have identified functional ontologies, molecular interactions and enriched canonical pathways that are distinct to the

stage-specific gene clusters of validated DETs. The IPA network analysis generated connections between validated DETs across various developmental stages in relation to well-established proteins or molecules and neurological disorders. In the study, members of the Wnt/β-catenin signaling pathway were enriched in networks 1 and 2 (embryonic-specific gene clusters). In neural development, Wnt/β-catenin signaling plays an important role in regulating regional specification of the cortex along the rostro-caudal and dorso-ventral axes, and proliferation of cortical progenitors [73]. IPA highlighted three genes (*Sox4*, *Sox11* and *Sfrp1*) associated with this pathway. In humans, the SFRP1 protein (secreted frizzled-related protein 1) stabilizes β-catenin and increases transcription from β-catenin-responsive promoters [74]. In β-catenin-deficient mouse mutants, expression of both *Sox4* and *Sox11* is downregulated [75,76]. Sox4 and Sox11 proteins play an important role in establishing neuronal properties, pan-neuronal gene expression and proper myelination of the central nervous system [65,66]. This suggests that the role of the Wnt/β-catenin signaling pathway in regulating neuronal development could be mediated, at least in part, by the Sox4 and Sox11 proteins.

The role of Wnt/β-catenin signaling in regulating DETs (*Btg1*, *Cdk4*, *Cdkn1c*, *Csrp2*, *Ezh2*, *Neurod1*, *Pcna*, and *Rps4x*) involved in cell cycle and proliferation remains unclear. Gain- and loss-of-function studies have established that Wnt/β-catenin signaling is essential to maintain the pool of precursors for proper development of the cerebral cortex [77,78]. To date, there is no direct evidence to show that *Ezh2*, *Pcna*, *Rps4x* and *Btg1* are involved in cell cycle regulation during early embryonic neurogenesis. But their expression in the ventricular/subventricular zone of the E15.5 developing mouse cerebral cortex [33,35,36] suggests that they may be involved in regulating cell proliferation during neurogenesis. The role of these DETs and their association with the Wnt/β-catenin signaling pathway remains unclear and requires further experimentation.

Networks 3 and 4 from the adult gene cluster were associated with various canonical pathways, in particular, synaptic long term potentiation (LTP) and calcium signaling. LTP is a process of synapse enhancement, which is thought to underlie some forms of learning and memory [79]. This process depends on Ca^{2+} and calmodulin, which are major components of the calcium signaling pathway. We identified four DETs, *Gria3*, *Itpr1*, *Ppp3ca* and *Camk2a*, in both canonical pathways. In particular, the Camk2a enzyme is highly expressed in the brain and regulates mainly glutamatergic synapses during LTP [79]. DET products such as Nrgn and Camk2n1 can directly or indirectly regulate Camk2a or Ca^{2+}/calmodulin and subsequently alter the outcome of the LTP pathway [80,81]. Therefore, they may serve as important candidate genes in the future analysis of the synaptic LTP pathway involving neurodegenerative diseases that cause the loss of cognitive function and memory.

Networks 5 and 6 are enriched for genes in the amyloid processing signaling pathway. *App* and *Mapt* are associated with this pathway. Under normal circumstances, the App protein is required for proper migration of neuronal precursors into the cortical plate in early embryonic corticogenesis [82]. The Mapt protein, on the other hand, plays an important role in maintaining the architecture of the neuronal cytoskeleton and intracellular trafficking. Overexpression of App protein and hyperphosphorylation of the Mapt protein have been implicated in the pathologies of Alzheimer's disease [83,84]. Interestingly, *Ctsd*, *Atp7a*, *Clcn2* and *Hprt1*, the genes responsible for other human neurological disorders such as neuronal ceroid lipofuscinosis (*Ctsd*), Menkes disease (*Atp7a*), epilepsy (*Clcn2*), and Lesch-Nyhan syndrome (*Hprt1*), are associated with *App* and *Mapt*. These candidate genes are also involved in cell morphogenesis, assembly and organization and could be linked to deterioration of neurons during the pathologic progression of these disorders.

Pathway analysis of DETs classified into the N, G and B groups showed DETs in the neuron (N and B groups) are associated with Huntington's disease and schizophrenia, which were not previously identified in networks 3 to 6. Our analysis showed that both disorders share three common DETs, namely *Rgs4*, *Ppp1r1b* and *Chgb*, whose expression is downregulated in humans with Huntington's disease or schizophrenia [85-88]. A proportion of patients with Huntington's disease also develop schizophrenia [89,90]. Taken together, downregulation of *Rgs4*, *Ppp1r1b* and *Chgb* expression in neurons may contribute to the common symptoms in these disorders. Our findings imply that many DETs (including *App*, *Hprt1* and *Sncb*) associated with both Huntington's disease and schizophrenia are also involved in neuronal/cell death processes [91-93]. Other DETs in the N group, not previously implicated in neuronal cell death, may serve as novel potential candidates during pathologic development in these disorders.

Regionalized development of the cerebral cortex involves the differential regulation of cell cycle exit, early migration and attainment of positional identity in neuronal fated cells. To date, only few genes have been associated with regionalized development of the cerebral cortex [3,94]. In the regionalization analysis, we identified the highest number of DETs in the comparison of the RL and CM libraries, which signifies that these two regions of the cerebral cortex are the most different. This finding supports the notion that the cerebral cortex is developed in a latero-medial axis followed by a rostro-caudal axis [7,8]. At E15.5, both *Actb* and *Tmsb4x* were expressed greater in the rostral cerebral cortex than in the caudal region. Both Actb and Tmsb4x proteins are involved in the actin cytoskeleton-signaling pathway [95]. In particular, the Tmsb4x protein has been shown to promote cardiomyocyte migration [96] and axonal tract growth in zebrafish [97]. Therefore, co-expression of *Actb* and *Tmsb4x* in the E15.5 mouse cortical plate suggests that they may have a synergistic

role in early cortical cell development. Conversely, *Blcap* and *Ankzf1* were expressed more highly in the caudal than in the rostral region of the E15.5 cerebral cortex. To date, the function of both *Blcap* and *Ankzf1* in the cerebral cortex remains uncharacterized. This study provides the first comprehensive expression profile of *Ankzf1* and suggests it could be an important transcription factor in cerebral corticogenesis.

At E15.5, *Sox4* sense transcripts were expressed in a high-rostral and high-caudal manner with lesser expression within the intermediate region. By E17.5, *Sox4* expression becomes obvious at the rostral cortical plate, which is similar to *Rorb*. But, we did not find that the regionalized expression of *Sox4* sense transcripts resembles that of the restricted *Rorb* expression at E15.5 or in the postnatal brain [1,62]. This finding could be caused by the combined expression profiles of different *Sox4* sense transcripts that are present across the rostro-caudal axis of the cortical plate. The regionalized expression of *Sox4* sense transcripts occurs only between E15.5 and E17.5. Because the thalamic axon innervates the cortical plate after E17.5 [98], the regionalization of *Sox4* sense transcripts in early cortical development could be an outcome of an intrinsic instead of an extrinsic mechanism that regulates early patterning of the cerebral cortex.

Genomic clustering of DETs identified the differentially regulated *Sox4* and *Sox11* gene loci. These genomic clusters imply that there are multiple overlapping sense and antisense transcripts surrounding the same gene locus that are co-transcribed simultaneously during cerebral cortex development. Both *Sox4* and *Sox11* are single exon genes and these transcript variants are therefore likely to be generated due to alternative polyadenylation. The 3' UTRs of both *Sox4* and *Sox11* have tandem terminal polyadenylation signals on both sense and antisense strands (data not shown), which supports the occurrence of multiple transcript forms or SAGE tags. Multiple mRNA forms with different 3' UTRs can lead to cell-specific regulation, different nuclear or cytoplasmic mRNA stability and translation rates [99,100]. The 3' UTR of *Sox4* and *Sox11* may contain AU-rich elements that play an important role in determining mRNA stability through deadenylation, decapping or 3' → 5' decay [101]. Besides, different 3' UTR lengths may be targeted by different miRNAs, thus interfering with the translation process. Both *Sox4* and *Sox11* transcripts may be targeted by various miRNAs at different predicted positions across the 3' UTR (Tables S7 and S9 and Figures S15 and S18 in Additional data file 1). Therefore, 3' UTR lengths of *Sox4* and *Sox11* may be an important feature in the regulation of their protein expression during cerebral corticogenesis.

In the study, NATs were found at both the *Sox4* and *Sox11* gene loci overlapping the sense transcripts. Overlapping NATs may function as templates for the generation of pre-miRNA and mature miRNA with exceptional high sequence conservation that complement the overlapping sense protein-coding transcripts [102]. To date, no mature or pre-miRNAs have been predicted on the *Sox4* and *Sox11* sense and antisense strands (data not shown) or have been reported in miRBase [103]. In addition, NATs can self-complement to form double stranded RNA or pair with sense transcripts and function as templates for the generation of endogenous small interfering RNAs, which could subsequently interfere with translation or transcription of multiple protein-coding transcripts [104]. Because both Sox4 and Sox11 proteins are highly expressed in the cerebral cortex, the overlapping NATs do not seem to be involved in the regulation of Sox4 and Sox11 through miRNA- or small interfering RNA-mediated translation repression mechanisms, but rather through antisense-regulated sense transcription within the nucleus.

Our ISH analysis showed complementary cellular expression profiles of *Sox4* sense and antisense transcripts at the piriform cortex, olfactory bulb and dentate gyrus. This finding implies that *Sox4* antisense transcripts may be essential in intracellular and interlocus negative feedback loop regulation of the *Sox4* sense transcripts. Similar expression profiles of *Sox4* sense and antisense transcripts in multiple mouse organs and brain regions, however, suggest that these transcripts may be co-expressed. This observation is also supported by the temporal co-expression of *Sox4* sense and antisense transcripts in the cortical plate or layers I to III of the cerebral cortex. Taken together, the sense and antisense transcripts of *Sox4* are co-expressed in some cells and expressed complementarily in other cells, suggesting crucial cell-type-specific regulation.

Sox4 and Sox11 have been shown to have redundant roles during mouse development [105], and Sox11 may play a compensatory role in the absence of Sox4 during brain development [106]. We demonstrated that *Sox4* and *Sox11* sense and antisense transcripts have a similar expression in the brain, but not in other organs, suggesting a compensatory role for Sox11 only in the brain. *Sox11* antisense transcripts were expressed in the brain, skin and stomach only, suggesting organ-specific regulation.

Our data show upregulation of *Sox4* and *Sox11* sense transcripts in differentiating P19 cells, consistent with the findings of others [107,108], and demonstrate upregulation of antisense transcripts as well. We also find both *Sox4* and *Sox11* sense transcripts expressed in the NSPCs cultured as neurospheres, which is in agreement with Dy *et al.* [109]. Furthermore, we identify upregulation of both *Sox4* sense and antisense transcripts but downregulation of *Sox11* sense transcripts in differentiating neurospheres. Taken together, our findings show that there are potentially common and distinct roles for *Sox4* and *Sox11* sense and antisense transcripts during neuronal and non-neuronal cell proliferation and differentiation. The underlying regulatory mechanism of these transcripts, particularly the antisense ones, remains unknown and requires further investigation.

Conclusions

This study provides avenues for future research focus in understanding the fundamental processes and development of neurological disorders related to the cerebral cortex. We confirm the regionalized expression of new candidate genes in the E15.5 cerebral cortex as well as differential regulation of multiple overlapping sense and novel antisense transcripts within *Sox4* and *Sox11* gene loci during cerebral corticogenesis. We also report for the first time the spatio-temporal regulation of *Sox4* antisense transcripts in the brain as well as differential regulation of novel *Sox4* and *Sox11* antisense transcripts in various mouse organs and in proliferating and differentiating NSPCs and P19 cells. The finding provides an insight for future investigations into the role of antisense transcripts during cerebral corticogenesis and neuronal differentiation.

Materials and methods

Serial analysis of gene expression

Handling of animals and dissection of the cerebral cortex

All experiments that involved animal breeding and handling were performed according to protocols approved by the Melbourne Health Animal Ethics Committee (Project numbers 2001.045 and 2004.041). All animals involved in the study were C57BL/6 mice unless specified otherwise. All mice were kept under conditions of a 12-h light/12-h dark cycle with unlimited access to food and water. All mice were culled by cervical dislocation prior to dissection. Cortical tissue was procured in the following fashion. For adult samples, after removal of the meninges, coronal cuts were used to excise the olfactory bulb from the rostral region, and the superior colliculus from the caudal region. A sagittal cut to separate the two cortical hemispheres was performed. The cortical pallium was dissected from the subpallial striatum and the septum. The neocortex was then dissected away from the cingulate cortex and the entorhinal cortex. For embryonic samples, the cortical tissue was dissected free from the underlying ganglionic eminences at the pallial-subpallial border. An orthogonal cut was made to remove the presumptive striatum and the overlying piriform cortex. On the medial aspect, the medial limbic cortex was included for analysis, but the adjacent hippocampal primordium, including the cortical hem, was excluded. For the E15.5 cerebral cortex, the resulting hemispheres containing cortical tissue only were placed on the bottom of the Petri dish and, using a fine scalpel, divided into four equal quadrants per hemisphere, namely RL, RM, CL and CM. Rostral and caudal quadrants from both hemispheres were pooled for SAGE library construction but separately tracked for RT-qPCR analysis. Procurement of other adult brain tissues and related mouse organs for *Sox4* and *Sox11* antisense transcript screening was carried out according to the standard mouse necropsy protocol accessible at the National Institute of Allergy and Infectious Diseases (NIAID) website [110].

SAGE libraries and analysis of tags

Ten SAGE libraries were constructed from the cerebral cortex of E15.5, E17.5 and 4- to 6-month-old (Ad) mice according to either one of the two methods described previously [25,26], using I-SAGE™ or I-SAGE™ Long Kits (Invitrogen, Mulgrave, Victoria, Australia). Additional libraries from E15.5 and P1.5 of the cerebral cortex described previously [27] were also included in the analysis. These libraries contain a total of 26,436 traces. SAGE tags were preprocessed - that is, TAGs were extracted and corrected for sequencing errors, and artifacts like SAGE linkers, ribosomal RNA and duplicated ditags were removed using the 'sagenhaft' package, which is available from the Bioconductor website [111,112]. To compare libraries that contain long tags with those that contain short tags, all short tags were mapped to the existing long tags from the other libraries. A table for all libraries containing the unique long or short tags was generated and redundant tags were removed. Only tags with a total count >2 (across all libraries) were considered for subsequent comparisons. Each unique tag was mapped to the mouse genome using ESTgraph, which employs ESTs and their genomic position information. ESTgraph was created by Tim Beissbarth (unpublished) [113]. Identity was assigned to these tags and they were further grouped into the following categories: matching to a gene, a genomic sequence, or an EST, or ambiguous matches or no alignment at all. All annotations were based on the latest mouse assembly (mm9 released in July 2007) accessible from the UCSC Genome Bioinformatics website [63].

Identification of differentially expressed tags

Library comparisons were performed using two methods. Fisher's exact test was used to compare two individual SAGE libraries. In the analysis, multiple testing correction [114] was carried out to control for false-discovery rate and adjusted *P*-value cutoffs (Q-values) were used to select DETs. In cases where several libraries were combined to focus on a specific biological comparison (for example, different stages of development), a Bayesian model, as described previously [115], was used to integrate multiple libraries in pairwise comparisons involving biological replicates of libraries. The model accounts for within-class variability by means of mixture distributions. The resulting E-values were used to select DETs. A table of all relevant comparisons, the comparison method and Q- or E-value cutoffs is provided in Table S1 in Additional data file 1.

Hierarchical clustering of SAGE tags

To identify co-regulated genes, the clustering of DETs was performed based on the log2 of normalized counts. Each library was normalized to 100,000 tags per library to account for size differences. A pseudocount of 0.5 was added before taking the log2 of the normalized tag counts. The tag-wise mean was subtracted from the log2 tag intensities before computing the Euclidean distance of the individual tag profiles. Hierarchical clustering was performed on the tags using

the 'hclust' function and complete linkage, which was implemented using the statistical computing environment of R [116].

Genomic clustering of SAGE tags

To assess whether there was any genomic clustering of tags, a method previously described [117] was adopted. In brief, first gene lists (based on all DETs in both pairwise and multiple library comparisons as well as gene lists from the hierarchical clustering analysis) were selected. The genomic clustering of either of these selections was compared to the total unique tag list (all 25,165 unique tags). The tags were mapped to the mouse genome. The number of selected tags in ten consecutive tag positions for each window of the chromosome was calculated. One thousand permutations were used to compute the null distribution of maximum tag counts per window. The method was implemented using the statistical computing environment of R [116].

Functional classification and characterization of DETs

Gene Ontology enrichment analysis

The DET lists generated from various comparisons were subjected to systematic functional annotation using the standardized Gene Ontology term analysis tools at the DAVID [30]. Functional clustering was performed using high stringency with a kappa similarity threshold of 0.85 and a minimum term overlap of 3. Classification was carried out using a multiple linkage threshold of 0.5 with both numbers of initial and final group members set to 3. A term was considered statistically significant when the computed *P*-value was < 0.05. All queries were performed in September 2009.

Molecular interactions and pathway analysis

Identification of molecular network interactions and pathway analysis of validated DETs or co-regulated genes was completed using the IPA [118] tools from Ingenuity Systems® (Redwood City, California, USA). Accession numbers for all genes with their corresponding fold changes or normalized counts were imported into the IPA software. No focus genes were set at the beginning of the analysis. To start building networks, the application queries the list of input genes and all other gene objects stored in the Ingenuity knowledge base. Networks with a maximum of 30 genes or proteins were constructed, and scores were computed based on the likelihood of the genes being connected together due to random chance. A score of 2 indicates that there is a 1/100 chance that these genes are connected in a network due to random chance. Therefore, any networks with a score of 2 or above are considered statistically significant (with >99% confidence). The most significant novel networks and their interactions with existing canonical pathways were investigated further.

Quantitative RT-PCR

Total RNA isolation and first strand cDNA synthesis

RT-qPCR was carried out to validate all selected candidate or co-regulated DETs. Biological triplicates from E15.5, E17.5, P1.5 and adult (5 to 6 months old) cerebral cortices were used. Genomic-free total RNA from independent mice was extracted using the RNeasy® Lipid Tissue Midi Kit (Qiagen, Doncaster, Victoria, Australia) according to the manufacturer's protocol (Section G in Additional data file 1). First strand cDNA was synthesized from high quality total RNA [119] using random hexamers (unless specified otherwise) and the SuperScript™ III Reverse Transcriptase Kit (Invitrogen, Mulgrave, Victoria, Australia), according to the manufacturer's protocol, under an RNAse-free environment.

Primer design and RT-qPCR

All primers were designed using ProbeFinder 2.35 at the Assay Design Center for Universal ProbeLibrary Assay provided by Roche Applied Science [120]. All RT-qPCR reactions were prepared in 10 μl volumes in a 384-well plate format consisting of LC480 Master Probe Mix (Roche Diagnostics, Castle Hill, New South Wales, Australia), Universal ProbeLibrary (Roche Diagnostics, Castle Hill, New South Wales, Australia) and forward and reverse primers (GeneWorks, Hindmarsh, South Australia, Australia or Bioneer Corporation, Daedeok-gu, Daejeon, Korea) according to the manufacturers' protocols. RT-qPCRs were performed using the LightCycler® 480 System (Roche Diagnostics, Castle Hill, New South Wales, Australia) according to the manufacturer's protocol (Section G in Additional data file 1). A full list of primers and probes used in this study is included in Additional data file 8.

Relative quantification using a standard curve method

The crossing point (Cp) from each signal was calculated based on the Second Derivative Maximum method [121]. A set of serially diluted cDNAs was used to construct a four-data point standard curve for every PCR system in each run. A total of three reference genes (from *Hprt1*, *Psmb2*, phosphoglycerate kinase 1 gene (*Pgk1*) *or Hmbs*) were used as endogenous controls. An estimated starting amount of each target gene was calculated and intra-samples multiple reference genes normalization was performed (Section G in Additional data file 1). A linear model was fitted to the time course of expression values for each gene. Genes differentially expressed between the various stages of development or regions were selected using empirical Bayesian moderated t-statistics, which borrow information between genes [122]. Standard errors for the mean expression at various developmental stages were obtained from the linear model. For each comparison, *P*-values were adjusted using the Benjamini and Hochberg [114] method to control the false discovery rate. See Section H in Additional data file 1 for the R code used to execute the analysis.

Validation of sense-antisense and multiple overlapping transcripts in genomic clusters

Strand specific RT-PCR

All RNA was prepared as described above. Total RNA from all developmental stages (N = 3 per developmental stage) was

equally pooled prior to cDNA synthesis. Four first strand cDNA synthesis reactions were prepared for each cluster as follows: with a primer complementary to the sense strand only; with a primer complementary to the antisense strand only; with oligo-d $[T]_{15}$ as a positive control; and without any primers as a negative control. In all four reactions, both primers were added in subsequent PCRs (Section G in Additional data file 1). PCR amplifications were carried out using FastStart PCR High Fidelity System (Roche Diagnostics, Castle Hill, New South Wales, Australia) according to the manufacturer's protocol. More than one primer set was used in the sense-antisense strand specific RT-PCR (Additional data file 8).

RACE

First strand cDNA synthesis was carried out using pooled total RNA extracted from three biological replicates of rostral and caudal E15.5 and whole E15.5, E17.5, P1.5 and adult (5 to 6 months old) cerebral cortices. Oligo-d $[T]_{15}$ with an adaptor sequence (5'-TACGACGTCTGCTAGGACTG-3') was used to prime the first strand cDNA synthesis. Second strand synthesis or PCR was then carried out using a strand-specific primer and the adaptor primers (Additional data file 8). All specific primers were designed to be complementary to the SAGE tags or their upstream sequences. PCR amplifications (Section G in Additional data file 1) were carried out using FastStart PCR High Fidelity System (Roche Diagnostics, Castle Hill, New South Wales, Australia) according to the manufacturer's protocol.

Southern blotting analysis

Amplified 3' RACE products were transferred to Hybond N^{+TM} (GE Healthcare, Rydalmere, New South Wales, Australia) nylon membrane using the neutral transfer method. Prehybridization and hybridization steps were performed in Rapid-Hyb buffer (GE Healthcare, Rydalmere, New South Wales, Australia) according to the manufacturer's protocol. All oligonucleotides were designed to be complementary to sequence between the specific primer-priming site and the tag of interest. Synthetic oligonucleotides were 5' end-labeled using T4 polynucleotide kinase (Promega, Alexandria, New South Wales, Australia) and [γ-^{32}P]ATP (GE Healthcare, Rydalmere, New South Wales, Australia) with modifications to the manufacturer's protocol. After the hybridization step, the membrane was washed with 5× sodium chloride sodium citrate solution (with 0.1% v/v sodium dodecyl sulphate (SDS)) and 1× sodium chloride sodium citrate solution (with 0.1% w/v SDS) (Section G in Additional data file 1). See Additional data file 8 for primer sequences and oligonucleotides used for detection.

Northern blotting analysis

Independent preparations of total RNA from the cerebral cortex of seven mice at E15.5 and E17.5, and three adult mice were equally pooled to a final concentration of 20 μg per developmental stage. These pooled total RNAs were electrophoresed overnight and capillary transferred onto Hybond N^{+TM} (GE Healthcare, Rydalmere, New South Wales, Australia) nylon membrane. Double-stranded DNA probes were radioactive-labeled using the Amersham Megaprime DNA Labeling System (GE Healthcare, Rydalmere, New South Wales, Australia) and [β-^{32}P]CTP, according to the manufacturer's protocol. Hybridization was carried out overnight (approximately 18 h) at 65°C in aqueous buffer (7% w/v SDS with 0.5 M phosphate). After hybridization, blots were washed using 1% w/v SDS at 65°C for 5 to 6 times until the background signal was low.

In situ *RNA hybridization*

ISH was carried out using paraffin sections (5 μm) of embryonic, postnatal and adult brains (E11.5, E13.5, E15.5, E17.5, P1.5 and P150) and a related [^{35}S]UTP-labeled complementary RNA probe (Additional data file 8) as described previously (Section G in Additional data file 1) [123].

Screening of *Sox4* and *Sox11* sense and antisense transcript expression in the adult mouse brain, organs and both the proliferating and differentiating P19 cells and neurospheres

Strand-specific RT-qPCR

Total RNA was extracted from harvested organs using the TRIzol®'s reagent (Invitrogen, Mulgrave, Victoria, Australia) according to the manufacturer's protocol. To avoid genomic DNA contamination, all isolated total RNA was treated with the recombinant DNAse I enzyme provided by the DNA-*free*™ kit (Applied Biosystems, Scoresby, Victoria, Australia) according to the manufacturer's protocol. First strand cDNA synthesis was carried out using strand-specific primers followed by qPCR analysis as described above.

Embryonic neural stem cells grown as neurospheres

Mouse used for generation of neurospheres had a mixed genetic background including MF1, 129SvEv, C57BL/6 and CBA. Cerebral cortices from E14 embryos were dissected out into calcium-magnesium-free phosphate-buffered saline. The tissue was mechanically dissociated and centrifuged. The cells were plated in complete neuroculture medium (Section G in Additional data file 1) for 4 days followed by induction of neuronal differentiation. These cells were then plated on poly-D-lysine (catalogue number P6407, Sigma Aldrich, Castle Hill, New South Wales, Australia) and laminin (catalogue number 23017-015, Invitrogen, Mulgrave, Victoria, Australia) coated culture dishes in neuroculture medium with the presence of 2% (v/v) fetal bovine serum but not epidermal growth factor and basic fibroblast growth factor. The differentiation was allowed to proceed for 5 days. Total RNA was extracted from both proliferating and differentiating cells using TRIzol® reagent as described above.

P19 embryonal carcinoma cells

P19 mouse embryonal carcinoma cells were cultured and differentiated into neurons as described previously [124].

Briefly, P19 cell cultures were maintained in P19GM complete medium (Section G in Additional data file 1). For induction of neuronal differentiation, 1×10^6 P19 cells were cultured in suspension form using bacteriological Petri dishes. The P19GM medium with additional supplementation of 5×10^{-7}M all-trans retinoic acid (catalogue number R-2625; Sigma Aldrich, Castle Hill, New South Wales, Australia) was used for the induction. After 4 days, P19 cells formed embryoid body stages. Embryoid bodies were collected from suspension cultures and re-plated in adherent culture flasks in the P19GM medium with only 5% (v/v) fetal bovine serum and without retinoic acid supplementation. The cells were allowed to differentiate for 5 days. Total RNA was extracted from both proliferating and differentiating cells using TRIzol® reagent as described in above.

Abbreviations

CL: caudo-lateral; CM: caudo-medial; DAVID: Database for Annotation, Visualization and Integrated Discovery; DET: differentially expressed transcript/tag; E: embryonic day; EST: expressed sequence tag; GEO: Gene Expression Omnibus; IPA: Ingenuity Pathway Analysis; ISH: *in situ* hybridization; LTP: long term potentiation; miRNA: microRNA; NAT: natural antisense transcript; NSPC: neural stem/progenitor cell; OMIM: Online Mendelian Inheritance in Man; P: postnatal day; PET: paired-end diTag; RACE: rapid amplification of cDNA ends; RL: rostro-lateral; RM: rostro-medial; RT-qPCR: quantitative RT-PCR; SAGE: serial analysis of gene expression; SDS: sodium dodecyl sulphate; UCSC: University of California Santa Cruz; UTR: untranslated region.

Authors' contributions

KHL performed all the SAGE validation experiments. CAH, PZC and SST procured the mouse cerebral cortex and constructed the SAGE libraries. KHL, TB, LH and GKS designed, performed and supervised the SAGE, RT-qPCR and IPA analyses. KHL and TT performed all the ISH studies. KHL, KB, PSC, CNH and PQT carried out the expression studies on *Sox4* and *Sox11* transcripts. KHL, CAH and CNH drafted the manuscript. CAH, GKS, TT and HSS conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Additional data files

The following additional data are available with the online version of this paper: analysis of SAGE, DETs, IPA, *Sox4* and *Sox11* genomic cluster analysis, and R script for implementing empirical Bayesian moderated *t*-test on multiple groups (Additional data file 1); SAGE tag information for 561 DETs (Additional data file 2); functional annotations clustering analysis using DAVID (Additional data file 3); RT-qPCR validation of DETs based on multiple comparisons between two developmental stages (E versus Ad, PN1.5 versus Ad and E15.5 versus PN1.5) (Additional data file 4); RT-qPCR validation of gene clusters based on hierarchical clustering analysis (Additional data file 5); RT-qPCR validation of DETs based on the rostral versus caudal E15.5 cerebral cortex comparison (Additional data file 6); statistically significant over-represented genomic loci based on genomic clustering of tags (Additional data file 7); list of primers, probes, clones and assays designed for RT-qPCR, RACE, Southern, Northern and ISH analysis (Additional data file 8).

Acknowledgements

We thank Felix Ng, Sarah King and Caitlin Collins for technical assistance; Frank Weissenborn for assistance with tissue dissections; Ken Simpson and Keith Satterley for computational assistance. The P19 cell line was kindly provided by Dr K Jensen, School of Molecular and Biomedical Science, University of Adelaide, South Australia, Australia. This work was supported by a National Health and Medical Research Council fellowships (171601 and 461204 to HSS); National Health and Medical Research Council Grants 219176, 257501, 215201, and 257529 (to HSS, GKS, and S-ST), a fellowship from the Deutsche Forschungsgemeinschaft (to TB). PQT is a Pfizer Australia Research Fellow. K-HL was a recipient of the Melbourne International Fee Remission Scholarship (MIFRS) and Universiti Putra Malaysia Staff Training Scholarship (UPMSTS), and Adelaide International Fee Scholarship (AIFS) equivalent. The majority of DNA sequencing in production of SAGE libraries was performed by the Australian Genome Research Facility, which was established through the Commonwealth-funded Major National Research Facilities program.

References

1. Miyashita-Lin EM, Hevner R, Wassarman KM, Martinez S, Rubenstein JL: **Early neocortical regionalization in the absence of thalamic innervation.** *Science* 1999, **285:**906-909.
2. Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD: **Neuronal subtype specification in the cerebral cortex.** *Nat Rev Neurosci* 2007, **8:**427-437.
3. O'Leary DD, Chou SJ, Sahara S: **Area patterning of the mammalian cortex.** *Neuron* 2007, **56:**252-269.
4. Gupta A, Tsai LH, Wynshaw-Boris A: **Life is a journey: a genetic look at neocortical development.** *Nat Rev Genet* 2002, **3:**342-355.
5. Marin O, Rubenstein JL: **A long, remarkable journey: tangential migration in the telencephalon.** *Nat Rev Neurosci* 2001, **2:**780-790.
6. Nadarajah B, Parnavelas JG: **Modes of neuronal migration in the developing cerebral cortex.** *Nat Rev Neurosci* 2002, **3:**423-432.
7. Smart IH: **Histogenesis of the mesocortical area of the mouse telencephalon.** *J Anat* 1984, **138:**537-552.
8. Uylings HBM, Van Eden CG, Parnavelas JG, Kalsbeek A: **The prenatal and postnatal development of rat cerebral cortex.** In *The Cerebral Cortex of the Rat* Edited by: Kolb B, Tees RC. Cambridge, MA: MIT Press; 1990:36-76.
9. Bishop KM, Goudreau G, O'Leary DD: **Regulation of area identity in the mammalian neocortex by Emx2 and Pax6.** *Science* 2000, **288:**344-349.
10. Mallamaci A, Muzio L, Chan CH, Parnavelas J, Boncinelli E: **Area identity shifts in the early cerebral cortex of Emx2-/- mutant mice.** *Nat Neurosci* 2000, **3:**679-686.
11. Armentano M, Chou SJ, Tomassy GS, Leingartner A, O'Leary DD, Studer M: **COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas.** *Nat Neurosci* 2007, **10:**1277-1286.
12. Sahara S, Kawakami Y, Izpisua Belmonte JC, O'Leary DD: **Sp8 exhibits reciprocal induction with Fgf8 but has an opposing effect on anterior-posterior cortical area patterning.** *Neural Dev* 2007, **2:**10.
13. Zembrzycki A, Griesel G, Stoykova A, Mansouri A: **Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain.** *Neural Dev* 2007, **2:**8.
14. Bulfone A, Puelles L, Porteus MH, Frohman MA, Martin GR, Rubenstein JL: **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* 1993, **13:**3155-3172.
15. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms.** *Nat Rev Genet* 2006, **7:**703-713.
16. Kadonaga JT: **Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines.** *Cell* 1998, **92:**307-313.
17. Ladd-Acosta C, Pevsner J, Sabunciyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP: **DNA methylation signatures within the human brain.** *Am J Hum Genet* 2007, **81:**1304-1315.
18. Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI: **Paraspeckles: a novel nuclear domain.** *Curr Biol* 2002, **12:**13-25.
19. Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL: **Regulating gene expression through RNA nuclear retention.** *Cell* 2005, **123:**249-263.
20. Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME: **A brain-specific microRNA regulates dendritic spine development.** *Nature* 2006, **439:**283-289.
21. Stark KL, Xu B, Bagchi A, Lai WS, Liu H, Hsu R, Wan X, Pavlidis P, Mills AA, Karayiorgou M, Gogos JA: **Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model.** *Nat Genet* 2008, **40:**751-760.
22. Bak M, Silahtaroglu A, Moller M, Christensen M, Rath MF, Skryabin B, Tommerup N, Kauppinen S: **MicroRNA expression in the adult mouse central nervous system.** *Rna* 2008, **14:**432-444.
23. Tochitani S, Hayashizaki Y: **Nkx2.2 antisense RNA overexpression enhanced oligodendrocytic differentiation.** *Biochem Biophys Res Commun* 2008, **372:**691-696.
24. Smalheiser NR, Lugli G, Torvik VI, Mise N, Ikeda R, Abe K: **Natural antisense transcripts are co-expressed with sense mRNAs in synaptoneurosomes of adult mouse forebrain.** *Neurosci Res* 2008, **62:**236-239.
25. Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE: **Using the transcriptome to annotate the genome.** *Nat Biotechnol* 2002, **20:**508-512.
26. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.
27. Gunnersen JM, Augustine C, Spirkoska V, Kim M, Brown M, Tan SS: **Global analysis of gene expression patterns in developing mouse neocortex using serial analysis of gene expression.** *Mol Cell Neurosci* 2002, **19:**560-573.
28. **NCBI Gene Expression Omnibus.** [http://www.ncbi.nlm.nih.gov/geo/]
29. Boon WM, Beissbarth T, Hyde L, Smyth G, Gunnersen J, Denton DA, Scott H, Tan SS: **A comparative analysis of transcribed genes in the mouse hypothalamus and neocortex reveals chromosomal clustering.** *Proc Natl Acad Sci USA* 2004, **101:**14972-14977.
30. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4:**P3.
31. Yuan X, Chittajallu R, Belachew S, Anderson S, McBain CJ, Gallo V: **Expression of the green fluorescent protein in the oligodendrocyte lineage: a transgenic mouse for developmental and physiological studies.** *J Neurosci Res* 2002, **70:**529-545.
32. **Allen Institute for Brain Science.** [http://mouse.brain-map.org/]
33. **Brain Gene Expression Map.** [http://www.stjudebgem.org/]
34. **Gene Expression Nervous System Atlas.** [http://www.gensat.org/]
35. **GenePaint.org.** [http://www.genepaint.org/]
36. Ino H, Chiba T: **Expression of proliferating cell nuclear antigen (PCNA) in the adult and developing mouse nervous system.** *Brain Res Mol Brain Res* 2000, **78:**163-174.
37. Lopez-Coviella I, Follettie MT, Mellott TJ, Kovacheva VP, Slack BE, Diesl V, Berse B, Thies RS, Blusztajn JK: **Bone morphogenetic protein 9 induces the transcriptome of basal forebrain cholinergic neurons.** *Proc Natl Acad Sci USA* 2005, **102:**6984-6989.
38. Bhat NR, Zhang P, Bhat AN: **The expression of myristoylated alanine-rich C-kinase substrate in oligodendrocytes is developmentally regulated.** *Dev Neurosci* 1995, **17:**256-263.
39. Calabrese B, Halpain S: **Essential role for the PKC target MARCKS in maintaining dendritic spine morphology.** *Neuron* 2005, **48:**77-90.
40. Lopez-Bendito G, Flames N, Ma L, Fouquet C, Di Meglio T, Chedotal A, Tessier-Lavigne M, Marin O: **Robo1 and Robo2 cooperate to control the guidance of major axonal tracts in the mammalian forebrain.** *J Neurosci* 2007, **27:**3395-3407.
41. Sher F, Rossler R, Brouwer N, Balasubramaniyan V, Boddeke E, Copray S: **Differentiation of neural stem cells into oligodendrocytes: involvement of the polycomb group protein Ezh2.** *Stem Cells* 2008, **26:**2875-2883.
42. Mermelstein F, Yeung K, Cao J, Inostroza JA, Erdjument-Bromage H, Eagelson K, Landsman D, Levitt P, Tempst P, Reinberg D: **Requirement of a corepressor for Dr1-mediated repression of transcription.** *Genes Dev* 1996, **10:**1033-1048.
43. Beiswanger CM, Diegmann MH, Novak RF, Philbert MA, Graessle TL, Reuhl KR, Lowndes HE: **Developmental changes in the cellular distribution of glutathione and glutathione S-transferases in the murine nervous system.** *Neurotoxicology* 1995, **16:**425-440.
44. Saha S, Datta K, Rangarajan P: **Characterization of mouse neuronal Ca2+/calmodulin kinase II inhibitor alpha.** *Brain Res* 2007, **1148:**38-42.
45. Wilson CA, Murphy DD, Giasson BI, Zhang B, Trojanowski JQ, Lee VM: **Degradative organelles containing mislocalized alpha- and beta-synuclein proliferate in presenilin-1 null neurons.** *J Cell Biol* 2004, **165:**335-346.
46. Bachoo RM, Kim RS, Ligon KL, Maher EA, Brennan C, Billings N, Chan S, Li C, Rowitch DH, Wong WH, DePinho RA: **Molecular diversity of astrocytes with implications for neurological disorders.** *Proc Natl Acad Sci USA* 2004, **101:**8384-8389.
47. Lencesova L, O'Neill A, Resneck WG, Bloch RJ, Blaustein MP: **Plasma membrane-cytoskeleton-endoplasmic reticulum complexes in neurons and astrocytes.** *J Biol Chem* 2004, **279:**2885-2893.
48. Zeng H, Chattarji S, Barbarosie M, Rondi-Reig L, Philpot BD, Miyakawa T, Bear MF, Tonegawa S: **Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory.** *Cell* 2001, **107:**617-629.
49. Kim PM, Aizawa H, Kim PS, Huang AS, Wickramasinghe SR, Kashani AH, Barrow RK, Huganir RL, Ghosh A, Snyder SH: **Serine racemase: activation by glutamate neurotransmission via glutamate receptor interacting protein and mediation of neuronal migration.** *Proc Natl Acad Sci USA* 2005, **102:**2105-2110.
50. Mandir AS, Poitras MF, Berliner AR, Herring WJ, Guastella DB, Feldman A, Poirier GG, Wang ZQ, Dawson TM, Dawson VL: **NMDA but not non-NMDA excitotoxicity is mediated by Poly(ADP-ribose) polymerase.** *J Neurosci* 2000, **20:**8005-8011.
51. Yang X, Skoff RP: **Proteolipid protein regulates the survival and differentiation of oligodendrocytes.** *J Neurosci* 1997, **17:**2056-2070.
52. Southwood CM, Peppi M, Dryden S, Tainsky MA, Gow A: **Microtubule deacetylases, SirT2 and HDAC6, in the nervous system.** *Neurochem Res* 2007, **32:**187-195.
53. Niciu MJ, Ma XM, El Meskini R, Ronnett GV, Mains RE, Eipper BA: **Developmental changes in the expression of ATP7A during a critical period in postnatal neurodevelopment.** *Neuroscience* 2006, **139:**947-964.
54. Sanosaka T, Namihira M, Asano H, Kohyama J, Aisaki K, Igarashi K, Kanno J, Nakashima K: **Identification of genes that restrict astrocyte differentiation of midgestational neural precursor cells.** *Neuroscience* 2008, **155:**780-788.
55. Mbebi C, See V, Mercken L, Pradier L, Muller U, Loeffler JP: **Amyloid precursor protein family-induced neuronal death is mediated by impairment of the neuroprotective calcium/calmodulin protein kinase IV-dependent signaling pathway.** *J Biol Chem* 2002, **277:**20979-20990.
56. Jinnah HA, Hess EJ, Wilson MC, Gage FH, Friedmann T: **Localization of hypoxanthine-guanine phosphoribosyltransferase messenger RNA in the mouse brain by *in situ* hybridization.** *Mol Cell Neurosci* 1992, **3:**64-78.
57. Poet M, Kornak U, Schweizer M, Zdebik AA, Scheel O, Hoelter S, Wurst W, Schmitt A, Fuhrmann JC, Planells-Cases R, Mole SE, Hubner CA, Jentsch TJ: **Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6.** *Proc Natl Acad Sci USA* 2006, **103:**13854-13859.
58. Palfi A, Kortvely E, Fekete E, Kovacs B, Varszegi S, Gulya K: **Differential calmodulin gene expression in the rodent brain.** *Life Sci* 2002, **70:**2829-2855.
59. Toyo-oka K, Shionoya A, Gambello MJ, Cardoso C, Leventer R, Ward HL, Ayala R, Tsai LH, Dobyns W, Ledbetter D, Hirotsune S, Wynshaw-Boris A: **14-3-3epsilon is important for neuronal migration by binding to NUDEL: a molecular explanation for Miller-Dieker syndrome.** *Nat Genet* 2003, **34:**274-285.
60. Kuo TY, Hsueh YP: **Expression of zinc finger transcription fac-**

tor Bcl11A/Evi9/CTIP1 in rat brain. *J Neurosci Res* 2007, **85:**1628-1636.
61. Li Q, Lee JA, Black DL: **Neuronal regulation of alternative pre-mRNA splicing.** *Nat Rev Neurosci* 2007, **8:**819-831.
62. Zhou C, Tsai SY, Tsai MJ: **COUP-TFI: an intrinsic factor for early regionalization of the neocortex.** *Genes Dev* 2001, **15:**2054-2059.
63. **UCSC Genome Bioinformatics.** [http://genome.ucsc.edu/]
64. **Ensembl Genome Browser.** [http://www.ensembl.org/]
65. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J: **The establishment of neuronal properties is controlled by Sox4 and Sox11.** *Genes Dev* 2006, **20:**3475-3486.
66. Potzner MR, Griffel C, Lutjen-Drecoll E, Bosl MR, Wegner M, Sock E: **Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system.** *Mol Cell Biol* 2007, **27:**5316-5326.
67. Velculescu VE, Madden SL, Zhang L, Lash AE, Yu J, Rago C, Lal A, Wang CJ, Beaudry GA, Ciriello KM, Cook BP, Dufault MR, Ferguson AT, Gao Y, He TC, Hermeking H, Hiraldo SK, Hwang PM, Lopez MA, Luderer HF, Mathews B, Petroziello JM, Polyak K, Zawel L, Kinzler KW, *et al.*: **Analysis of human transcriptomes.** *Nat Genet* 1999, **23:**387-388.
68. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* 2000, **18:**630-634.
69. **The Cancer Genome Anatomy Project: SAGE Genie.** [http://cgap.nci.nih.gov/SAGE]
70. Ng P, Tan JJ, Ooi HS, Lee YL, Chiu KP, Fullwood MJ, Srinivasan KG, Perbost C, Du L, Sung WK, Wei CL, Ruan Y: **Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes.** *Nucleic Acids Res* 2006, **34:**e84.
71. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y: **Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.** *Proc Natl Acad Sci USA* 2003, **100:**15776-15781.
72. Wei CL, Ng P, Chiu KP, Wong CH, Ang CC, Lipovich L, Liu ET, Ruan Y: **5' Long serial analysis of gene expression (LongSAGE) and 3' LongSAGE for transcriptome characterization and genome annotation.** *Proc Natl Acad Sci USA* 2004, **101:**11701-11706.
73. Chenn A: **Wnt/beta-catenin signaling in cerebral cortical development.** *Organogenesis* 2008, **4:**76-80.
74. Bafico A, Gazit A, Pramila T, Finch PW, Yaniv A, Aaronson SA: **Interaction of frizzled related protein (FRP) with Wnt ligands and the frizzled receptor suggests alternative mechanisms for FRP inhibition of Wnt signaling.** *J Biol Chem* 1999, **274:**16180-16187.
75. Morkel M, Huelsken J, Wakamiya M, Ding J, Wetering M van de, Clevers H, Taketo MM, Behringer RR, Shen MM, Birchmeier W: **Beta-catenin regulates Cripto- and Wnt3-dependent gene expression programs in mouse axis and mesoderm formation.** *Development* 2003, **130:**6283-6294.
76. Fevr T, Robine S, Louvard D, Huelsken J: **Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells.** *Mol Cell Biol* 2007, **27:**7551-7559.
77. Machon O, Backman M, Machonova O, Kozmik Z, Vacik T, Andersen L, Krauss S: **A dynamic gradient of Wnt signaling controls initiation of neurogenesis in the mammalian cortex and cellular specification in the hippocampus.** *Dev Biol* 2007, **311:**223-237.
78. Woodhead GJ, Mutch CA, Olson EC, Chenn A: **Cell-autonomous beta-catenin signaling regulates cortical precursor proliferation.** *J Neurosci* 2006, **26:**12620-12630.
79. Lisman J, Schulman H, Cline H: **The molecular basis of CaMKII function in synaptic and behavioural memory.** *Nat Rev Neurosci* 2002, **3:**175-190.
80. Pak JH, Huang FL, Li J, Balschun D, Reymann KG, Chiang C, Westphal H, Huang KP: **Involvement of neurogranin in the modulation of calcium/calmodulin-dependent protein kinase II, synaptic plasticity, and spatial learning: a study with knockout mice.** *Proc Natl Acad Sci USA* 2000, **97:**11232-11237.
81. Lepicard EM, Mizuno K, Antunes-Martins A, von Hertzen LS, Giese KP: **An endogenous inhibitor of calcium/calmodulin-dependent kinase II is up-regulated during consolidation of fear memory.** *Eur J Neurosci* 2006, **23:**3063-3070.
82. Young-Pearse TL, Bai J, Chang R, Zheng JB, LoTurco JJ, Selkoe DJ: **A critical function for beta-amyloid precursor protein in neuronal migration revealed by *in utero* RNA interference.** *J Neurosci* 2007, **27:**14459-14469.
83. Alonso A, Zaidi T, Novak M, Grundke-Iqbal I, Iqbal K: **Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments.** *Proc Natl Acad Sci USA* 2001, **98:**6923-6928.
84. Rapoport M, Dawson HN, Binder LI, Vitek MP, Ferreira A: **Tau is essential to beta-amyloid-induced neurotoxicity.** *Proc Natl Acad Sci USA* 2002, **99:**6364-6369.
85. Carter CJ: **eIF2B and oligodendrocyte survival: where nature and nurture meet in bipolar disorder and schizophrenia?** *Schizophr Bull* 2007, **33:**1343-1353.
86. Craddock N, O'Donovan MC, Owen MJ: **The genetics of schizophrenia and bipolar disorder: dissecting psychosis.** *J Med Genet* 2005, **42:**193-204.
87. Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D, Hollingsworth ZR, Collin F, Synek B, Holmans PA, Young AB, Wexler NS, Delorenzi M, Kooperberg C, Augood SJ, Faull RL, Olson JM, Jones L, Luthi-Carter R: **Regional and cellular gene expression changes in human Huntington's disease brain.** *Hum Mol Genet* 2006, **15:**965-977.
88. Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, Frey AS, Spektor BS, Penney EB, Schilling G, Ross CA, Borchelt DR, Tapscott SJ, Young AB, Cha JH, Olson JM: **Decreased expression of striatal signaling genes in a mouse model of Huntington's disease.** *Hum Mol Genet* 2000, **9:**1259-1271.
89. Lovestone S, Hodgson S, Sham P, Differ AM, Levy R: **Familial psychiatric presentation of Huntington's disease.** *J Med Genet* 1996, **33:**128-131.
90. Shiwach R: **Psychopathology in Huntington's disease patients.** *Acta Psychiatr Scand* 1994, **90:**241-246.
91. Chandra S, Gallardo G, Fernandez-Chacon R, Schluter OM, Sudhof TC: **Alpha-synuclein cooperates with CSPalpha in preventing neurodegeneration.** *Cell* 2005, **123:**383-396.
92. Evert BO, Wullner U, Klockgether T: **Cell death in polyglutamine diseases.** *Cell Tissue Res* 2000, **301:**189-204.
93. Morishima Y, Gotoh Y, Zieg J, Barrett T, Takano H, Flavell R, Davis RJ, Shirasaki Y, Greenberg ME: **Beta-amyloid induces neuronal apoptosis via a mechanism that involves the c-Jun N-terminal kinase pathway and the induction of Fas ligand.** *J Neurosci* 2001, **21:**7551-7560.
94. Mallamaci A, Stoykova A: **Gene networks controlling early cerebral cortex arealization.** *Eur J Neurosci* 2006, **23:**847-856.
95. Philp D, Huff T, Gho YS, Hannappel E, Kleinman HK: **The actin binding site on thymosin beta4 promotes angiogenesis.** *Faseb J* 2003, **17:**2103-2105.
96. Bock-Marquette I, Saxena A, White MD, Dimaio JM, Srivastava D: **Thymosin beta4 activates integrin-linked kinase and promotes cardiac cell migration, survival and cardiac repair.** *Nature* 2004, **432:**466-472.
97. Roth LW, Bormann P, Bonnet A, Reinhard E: **beta-thymosin is required for axonal tract formation in developing zebrafish brain.** *Development* 1999, **126:**1365-1374.
98. Lopez-Bendito G, Molnar Z: **Thalamocortical development: how are we going to get there?** *Nat Rev Neurosci* 2003, **4:**276-289.
99. Winter J, Kunath M, Roepcke S, Krause S, Schneider R, Schweiger S: **Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1.** *BMC Mol Biol* 2007, **8:**105.
100. Abdel Wahab N, Gibbs J, Mason RM: **Regulation of gene expression by alternative polyadenylation and mRNA instability in hyperglycaemic mesangial cells.** *Biochem J* 1998, **336:**405-411.
101. Wilusz CJ, Wilusz J: **Bringing the role of mRNA decay in the control of gene expression into focus.** *Trends Genet* 2004, **20:**491-497.
102. Lu C, Jeong D-H, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, Liu B, Cao X, Meyers BC, Green PJ: **Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs).** *Proc Natl Acad Sci USA* 2008, **105:**4951-4956.
103. **miRBase.** [http://microrna.sanger.ac.uk/]

104. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.** *Nature* 2008, **453:**539-543.
105. Hoser M, Potzner MR, Koch JM, Bosl MR, Wegner M, Sock E: **Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors.** *Mol Cell Biol* 2008, **28:**4675-4687.
106. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ: **Roles of Sox4 in central nervous system development.** *Brain Res Mol Brain Res* 2000, **79:**180-191.
107. Jin Z, Liu L, Bian W, Chen Y, Xu G, Cheng L, Jing N: **Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development.** *J Biol Chem* 2009, **284:**8160-8173.
108. Lee MS, Jun DH, Hwang CI, Park SS, Kang JJ, Park HS, Kim J, Kim JH, Seo JS, Park WY: **Selection of neural differentiation-specific genes by comparing profiles of random differentiation.** *Stem Cells* 2006, **24:**1946-1955.
109. Dy P, Penzo-Mendez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V: **The three SoxC proteins - Sox4, Sox11 and Sox12 - exhibit overlapping expression patterns and molecular properties.** *Nucleic Acids Res* 2008, **36:**3101-3117.
110. **Virtual Mouse Necropsy.** [http://www3.niaid.nih.gov/labs/aboutlabs/cmb/InfectiousDiseasePathogenesisSection/mouseNecropsy/]
111. Beissbarth T, Hyde L, Smyth GK, Job C, Boon WM, Tan SS, Scott HS, Speed TP: **Statistical modeling of sequencing errors in SAGE libraries.** *Bioinformatics* 2004, **20(Suppl 1):**I31-I39.
112. **Bioconductor.** [http://www.bioconductor.org/]
113. **Mapping SAGE Tags to Genes Using EST and Genomic Information.** [http://www.mbgproject.org/tagmapping/]
114. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc* 1995, **57:**289-300.
115. Vencio RZ, Brentani H, Patrao DF, Pereira CA: **Bayesian model accounting for within-class biological variability in serial analysis of gene expression (SAGE).** *BMC Bioinformatics* 2004, **5:**119.
116. **The R Project for Statistical Computing.** [http://www.r-project.org/]
117. Gotter J, Brors B, Hergenhahn M, Kyewski B: **Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters.** *J Exp Med* 2004, **199:**155-166.
118. **Ingenuity Pathway Analysis.** [http://www.ingenuity.com]
119. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7:**3.
120. **Assay Design Center for Universal ProbeLibrary Assay.** [http://www.roche-applied-science.com/sis/rtpcr/upl/ezhome.html]
121. Luu-The V, Paquet N, Calvo E, Cumps J: **Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction.** *Biotechniques* 2005, **38:**287-293.
122. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3:**. Article3
123. Thomas T, Voss AK, Chowdhury K, Gruss P: **Querkopf, a MYST family histone acetyltransferase, is required for normal cerebral cortex development.** *Development* 2000, **127:**2537-2548.
124. Rudnicki MA, McBurney MW: **Cell culture methods and induction of differentiation of embryonal carcinoma cell lines.** In *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* Edited by: Robertson EJ. Oxford: IRL Press; 1987:19-49.

CHAPTER 4

Spatiotemporal regulation of multiple overlapping sense and novel natural antisense transcripts at the Nrgn and Camk2n1 gene loci during mouse cerebral corticogenesis

King-Hwa Ling, Chelsee A Hewitt, Tim Beissbarth, Lavinia Hyde, Pike-See Cheah, Gordon K Smyth, Seong-Seng Tan, Christopher N Hahn, Tim Thomas, Paul Q Thomas and Hamish S Scott.

Published in: *Cerebral Cortex*, published online on August 6, 2010:
doi: 10.1093/cercor/bhq141

4.1 Summary

After the discovery of the unusual organization of multiple overlapping sense and NATs in 2 embryonic specific genomic clusters at *Sox4* and *Sox11* gene loci, this study continues to discover similar genomic clusters that are significantly expressed in the adult cerebral cortex. This chapter describes the finding of 2 adult brain specific genomic clusters at the neurogranin (*Nrgn*) and calcium/calmodulin-dependent protein kinase II inhibitor 1 (*Camk2n1*) gene loci. The finding was based on the same SAGE datasets for the global transcriptome analysis presented in Chapter 3.

The study describes the identification and validation of various spatiotemporal expression profiles of multiple *Nrgn* and *Camk2n1* sense and novel overlapping NATs in the developing and adult mouse brains. Further characterisation of these sense and NATs showed that they utilize alternative polyadenylation sites during transcription, thus leading to various transcript variants with different 3' UTR lengths. Different expression profiles of transcript variants (both sense and NATs) were observed in the cerebral cortex during development. NATs were also differentially expressed in various adult mouse organs and differentiating neuronal/glial cells suggesting their potential roles in regulating the development or function of these organs/cells.

The study confirmed that the *Nrgn* and *Camk2n1* genomic clusters have similar features as *Sox4* and *Sox11* genomic clusters characterised with multiple sense and overlapping transcript variants with different 3' UTR lengths due to the utilisation of alternative polyadenylation sites. Besides, the study also reported a landmark finding on the ability of sense and NATs from *Nrgn* and *Camk2n1* gene loci to interact with one and another. RNA fluorescent in situ hybridization (FISH) showed that these overlapping NATs form double-stranded RNA aggregates with their sense counterparts in the cytoplasm implying that they could be playing a role in posttranscriptional regulation. Consequently, more effort was focused in the next chapter to elucidate the molecular role of NATs on the transcription and translation of their corresponding sense transcripts.

4.2 Notes

The published article and 3 supplementary information files have been included as Appendix B according to the following order:

1. Authors' declaration.
2. Supplementary information 1: (Original file is accessible at http://cercor.oxfordjournals.org/content/suppl/2010/07/14/bhq141.DC1/cercor-2010-00377-File009.pdf)
3. Supplementary information 2: (Original file is accessible at http://cercor.oxfordjournals.org/content/suppl/2010/07/14/bhq141.DC1/cercor-2010-00377-File010.doc)
4. Supplementary information file 3 contains movie files, which are provided in an electronic form in a compact disc (CD). Original file is accessible at the publisher's website (http://cercor.oxfordjournals.org/content/suppl/2010/07/14/bhq141.DC1/cercor-2010-00377-File011.tar).

4.3 Permission to Reuse Published Materials

The authors grant an exclusive license to Oxford Journals or the society of ownership when the manuscript was accepted to publication. Permission to reproduce the abstract, main text and all the figures were granted on 10th of August 2010, in three separate license numbers; 1) 2485421390870 (for reproduction of the abstract), 2) 2485421026254 (for reproduction of the main text) and 3) 2485421288763 (for reproduction of all the figures within). These licenses are provided as Appendix B in Chapter 8.

4.4 The published article

Please refer to the next page for the following published article:

Ling KH, Hewitt CA, Beissbarth T, Hyde L, Cheah PS, Smyth GK, Tan SS, Hahn CN, Thomas T, Thomas PQ, Scott HS: **Spatiotemporal regulation of multiple overlapping sense and novel natural antisense transcripts at the *Nrgn* and *Camk2n1* gene loci during mouse cerebral corticogenesis.** Cereb Cortex 2010. Published online on August 6, 2010, doi: 10.1093/cercor/bhq141.

Ling, K.H., Hewitt, C.A., Beissbarth, T., Hyde, L., Cheah, P.S., Smyth, G.K., Tan, S.S., Hahn, C.N., Thomas, T., Thomas, P.Q. & Scott, H.S. (2010) Spatiotemporal regulation of multiple overlapping sense and novel natural antisense transcripts at the *Nrgn* and *Camk2n1* gene loci during mouse cerebral corticogenesis.
Cerebral Cortex, v. 21 (3), pp. 683-697

NOTE:
This publication is included on pages 132-146 in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1093/cercor/bhq141

CHAPTER 5

Sense and overlapping natural antisense transcripts form double stranded RNA to produce a novel endogenous small interfering RNA during brain development

King-Hwa Ling, Peter J Brautigan, Sarah Moore, Rachel Fraser, Pike-See Cheah, Joy M Raison, Milena Stankovic, Tasman Daish, Jeffrey Mann, David L Adelson, Paul Q Thomas, Christopher N Hahn and Hamish S Scott.

Submitted to: *Nucleic Acids Research*
(Manuscript ID: NAR-00701-C-2011)

5.1 Summary

Following previous chapters on transcriptomic analysis and validations of all the 4 differentially regulated genomic clusters at *Sox4, Sox11, Nrgn* and *Camk2n1* gene loci, this study aims to functionally characterise the role of NATs during the development of the cerebral cortex. Of the 4 genes, *Sox4* is the most studied and well comprehended and therefore the *Sox4* genomic cluster was selected as a model for further characterisation.

The study began with both 5' and 3' Rapid Amplification of cDNA Ends (RACE) to characterise the NAT organisation within the *Sox4* gene locus. By employing Paired-End Ditags sequences from ENSEMBL (http://www.ensembl.org/) [1], 6 full-length *Sox4* NATs were identified. Of these, 4 NATs were amplified and cloned into an expression vector system. The study led to the discovery of the following findings:

a. Sense and NATs formed dsRNA aggregates in the cytoplasm of trypsinised adult brain cells.
b. NATs did not affect the transcription and translation of *Sox4* sense transcripts *in vitro*.
c. Sense and NATs formed dsRNA templates for the generation of a novel endogenous small interfering RNA (endo-siRNA).
d. The novel endo-siRNA is known as *Sox4_sir3* and is spatiotemporally regulated in the central nervous system especially within germinal cell layers of the developing cerebral cortex and liver.
e. The *Sox4_sir3* may repress *Creb1* translation.

Similarly to *Nrgn* and *Camk2n1*, the study demonstrated interaction between *Sox4* sense and antisense transcripts in the cytoplasm of different cells isolated from the mouse brain. Surprisingly, the interaction between sense-NATs forms dsRNAs that did not affect both transcription and translation processes. Instead, they serve as templates for *Sox4_sir3* biogenesis. This type of phenomenon is common in viral or plant systems but is the first demonstration in the mammalian system adding to our understanding of this long-debated and controversial role of non-coding RNAs.

5.2 Notes

The submitted manuscript and supplementary information files have been included as Appendix C according to the following order:

1. Authors' declaration.
2. Supplementary information 1.
3. Supplementary information 2.
4. Supplementary information 3.

5.3 The submitted manuscript

Please refer to the next page for the following submitted manuscript:

Ling KH, Brautigan PJ, Moore S, Fraser R, Cheah PS, Raison JM, Stankovic M, Daish T, Mattiske DM, Mann JR, Adelson DL, Thomas PQ, Hahn CN and Scott HS: **Sense and overlapping natural antisense transcripts form double stranded RNA to produce a novel endogenous small interfering RNA during brain development.** *Manuscript submitted to Nucleic Acids Research (Manuscript ID: NAR-00701-C-2011).*

Sense and overlapping natural antisense transcripts form double stranded RNA to produce a novel endogenous small interfering RNA during mammalian brain development

King-Hwa Ling[1,2,3], Peter J Brautigan[1], Sarah Moore[1], Rachel Fraser[1], Pike-See Cheah[4,5], Joy M Raison[6], Milena Stankovic[1], Lee YK[1], Tasman Daish[4], Deidre M Mattiske[7], Jeffrey R Mann[7], David L Adelson[4], Paul Q Thomas[4], Christopher N Hahn[1] and Hamish S Scott[1,2,*].

[1] Department of Molecular Pathology, The Institute of Medical and Veterinary Science and The Hanson Institute, P.O. Box 14 Rundle Mall Post Office, Adelaide, SA 5000, Australia.

[2] School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

[3] Department of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia.

[4] School of Molecular and Biomedical Science, Faculty of Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

[5] Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia.

[6] eResearchSA, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

[7] Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, VIC 3052, Australia.

* Corresponding author; Email: hamish.scott@health.sa.gov.au, Telephone: +618 8222 3651, Fax: +618 8222 3146.

Abstract

Long noncoding RNAs (lncRNAs) are now accepted as being functional transcripts involved in cellular development and regulatory processes. To date, only a few lncRNAs have been functionally characterised as having a crucial role in gene regulation mainly via transcriptional interference, chromatin remodelling and protein trafficking within the nucleus. However, the majority of lncRNAs are located in the cytoplasm and their regulatory roles remain elusive. We have previously described the existence of multiple overlapping natural antisense transcripts (NATs) at the *Sox4* gene locus, which are spatiotemporally regulated throughout cerebral corticogenesis. In this study, we characterised *Sox4* NATs and their role in *Sox4* sense transcript regulation. Our study showed *Sox4* sense and NATs formed double stranded RNA aggregates in the cytoplasm. Overexpression of selected *Sox4* NATs did not alter the level of *Sox4* sense expression or protein translation. Instead, overexpression of one of the NATs, *Sox4ot1,* produced a novel endogenous small interfering RNA (endo-siRNA) known as *Sox4_sir3*. *Sox4_sir3* originates from the *Sox4* sense transcript and its biogenesis is Dicer1-dependent. *Sox4_sir3* is expressed in the marginal and germinal zones of the developing and postnatal brains suggesting a specific role of this novel endo-siRNA in the regulation of brain development. To our knowledge, this is the first demonstration in the mammalian system that sense-NATs serve as Dicer1-dependent templates in the cytoplasm to produce an endo-siRNA. These findings add to our understanding of the versatility of NAT function in mammalian biology.

Introduction

Less than 2% of the 3 billion bases in the human genome consist of protein coding sequences [2, 3]. Recently, the Encyclopedia of DNA Elements (ENCODE) project showed as much as 90% of the studied 30 million bases of the human genome were transcribed, mostly into non-protein coding RNA [4]. This finding is in agreement with previously reported transcriptional landscapes of the mammalian genome [5-7] suggesting the previously termed 'junk' DNA contributes far more to the transcriptional profile of an organism. These non-protein coding transcripts differ from the canonical protein-coding transcripts and other structural RNAs (e.g. rRNA and tRNA) in terms of their gene distribution, abundance of regulatory motifs and transcription start sites, and evolutionary constraint [4, 8, 9]. As a consequence, these transripts are difficult to characterise functionally.

One type of non-protein coding transcript has sequences that are partially complementary to a protein-coding gene or RNA. These transcripts are known as natural antisense transcripts (NATs) and are either transcribed from the opposite DNA strand of the same protein-coding gene locus (known as *cis*-NATs) or from different loci within the genome (known as *trans*-NATs). To date, numerous NATs have been discovered in the mammalian genome and up to 20% of protein coding genes in human and mouse have at least one overlapping NAT [10-13]. Although numerous sense-NAT pairs have been identified *in silico* in the mammalian genome, only a handful of NATs such as *Evf-2* [14], *Air* [15], *HOTAIR* [16] and *Kcnq1ot1* [17] are implicated in a significant biological mechanism. These NATs are mainly nuclear localised and regulate gene expression during development through transcriptional activation (*Evf-2*), transcriptional repression (*HOTAIR* and *Kcnq1ot1*) or via chromatin modification (*Air*). Differential regulation of NATs expression has been implicated in various human disorders such as fragile X-associated tremor and ataxia syndrome [18], breast, renal and colon cancer [19-21], human follicular lymphoma [22], Beckwith-Wiedemann syndrome [23] and alpha-thalassemia [24], implicating the involvement of NATs in disease development. However, the mechanism underlying the role of NATs in disease progression remains unknown.

We recently reported that the *Sox4* (Sry-related HMG box 4 gene) gene locus produced multiple overlapping protein-coding mRNAs and NATs during mouse cerebral corticogenesis [25]. *Sox4* is a single exon gene encoding a 47kDa transcription factor containing a high-mobility group (HMG) domain, which functions in DNA binding, DNA bending, protein interactions and nuclear import or export [26, 27]. Sox4 binds the DNA sequence (A/T)(A/T)CAA(A/T)G (or IUPAC code WWCAAW) and in the presence of one or more specific co-factors, can function as either a transcription repressor or activator [26, 28] that regulates various developmental processes such as lymphocyte differentiation and bone, heart and brain development [29-33]. Its expression has also been implicated in the progression or transformation of various tumours and cancerous cells [34-39]. In the central nervous system, Sox4 regulates pan-neuronal gene expression that involves in the establishment of neuronal properties [31], and must be downregulated for proper myelination to occur [33]. The expression of multiple *cis*-NATs at the *Sox4* gene locus may be required for proper Sox4 protein synthesis through regulation of protein-coding transcripts at either the transcription, post-transcription or translation level. To investigate the role of *Sox4 cis*-NATs during brain development, we characterised NATs at the *Sox4* gene locus, and their underlying regulatory effect on the Sox4 protein-coding transcripts.

In this study, we demonstrated that *Sox4 cis*-NATs form double stranded RNA with the *Sox4* protein-coding (sense) transcripts to produce a novel endogenous small interfering RNA found specifically in the germinative layers of the mouse brain. To our knowledge, this is the first demonstration in the mammalian system that sense-NATs serve as Dicer1-dependent templates in the cytoplasm to produce a functional RNA. Our findings add to our understanding of the versatility of NAT function in the mammalian biology.

Materials and Methods

Breeding and handling of animals

Breeding and handling of animals were carried out according to guidelines approved by the Melbourne Health Animal Ethics Committee (Project numbers 2001.045 and 2004.041) and the University of Adelaide Animal Ethics Committee (S-086-2005). All mice used in the study were C57BL/6 unless otherwise specified. Mice were kept under a 12 hour light/12 hour dark cycle with unlimited access to food and water and were sacrificed by CO_2 asphyxiation. Brain tissues and mouse organs were dissected as previously described [25].

Rapid amplification of cDNA ends (RACE)-southern analysis

Total RNA was extracted from the E15.5 cerebral cortex using RNeasy Lipid Tissue Mini Kit (QIAGEN) according to the manufacturer's protocol. One thousand nanogram of pooled total RNA (n=3) was used for 5' and 3' RACE analyses of *Sox4* NATs. 5' and 3' RACE were carried out using SMART™ RACE Amplification Kit (Clonetech) according to the manufacturer's protocol. For both 5' and 3' RACE, up to seven *Sox4* specific primers were included for first strand cDNA synthesis. These primers were located over the ~5kb *Sox4* gene locus. Following cDNA synthesis, PCR was individually performed using all seven primers and a 5' universal primer for 5' RACE or 3' adaptor primer for 3' RACE (Clontech). Amplified products were blotted onto Hybond-N+ nylon membrane (GE Healthcare) and probed using independent oligonucleotides designed across the ~5kb *Sox4* gene locus to determine the specific *Sox4* amplicons. Independent oligonucleotide probes were end-labelled using [γ-^{32}P]-dATP and T4 polynucleotide kinase (Promega) according to the manufacturer's protocol. Five and six probes were used in 5' and 3' RACE analysis, respectively. See Supplementary Information 1 (SI-1) for primers and probe sequences.

The membrane was pre-hybridised in Amersham Rapid-hyb™ Buffer (GE Healthcare) with 100μg/ml of herring sperm DNA (Promega) at 42°C. Approximately $2x10^6$ dpm/ml labelled probe was added to membrane and hybridised for 2 hours. After hybridisation, the membrane was removed and washed in 5X sodium chloride/sodium citrate (SSC) with 0.1% (w/v) sodium dodecyl sulfate (SDS) (20 minutes at 37°C) followed by 1X SSC with 0.1% (w/v)

SDS (2x for 15 minutes, each at 65ºC) or until the recorded radioactivity (with a Geiger counter) was below 10 CPM. The membrane was exposed to Amersham Hyperfilm MP (GE Healthcare) in an intensifying cassette at -80ºC followed by standard autoradiography procedures.

Real-time quantitative polymerase chain reaction (RT-qPCR)

Total RNA was isolated from various mouse organs, brain regions and P19 cells using TRIzol® Reagent (Invitrogen). Contaminating genomic DNA (gDNA) was removed from the total RNAs using the DNA-*free*™ Kit (Applied Biosystems) according to the manufacturers' protocols. Reverse transcription was performed using the Superscript® III Reverse Transcriptase Kit (Invitrogen) according to the manufacturer's protocol. RT-qPCR was carried out in 10µl reaction volume using 1X LightCycler 480 (LC480) Probe Master mix (Roche Diagnostics), 0.1µM of a relevant Universal ProbeLibrary probe (Roche Diagnostics), 0.25µM of each forward and reverse primers and 1µl of 0.1X synthesised cDNA. RT-qPCR was performed with an initial denaturation at 95ºC for 10 minutes followed by 45 cycles at 95ºC for 10 seconds, 60ºC for 30 seconds and 72ºC for 10 seconds with a final step at 40ºC for 1 second.

Real-Time amplification signals were acquired during the elongation step and recorded live using LightCycler® 480 Software version 1.5 (Roche Diagnostics). The cycle threshold or crossing point (Cp) from each signal was calculated based on the Second Derivative Maximum method [40]. Relative quantification using a standard curve was employed to estimate the differences between normalised expression values according to the method as described [25]. T-tests were used to test the differences in normalised expression levels between two groups with a *P* value <0.05 considered statistically significant.

Stemloop RT-qPCR

Reverse transcription of small RNA was performed based on modified methods [41, 42]. cDNA was synthesised from 150ng-2.5µg of small RNA enriched total RNA using 0.05µM of an in-house designed stem loop primer (5'-GTTGGCTCT GGTAGGATG CCGCTCTCA GGGCATCCT ACCAGAGCCA AACGGAATC-3', GeneWorks), and the Superscript® III Reverse Transcriptase Kit (Invitrogen) with modifications to the manufacturer's protocol. The stem loop primer was

added after a denaturation step at 65°C for 5 minutes. The last 6nt at the 3' end of the stem loop primer complements with the last 6nt of the 3' end of *Sox4_sir3* small RNA. The stem loop RT primer contains a target site for a universal reverse primer (5'-GTAGGATGCC GCTCTCAGG-3', GeneWorks) and a target site for UniversalProbe Library (UPL) Probe #21 (Roche Diagnostics), which were used in subsequent cDNA amplication process together with a specific forward primer for *Sox4_sir3* (5'-TCTGACTCAA GGACAGCGAC AA-3', GeneWorks). Briefly, cDNA synthesis was performed at 16°C for 30 minutes followed by 60 cycles of 20°C for 30 seconds, 42°C for 30 seconds and 50°C for 1 second. Final incubation at 75°C for 15 minutes was performed to inactivate the reverse transcriptase enzyme.

Prior to qPCR, pre-PCR of *Sox4_sir3* was performed in a 10µl reaction volume containing 1X LC480 Probe Master mix (Roche Diagnostics), 50nM of each forward and universal reverse primers and 0.2X of synthesised cDNA. Pre-PCR was initially carried out at 95°C for 10 minutes, 55°C for 2 minutes and 75°C for 2 minutes followed by additional 14 cycles of 95°C for 15 seconds and 60°C for 4 minutes. After pre-PCR, 0.01X of amplicons were used for RT-qPCR according to the conditions above.

Small RNA northern analysis

Approximately 30µg of total RNA from each sample was denatured in 1X Ambion Gel Loading Buffer II (Ambion®) at 85°C for 3 minutes. RNAs were electrophoresed in 15% acrylamide/urea gel (48% (w/v) urea, 15% (v/v) acrylamide, 0.05% (w/v) ammonium persulfate and 0.1% (v/v) tetramethylethylenediamine prepared in 1X TBE) in 1X TBE buffer at 300V for 90 minutes followed by blotting of small RNAs in the gel onto Hybond-N+ nylon membrane (GE Healthcare). Pre-hybridisation, hybridisation, washing and visualisation steps were performed according to the protocols mentioned in the previous RACE section with the washing conditions changed to 5 minutes each time at room temperature. Long dsDNA probes used in the analysis were labelled using DECAprime™ II Random Priming DNA Labelling Kit (Ambion®) and [α-^{32}P]-dCTP (GE Healthcare) according the manufacturers' protocol.

RNA fluorescence in situ hybridisation (RNA FISH)

Preparation of cells from 56-day old (P56) adult mouse cerebral cortex, hippocampus, olfactory bulbs and cerebellum, cell fixation, RNAse A digestion, preparation of non-overlapping complementary RNA (cRNA) probes for *Sox4* sense and antisense transcripts (see SI-1), hybridisation, washing steps, staining and mounting of slides were carried out according to a method as described previously [43].

Locked nucleic acid-in situ hybridisation (LNA-ISH)

Paraffin embedded sections (8µm) were used for LNA-ISH. Sections were prepared according to the following steps: de-paraffinised with washes in xylene (3x for 5 minutes each), rehydrated, fixed in 4% (w/v) PFA (pH7.0) (10 minutes), digested in Proteinase K buffer (6.7µg/ml of Proteinase K, 50mM of Tris HCl pH7.5, 5mM of EDTA) (30 minutes), re-fixed in 4% (w/v) PFA (5 minutes) and acetylated (0.1M of triethanolamine, 0.178% (v/v) of concentrated HCl and 0.25% (v/v) of acetic anhydride) (10 minutes).

Pre-hybridisation step was carried out in a humidified chamber (50% (v/v) formamide, 5X SSC) at 60^{o}C. Amersham Rapid-hyb™ Buffer (GE Healthcare) was used for pre-hybridisation with additional *Escherichia coli* tRNA (Sigma Aldrich) and Herring Sperm DNA (Promega) to a final concentration of 100µg/ml each. After 1-2 hours of pre-hybridisation, custom-made *Sox4_sir3* LNA probes (Cat. no: EQ-70537, Exiqon) were added to the buffer to a final concentration of 0.020pmol/µl. Hybridisation was carried out in the oven for 16-20 hours.

After hybridisation step, sections were washed in 5X SSC (20 minutes at hybridisation temperature), then in 0.2X SSC (3 hours at hybridisation temperature), rinsed in fresh 0.2X SSC (5 minutes) and incubated in pre-blocking buffer (0.1M of Tris HCl pH7.5, 0.15M of NaCl and 240µg/ml of levamisole) (5 minutes). In a humidified chamber, sections were blocked in 20% (v/v) foetal calf serum (Sigma Aldrich) and 2% (w/v) blocking powder (Roche Diagnostics) in maleate buffer for 1 hour, then incubated with 0.0002X (0.00015U) anti-DIG antibody with alkaline phosphatase, Fab fragments (Roche Diagnostics) in blocking buffer (16 hours) in dark. Consequently, sections were washed in NTMT buffer (3x for 10 minutes each; 0.1M Tris HCl pH9.5, 0.1M NaCl, 0.05M $MgCl_2$,

1% (v/v) Tween-20 and 240μg/ml levamisole) and then in nitro blue tetrazolium chloride (NBT)/ 5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP) colour reaction (0.375mg/ml of NBT and 0.188mg/ml of BCIP in NTMT buffer) for 3 hours to 5 days. After colour reaction step, sections were washed with Tris EDTA buffer pH8.0 (0.01M of Tris HCl pH7.5 and 0.001M EDTA pH8.0) for 10 minutes and were mounted in Entellan® media (ProSciTech).

Mouse embryonic stem (mES) cells with Dicer1^{C} allele

Mouse embryonic stem (mES) cells with DICER1 activity were obtained from a line heterozygous for a conditionally mutant *Dicer1* allele (*Dicer1*^{c}) and a null *Dicer1* allele (*Dicer1*$^{-}$). These genetic modifications were described previously [44]. mES cells without DICER1 activity were produced by transient transfection of this *Dicer1*$^{c/-}$ line with Cre recombinase to produce *Dicer1*$^{-/-}$ subclones (Jeffrey R Mann and Deidre M Mattiske, unpublished data).

NIH/3T3 mouse fibroblast cell line

NIH/3T3 cells were obtained from American Type Culture Collection (www.atcc.org) and maintained in Dulbecco's Modified Eagle's Medium (Sigma Aldrich) supplemented with 10% (v/v) heat-inactivated foetal calf serum (FCS; Invitrogen), 100 units/ml penicillin, 100 μg/ml streptomycin and 2mM L-glutamine. Cells were subcultured into new dishes when they reached 80% confluence or less using approximately 3-5x10^{3} cells/cm^{2} inoculums.

P19 teratocarcinoma cell line

Propagation and differentiation of P19 cells were carried out according to protocols described previously [25].

Overexpression of Sox4 NATs

Full length *Sox4* NATs were amplified from mouse gDNA using the paired-end ditags sequences as primers (see SI-1). Proofreading polymerase enzyme from Expand Long Template PCR System Kit (Roche Diagnostics) was used to amplify *Sox4* NATs (with sizes between 0.8-3kb) from gDNA. Amplicons were blunt-end-cloned into pcDNA3 vector (Invitrogen) at EcoRV restriction site. The right clones were screened by orientation- and *Sox4*-specific PCR. The sizes of the constructs were validated by gel electrophoresis. Confirmed *Sox4* NAT

constructs, namely, PET2, PET3, PET5 and PET6, were individually co-transfected into NIH/3T3 cells together with pcDNA3-eGFP constructs using Lipofectamine™ 2000 Transfection Reagent (Invitrogen) according to the manufacturer's protocol. PET construct to pcDNA-eGFP construct ratio used during transfection was 3:2. Equal number of molecules for PET2, PET3, PET5 and PET6 were considered in each transfection and pUC18 vector was added to normalize the total amount of DNA transfected (6μg per 2 million cells) into NIH/3T3 cells. Each transfection for PET construct was controlled by a mock transfection using pcDNA3 empty vector and without any DNA.

Immunoblotting analysis

Approximately 24 hours after transfection, cells were lysed in ice-cold lysis buffer (50mM Tris.HCl, pH 7.4, with 1% NP-40, 150mM NaCl, 2mM ethylene glycol-bis(beta-aminoethyl ether)-N,N,N,N-tetraacetic acid, 1mM $NaVO_4$, 100mM NaF, 10mM $Na_4P_2O_7$, and protease inhibitor cocktail). Protein concentrations were assayed using Bradford Reagent (BioRad) according to the manufacturer's protocol. Equal amounts of protein (~20μg) were loaded onto 10% acrylamide gels, separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Electrophoresed proteins were transferred onto Amersham Hybond™-P hydrophobic polyvinylidene difluoride (PVDF) membrane followed by blocking with 5% (w/v) skim milk powder with 0.1% (v/v) Triton-X100 prepared in 1X PBS. The membrane was probed with a primary polyclonal rabbit antibody directed against Sox4 (Cat. No.: AB80261, Abcam) or polyclonal goat antibody directed against actin (Cat. No.: SC-1616, Santa Cruz) at 4^{o}C overnight followed by ImmunoPure® Goat Anti-Rabbit IgG, (H+L) Peroxidase Conjugate (Cat. No.: 31460, Pierce) or Polyclonal Rabbit Anti-Goat Immunoglobulins/HRP (Cat. No.: P0449, Dako) secondary antibody. Reactive bands were detected using Amersham ECL Plus™ Western Blotting Detection Reagents (GE Healthcare) according to the manufacturer's protocol.

Results

Sox4 NATs have multiple transcription start sites and polyadenylation sites

To characterise *Sox4* NATs that were expressed during brain development, we performed 5' and 3' RACE-southern analyses using E15.5 mouse cerebral cortex total RNA and combinations of primers/probes designed across the *Sox4* gene locus (Figures 1A-1D). To aid identification of full length *Sox4* NATs, we mapped previously described *Sox4* SAGE tags generated from 4 developmental stages of the mouse cerebral cortex to both sense and antisense strands of the *Sox4* gene locus [25]. Additional information from the 3' RACE analysis of *Sox4* NATs in the same study was also mapped to our data (Figure 1D). Next, we compared our results to the FANTOM Paired-End Ditags (PET) sequences, which were obtained from the Ensembl website (www.ensembl.org) and mapped to the *Sox4* gene locus (Figure 1E). All *Sox4* gene locus annotations are provided in one GenBank file (Supplementary information 1).

Surprisingly, we identified multiple transcription start sites (TSSs) as well as polyadenylation sites for *Sox4* NATs across the entire *Sox4* gene locus. 5' RACE analysis showed 19 different TSSs for *Sox4* NATs with 9 potential TSSs based on the prominent bands in southern analysis (Figures 1A and 1C). A total of 12 polyadenylation sites were found for *Sox4* NATs based on 3' RACE analysis, with four of them represented by prominent bands in southern analysis (Figures 1B and 1D). This suggests that transcripts with different TSSs or polyadenylation sites have different expression levels.

We compared our RACE results with the mapped PET sequences (Figure 1E). PET1-5 have TSSs that corresponded well (within ±100nt) to our mapped TSSs of *Sox4* NATs based on 5' RACE analysis, whereas the TSS for PET6 was found >100nt upstream to the TSS of our dataset. SAGE tags generated from the most 3' regions of PolyA+ RNAs confirmed 8 of the mapped polyadenylation sites for *Sox4* NATs. Of these, 3 polyadenylation sites from the previous study [25] matched with PET3. Other polyadenylation sites without any SAGE tags (beyond L6 in Figure 1D) also matched (within ±100nt) with the polyadenylation sites for PET1 and PET6. We did not find any polyadenylation sites from our 3' RACE analysis that matched the polyadenylation sites of PET2, PET4 and PET5. This

result could be attributed to the primer and/or probe used in our RACE experiments, which determine the specificity.

Next, we searched and mapped all the TATA box sequences, and 12 polyadenylation signal variants [45] across the antisense strand of the *Sox4* gene locus (Figure 1E). Only six possible TATA boxes were found with two containing a very conserved TATA box sequence, TATAAAAAA. The TATA box was located upstream of the TSSs of *Sox4* NATs from our 5' RACE analysis (between L1 and L3 in Figure 1C) and PET2, PET3 and PET4. On the other hand, polyadenylation signal sequences were found in all the mapped polyadenylation sites of *Sox4* NATs and in all the PETs except for PET4 (>500nt from polyA site) and PET5 (no signal). The analysis shows that these *Sox4* NATs utilised alternative core promoter sequences and polyadenylation signals during transcription initiation and polyadenylation, respectively.

Sox4 sense and NATs form cytoplasmic dsRNA aggregates

To determine whether *Sox4* NATs are functional, we first evaluated the cellular localisation of these transcripts in relation to sense transcripts. First, we performed RNA FISH analysis using non-overlapping RNA probes for both sense and antisense transcripts on cells obtained from the adult cerebral cortex, hippocampus, olfactory bulbs and cerebellum (Figure 2). The analysis showed cytoplasmic co-localisation of sense and NATs as aggregates in 5-10% of cells assessed. No sense and NATs aggregates or signals were found in the nucleus suggesting *Sox4* NATs are not playing a direct role in regulating *Sox4* gene transcription or chromatin modification. Next, we asked whether NATs formed double stranded RNA with the sense transcript in the cytoplasm, an event that is important in RNA interference and translation regulation. We performed RNA FISH on RNase A treated cells from the same mouse brain regions (Figure 2). Additional RNA FISH experiments targeting the *Hmbs* housekeeping gene were performed as a control for the RNase A digestion step. RNA FISH on RNase A treated cells confirmed that *Sox4* NATs indeed formed dsRNA with the sense transcripts. These observations suggest that *Sox4* NATs may regulate the translation of *Sox4* in the cytoplasm.

Sox4 NATs did not regulate the transcription or translation of Sox4 mRNA

To determine whether *Sox4* NATs affect the translation of Sox4 protein, we overexpressed selected PET2 (3.216kb), PET3 (1.919kb), PET5 (0.807kb) and PET6 (1.824kb) NATs in NIH/3T3 cells and assessed the level of Sox4 protein expression. The selected PETs were mapped to different regions of the *Sox4* protein-coding (sense) transcript or gene locus. PET2, PET3, PET5 and PET6 were overexpressed at 24, 16, 32 and 381 times greater than the pcDNA3-empty control in NIH/3T3 cells, respectively (Figures 3A-3D). In all the cells that overexpressed PETs, we did not observed any changes in *Sox4* sense transcript levels, except for those cells that overexpressed PET2, which exhibited about 2.6 times upregulation of the sense transcripts (Figure 3A). These findings support our previous RNA FISH analysis that demonstrated *Sox4* NATs are cytoplasm-localised and not involved in regulating nuclear *Sox4* protein-coding transcript expression. In addition, immunoblotting of protein lysates isolated from these cells showed no differential regulation of Sox4 proteins suggesting none of the *Sox4* NATs inhibit the protein translation process (Figure 3E).

A novel small RNA, Sox4_sir3, originates from the Sox4 sense transcript

Long dsRNAs can serve as templates for the biogenesis of small RNAs, which function via the RNA interference (RNAi) machinery. To answer the question as to whether *Sox4* sense and NATs dsRNA generate functional small RNAs, we screened the entire *Sox4* gene locus using 6 different dsDNA probes (Figure 4A). Small RNA northern analysis on the E15.5 and P150 whole brain total RNAs revealed 5 small RNA fragments (70nt to 140nt) within the *Sox4* gene locus suggesting the presence of intermediate precursor RNAs, which could serve as templates for small RNA production (Figure 4B). The strand specificity of these small RNA fragments, however, remains unknown. These findings led us to generate ~3.7 million small RNA sequences (36nt) from a mouse E15.5 whole brain using a massively parallel sequencing platform, the Illumina Genome Analyzer II (GSE22653). Based on the screening of these small RNA sequences, we mapped 7 small RNA sequences with a single unique hit to the *Sox4* gene locus where both sense and NATs were expressed (Table 1; Figure 4A). Interestingly, all 7 small RNA sequences were mapped to the *Sox4* sense transcript and were encountered only once in the dataset and could be generated by *Sox4* sense transcript degradation.

To ascertain whether these small RNAs were not *Sox4* sense transcript degradation by-products, we performed stem loop RT-PCR to evaluate the existence of these small RNAs in the E15.5 mouse brain. The analysis showed only *Sox4*_sir3 small RNA (5'-TCAAGGACAG CGACAAGATT CCGT-3'; GenBank Accession: HM596744) was specifically amplified, suggesting that this is the only genuine small RNA (Figure 4C). To further validate this small RNA, we performed another northern analysis using a radioactively labelled oligonucleotide probe that complements *Sox4*_sir3. We did not find any signals from northern analysis performed on the E15.5 and P150 whole brain small RNAs (data not shown) and therefore repeated the analysis using small RNAs isolated from proliferating and differentiated P19 teratocarcinoma cells. Differentiated P19 cells have characteristics resembling astrocytes and neurones in the brain [46]. The analysis showed a band with approximately 120nt in both the proliferating and differentiated P19 cells (Figure 4D) confirming our previous analysis using a long dsDNA probe (Figure 4B). In both northern analyses, however, we did not detect the mature small RNA sequence suggesting *Sox4*_sir3 expression is specific to a particular cell type, has a very short half life and is beyond the detectable limits of northern blotting. We later employed stemloop RT-qPCR to confirm *Sox4*_sir3 expression in both proliferating and differentiating P19 cells. This analysis confirmed the existence of *Sox4*_sir3 in P19 cells as well as a significant upregulation of the small RNA (~2-fold) in differentiating cells compared to proliferating cells (Figure 4D).

Next, we also searched *Sox4*_sir3 in other high-throughput mouse small RNA sequencing datasets available in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). We screened 53 high-throughput small RNA sequencing datasets generated from various studies using *Mus musculus* as a model (Series records: GSE20384, GSE19172, GSE17319, GSE7414, GSE5026 and GPL7059). Of these datasets, we found *Sox4*_sir3 was sequenced in only 2 datasets, GSM433295 and GSM475280, which were generated from E18.5 mouse testis small RNAs and Mili-immunoprecipitated adult mouse testis small RNAs, respectively. Similarly to our dataset, *Sox4_sir3* from these datasets has a low count number suggesting that the small RNA may have been subjected to nucleotide modifications at 5' or 3' ends or both. By considering both laboratory-

based and *in silico* analyses of *Sox4*_sir3, we demonstrate that the *Sox4* sense transcript is the origin for the generation of *Sox4*_sir3 novel small RNAs. However, the role of NATs and the mechanism involved in the biogenesis of this small RNA remains unknown.

To determine whether *Sox4*_sir3 small RNA is a novel microRNA (miRNA), we used 150nt upstream and downstream of *Sox4*_sir3 to look for potential hairpin stem loop structures using the RNAfold program [47]. Using default thermodynamic prediction parameters [48] and based on the uniform miRNA annotation criteria [49], we did not find any potent hairpin stem loop structures that may have functioned as the *Sox4*_sir3 precursor (Figure 4E). Therefore, we conclude that *Sox4_sir3* is not a miRNA.

Sox4_sir3 is an endogenous small interfering RNA

Thus far, we have confirmed that both *Sox4* sense and NATs can form dsRNA and *Sox4_sir3* originates from the *Sox4* gene locus, where both sense and NATs are expressed. Since *Sox4*_sir3 is not a miRNA, we speculated that this small RNA could be an endogenous small interfering RNA (endo-siRNA) that is generated from a long dsRNA via a Dicer1-mediated mechanism. To prove our hypothesis, we evaluated the expression of *Sox4*_sir3, *Sox4* sense and NATs in mouse embryonic stem (mES) cells with conditional alleles for *Dicer1* (Figure 4F). The analysis showed approximately 32-fold down-regulation of *Sox4*_sir3 in Dicer1 null mES cells, compared to cells expressing Dicer1 (Figure 4F). In addition, we observed no difference in *Sox4* sense transcripts and slight downregulation of NATs expression levels between the two cell types, confirming that *Sox4*_sir3 biogenesis is in fact Dicer1-dependent.

Sox4 NATs are required for the biogenesis of Sox4_sir3

Next, we wanted to determine whether *Sox4* NATs are required for the biogenesis of *Sox4*_sir3. *Sox4*_sir3 is mapped to the coding sequence (CDS) of the *Sox4* sense transcript, and PET6 is the NAT that overlaps this region (Figure 5A). We transfected NIH/3T3 cells with PET6 and assessed the expression level of *Sox4*_sir3. PET3-pcDNA3, a *Sox4* NAT that overlaps the 3'UTR of *Sox4* sense transcript, and pcDNA3-empty were transfected as controls (Figure 5A). We observed significant upregulation of both PET3 and PET6 ($P<0.001$) in the cells

and these transcripts did not affect the level of *Sox4* sense transcripts supporting our findings in the previous transfection analyses (Figures 5B-C). Cells overexpressing PET6 showed approximately 63-fold upregulation of *Sox4*_sir3 compared to the pcDNA3-empty vector control (Figure 5D). However, we did not observe any significant difference in *Sox4*_sir3 expression between cells overexpressing the PET3, and the pcDNA3-empty vector control (Figure 5D). In addition, *Dicer1* expression was not significantly different between transfection groups (Figure 5E), confirming *Sox4*_sir3 biogenesis is dependent on the expression of PET6 and the ability of *Sox4* sense and NATs to form dsRNA.

Full-length sequencing of PET6

In general, NATs are expressed in a specific cell type and stage during development. Full-length sequencing of NATs has been difficult due to their low expression level. To characterize the structure of the PET6 NAT expressed in the brain, we performed full-length sequencing of PET6 NAT isolated from transfected NIH/3T3 cells. We predicted that the cloned PET6 would be expressed and subjected to post-transcriptional processing similar to the naturally occurring PET6. We perform RT-PCR using the primers initially used to amplify PET6 from mouse gDNA. RT-PCR analysis showed amplification of two amplicons, approximately 1.8kb and 0.6kb in size (Figure 6). Full-length sequencing confirmed the smaller amplicon as a 0.637kb spliced transcript variant of PET6. The transcript was spliced from +354 to +1545 (1.187kb) at the canonical AG…GT acceptor and donor site. A BLASTp homology search and protein domain analysis using Simple Modular Architecture Research Tool (SMART) [50] for open reading frames with greater than 100aa within both transcript variants of PET6 did not show any significant protein homologues or functional domains within these transcripts (Supplementary information 1). These transcripts were not translated into peptides/proteins of known function to date. These observations led us to ask which transcripts are involved in the generation *Sox4*_sir3. To address this, we overexpressed both spliced and unspliced PET6 variants in NIH/3T3 cells and observed upregulation of *Sox4_sir3* only in cells overexpressing the unspliced variant suggesting the overlapping portion of the transcript is required for *Sox4_sir3* production (see SI-2). We suspected that the spliced variant of PET6 was an *in vitro* artefact, due to the overexpression of PET6 transcript. In addition, we did not find any RT-qPCR amplification of the

spliced variant in various mouse tissues such as brain, heart, spleen, pancreas and skeletal muscle using a pair of intron-spanning primers (data not shown). The unspliced PET6 sequence was termed *Sox4ot1* and was submitted to GenBank with the accession number HM596742.

Expression pattern of Sox4_sir3

To investigate the role of *Sox4*_sir3 during brain development, we performed *in situ* hybridisation (ISH) using Locked-Nucleic Acid (LNA) probes on sections obtained from E11.5, E13.5 and E15.5 whole mouse embryos, and E17.5 and P1.5 whole brains. We observed *Sox4*_sir3 expression in the telencephalon and mesencephalon of E11.5 and E13.5 mouse embryos (Figures 7A-B). The expression was gradually downregulated and confined to the ventricular and marginal zones (known as layer I after birth) of the cerebral cortex, ventricular zone, cerebellar anlarge and the granule layer of the olfactory bulb at E15.5 and E17.5 (Figures 7C-G). At P1.5, *Sox4_sir3* expression was observed in the diminishing ventricular zone of the cerebral cortex, subventricular zone of the lateral ventricle, layer I of the cerebral cortex, pyramidal layer of the hippocampus, granule cell layer of the dentate gyrus, granule layer of the olfactory bulb and Purkinje cell layers of the cerebellum (Figures 7H-J). Besides the developing and postnatal brains, *Sox4*_sir3 was strongly expressed in the developing liver between E11.5 and E15.5 (Figures A-C), and the developing lungs at E15.5 (Figure 7C). These findings show that *Sox4*_sir3 is expressed mainly in germinative zones and specialised neuronal cell layers in the brain as well as the developing lungs and liver suggesting that this endo-siRNA plays a role in the development or function of these cells or organs.

We also performed stemloop RT-qPCR to quantitatively analyse the expression profile of *Sox4_sir3* in whole brains at different developmental stages, different adult brain regions and different organs of adult mice (Figure 8). We showed *Sox4_sir3* expression in the whole brain decreases as embryos developed from E11.5 to E17.5. At P1.5, we observed a sudden surge of expression in the whole brain with about a 9-fold increase from E17.5, which then decreased (~6-fold) through to the adult stage at P150 (Figure 8A). When we analysed different brain regions of adult mice, *Sox4_sir3* expression was lower in the cerebellum compared to other brain regions (Figure 8B). When we compared the adult whole

brain at P150 with other adult organs, we observed significantly higher *Sox4_sir3* expression in the heart, kidney and pancreas (Figure 8C). Interestingly, *Sox4_sir3* that was found highly expressed in the liver during embryonic development was least expressed among the adult organs screened suggesting that *Sox4_sir3* is important in embryonic but not adult liver development.

Downstream targets of Sox4_sir3

Given the specific expression of *Sox4*_sir3 at mainly germinative zones and specialised neuronal cell layers in the brain, we asked whether *Sox4*_sir3 has any downstream targets that are involved in neurogenesis or neuronal cell development and function. To predict downstream targets for *Sox4*_sir3, we submitted the sequence to the DIANA - microT 3.0 program [51]. The analysis revealed 151 predicted target sites in the 3' UTRs of 40 genes (SI-3). We subjected all 40 genes to functional ontology analysis using Ingenuity Pathway Analysis (www.ingenuity.com) and found 6 candidate genes (*Creb1*, *Itsn1*, *Slc6a2*, *Xpo7*, *Kcnh1* and *Eya3*) that were associated with various neurological conditions such as neurodegenerative disorders, bipolar disorder and depression (SI-2). Based on the conservation analysis of target sites in these transcripts we showed 2 sites in *Creb1* (+598 to +626 and +4182 to +4210 of ENSMUSG00000025958) were conserved across 8 and 11 organisms.

Discussion

In this study, we characterised a cluster of NATs that overlap the *Sox4* sense transcript and described a new mechanism related to the ability of these transcripts to form dsRNAs. These dsRNAs served as templates for the production of a novel endogenous siRNA, *Sox4*_sir3, via a Dicer1-mediated mechanism in the mouse. Biogenesis of small RNAs from dsRNAs is common in viruses and plants where RNA-dependent RNA polymerase (RdRP) plays a pivotal role in catalyzing the RNA replication process [52, 53]. In mammals, where no RdRP activity has been reported, the mechanism responsible for the generation of dsRNAs remains unclear. In the absence of RdRP, the formation of mammalian dsRNAs have been proposed to ensue via other means such as pairing of partially or fully overlapping sense and NATs. This postulation, however, has never been proven or demonstrated in a mammalian system. The generation of functional endogenous siRNAs from dsRNAs in mammals is rare. To date, only limited numbers of endogenous siRNAs have been reported as derived from naturally occurring long dsRNAs or retrotransposons found in cultured human cells and murine germ cells during gonadal development [54-56]. Therefore, this is the first study to confirm such a phenomenon in the mammalian system and strengthens the role of NATs in eukaryotes especially the mouse or human.

Unlike other well-characterised NATs such as *Air*, *HOTAIR*, *Evf-2*, *Kcnq1ot1* and *NRON*, [14-17, 57], *Sox4* NATs have a unique role in the cytoplasm instead of directly regulating gene expression via chromatin modification, transcription activation or repression, or mRNA trafficking within the nucleus. In various incidences, sense-NATs formation in the cytoplasm can inhibit protein translation such as NOS2A in mollusks [58] and FGF-2 in human [59]. In the central nervous system, sense-NATs pairs or NATs alone have been described in the cytoplasm or synaptoneurosomes of neuronal cells with specialised function [60-62]. It has been suggested that NATs could exert post-transcriptional regulation on sense transcripts in the brain [63, 64]. A recent report showed cytoplasmic expression of *BACE1* NATs in brain samples from Alzheimer's disease patients stabilizes the sense transcript by masking the miR-485-5p binding site thus preventing miRNA-induced translational repression [60]. However, the functional role of these NATs in the development or function of neuronal cells has not been conclusively

proven. In our context, *Sox4* NATs are cytoplasm-localised and form dsRNAs with the sense transcript *in vivo*. Unlike NATs overlapping *NOS2A*, *FGF-2* and *BACE1*, when overexpressed *in vitro*, *Sox4* NATs (except for PET2) did not affect the *Sox4* sense transcript or protein levels but gave rise to *Sox4*_sir3 (via PET6), which origin is the sense transcript. However, the ectopic expression of artificially induced *Sox4* NATs in the cytoplasm may lead to the capturing of only downstream targets or functional properties. The method employed in the study will not explain any *cis* regulation features of *Sox4* NATs at transcription or chromatin level beyond *Sox4* gene locus.

Cytoplasmic generation of *Sox4*_sir3 from the sense transcript prevents self-targeting and the low expression of *Sox4* NATs in the brain [25] would not significantly alter the level of *Sox4* mRNA or protein levels *in vivo*. Instead, the small amount of *Sox4*_sir3 generated from the dsRNAs may have regulatory effects on other transcripts. This mechanism has been overlooked in the mammalian system and the small RNAs generated from the sense-NATs pairs could explain the existence of a large proportion of uncharacterised small RNAs, which are expressed in low abundance in the CNS [65, 66]. In this study, we demonstrate for the first time in a mammalian system that the function of NATs in the cytoplasm can contribute to the biogenesis of a small RNA that may regulate other transcripts.

Expression of *Sox4_sir3* is spatiotemporally regulated during embryogenesis between E11.5 and E17.5 with specific expression in the budding liver and the telencephalon of the E11.5 embryo. Between E11.5 and E15.5, the developing liver bud is a major site for foetal haematopoiesis. During these stages of development, haematopoietic stem cells (HSCs) increased exponentially in the foetal liver [67, 68] followed by mobilisation of HSCs to spleen and bone marrow after E16 [68]. Although *Sox4_sir3* expression was very specific to the liver increasing at E15.5, its role in liver development or foetal haematopoiesis remains unclear and requires further experimental validation. Based on the stemloop RT-qPCR results, we observed a sudden surge of *Sox4_sir3* expression in the P1.5 whole brain. When we compared the sagittal sections of E17.5 and P1.5 whole brains, we did not find any obvious differences in term of their spatial expression profiles suggesting that the surge may have been contributed by increased

expression of *Sox4_sir3* in the subventricular/ventricular zones associated with gliogenesis, which is at peak afterbirth. The *Sox4_sir3* expression was also specifically found in the marginal zone and layer I of the embryonic and postnatal mouse cerebral cortices, respectively. Interestingly, these regions are predominated by Cajal-Retzius neurones that secrete Reelin, a protein involved in establishing early neuronal circuitry, cortical lamination and cortical evolution [69, 70]. Therefore, the role of *Sox4_sir3* in the development and function of Cajal-Retzius neurones should be further evaluated.

Sox4 sense transcripts are highly expressed throughout the developing mouse brain except at the ventricular or subventricular zones of the cerebral cortex [25, 32, 71]. Interestingly, *Sox4_sir3* was observed mainly in these germinal layers of the brain after E13.5 where *Sox4* sense was not found. If *Sox4_sir3* expression *in vivo* is the outcome of *Sox4* sense-NATs interaction outlined by our model, then it is not possible to detect both sense and NATs transcripts in these regions because they have been processed into *Sox4_sir3*. When we compared our findings with *Sox4* antisense expression profiles describe previously [25], we found similarities between *Sox4_sir3* and *Sox4* antisense expression profiles. Both expression profiles showed lowest expression in the adult cerebellum and liver among the organs screened. The finding supports our overexpression studies that showed *Sox4_sir3* biogenesis as *Sox4ot1*-dependent. The balance between sense and antisense expression levels influences the level of *Sox4_sir3* expression. The unequal expression levels observed in the overexpression studies could explain why magnitude of changes for *Sox4_sir3* was lower than the overexpressed *Sox4ot1* transcripts. It is worth noting that the *Sox4_sir3* biogenesis is Dicer1-dependent and therefore the amount of *Sox4_sir3* production is limited by the intracellular Dicer1 activities. Additionally, *Sox4_sir3* originates from *Sox4* sense transcripts with perfect sequence complementarity to *Sox4ot1* thus directing transcript degradation. The existence of multiple overlapping *Sox4* antisense transcripts and the above-mentioned factors make the relationship of expression profiles between *Sox4* sense, antisense and *Sox4_sir3* difficult to interpret independently. After considering the expression at both the protein and transcript levels, we, therefore, propose that these transcripts exist in different levels as an outcome of an ‘equilibrium effect’ between each of these transcripts in the cell (Figure 9).

Sox4_sir3 was predicted to target a set of genes that were implicated in neurological disorders especially at a highly conserved target site within the 3'UTR of the *Creb1* mRNA. Mice lacking Creb1 feature extensive apoptosis of postmitotic neurones during brain development, linked with progressive neurodegeneration of the hippocampus and the dorsolateral striatum of postnatal forebrains [72]. Due to the type of sequence complementary between the two [73], it is likely that *Sox4_sir3* may cause translational repression of Creb1. This prediction, however, requires experimental validations of the *Sox4_sir3* effect on Creb1 at both mRNA and protein levels.

We have demonstrated a unique example of how *Sox4* NATs could diversify the post-transcriptional regulation of other genes without disrupting *Sox4* expression. Our findings, when further proven, will have significant implications on our understanding of NATs role in fine-tuning post-transcriptional regulation of gene expression within the cytoplasm. About 20% of well-defined proteins have at least one overlapping transcript [74], thus interactions of these sense-antisense RNA pairs may lead to alternative gene regulation pathways. Such unorthodox pathways can aid in explaining the diverse regulation of gene expression, hence explaining the missing link and the wide-spectrum transcript-protein relationship in various complex developmental processes.

Authors' contributions

KHL, PJB and CNH supervised, designed and carried out the overexpression studies and all qPCR analysis. KHL, SM and RF participated in the design and RNA FISH experiment. PQT supervised whereas KHL and PSC performed the LNA-ISH analysis. KHL, JMR and DLA participated in the small RNA sequence analysis, annotations and statistical analysis. KHL and MS performed the immunoblotting analysis. KHL and TD performed the small RNA northern analysis. DMM and JRM cultured and provided the mES cells with DICER1 conditional allele. KHL drafted the manuscript. PQT, CNH and HSS conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Health and Medical Research Council fellowships (171601 and 461204 to H.S.S.); National Health and Medical Research Council Grants 219176, 257501 and 257529 (to H.S.S.) and a fellowship from Pfizer Australia (to P.Q.T.). K.-H.L. was a recipient of the Melbourne International Fee Remission Scholarship (MIFRS) and Universiti Putra Malaysia Staff Training Scholarship (UPMSTS) and Adelaide Fees Scholarship International equivalent.

Table 1: Mapped small RNA sequences at the *Sox4* gene locus

ID	*Sequence*	*nt*	*%GC*	*Mapping*	*Sox4 strand*	*Sox4 region*
*Sox4*_sir1	aggcggagagtagacggg	18	67	chr13:29043245-	sense	3'UTR
*Sox4*_sir2	ccactggggttgtacgaa	18	56	chr13:29044007-	sense	CDS
*Sox4*_sir3	tcaaggacagcgacaagattccgt	24	50	chr13:29044567-	sense	CDS
*Sox4*_sir4	tcagggaaaggggtgggggga	20	65	chr13:29045181-	sense	5'UTR
*Sox4*_sir5	agacgatgtcgctttcctga	20	50	chr13:29045235-	sense	5'UTR
*Sox4*_sir6	ggacttaggcgctagag	17	59	chr13:29045252-	sense	5'UTR
*Sox4*_sir7	aggcgctagagacgatgt	18	56	chr13:29045246-	sense	5'UTR

Figures

Figure 1 **RACE-southern analysis of *Sox4* antisense transcripts expressed in the E15.5 cerebral cortex**

5’ RACE for *Sox4* antisense transcripts were independently carried out using the universal primer for 5’ and each of the 7 gene specific primers designed across the *Sox4* gene locus (A). Amplicons for each reaction was electrophoresed, blotted and specific *Sox4* antisense amplicons were detected using independent oligonucleotide probes designed downstream to each of the original primer used. Similar approach was performed for 3’ RACE analysis (B). The oligonucleotide probes used for detection are given at the lower right corner of each gel photo and represented by a coloured arrow at the lower left corner. These colour arrows denote all the corresponding amplicons in each gel and are schematically represented in figures (C) and (D) for both 5’ and 3’ RACE analyses, respectively. FANTOM Paired Ends di-Tag (PET) sequences for *Sox4* antisense transcripts, which were obtained from Ensembl website (www.ensembl.org) were mapped to the *Sox4* gene locus (E). Previously reported SAGE tags [25], selected TATA box and polyadenylation signal sequences were also mapped to the gene locus. Detail legend descriptions are provided in the bottom panel of the figure.

Figure 1

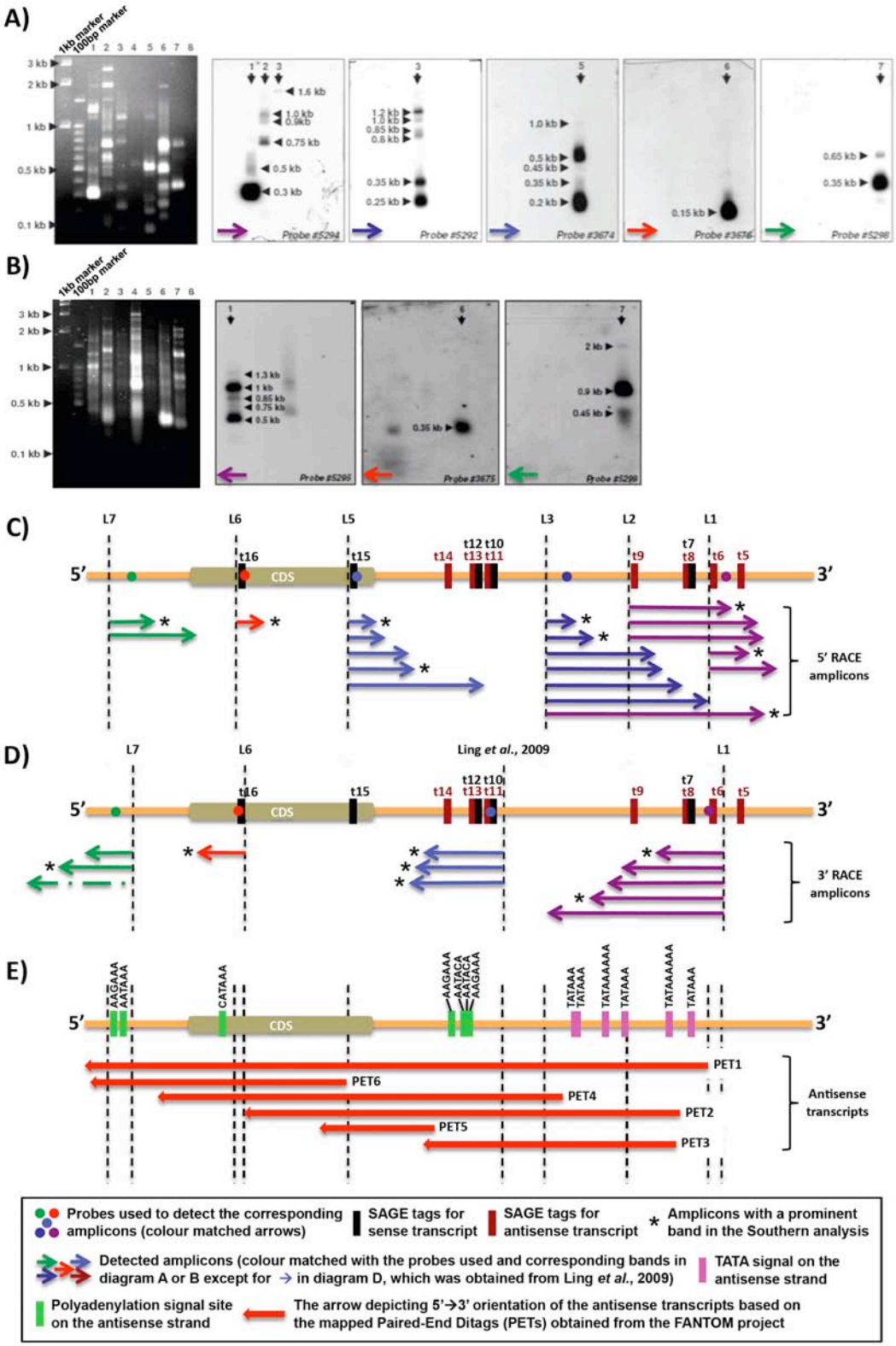

A)
1kb marker
100bp marker
1 2 3 4 5 6 7 8
3 kb
2 kb
1 kb
0.5 kb
0.1 kb
1.6 kb
1.0 kb
0.9kb
0.75 kb
0.5 kb
0.3 kb
Probe #5294
1.2 kb
1.0 kb
0.85 kb
0.8 kb
0.35 kb
0.25 kb
Probe #5292
1.0 kb
0.5 kb
0.45 kb
0.35 kb
0.2 kb
Probe #3674
0.15 kb
Probe #3676
0.65 kb
0.35 kb
Probe #5298
B)
1kb marker
100bp marker
1 2 3 4 5 6 7 8
3 kb
2 kb
1 kb
0.5 kb
0.1 kb
1.3 kb
1 kb
0.85 kb
0.75 kb
0.5 kb
Probe #5295
0.35 kb
Probe #3675
2 kb
0.9 kb
0.45 kb
Probe #5299
C)
L7
L6
L5
L3
L2
L1
5'
3'
t16
CDS
t15
t14
t12 t10
t13 t11
t9
t7
t8
t6
t5
5' RACE amplicons
D)
Ling et al., 2009
3' RACE amplicons
E)
AAGAAA
AATAAA
CATAAA
AAGAAA
AATACA
AATACA
AAGAAA
TATAAA
TATAAA
TATAAAAAA
TATAAA
TATAAAAAA
TATAAA
PET1
PET6
PET4
PET2
PET5
PET3
Antisense transcripts
Probes used to detect the corresponding amplicons (colour matched arrows)
SAGE tags for sense transcript
SAGE tags for antisense transcript
* Amplicons with a prominent band in the Southern analysis
Detected amplicons (colour matched with the probes used and corresponding bands in diagram A or B except for → in diagram D, which was obtained from Ling et al., 2009)
TATA signal on the antisense strand
Polyadenylation signal site on the antisense strand
The arrow depicting 5'→3' orientation of the antisense transcripts based on the mapped Paired-End Ditags (PETs) obtained from the FANTOM project

Figure 2 **RNA FISH of *Sox4* sense and NATs**

RNA FISH of *Sox4* sense and NATs was performed on trypsinised cells obtained from different regions of the adult mouse brain. The type of transcripts analysed is shown at the top of the figure and the origin of cells are shown to the left of the micrographs.

Figure 2

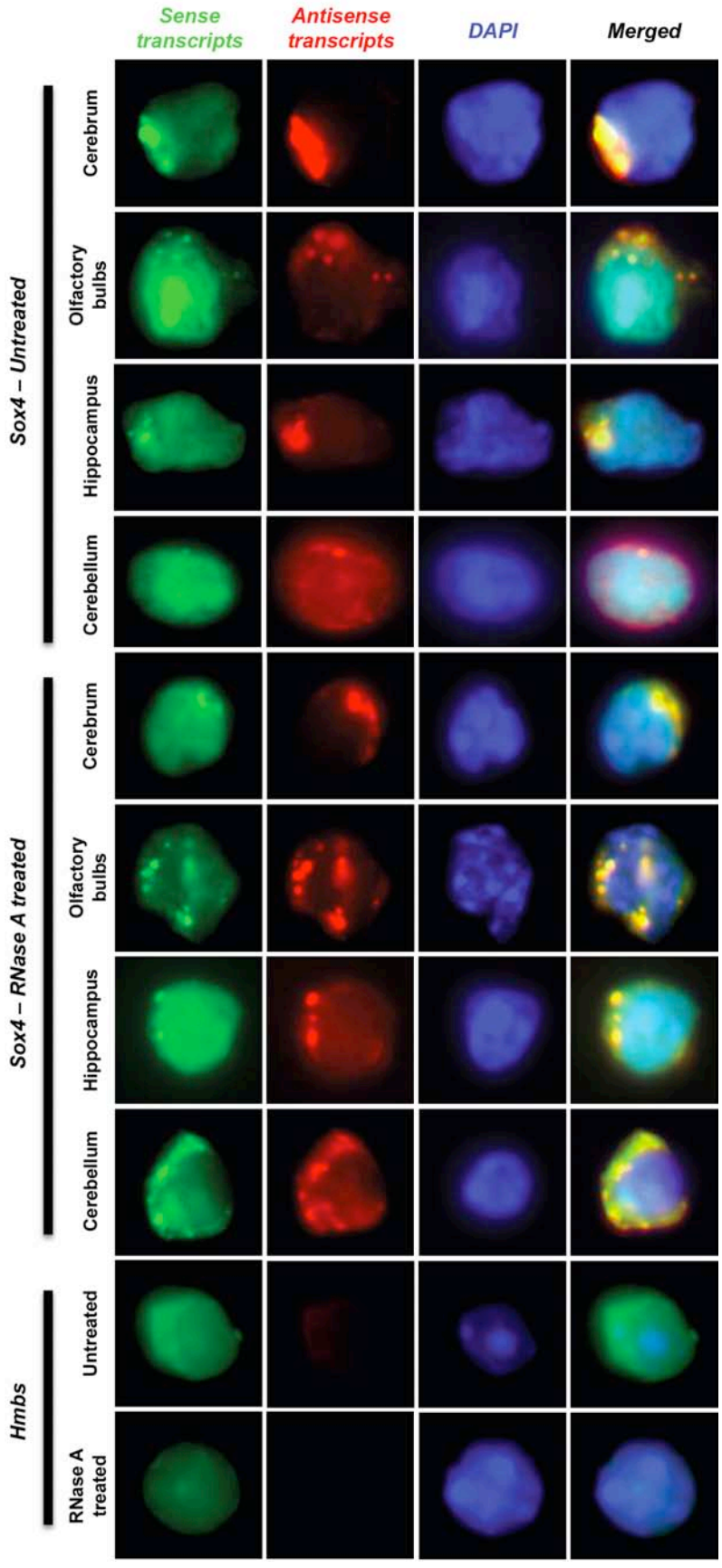

Sense transcripts
Antisense transcripts
DAPI
Merged
Sox4 – Untreated
Cerebrum
Olfactory bulbs
Hippocampus
Cerebellum
Sox4 – RNase A treated
Cerebrum
Olfactory bulbs
Hippocampus
Cerebellum
Hmbs
Untreated
RNase A treated

Figure 3 **Overexpression study of PETs 2, 3, 5 and 6 using NIH 3T3 cells**

Normalised log2 expression level of *Sox4* sense and NATs in NIH 3T3 cells transfected with reagent only (control), pcDNA3 empty vector (pcDNA3) and individual pcDNA3-PET construct is illustrated in (A) for PET2, (B) for PET3, (C) for PET5 and (D) for PET6. Western blot analysis using antibody against Sox4 and actin proteins is shown in (E) with additional control lysates from HeLa cells, P1.5 and P150 cerebral cortices (CC). For (A)-(D), N=3 per group and asterisks denote significant level at * P <0.05, ** P <0.01 and *** P <0.001. Error bars denote standard error of mean.

Figure 3

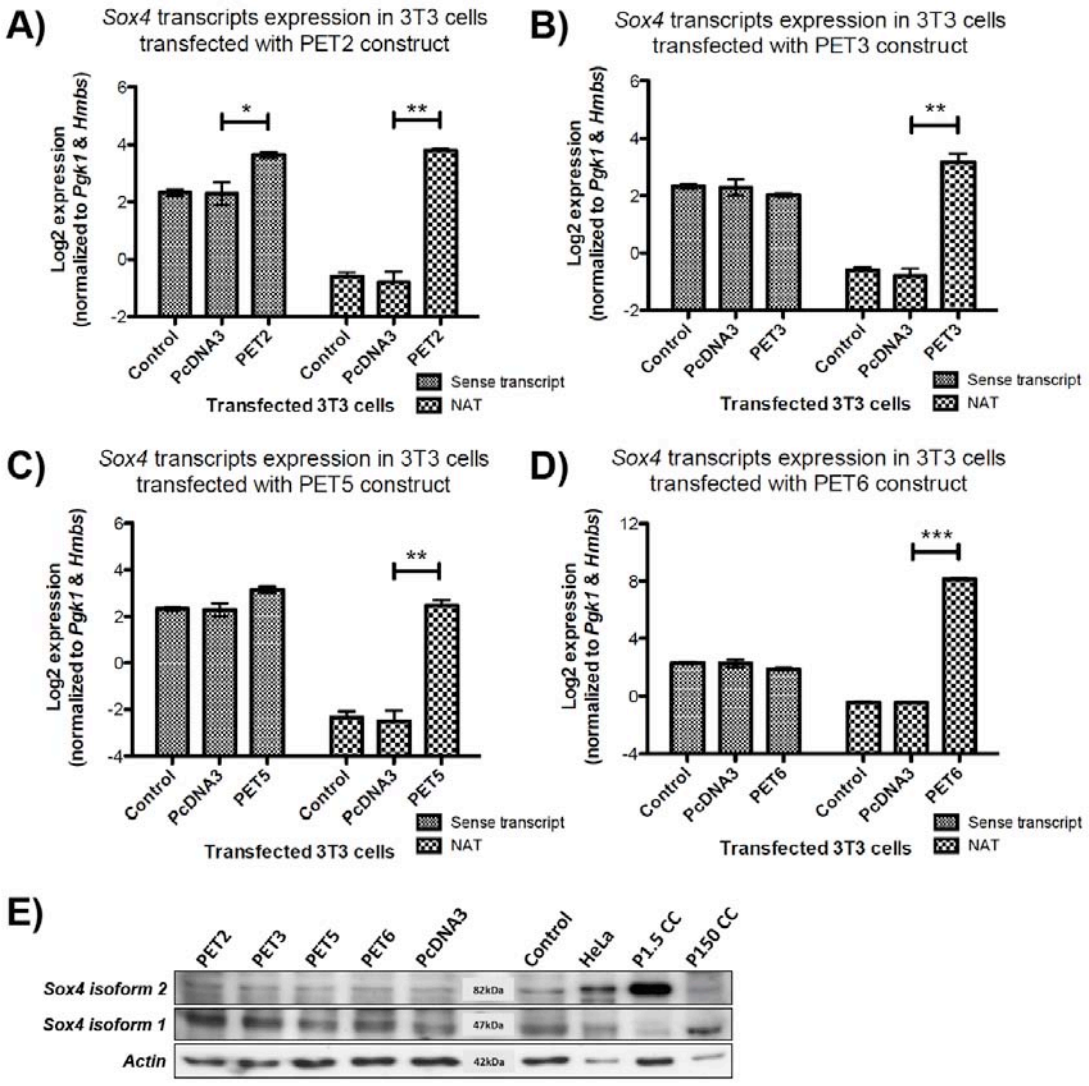

A)
Sox4 transcripts expression in 3T3 cells transfected with PET2 construct
Log2 expression (normalized to Pgk1 & Hmbs)
Control
PcDNA3
PET2
Transfected 3T3 cells
Sense transcript
NAT
B)
Sox4 transcripts expression in 3T3 cells transfected with PET3 construct
PET3
C)
Sox4 transcripts expression in 3T3 cells transfected with PET5 construct
PET5
D)
Sox4 transcripts expression in 3T3 cells transfected with PET6 construct
PET6
E)
PET2
PET3
PET5
PET6
PcDNA3
Control
HeLa
P1.5 CC
P150 CC
Sox4 isoform 2
Sox4 isoform 1
Actin
82kDa
47kDa
42kDa

Figure 4 **Identification of *Sox4_sir3* small RNA**

(A) Mapping of small RNA sequences that were originated from the *Sox4* gene locus. The probes used to screen for small RNAs in E15.5 and P150 whole brain (WB) samples using small RNA northern approach (B) are also illustrated in the diagram. P1-P6 in (B) corresponds to probes 1-6 in (A). Small RNA northern analysis of *Rnu6* was performed to serve as a positive/loading control. (C) Validation of small RNAs (S4_sir denotes *Sox4_sir* and NTC denotes no template control) using stemloop RT-PCR method identified *Sox4_sir3* as the only specific amplicon. (D) Small RNA northern analysis of *Sox4_sir3* using an oligonucleotide probe on total RNAs isolated from proliferating (prolif.) and differentiating (diff.) P19 teratocarcinoma cells. (E) RNA fold prediction of sequences 150nt upstream and downstream of *Sox4_sir3*. The colour in the scale bar presented below to the predicted structure denotes the possibility of base-pairing between nucleotides. The minimum free energy (MFE) structure, the thermodynamic ensemble of RNA structures (pf), the centroid structure (centroid) and the positional entropy for each position are presented in two separated graphs to right of the predicted structure. (F) Normalised log2 expression of *Sox4* sense and NATs, *Dicer1* and *Sox4_sir3* in mouse embryonic stem (mES) cells with conditional allele for *Dicer1*. Dicer c/- denotes mES cells with Dicer1 activity and Dicer -/- denotes mES cells without Dicer1 activity. For (F), N=3 per group and asterisks denote significant level at * P <0.05, ** P <0.01 and *** P <0.001. Error bars denote standard error of mean.

Figure 4

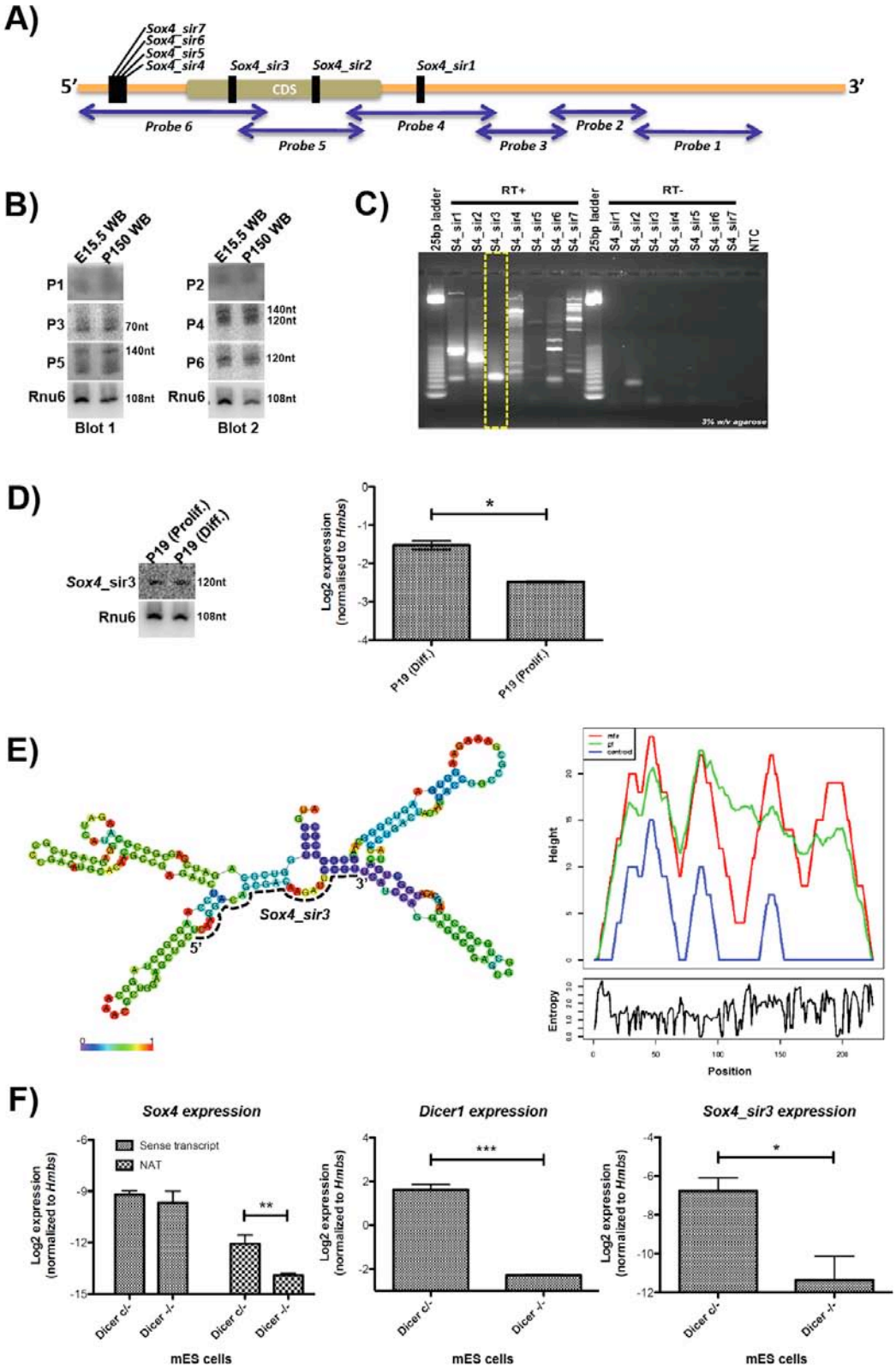

A)
Sox4_sir7
Sox4_sir6
Sox4_sir5
Sox4_sir4
Sox4_sir3
Sox4_sir2
Sox4_sir1
5'
CDS
3'
Probe 6
Probe 5
Probe 4
Probe 3
Probe 2
Probe 1
B)
E15.5 WB
P150 WB
P1
P3
70nt
140nt
P5
Rnu6
108nt
Blot 1
P2
140nt
120nt
P4
120nt
P6
Rnu6
108nt
Blot 2
C)
RT+
RT-
25bp ladder
S4_sir1
S4_sir2
S4_sir3
S4_sir4
S4_sir5
S4_sir6
S4_sir7
NTC
3% w/v agarose
D)
P19 (Prolif.)
P19 (Diff.)
Sox4_sir3
120nt
Rnu6
108nt
Log2 expression (normalised to Hmbs)
*
P19 (Diff.)
P19 (Prolif.)
E)
Sox4_sir3
5'
3'
Height
Entropy
Position
F)
Sox4 expression
Sense transcript
NAT
Log2 expression (normalized to Hmbs)
**
Dicer c/-
Dicer -/-
mES cells
Dicer1 expression
***
Sox4_sir3 expression
*

Figure 5 The effect of PET6 overexpression on *Sox4_sir3* expression

(A) A schematic diagram represents the overlapping regions between the *Sox4* sense transcript, and the PET3 and PET6 NATs. *Sox4_sir3* and primers used (Primers 1-6) in the study are also mapped. Normalised log2 expression level of *Sox4* sense (assessed by primers 3 and 4) and NATs in NIH 3T3 cells transfected with reagent only (control), pcDNA3 empty vector (pcDNA3) and individual pcDNA3-PET construct is illustrated in (B) for PET3 and (C) for PET6. (D) Normalised log2 expression of *Sox4_sir3* small RNA in NIH 3T3 cells transfected with reagent only (control), pcDNA3 empty vector (pcDNA3), pcDNA3-PET3 (PET3) and pcDNA3-PET6 (PET6) constructs. (E) Normalised log2 expression of the *Dicer1* transcript in all the 3T3-transcfected cells. Primers used during the sense- (S-RT) or antisense reverse-transcription (AS-RT) and sense- (S-qPCR) or antisense-quantitative PCR (AS-qPCR) are given in parentheses located below each graph. For (B)-(E), N=3 per group and asterisks denote significant level at ** P <0.01 and *** P <0.001. Error bars denote standard error of mean.

Figure 5

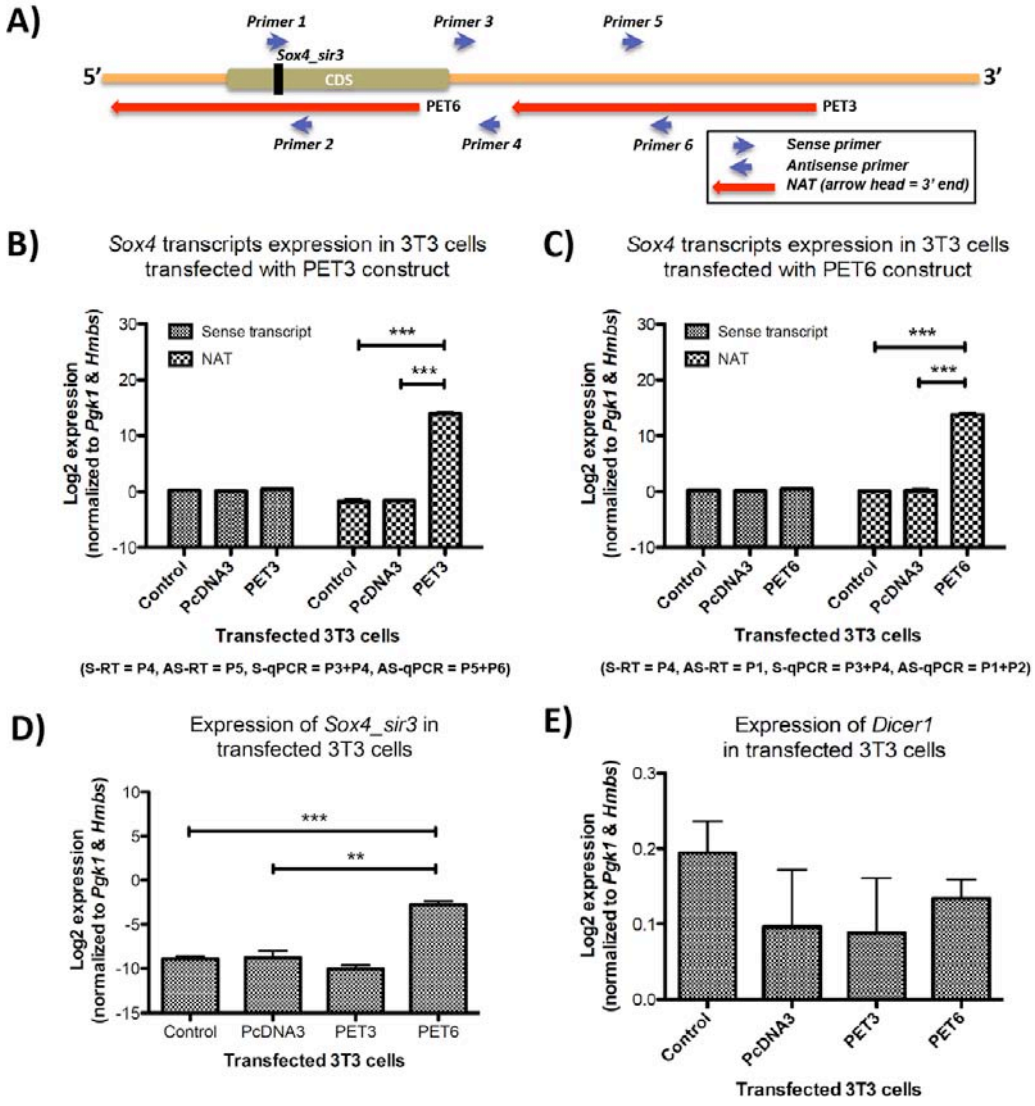

A)
Primer 1
Primer 3
Primer 5
Sox4_sir3
5'
CDS
3'
PET6
PET3
Primer 2
Primer 4
Primer 6
Sense primer
Antisense primer
NAT (arrow head = 3' end)
B)
Sox4 transcripts expression in 3T3 cells transfected with PET3 construct
Log2 expression (normalized to Pgk1 & Hmbs)
Sense transcript
NAT
***
***
Control
PcDNA3
PET3
Transfected 3T3 cells
(S-RT = P4, AS-RT = P5, S-qPCR = P3+P4, AS-qPCR = P5+P6)
C)
Sox4 transcripts expression in 3T3 cells transfected with PET6 construct
PET6
(S-RT = P4, AS-RT = P1, S-qPCR = P3+P4, AS-qPCR = P1+P2)
D)
Expression of Sox4_sir3 in transfected 3T3 cells
**
E)
Expression of Dicer1 in transfected 3T3 cells

Figure 6 **Sequencing of PET6 transcripts expressed in NIH 3T3-transfected cells**

RT-PCR of PET6 NATs expressed in NIH 3T3-transfected cells revealed 2 transcript variants, which is schematically illustrated in the diagram next to the gel. RT+ denotes full RT-PCR reaction performed on the total RNA isolated from 3T3-transfected cells, RT- denotes a reaction without reverse transcriptase performed on the same sample during RT step (genomic DNA contamination control), gDNA denotes RT-PCR performed on ~100ng mouse genomic DNA (positive control) and no template control (NTC).

Figure 6

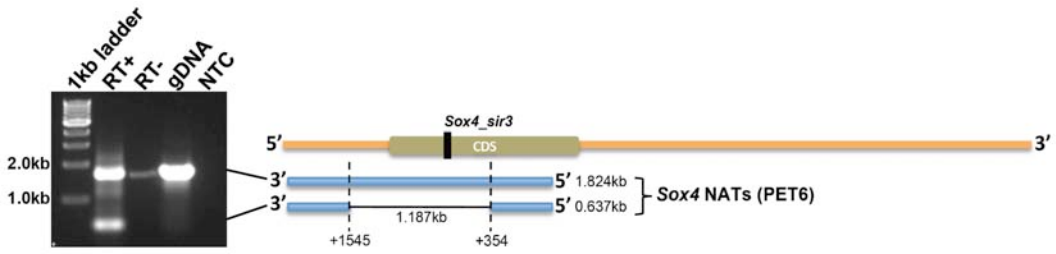

1kb ladder
RT+
RT-
gDNA
NTC
2.0kb
1.0kb
Sox4_sir3
CDS
5′
3′
5′ 1.824kb
5′ 0.637kb
Sox4 NATs (PET6)
1.187kb
+1545
+354

Figure 7 LNA-ISH of *Sox4_sir3* in whole mouse embryo and brain sections

LNA-ISH of *Sox4_sir3* was performed on whole embryo sections obtained from (A) E11.5, (B) E13.5 and (C) E15.5 embryos. For E17.5 developmental stage, only (D) coronal and (E-G) sagittal whole brain sections were analysed whereas for (H-J) P1.5, sagittal whole brain sections were analysed. 'bLV' = budding liver, 'CB' = cerebellum, 'CbAn' = cerebellar anlage, 'CC' = cerebral cortex, 'DG' = dentate gyrus, 'GrOB' = granule cell layer of the olfactory bulb, 'Hipp' = hippocampus, 'LG' = lungs, 'LI' = layer I of the cerebral cortex, 'LV' = liver, 'mes' = mesencephalon, 'MZ' = marginal zone, 'SVZ' = subventricular zone, 'tel' = telencephalon, 'VZ' = ventricular zone.

Figure 7

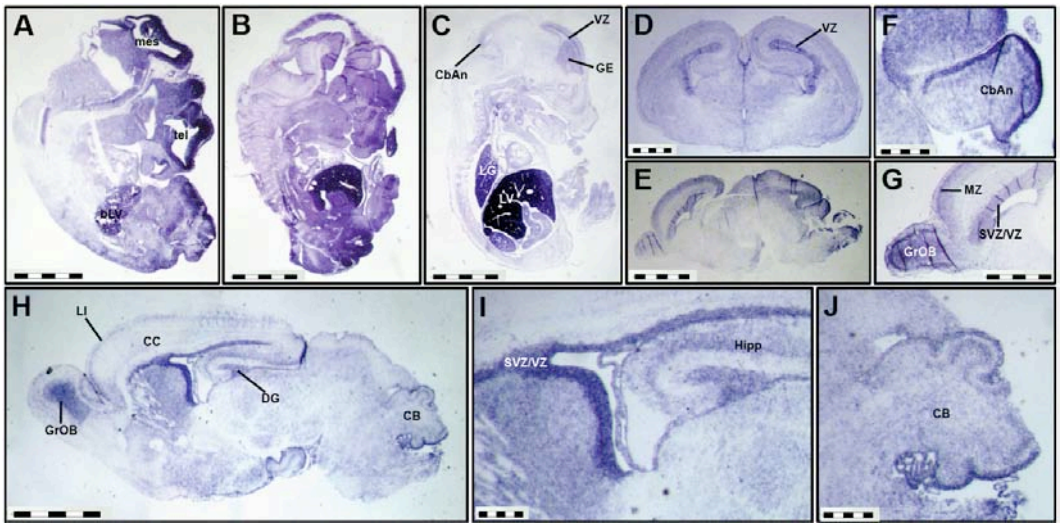

A
mes
tel
dLV
B
C
VZ
CbAn
GE
LG
LV
D
VZ
E
F
CbAn
G
MZ
SVZ/VZ
GrOB
H
LI
CC
DG
GrOB
CB
I
SVZ/VZ
Hipp
J
CB

Figure 8 Stemloop RT-qPCR characterisation of *Sox4_sir3* expression

The characterisation of *Sox4_sir3* expression was performed on (A) whole brains obtained from different developmental stages, (B) different brain regions of adult mice and (C) different adult mouse organs. Data were presented as log2 of normalised expression (to *Hmbs* housekeeping gene). For all analyses, N=2 per sample was used except for skeletal muscle where N=3. Asterisks denote significant level at * P <0.05 and *** P <0.001. Error bars denote standard error of mean.

Figure 8

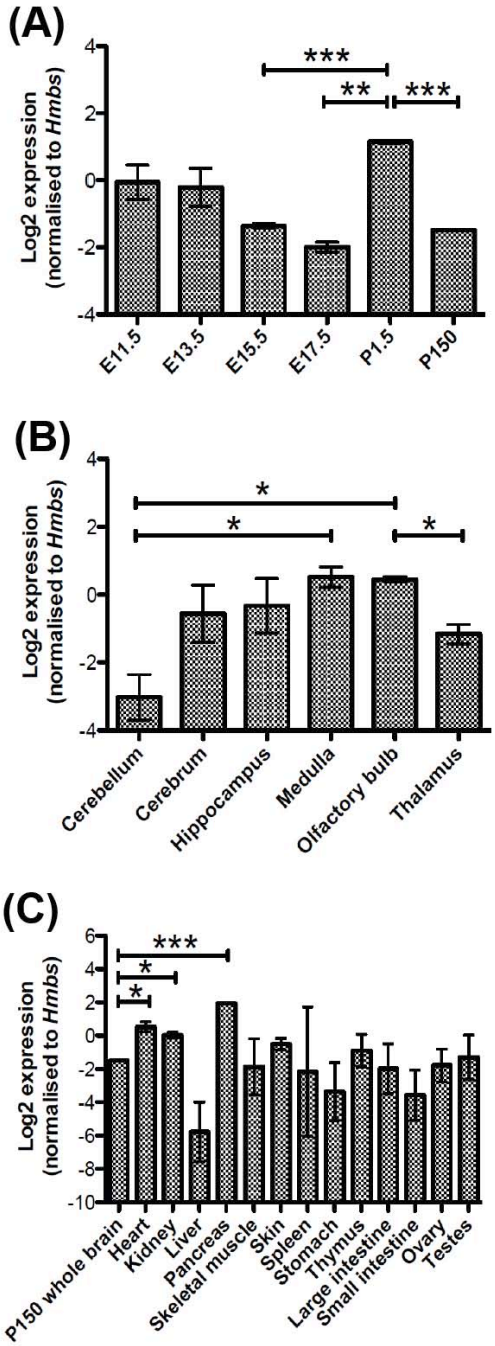
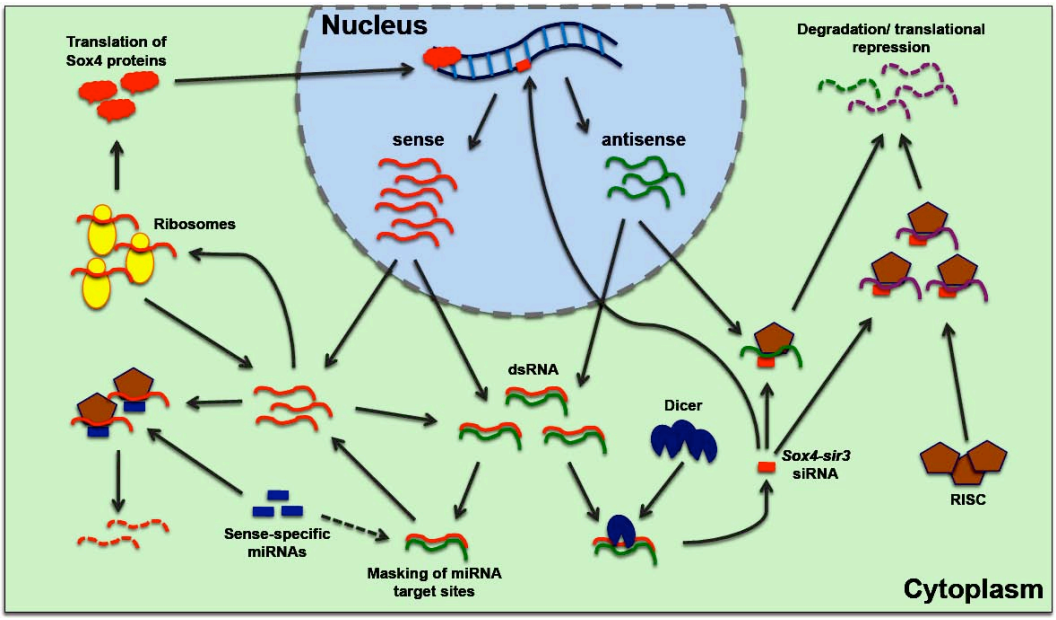

(A)
Log2 expression (normalised to Hmbs)
4
2
0
-2
-4
***
**
***
E11.5
E13.5
E15.5
E17.5
P1.5
P150
(B)
Log2 expression (normalised to Hmbs)
4
2
0
-2
-4
*
*
*
Cerebellum
Cerebrum
Hippocampus
Medulla
Olfactory bulb
Thalamus
(C)
Log2 expression (normalised to Hmbs)
6
4
2
0
-2
-4
-6
-8
-10
***
*
*
P150 whole brain
Heart
Kidney
Liver
Pancreas
Skeletal muscle
Skin
Spleen
Stomach
Thymus
Large intestine
Small intestine
Ovary
Testes

Figure 9 **Proposed relationship between *Sox4* sense transcripts, NATs, *Sox4_sir3* and protein expression**

Sox4 NATs are lowly expressed in the cell and form dsRNAs with the sense transcripts to produce *Sox4_sir3*. Since *Sox4_sir3* is originated from the *Sox4* sense transcript, it has perfect complementarities with *Sox4* NATs, hence causes RNA degradation via RNA-induced silencing complexes (RISC). These interrelationships seem to be very inefficient in regulating *Sox4* protein expression but in fact very useful in diversifying the downstream targets of *Sox4* via the action of *Sox4_sir3*. At a state of equilibrium, *Sox4_sir3* is capable of regulating both *Sox4* NATs and other transcripts. At the same time, the amount of *Sox4* sense transcripts degraded as a result of *Sox4_sir3* production is low and insignificant, thus do not affect the level of Sox4 protein expression. In a different context, *Sox4* NATs (such as PET2) overlaps 3'UTR of *Sox4* sense transcript can mask the miRNA target sites, hence stabilising the mRNA within the cytoplasm. The possibility of *Sox4_sir3* being transported into the nucleus and causes transcriptional regulation of other genes via chromatin remodelling in the nucleus remains unknown and required further experimental validations.

Figure 9


Nucleus
Translation of Sox4 proteins
sense
antisense
Degradation/ translational repression
Ribosomes
dsRNA
Dicer
Sox4-sir3 siRNA
RISC
Sense-specific miRNAs
Masking of miRNA target sites
Cytoplasm

References

1. **Ensembl Genome Browser** [http://www.ensembl.org/]
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**(6822):860-921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al*: **The sequence of the human genome**. *Science* 2001, **291**(5507):1304-1351.
4. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE *et al*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project**. *Nature* 2007, **447**(7146):799-816.
5. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G *et al*: **Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution**. *Science* 2005, **308**(5725):1149-1154.
6. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S *et al*: **Global identification of human transcribed sequences with genome tiling arrays**. *Science* 2004, **306**(5705):2242-2246.
7. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C *et al*: **The transcriptional landscape of the mammalian genome**. *Science* 2005, **309**(5740):1559-1563.
8. Trinklein ND, Karaoz U, Wu J, Halees A, Force Aldred S, Collins PJ, Zheng D, Zhang ZD, Gerstein MB, Snyder M *et al*: **Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome**. *Genome Res* 2007, **17**(6):720-731.
9. Margulies EH, Cooper GM, Asimenos G, Thomas DJ, Dewey CN, Siepel A, Birney E, Keefe D, Schwartz AS, Hou M *et al*: **Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome**. *Genome Res* 2007, **17**(6):760-774.
10. Fahey ME, Moore TF, Higgins DG: **Overlapping antisense transcription in the human genome**. *Comp Funct Genomics* 2002, **3**(3):244-253.
11. Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir Y, Khosravi R *et al*: **Widespread occurrence of antisense transcription in the human genome**. *Nat Biotechnol* 2003, **21**(4):379-386.
12. Shendure J, Church GM: **Computational discovery of sense-antisense transcription in the human and mouse genomes**. *Genome Biol* 2002, **3**(9):RESEARCH0044.
13. Kiyosawa H, Yamanaka I, Osato N, Kondo S, Hayashizaki Y: **Antisense transcripts with FANTOM2 clone set and their implications for gene regulation**. *Genome Res* 2003, **13**(6B):1324-1334.
14. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD: **The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator**. *Genes Dev* 2006, **20**(11):1470-1484.

15. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes**. *Nature* 2002, **415**(6873):810-813.
16. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al*: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.
17. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: **Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation**. *Mol Cell* 2008, **32**(2):232-246.
18. Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS, Hagerman RJ, Tassone F, Tapscott SJ, Filippova GN: **An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals**. *Hum Mol Genet* 2007, **16**(24):3174-3187.
19. Berteaux N, Aptel N, Cathala G, Genton C, Coll J, Daccache A, Spruyt N, Hondermarck H, Dugimont T, Curgy JJ *et al*: **A Novel H19 Antisense Rna Overexpressed in Breast Cancer Contributes to Paternal Igf2 Expression**. *Mol Cell Biol* 2008, **28**(22):6731-6745.
20. Yamamoto T, Manome Y, Nakamura M, Tanigawa N: **Downregulation of survivin expression by induction of the effector cell protease receptor-1 reduces tumor growth potential and results in an increased sensitivity to anticancer agents in human colon cancer**. *Eur J Cancer* 2002, **38**(17):2316-2324.
21. Thrash-Bingham CA, Tartof KD: **aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia**. *J Natl Cancer Inst* 1999, **91**(2):143-151.
22. Capaccioli S, Quattrone A, Schiavone N, Calastretti A, Copreni E, Bevilacqua A, Canti G, Gong L, Morelli S, Nicolin A: **A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t(14;18) cell lines**. *Oncogene* 1996, **13**(1):105-115.
23. Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, Smallwood AC, Joyce JA, Schofield PN, Reik W *et al*: **A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome**. *Proc Natl Acad Sci U S A* 1999, **96**(14):8064-8069.
24. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR: **Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease**. *Nat Genet* 2003, **34**(2):157-165.
25. Ling KH, Hewitt CA, Beissbarth T, Hyde L, Banerjee K, Cheah PS, Cannon PZ, Hahn CN, Thomas PQ, Smyth GK *et al*: **Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling**. *Genome Biol* 2009, **10**(10):R104.
26. Lefebvre V, Dumitriu B, Penzo-Mendez A, Han Y, Pallavi B: **Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors**. *Int J Biochem Cell Biol* 2007, **39**(12):2195-2214.
27. Kiefer JC: **Back to basics: Sox genes**. *Dev Dyn* 2007, **236**(8):2356-2366.

28. Kamachi Y, Uchikawa M, Kondoh H: **Pairing SOX off: with partners in the regulation of embryonic development**. *Trends Genet* 2000, **16**(4):182-187.
29. Schilham MW, Oosterwegel MA, Moerer P, Ya J, de Boer PA, van de Wetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumano A *et al*: **Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4**. *Nature* 1996, **380**(6576):711-714.
30. van de Wetering M, Oosterwegel M, van Norren K, Clevers H: **Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes**. *Embo J* 1993, **12**(10):3847-3854.
31. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J: **The establishment of neuronal properties is controlled by Sox4 and Sox11**. *Genes Dev* 2006, **20**(24):3475-3486.
32. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ: **Roles of Sox4 in central nervous system development**. *Brain Res Mol Brain Res* 2000, **79**(1-2):180-191.
33. Potzner MR, Griffel C, Lutjen-Drecoll E, Bosl MR, Wegner M, Sock E: **Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system**. *Mol Cell Biol* 2007, **27**(15):5316-5326.
34. Liu P, Ramachandran S, Ali Seyed M, Scharer CD, Laycock N, Dalton WB, Williams H, Karanam S, Datta MW, Jaye DL *et al*: **Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells**. *Cancer Res* 2006, **66**(8):4011-4019.
35. Aaboe M, Birkenkamp-Demtroder K, Wiuf C, Sorensen FB, Thykjaer T, Sauter G, Jensen KM, Dyrskjot L, Orntoft T: **SOX4 expression in bladder carcinoma: clinical aspects and in vitro functional characterization**. *Cancer Res* 2006, **66**(7):3434-3442.
36. Bangur CS, Switzer A, Fan L, Marton MJ, Meyer MR, Wang T: **Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis**. *Oncogene* 2002, **21**(23):3814-3825.
37. Boyd KE, Xiao YY, Fan K, Poholek A, Copeland NG, Jenkins NA, Perkins AS: **Sox4 cooperates with Evi1 in AKXD-23 myeloid tumors via transactivation of proviral LTR**. *Blood* 2006, **107**(2):733-741.
38. Lee CJ, Appleby VJ, Orme AT, Chan WI, Scotting PJ: **Differential expression of SOX4 and SOX11 in medulloblastoma**. *J Neurooncol* 2002, **57**(3):201-214.
39. Sinner D, Kordich JJ, Spence JR, Opoka R, Rankin S, Lin SC, Jonatan D, Zorn AM, Wells JM: **Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells**. *Mol Cell Biol* 2007, **27**(22):7802-7815.
40. Luu-The V, Paquet N, Calvo E, Cumps J: **Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction**. *Biotechniques* 2005, **38**(2):287-293.
41. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR *et al*: **Real-time quantification of microRNAs by stem-loop RT-PCR**. *Nucleic Acids Res* 2005, **33**(20):e179.
42. Tang F, Hajkova P, Barton SC, Lao K, Surani MA: **MicroRNA expression profiling of single whole embryonic stem cells**. *Nucleic Acids Res* 2006, **34**(2):e9.

43. Ling KH, Hewitt CA, Beissbarth T, Hyde L, Cheah PS, Smyth GK, Tan SS, Hahn CN, Thomas T, Thomas PQ *et al*: **Spatiotemporal Regulation of Multiple Overlapping Sense and Novel Natural Antisense Transcripts at the Nrgn and Camk2n1 Gene Loci during Mouse Cerebral Corticogenesis**. *Cereb Cortex* 2010.
44. Mattiske DM, Han L, Mann JR: **Meiotic maturation failure induced by DICER1 deficiency is derived from primary oocyte ooplasm**. *Reproduction* 2009, **137**(4):625-632.
45. Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D: **Patterns of variant polyadenylation signal usage in human genes**. *Genome Res* 2000, **10**(7):1001-1010.
46. McBurney MW, Reuhl KR, Ally AI, Nasipuri S, Bell JC, Craig J: **Differentiation and maturation of embryonal carcinoma-derived neurons in cell culture**. *J Neurosci* 1988, **8**(3):1063-1073.
47. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W70-74.
48. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure**. *J Mol Biol* 1999, **288**(5):911-940.
49. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M *et al*: **A uniform system for microRNA annotation**. *Rna* 2003, **9**(3):277-279.
50. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments**. *Nucleic Acids Res* 2009, **37**(Database issue):D229-232.
51. Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K *et al*: **DIANA-microT web server: elucidating microRNA functions through target prediction**. *Nucleic Acids Res* 2009, **37**(Web Server issue):W273-276.
52. Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SI: **RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production**. *Proc Natl Acad Sci U S A* 2005, **102**(1):152-157.
53. Ahlquist P: **RNA-dependent RNA polymerases, viruses, and RNA silencing**. *Science* 2002, **296**(5571):1270-1273.
54. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T *et al*: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes**. *Nature* 2008.
55. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H: **Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes**. *Genes Dev* 2006, **20**(13):1732-1743.
56. Yang N, Kazazian HH, Jr.: **L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells**. *Nat Struct Mol Biol* 2006, **13**(9):763-771.
57. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG: **A strategy for probing the function of noncoding RNAs finds a repressor of NFAT**. *Science* 2005, **309**(5740):1570-1573.

58. Korneev SA, Park JH, O'Shea M: **Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene**. *J Neurosci* 1999, **19**(18):7711-7720.
59. Li AW, Murphy PR: **Expression of alternatively spliced FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation**. *Mol Cell Endocrinol* 2000, **162**(1-2):69-78.
60. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, Cookson MR, St-Laurent G, 3rd, Wahlestedt C: **Evidence for natural antisense transcript-mediated inhibition of microRNA function**. *Genome Biol* 2010, **11**(5):R56.
61. Chiba M, Kubo M, Miura T, Sato T, Rezaeian AH, Kiyosawa H, Ohkohchi N, Yasue H: **Localization of sense and antisense transcripts of Prdx2 gene in mouse tissues**. *Cytogenet Genome Res* 2008, **121**(3-4):222-231.
62. Smalheiser NR, Lugli G, Torvik VI, Mise N, Ikeda R, Abe K: **Natural antisense transcripts are co-expressed with sense mRNAs in synaptoneurosomes of adult mouse forebrain**. *Neurosci Res* 2008, **62**(4):236-239.
63. Parenti R, Paratore S, Torrisi A, Cavallaro S: **A natural antisense transcript against Rad18, specifically expressed in neurons and upregulated during beta-amyloid-induced apoptosis**. *Eur J Neurosci* 2007, **26**(9):2444-2457.
64. Spigoni G, Gedressi C, Mallamaci A: **Regulation of Emx2 expression by antisense transcripts in murine cortico-cerebral precursors**. *PLoS ONE* 2010, **5**(1):e8658.
65. Cao X, Yeo G, Muotri AR, Kuwabara T, Gage FH: **Noncoding RNAs in the mammalian central nervous system**. *Annu Rev Neurosci* 2006, **29**:77-103.
66. Linsen SE, de Wit E, de Bruijn E, Cuppen E: **Small RNA expression and strain specificity in the rat**. *BMC Genomics* 2010, **11**:249.
67. Ikuta K, Weissman IL: **Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation**. *Proc Natl Acad Sci U S A* 1992, **89**(4):1502-1506.
68. Morrison SJ, Hemmati HD, Wandycz AM, Weissman IL: **The purification and characterization of fetal liver hematopoietic stem cells**. *Proc Natl Acad Sci U S A* 1995, **92**(22):10302-10306.
69. Aguilo A, Schwartz TH, Kumar VS, Peterlin ZA, Tsiola A, Soriano E, Yuste R: **Involvement of cajal-retzius neurons in spontaneous correlated activity of embryonic and postnatal layer 1 from wild-type and reeler mice**. *J Neurosci* 1999, **19**(24):10856-10868.
70. Bar I, Lambert de Rouvroit C, Goffinet AM: **The evolution of cortical development. An hypothesis based on the role of the Reelin signaling pathway**. *Trends Neurosci* 2000, **23**(12):633-638.
71. **Brain Gene Expression Map** [http://www.stjudebgem.org/]
72. Mantamadiotis T, Lemberger T, Bleckmann SC, Kern H, Kretz O, Martin Villalba A, Tronche F, Kellendonk C, Gau D, Kapfhammer J *et al*: **Disruption of CREB function in brain leads to neurodegeneration**. *Nat Genet* 2002, **31**(1):47-54.
73. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.

74. Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD: **Over 20% of human transcripts might form sense-antisense pairs**. *Nucleic Acids Res* 2004, **32**(16):4812-4820.

CHAPTER 6

Deep sequencing analysis of the developing mouse brain reveals a novel microRNA

King-Hwa Ling, Peter J Brautigan, Christopher N Hahn, Tasman Daish, John R Rayner, Pike-See Cheah, Joy M Raison, Sandra Piltz, Jeffrey R Mann, Deidre M Mattiske, Paul Q Thomas, David L Adelson and Hamish S Scott.

Published in: *BMC Genomics* 2011, 12(1):176.

6.1 Summary

In the last decade, noncoding RNAs have emerged as important regulators of various developmental mechanisms, disease onset and progression in mammals. This chapter reports an analysis of about 3.7 million small RNA sequences (36nt) generated from an E15.5 mouse whole brain using deep sequencing approach. The study reported here did not have direct link with the preceding chapters, but it utilises the next-generation sequencing dataset produced in Chapter 5 for novel miRNA discovery. Owing to the growing important of miRNAs in various mammalian developmental events, the dataset was re-analysed to identify novel miRNAs that could be pertinent to the brain development particularly in the cerebral cortex.

Analysis of the ~3.7 million sequences revealed putative miRNAs, which were validated using northern analysis. Further characterisation of one of the novel miRNAs, namely *miR-3099*, showed that this miRNA biogenesis as Dicer1-dependent. *MiR-3099* expression was observed in as early as E3.5 blastocycts and E7.5 embryos. Whole mount *in situ* hybridisation analysis showed embryo wide expression of *miR-3099* in E9.5 and E11.5 embryos. Interestingly, *miR-3099* expression was generally down-regulated and restricted to specific layers within the developing (E13.5, E15.5 and E17.5) and postnatal (P1.5) brains. *MiR-3099* expression was also upregulated in neurodifferentiating as compared to proliferating P19 cells. Taken together, *in vivo* and *in vitro* expression profiles for *miR-3099* suggest that this novel miRNA has potential regulatory effects in early embryogenesis, during brain development and postnatal brain function in the mouse.

The novel miRNA adds to the very limited number of known mouse miRNA suggesting that many miRNAs encoded in the genome are yet to be characterised due to their very specific spatiotemporal expression. The study concludes the last objective of the thesis leading to the identification of novel targets (ranging from mRNAs, NATs and small RNAs) involved in the proliferation, differentiation and developmental networks of the cerebral cortex.

6.2 Notes

The authors's declaration and 14 additional data files have been included as Appendix D according to the following order:

1. Authors' declaration.
2. Additional file 1: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s1.zip).
3. Additional file 2: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s2.txt).
4. Additional file 3: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s3.txt).
5. Additional file 4: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s4.txt).
6. Additional file 5: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s5.txt).
7. Additional file 6: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s6.txt).
8. Additional file 7: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s7.txt).
9. Additional file 8: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s8.txt).
10. Additional file 9: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s9.txt).

11. Additional file 10: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s10.txt).
12. Additional file 11: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s11.txt).
13. Additional file 12: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s12.txt).
14. Additional file 13: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s13.xls).
15. Additional file 14: (Original file is accessible at http://www.biomedcentral.com/content/supplementary/1471-2164-12-176-s14.doc).

6.3 Permission to Reuse Published Materials

The published manuscript is an open access article distributed under the terms of the Creative Common Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited (http://creativecommons.org/license/by/2.0/).

6.4 The published article

Please refer to the next page for the following published article:

Ling KH, Brautigan PJ, Hahn CN, Daish T, Rayner JR, Cheah PS, Raison JM, Piltz S, Mann JR, Mattiske DM, Thomas PQ, Adelson DL and Scott HS: **Deep sequencing analysis of the developing mouse brain reveals a novel microRNA.** *BMC Genomics* 2011, 12(1):176.

BMC Genomics

RESEARCH ARTICLE Open Access

Deep sequencing analysis of the developing mouse brain reveals a novel microRNA

King-Hwa Ling[1,2,3], Peter J Brautigan[1], Christopher N Hahn[1,2], Tasman Daish[4], John R Rayner[1], Pike-See Cheah[4,5], Joy M Raison[6], Sandra Piltz[4], Jeffrey R Mann[7], Deidre M Mattiske[7], Paul Q Thomas[4], David L Adelson[4] and Hamish S Scott[1,2*]

Abstract

Background: MicroRNAs (miRNAs) are small non-coding RNAs that can exert multilevel inhibition/repression at a post-transcriptional or protein synthesis level during disease or development. Characterisation of miRNAs in adult mammalian brains by deep sequencing has been reported previously. However, to date, no small RNA profiling of the developing brain has been undertaken using this method. We have performed deep sequencing and small RNA analysis of a developing (E15.5) mouse brain.

Results: We identified the expression of 294 known miRNAs in the E15.5 developing mouse brain, which were mostly represented by *let-7* family and other brain-specific miRNAs such as *miR-9* and *miR-124*. We also discovered 4 putative 22-23 nt miRNAs: mm_br_e15_1181, mm_br_e15_279920, mm_br_e15_96719 and mm_br_e15_294354 each with a 70-76 nt predicted pre-miRNA. We validated the 4 putative miRNAs and further characterised one of them, mm_br_e15_1181, throughout embryogenesis. Mm_br_e15_1181 biogenesis was Dicer1-dependent and was expressed in E3.5 blastocysts and E7 whole embryos. Embryo-wide expression patterns were observed at E9.5 and E11.5 followed by a near complete loss of expression by E13.5, with expression restricted to a specialised layer of cells within the developing and early postnatal brain. Mm_br_e15_1181 was upregulated during neurodifferentiation of P19 teratocarcinoma cells. This novel miRNA has been identified as *miR-3099*.

Conclusions: We have generated and analysed the first deep sequencing dataset of small RNA sequences of the developing mouse brain. The analysis revealed a novel miRNA, *miR-3099*, with potential regulatory effects on early embryogenesis, and involvement in neuronal cell differentiation/function in the brain during late embryonic and early neonatal development.

Background

A class of small non-coding RNA (19-25 nt in length) known as microRNA (miRNA) [1-3] can exert multilevel inhibition/repression processes during post-transcriptional or protein synthesis stages [4,5]. miRNAs are transcribed in the nucleus into long polyadenylated RNAs known as primary (pri)-miRNAs that contain ~60-90 nt secondary hairpin structures termed precursor (pre)-miRNAs. The RNase III enzymes Rnasen and Dgcr8 then excise the pre-miRNA from the pri-miRNA [1,6-9]. The pre-miRNA hairpin is transported into the cytoplasm via the nuclear transport receptor, Xpo5, and further processed by another RNase III enzyme, Dicer, into a small RNA duplex containing the functional mature miRNA and a passenger strand known as miRNA star [9-11]. The majority of the miRNA star are non-functional and are rapidly degraded, but a small proportion have conserved seed regions, potentially with regulatory roles [12]. The mature miRNA forms a component of the RNA-induced silencing complexes (miRISC) and guides these complexes to mRNA targets via sequence-specific pairing between the miRNA seed sequence (the first 7 nt of the miRNA starting from position 2) and the mRNA. Typically, miRNAs guide the RISC complex to the target mRNA 3' UTR, but incidences where 5' UTR and coding-sequences were targeted have been reported [13-15]. In mammals, miRISC normally effects translational repression and, depending

* Correspondence: hamish.scott@health.sa.gov.au
[1]Department of Molecular Pathology, SA Pathology and Centre for Cancer Biology, P.O. Box 14 Rundle Mall Post Office, Adelaide, SA 5000, Australia
Full list of author information is available at the end of the article

BioMed Central

© 2011 Ling et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

on the degree of miRNA:mRNA sequence complementation, can direct mRNA degradation [5,16]. Another intriguing regulatory role of miRNAs is the silencing of gene transcription which has been observed in plants [17], but has not yet been reported in the mammalian system.

Mammalian brain development requires meticulous spatio-temporal regulation of gene/protein expression, from the transcription of DNA within the nucleus to translation of mRNA in the cytoplasm [18,19]. At embryonic day 15.5 (E15.5), the mouse brain undergoes rapid cellular and anatomical changes involving neuronal migration in the cerebral cortex, proliferation of neural progenitor/stem cells at germinative zones, gliogenesis, axonogenesis and rostro-lateral to caudo-medial structure patterning [20-22]. MiRNAs play crucial roles during brain development and function. *MiR-134*, for example, is localised to the synapto-dendritic compartment of rat hippocampal neurones and has been linked to synaptic development, maturation and plasticity [23]. *MiR-9* regulates the patterning activities and neurogenesis at the midbrain-hindbrain boundary in zebrafish [24] and *miR-124* triggers brain-specific alternative pre-mRNA splicing leading to neuronal differentiation in the mouse [25]. MiRNAs are also associated with neurological disorders such as schizophrenia [26] and Huntington's disease [27]. To date, there are only 672 mature miRNAs in the mouse genome and 1048 in the human genome (miRBase release 16.0, September 2010) [28] in the mouse and human genomes, respectively. These figures are likely to be a gross underestimate of the actual number of miRNAs expressed. Most miRNAs are short lived, expressed in low abundance and found in specialised cell types during a specific developmental stage, and are therefore likely to remain uncharacterised due to technical limitations or the biological complexity of the tissues and cells of interest.

The emergence of next-generation sequencing technologies based on the massively parallel sequencing (MPS) concept has revolutionised the field of genomics and transcriptomics [29,30]. High-throughput generation of sequences from DNA or RNA has enabled the discovery of rare transcripts, such as alternatively spliced or fusion transcripts, as well as transcripts with low abundance [31,32]. Many next-generation sequencing datasets for small RNAs have been generated from the adult rodent and human brains [33-38]. However, to date, no small RNA profiling of the developing rodent or human brain has been performed using these methods. In this study, we performed deep sequencing of small RNAs prepared from an E15.5 mouse brain. *In silico* and laboratory based analyses led us to the discovery of 4 putative miRNAs; mm_br_e15_1181, mm_br_e15_279920, mm_br_e15_96719 and mm_br_e15_294354. Of these, mm_br_e15_1181 is novel and potentially involved in mouse embryogenesis, and brain development and function. This novel miRNA has been identified as *miR-3099*.

Results and Discussion

High-throughput sequencing and annotation of small RNA sequences

A total of 3,763,491 36 nt sequence reads were generated from a cDNA library constructed from mouse E15.5 whole brain small RNAs. The dataset was deposited into NCBI Gene Expression Omnibus GSE22653 [39]. Clustering of these sequence reads revealed 413,494 unique tags (Additional file 1). Screening for adaptor sequences (both 5' and 3') using a local blastn program showed 105,993 unique tags (6.9% or 259,681 sequence reads) did not have adaptor sequences indicating contamination of larger RNA transcripts during library construction (Figure 1A). Cloning errors resulted in 40,622 unique tags (11.0% or 413,837 sequence reads) consisting of only 5'and 3' adaptor sequences. The remaining 266,879 unique tags (82.1% or 3,089,973 sequence reads) were considered legitimate as they contained partial adaptor sequences at 5' or 3' or both ends. Of the legitimate unique tags, 59,710 (6.5% or 245,722 sequence reads) belonged to the 26-29 nt category, whereas 131,383 unique tags (61.5% or 2,314,244 sequence reads) of 20-25 nt were discovered, and therefore formed the majority of the small RNAs found in the cDNA library (Figure 1B). A total of 48,902 unique tags (3.8% or 141,783 sequence reads) were classified into the 16-19 nt category and 26,884 unique tags (10.3% or 388,224 sequence reads) of 16 nt or shorter were generated from either a pool of very small RNAs with unknown function or random RNA degradation by-products. The recent identification of tiny RNAs (~17-18 nt) shows that these small RNAs are associated with transcription initiation and splice sites specific to metazoans [40,41] suggesting that these tiny RNAs could be functional and represent another level of regulation during gene transcription in the nucleus.

Bowtie analyses, allowing only perfect matches, were performed on both the 5' and 3' end of each of the unique tags resulting in 339,201 tags (42% or 1,579,209 sequence reads) not finding a match in the mouse genome. This large proportion of unmatched unique tags included adaptors and low quality tags with errors in sequencing/base-calling. In exceptional circumstances, these unique tags could be derived from intron/exon or exon/exon boundaries, fusion transcripts or uncharacterised genomic regions. These unique tags with their corresponding sequence reads were not included for further analysis. The number of unmatched sequences varies from one study to another. Morin and colleagues

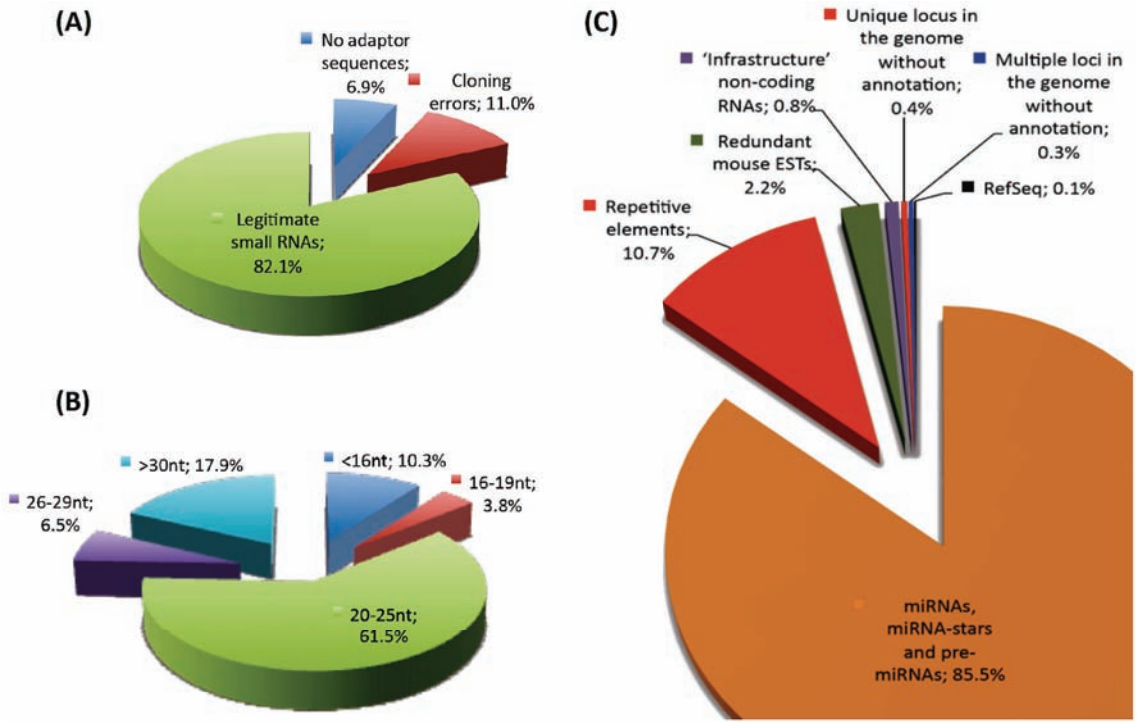


Figure 1 Generation and analysis of small RNAs sequences. There were 3,763,491 sequence reads generated. (A) Distribution of the small RNA sequences based on the analysis of 5' and 3' adaptor sequences. (B) Distribution of the small RNA sequences based on their size group. (C) Distribution of the small RNAs according to their annotations. All values presented in the figure were calculated based on the total sequence reads.

reported 29-35% of their total sequence reads generated from human embryonic stem cells and embryoid bodies small RNA libraries either consisted of errors or were not perfectly matched to the human genome [42]. In a different study, deep sequencing of small RNA libraries generated from cold-treated and untreated *Brachypodium* monocot plants resulted in only 49-54% of total sequence reads matching perfectly to the genome [43]. These studies suggested that a large proportion of the total sequence reads produced by deep sequencing are discarded from further analysis due to the quality of the sequence reads and stringency imposed during sequence alignment.

A total of 74,293 unique tags (58% or 2,184,282 sequence reads) were perfectly matched to the mouse genome. Of these, 7,136 (6.2% or 234,381 sequence reads) were matched to repetitive elements, and 6,929 (0.5% or 17,853 sequence reads) were matched to 'infrastructure' non-coding RNAs such as tRNA, rRNA, scRNA, snRNA or snoRNA (Table 1; Additional files 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12). These unique tags and their corresponding sequence reads were also excluded from further analysis. A total of 45,623 unique tags (49.6% or 1,867,113 sequence reads) were matched to either mature miRNA, miRNA star or pre-miRNA from miRBase, 2,448 (0.1% or 2,775 sequence reads) were matched to RefSeq, 6,584 (1.3% or 48,465 sequence reads) were matched to redundant mouse EST sequences, 1,752 (0.2% or 7,656 sequence reads) mapped to a single genomic locus and 3,821 (0.2% or 6,039 sequence reads) mapped to multiple loci within the genome (Figure 1C). Intriguingly, a large number of mapped unique tags in unique genomic loci have low abundance and lack association with any known mouse mRNAs, ESTs or miRNAs suggesting that these small RNAs could be generated from specific type of cells at specific stages of development and therefore have not been characterised to date.

The most abundantly expressed known miRNAs

To assess the expression of known miRNAs in the developing mouse brain at E15.5, we analysed all 294 mapped miRNAs in the dataset. Their counts ranged from 1 to 487,654 sequence reads or 0.27 to 129,575 per 1,000,000 sequence reads (CPM). The top 10% of the most abundantly expressed miRNAs are presented in Table 2 (see full list of known miRNAs in Additional file 13). The most abundantly expressed miRNA in the E15.5 developing mouse brain is *let-7c-1* with its 7 family members (*let-7a-2*, *let-7b*, *let-7d*, *let-7e*, *let-7f-2*, *let-7g* and *let-7i*) having a combined 335,288 CPM. Our finding agrees with the first report by Lagos-Quintana and colleagues [44] regarding

Table 1 Annotation of unique tags

Annotation of unique tags	Unique tags			Total combined counts^	Additional file(s)[#]
	22 nt of 3' end*[@]	22 nt of 5' end*[@]	Combined non-redundant^		
Repetitive elements	4,651	3,266	7,136	234,381	2 and 3
'Infrastructure' non-coding RNAs	6,907	30	6,929	17,853	2 and 3
miRNAs, miRNA stars and pre-miRNAs	45,623	0	45,623	1,867,113	4
RefSeq	2,431	22	2,448	2,775	5 and 6
Redundant mouse ESTs	5,954	737	6,584	48,465	7 and 8
Unique locus in the genome without annotation	1,377	439	1,752	7,656	9 and 10
Multiple loci in the genome without annotation	3,761	241	3,821	6,039	11 and 12
Total	70,704	4,735	74,293	2,184,282	

* No mismatch was allowed during Bowtie analysis.

^ Combined non-redundant values.

[@] Redundant values are presented. Redundant values were due to the same unique tag being analysed twice in both 5' and 3' Bowtie analysis.

[#] Annotation of unique tags based on 3' or 5' end sequences are presented in the additional files 2, 4, 5, 7, 9, 11 and 3, 6, 8, 10, 12, respectively.

Table 2 Top 10% of the most abundantly expressed known miRNAs

Small RNA ID	Accession ID	miRNA ID	Count per million	Chromosome	Start locus	Stop locus	Strand
mm_br_e15_1	MI0000559	mmu-let-7c-1	129574.91	16	77599901	77599995	+
mm_br_e15_1010	MI0000563	mmu-let-7f-2	59507.25	X	148346888	148346971	+
mm_br_e15_1001	MI0000557	mmu-let-7a-2	56984.06	9	41344798	41344894	+
mm_br_e15_10749	MI0000721	mmu-mir-9-3	27058.39	7	86650149	86650239	+
mm_br_e15_103211	MI0000137	mmu-let-7 g	25511.42	9	106081170	106081258	+
mm_br_e15_10459	MI0000561	mmu-let-7e	21824.95	17	17967315	17967408	+
mm_br_e15_1036	MI0000558	mmu-let-7b	19422.39	15	85537748	85537833	+
mm_br_e15_10	MI0000588	mmu-mir-103-2	16537.04	2	131113787	131113873	+
mm_br_e15_101787	MI0000138	mmu-let-7i	13005.48	10	122422695	122422780	-
mm_br_e15_10133	MI0000157	mmu-mir-9-2	11269.06	13	83878418	83878490	+
mm_br_e15_106	MI0000720	mmu-mir-9-1	9653.54	3	88019519	88019608	+
mm_br_e15_10266	MI0000405	mmu-let-7d	9457.18	13	48631380	48631483	-
mm_br_e15_10166	MI0000689	mmu-mir-25	7797.55	5	138606548	138606632	-
mm_br_e15_10031	MI0000155	mmu-mir-128-1	7303.33	1	130098937	130099007	+
mm_br_e15_1011	MI0000147	mmu-mir-99b	6712.12	17	17967151	17967221	+
mm_br_e15_10023	MI0000152	mmu-mir-125b-2	5810.83	16	77646517	77646588	+
mm_br_e15_1017	MI0000146	mmu-mir-99a	5567.70	16	77599180	77599245	+
mm_br_e15_13198	MI0000150	mmu-mir-124-3	3957.50	2	180628744	180628812	+
mm_br_e15_10339	MI0000144	mmu-mir-30a	3903.82	1	23279107	23279178	+
mm_br_e15_10279	MI0000165	mmu-mir-140	2629.74	8	110075143	110075213	+
mm_br_e15_1000	MI0005450	mmu-mir-181d	2452.78	8	86702614	86702686	-
mm_br_e15_10303	MI0000697	mmu-mir-181a-1	2322.84	1	139863031	139863118	+
mm_br_e15_11367	MI0000148	mmu-mir-101a	2237.28	4	101019549	101019632	-
mm_br_e15_10306	MI0000704	mmu-mir-320	2137.64	14	70843316	70843398	+
mm_br_e15_10234	MI0000684	mmu-mir-107	2068.03	19	34895176	34895263	-
mm_br_e15_10302	MI0000723	mmu-mir-181b-1	1851.47	1	139863215	139863295	+
mm_br_e15_11023	MI0000549	mmu-mir-30d	1836.86	15	68172769	68172851	-
mm_br_e15_10013	MI0000154	mmu-mir-127	1646.61	12	110831055	110831125	+
mm_br_e15_11551	MI0000729	mmu-mir-7a-2	1597.19	7	86033162	86033259	+
mm_br_e15_100	MI0000719	mmu-mir-92a-1	1563.97	14	115443648	115443728	+

the high representation of *let-7* family members in the mouse brain, which was also later found in the primate brain [45]. Despite their high level of expression in the brain, the functional role of *let-7* in the development of the central nervous system is poorly characterised. However, the expression of *let-7* has been associated with neural differentiation and lineage specification processes in early brain development [46].

Other miRNAs or miRNA families that were abundantly expressed in the E15.5 developing mouse brain include *miR-124* (3,958 CPM), which promotes and regulates neuronal differentiation [25] and *miR-9* (47,981 CPM), which has a role in the patterning activities and neurogenesis of the central nervous system [24]. *MiR-128* (7,303 CPM) was highly expressed in our dataset and the finding is in agreement with a previous study [47]. Down-regulation of *miR-128* expression has been associated with glioblastoma multiforme [48] whereas its up-regulation has been implicated with reduced neuroblastoma cell motility, invasiveness and cell growth [49]. In addition, both *miR-128* and *miR-9* are highly expressed in the foetal hippocampus and differentially regulated in the normal adult hippocampus as well as the hippocampus of Alzheimer's disease sufferers [50]. *MiR-125* (5,811 CPM) and *miR-99* (12,280 CPM) were also expressed highly in the developing mouse brain. Together with *let-7c*, both *miR-125* and *miR-99* are over-expressed by at least 50% in the foetal hippocampus of individuals with Down syndrome compared to age and sex matched controls suggesting that miRNAs are playing an important role in this brain region, which is pertinent for learning and long-term memory formation [51]. Interestingly, the *miR-103-2* (16,537 CPM), *miR-107* (2,068 CPM), *miR-181* (6,627 CPM) and *miR-30* (5,740 CPM) families have not previously been associated with the development of the brain, but were found to be highly expressed in our dataset. Both *miR-103* and *miR-107* are paralogous miRNAs and have been associated with lipid metabolism [52]. *MiR-181* plays a crucial role in modulating haematopoietic lineage differentiation [53] whereas *miR-30* has been strongly implicated with kidney development and nephropathies [54].

The identification of brain-related miRNAs by our deep sequencing analysis shows that the dataset is reliable not only for characterising expression profiles of known miRNAs but also for discovery of novel miRNAs. Further investigation of these miRNAs may shed light on their regulatory roles in various molecular pathways underlying the development of the embryonic brain.

Screening and validation of putative miRNAs and pre-miRNAs

To identify putative miRNAs, we analysed unique tags with a single match to the genome that were annotated as matched to RefSeq or redundant mouse EST sequences or were without annotation. A total of 10,784 unique tags (1.6% or 58,896 sequence reads) were selected under these criteria. We included all sequences with 1-2 counts into the analysis because we had found 34 known miRNAs residing in a similar range of expression within the dataset (see Additional File 13), suggesting some of the single count unique tags might be true positives. Pre-miRNA sequences were predicted using the RNA22 program, a pattern-based method reported previously [55]. The program predicted 8 putative miRNAs with pre-miRNA sequences; mm_br_e15_1181, mm_br_e15_279920, mm_br_e15_96719, mm_br_e15_294354, mm_br_e15_276138, mm_br_e15_331608, mm_br_e15_255873 and mm_br_e15_363469 (see Additional File 14). The resulting candidate pre-miRNA sequences were subjected to hairpin structure or fold prediction using the RNAfold program [56]. Of all the candidate putative miRNAs, only 4 fulfilled the criteria outlined for mature miRNA and pre-miRNA [2]. These were mm_br_e15_1181 (chr7:6756349-6756370), mm_br_e15_279920 (chr2:29597247-2959768), mm_br_e15_96719 (chr7:68982209-68982231), and mm_br_e15_294354 (chr7:68935407-68935429) which featured a 22-23nt mature miRNAs and a 70-76nt predicted pre-miRNAs (Figure 2A B and 2D). The other 4 putative miRNAs, mm_br_e15_276138, mm_br_e15_331608, mm_br_e15_255873 and mm_br_e15_363469 contained a large internal loop, branching stem or oversized pre-miRNA structural properties (see Additional file 14). These putative miRNAs were excluded from further analysis.

Mm_br_e15_1181 was matched to the second intron of the ubiquitin specific peptidase 29 (*Usp29*) gene. Mm_br_e15_279920 was matched to a single locus within the mouse genome without any annotations, whereas both mm_br_e15_96719 and mm_br_e15_294354 miRNAs were matched to two different introns of the same EST, BU505171. We performed a small RNA northern analysis on the E15.5 whole brain small RNAs to validate all the 4 predictions. We also included mm_br_e15_276138, mm_br_e15_331608, mm_br_e15_255873 and mm_br_e15_363469 in our northern analysis to serve as negative controls. The analysis confirmed all 4 predictions at the mature miRNA level for mm_br_e15_1181, mm_br_e15_96719 and mm_br_e15_294354, and at the pre-miRNA level for mm_br_e15_1181 and mm_br_e15_279920 (Figure 2C). As expected, the northern analysis of negative controls showed no detectable signals for mm_br_e15_276138 and mm_br_e15_363469, and multiple bandings for mm_br_e15_331608 and mm_br_e15_255873, signifying random by-products due to RNA degradation (see Additional file 14). Depending on the biological context of the assessed tissue, miRNA may be preserved or

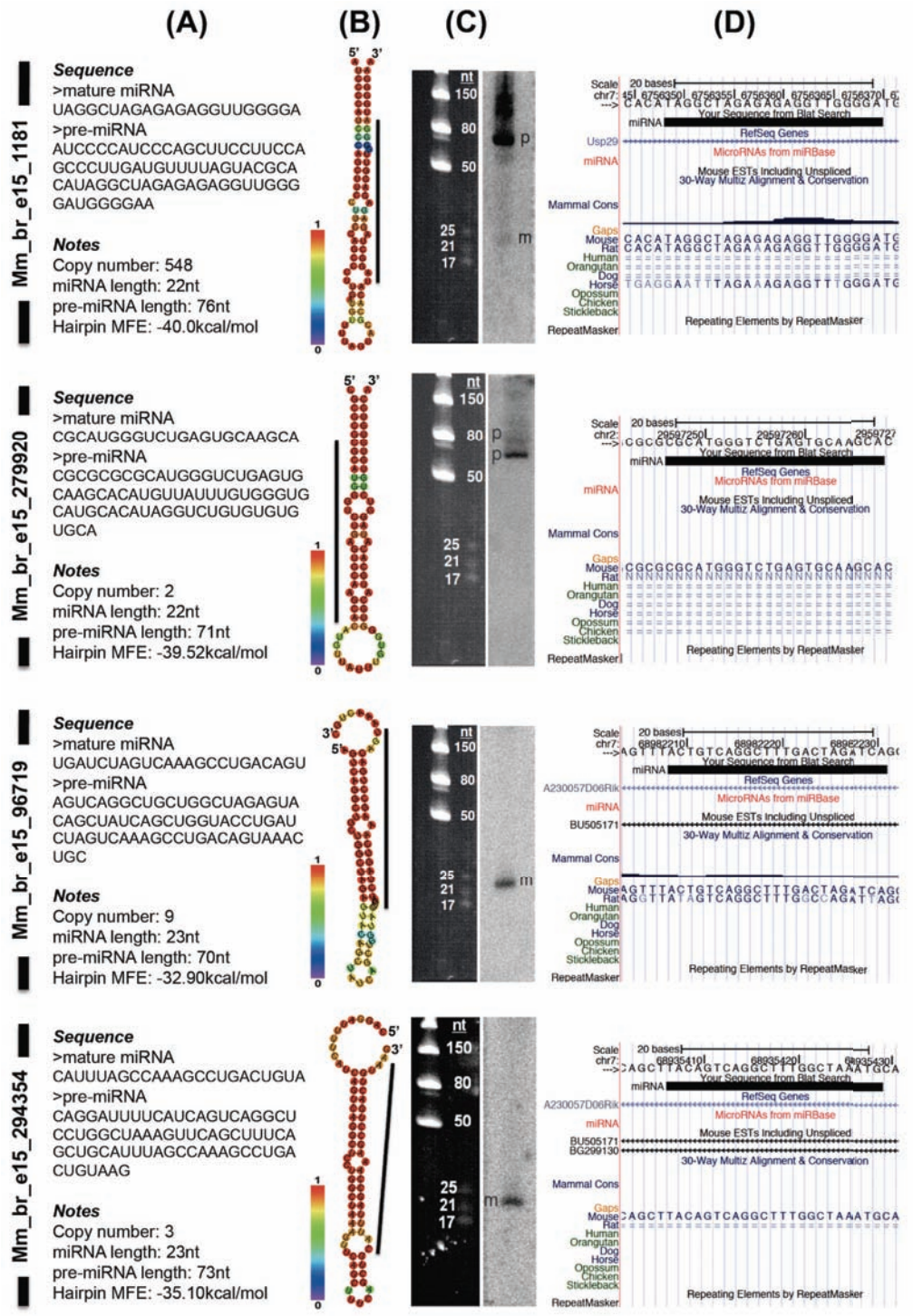


Figure 2 Validated putative miRNAs. (A) Sequences for both mature miRNA and predicted pre-miRNA. Copy number refers to the occurrences of the mature sequences in the E15.5 whole brain small RNA sequencing analysis. (B) RNAfold prediction of the stemloop hairpin structure. The colours in the vertical bar denote the base-pairing probability between two nucleotides within the structure. The black line located next to the hairpin structure denotes the position of the small RNA within the pre-miRNA. (C) Small RNA northern analysis using radiolabeled oligonucleotide probes. 'p' and 'm' refer to pre-miRNA and mature miRNA, respectively. Four independent small RNA northern blots were used to validate the putative miRNA. After hybridization and washing steps, mm_br_e15_1181 blot was exposed to phosphor screen for 1 day whereas the other 3 blots for mm_br_e15_279920, mm_br_e15_96719 and mm_br_e15_294354 were exposed for 8 days. (D) Mapping of the mature miRNA to the mouse genome and other corresponding features such as RefSeq genes, miRNAs from miRBase, mouse ESTs, mammalian conservation information and repeating elements.

accumulated at the pre-miRNA level due to specific factors such as the activity levels of dicer, argonaute or nuclear export receptors [57-59]. Therefore, we considered the existence of these small RNAs validated when either the mature or precursor miRNA with specific size was detected using the northern analysis.

Further analysis using the University of California, Santa Cruz (UCSC) genome browser [60] showed that mm_br_e15_1181 was mapped to a region within the mouse genome that is homologous to the rat and horse genomes. Other putative miRNAs were mapped either to a region specific to the mouse genome (mm_br_e15_294354) or a region homologous to the rat only (mm_br_e15_279920 and mm_br_e15_96719) (Figure 2D). By using both the full-length and seed sequences of all the 4 putative miRNAs, we performed homology searches against all the known miRNA sequences and were unable to find any orthologous miRNAs, indicating that these putative miRNAs could be specific to the mouse or rat especially mm_br_e15_1181 and mm_br_e15_96719. Sequence conservation of miRNAs is relatively common among vertebrates as well as invertebrates. For example *miR-263* (consisting of *miR-263a* and *miR-263b*) and *miR-183* (consisting of *miR-96, miR-182* and *miR-183*) families are found in many organisms including human, mouse, chicken, zebrafish, frog, worm and fruit fly, with high sequence and expression profile similarity particularly in sensory organs [61,62]. However, lack of sequence homology among miRNAs from different organisms does not negate the possibility of functional conservation among them. For example, both *lin-4* and *let-7* target multiple sequence motifs at the 3' UTR of *Caenorhabditis elegans* hunchback homolog mRNA, *hbl-1*, and regulate its expression in the ventral nerve cord neurones [63]. In addition, different miRNAs with similarity at the seed region may exert the same effect on a same mRNA. *Drosophila* bearded (*Brd*) gene has motifs that are complementary to two different miRNAs, *miR-4* and *miR-79*, which bear the same seed sequence. Both the miRNAs target the motifs based entirely on the seed sequence with little or no base-pairing to the 3' region [64]. Although this phenomenon is rare across different organisms, it proves that functional conservation between non-conserved miRNAs may lie within the seed region alone.

Mm_br_e15_1181 biogenesis is Dicer1-dependent

Of the 4 putative miRNAs, we selected mm_br_e15_1181 for further characterisation due to its high copy number. First, we evaluated mm_br_e15_1181 expression in mouse embryonic stem (mES) cells, with and without Dicer1 enzyme activity using the stemloop RT-qPCR technique (Figure 3A). Mm_br_e15_1181 was expressed in mES cells with Dicer1 activity, however its expression was not detected or was weak in cells lacking Dicer1 activity confirming that mm_br_e15_1181 biogenesis is Dicer1-dependent ($P < 0.01$). The evaluation of Dicer1-dependency using the mES cell model is limited to miRNAs that are expressed in this cell type. It is worth noting that Dicer1-dependency is not a definitive property for defining mm_br_e15_1181 as a novel miRNA because endogenous small siRNAs are also subjected to the same dicing mechanism in the cytoplasm [1]. A recent study reported the Dicer1-independent biogenesis of *miR-451*, in which the catalytic activity of Argonaute2 was responsible for the *pre-mir-451* hairpin cleavage process [65].

In this study, we used a number of validation analyses for mm_br_e15_1181: Dicer1-dependence, pre-miRNA structure prediction and northern analysis to define mm_br_e15_1181 as a novel miRNA. This novel miRNA has been identified as *miR-3099*.

Expression profiling of *miR-3099* throughout embryogenesis

The expression of *miR-3099* in mES cells led us to hypothesize that this miRNA may play a role in early embryogenesis and therefore we characterised its expression profile throughout development. Using stemloop RT-qPCR, we showed that *miR-3099* was expressed in E3.5 blastocysts (Figure 3B). The expression of *miR-3099* reduced (by ~9-fold; $P < 0.001$) as the blastocysts developed into an early stage embryo at day 7 (E7), suggesting that *miR-3099* was either expressed in a spatially restricted manner or generally down-regulated at this stage. To specifically locate the expression of *miR-3099* during embryogenesis, we performed whole mount *in situ* hybridisation on E9.5 embryos (n = 3) and showed that *miR-3099* was expressed throughout the embryo with the exception of the developing heart (Figure 3C). Stronger expression was observed in the telencephalon, somites, branchial arches, and both forelimb and hindlimb buds. Cross sectional analysis of the telencephalon confirmed that *miR-3099* was expressed in the neuroepithelium (Figure 3D). Whole mount ISH analysis on embryos of the same age was performed using *miR-scrambled* LNA probe to serve as the background control (n = 2) (Figure 3E &3F).

To evaluate the expression profile of *miR-3099* in the later stages of embryogenesis, we performed section ISH. Section ISH of the E11.5 whole embryos showed that *miR-3099* was expressed throughout the embryo, especially in the preplate of the telencephalon, somites and hindlimb region (Figure 4). By E13.5, *miR-3099* expression was restricted to the cortical plate of

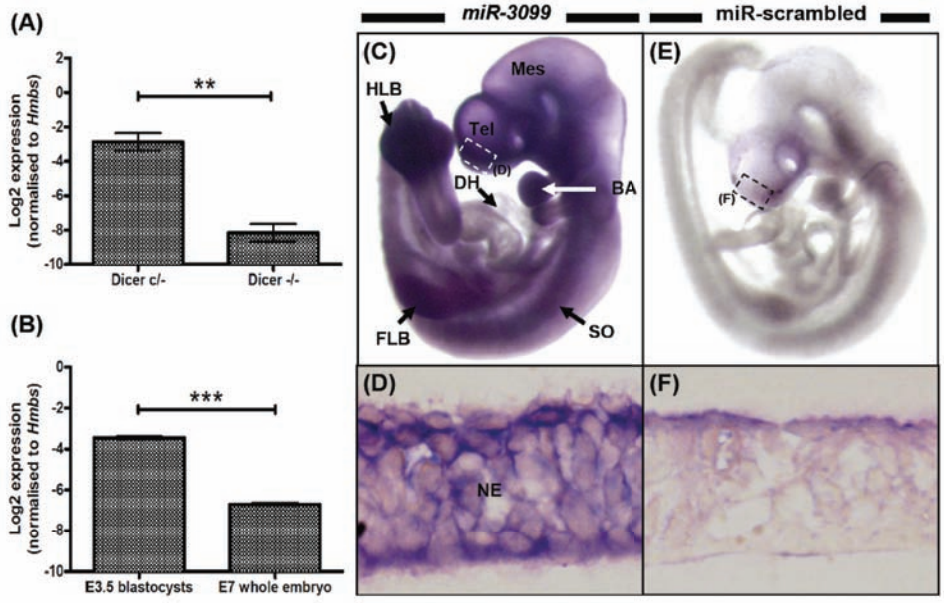


Figure 3 Expression profiling of *miR-3099* novel miRNA in mouse embryonic stem cells (with conditional allele for Dicer), E3.5 blastocysts, E7 and E9.5 embryos. (A) Stemloop RT-qPCR analysis of *miR-3099* novel miRNA in mouse embryonic stem (mES) cells with conditional allele for *Dicer1*. mES cells with and without Dicer1 activity are denoted by Dicer c/- and Dicer -/-, respectively (n = 3 per group). (B) Expression of *miR-3099* in E3.5 blastocysts (n = 14; pooled) and the E7 whole embryo (n = 3). (C-F) Whole mount ISH of E9.5 embryos using DIG-labeled LNA probes for *miR-3099* (C and D) (n = 3) and miR-scrambled (E and F) (n = 2). Cryosection of the stained embryos shows expression of *miR-3099* in the neuroepithelium of the telencephalon (D, inset in C). BA = branchial arches, DH = developing heart, FLB = forelimb bud, HLB = hindlimb bud, Mes = mesencephalon, NE = neuroepithelium, SO = somite, Tel = telencephalon. The mean ± SE for each tissue is presented in the bar graphs. Asterisks denote the statistical significance level at $P < 0.01$ (**) and $P < 0.001$ (***) based on the one-way ANOVA test (see Additional file 14 for analysis details).

the cortical neuroepithelium, striatum, medial pallium (hippocampal allocortex) and subventricular/ventricular zone of the superior and inferior colliculi. In E15.5 embryos, *miR-3099* expression was observed primarily in the cortical plate of the cerebral cortex. In E17.5 whole brains, *miR-3099* expression was prominent in the cortical plate, piriform cortex and at lower levels, in the hippocampal formation. Embryo-wide expression of *miR-3099* during early embryogenesis suggests a pan-regulatory role, possibly functioning as a 'housekeeping' miRNA in basic cellular processes. This feature has been described in a few clusters of miRNAs expressed in the mouse retina, brain and heart [66]. Many miRNAs have ubiquitous expression patterns and their function remains unclear as they may have roles in subtle miRNA networks, which exert combinatorial effects during development [67,68]. Contrasting with the almost ubiquitous expression profile in early development, *miR-3099* was not detected in a few regions such as the E9.5 developing heart and the ventricular zone of the telencephalon/developing cerebrum. This suggests that the function of *miR-3099* may be tissue or cell-specific, especially after E11.5, this warrants further characterisation.

We also performed stemloop RT-qPCR expression analysis of *miR-3099* in various regions of the mouse brain and organs. Using the mouse whole brain, there was a significant difference ($P = 0.02$) in the *miR-3099* expression among E11.5, E13.5, E15.5, E17.5, postnatal day (P) 1.5 and P150 samples (Figure 5A). *MiR-3099* expression was found to be increased after E11.5 and was maintained in postnatal day 1.5 (P1.5) and P150 whole brains. The qPCR analysis supports the previous section ISH analysis. No significant differences ($P = 0.45$) in *miR-3099* expression were observed among cerebellum, cerebrum, hippocampus, medulla, olfactory bulb and thalamus (Figure 5B). When we compared the expression of *miR-3099* in various adult mouse organs to the P150 whole brain, we found significant differences in the expression levels among the organs ($P < 0.001$) (Figure 5C). *MiR-3099* was found to be expressed at the highest level in the pancreas, followed by the thymus, large intestine, heart, small intestine, kidney, brain, testis, ovary, skin, skeletal muscle, liver, stomach and spleen. Similar to the embryonic expression profiles, the diverse expression profile of *miR-3099* in multiple organs of the adult mouse further supports a widespread role in the development and function of these organs.

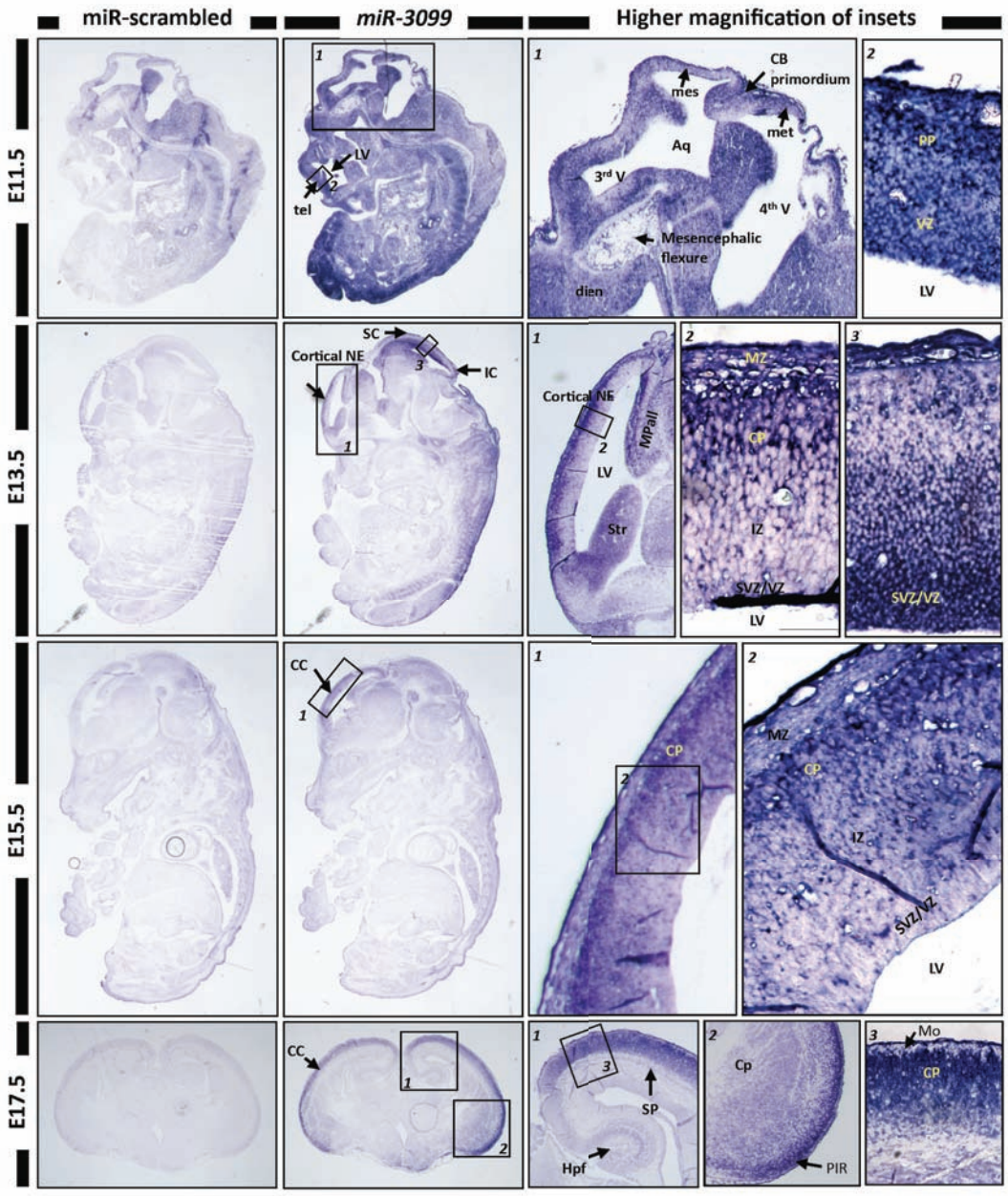


Figure 4 Expression profiling of *miR-3099* novel miRNA in E11.5-E15.5 whole embryos and the E17.5 whole brain. *In situ* hybridisation analysis using LNA probes for miR-scrambled and *miR-3099* was performed on E11.5-E15.5 developing embryos and E17.5 whole brain paraffin sections. Strong expression of *miR-3099* was detected in the E11.5 embryo. From E13.5 onwards, the expression was retained only in the neuroepithelium (NE) or cerebral cortex (CC). Under high magnification, *miR-3099* was found to express specifically in the preplate (PP) of telencephalon (tel) (E11.5), cortical plate (CP) of the CC (E13.5-E17.5) and the germinal layer of mesencephalon (mes) (E11.5-E13.5). Aq = aqueduct, CB = cerebellum, Cp = caudo-putamen, dien = diencephalon, Hpf = hippocampal formation, IC = inferior colliculus, IZ = intermediate zone, LV = lateral ventricle, met = metencephalon, Mo = molecular layer, MPall = medial pallium (hippocampal allocortex), MZ = marginal zone, PIR = piriform cortex, SC = superior colliculus, SP = subplate, Str = striatum, SVZ = subventricular zone, V = ventricle.

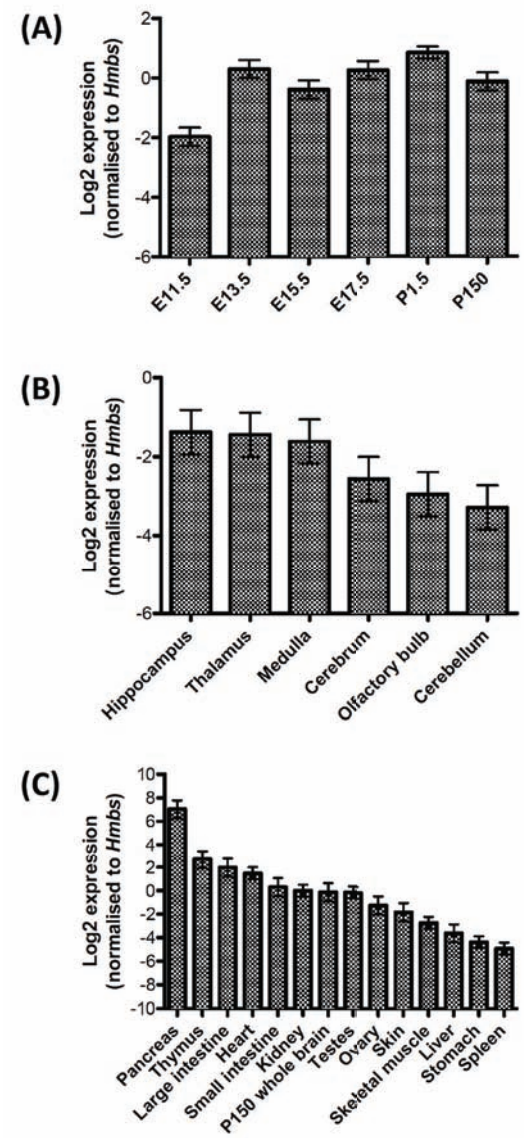


Figure 5 Expression profiling of *miR-3099* novel miRNA in the whole brain of different developmental stages, different adult mouse brain regions and organs. Stemloop RT-qPCR analysis of *miR-3099* in E11.5-P150 whole brain (A), brain regions in P150 whole brain (B) (n = 2 for each group) and various mouse organs harvested from P150 adult mouse (C) (n = 2 for all except P150 whole brain, skeletal muscle, spleen, stomach and testes, where n = 3). The mean ± SE for each organ is presented in the bar graphs. The one-way ANOVA test is significant at $P < 0.05$ for (A), not significant for (B) and $P < 0.001$ for (C) (see Additional file 14 for analysis details).

Expression of *miR-3099* is upregulated in differentiating neuronal/glial cells

Expression of *miR-3099* was observed in the preplate of the E11.5 telencephalon and later in the cortical plate of the E13.5-E17.5 cerebral cortex, by which time the majority of the cells in these structures are committed to their respective neuronal lineages. This finding further suggests that *miR-3099* may play an important regulatory role during neurogenesis or in neuronal function. To further test this idea, we used P19 teratocarcinoma cells as an *in vitro* model. Upon retinoic acid induction and under reduced serum concentration, P19 cells differentiate into glutamatergic and glutamate-responsive neurones, glial and fibroblast-like cells [69-72]. We analysed the expression level of *miR-3099* in P19 cells (Figure 6A) and found a statistically significant ($P = 0.04$) ~2-fold upregulation of *miR-3099* in ~50% differentiated P19 cells compared to the proliferating cells (Figure 6B). Various miRNAs have been found to be upregulated during neural differentiation and some of their expression could be negatively regulated by important transcription factors such as Oct4 and Sox2, the expression levels of which gradually diminish as cells differentiate into neurones [73]. Therefore, increased *miR-3099* expression during P19 differentiation raises the possibility that this miRNA may

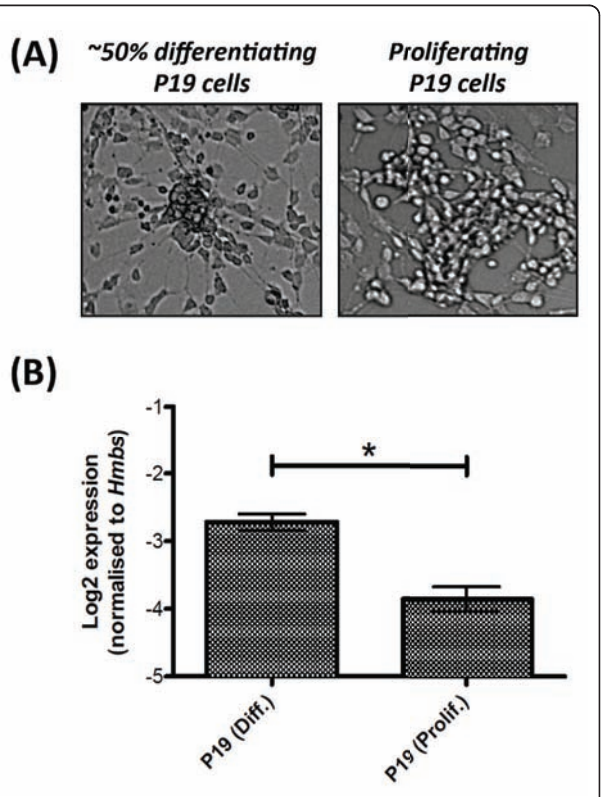


Figure 6 Expression of *miR-3099* in P19 teratocarcinoma cells. (A) Phase contrast micrographs of differentiating and proliferating P19 cells. (B) Stemloop RT-qPCR analysis of *miR-3099* expression in differentiating (diff.) (n = 3) and proliferating (prolif.) (n = 2) P19 cells. The mean ± SE for each cell type is presented in the bar graph. The asterisk (*) denotes statistical significance at $P < 0.05$ based on the one-way ANOVA test (see Additional file 14 for analysis details).

have a functional role during neural differentiation or neuronal cell function.

Conclusions

In this study, we have reported the first deep sequencing analysis of small RNAs of a developing mouse brain. We have identified and validated 4 putative miRNAs from the analysis and further characterised one of them, *miR-3099*, during embryogenesis. A significant finding of the study was the embryo-wide expression profile of *miR-3099* in mid-gestation embryos, which became restricted to the central nervous system, suggesting a role for this miRNA in neural differentiation or function.

Methods

Animals and dissections

The Melbourne Health Animal Ethics Committee and the University of Adelaide Animal Ethics Committee approved procedures involved in the breeding and handling of animals. Mice were housed under a 12-hour light and 12-hour dark cycle with access to unlimited food and water. Mice were culled by CO_2 inhalation and all dissections of mouse embryos, brains and organs were carried out according to the methods described previously [18].

Deep sequencing and analysis

Total RNA was isolated from a whole brain dissected from an E15.5 embryo of C57BL/6 background using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Small RNAs with sizes ranging from 16-30nt were isolated from 10 μg total RNA using polyacrylamide gel electrophoresis. The complementary small RNA library was constructed using the Small RNA Sample Prep Kit version 1.0 (Illumina) according to the manufacturer's protocol with 5'-GTTCAGAGTT CTACAGTCCG ACGATC-3' and 5'-TCGTATGCCG TCTTCTGCTT GT-3' adapters at the 5' and 3' ends, respectively. Sequencing was carried out using a Genome Analyzer II (Illumina). Image data was generated by the Genome Analyzer II and was processed using the Illumina pipeline software (Pipeline version 1.0 was used for the FASTQ data). This consists of an image analysis module (Firecrest), followed by basecalling using the BUSTARD module and finally production of a data file in FASTQ format using the GERALD module.

Sequence annotation pipeline

The FASTQ data was ranked according to decreasing abundance of the unique tags. This file was created using a PERL script in Linux without taking into consideration any filters (adapter sequences) or quality. A file with unique tags and their corresponding counts was generated. All unique tags (including those with a single count) were mapped to the NCBI Mouse Assembly Build 37.1 using the Bowtie program [74]. Two sets of alignments were carried out: one stripping off 14 bases from the 5' end of unique tags and the other stripping off 14 bases from the 3' end. In both alignments, no mismatches are allowed and unique tags that hit more than one locus within the mouse genome were discarded. Unique tags with a single hit within the genome were further annotated using various databases such as RepeatMasker (analysis was performed on NCBI Mouse Assembly build 37.1 and the output was downloaded from UCSC genome browser on the 28th of November, 2008), mouse RefSeq in release 32, mouse miRNA in miRBase release 12.0 and redundant mouse EST database (downloaded from UCSC mm9 on 27th January, 2009).

Identification of candidate novel miRNAs

Unique tags that mapped to a genomic locus with a RefSeq, redundant EST or no annotations were subjected to pre-miRNA prediction using the RNA22 program [55]. Sequences encompassing 100- to 200-nt upstream and downstream of these unique sequences were used to predict any potential pre-miRNAs with hairpin structures. The minimum number of patterns that should support a pre-miRNA before it can get reported was set to 60, and the minimum and maximum pre-miRNA lengths were set to 60nt and 150nt, respectively. All predicted pre-miRNA sequences based on these settings were used to determine the hairpin fold structure using RNAfold program [56]. The predicted hairpin fold structure with the lowest minimum free energy (MFE) (cut off at -30 kcal/mol or lower) and conforming to the annotation criteria for pre-miRNA [2] was selected as the final predicted pre-miRNA. Briefly, the predicted precursor structure must be between 60-80 nt in size and must not have a large internal loop or any asymmetric bulges. The predicted pre-miRNA must contain the aligned unique sequence within one arm of the hairpin and include at least 16 bp from the 5' end of the unique sequence and the other arm of the hairpin.

Small RNA northern analysis

Eight blots were prepared from four independent E15.5 whole brains. Approximately 30 μg of total RNA was denatured in 1X Ambion Gel Loading Buffer II (Ambion®) at 85°C for 3 minutes. RNAs were electrophoresed in 15% acrylamide/urea gels (48% (w/v) urea, 15% (v/v) acrylamide, 0.05% (w/v) ammonium persulfate and 0.1% (v/v) tetramethylethylenediamine prepared in 1X TBE) in 1X TBE buffer at 300 V for 90 minutes. Separated small RNAs in the gel were then transferred onto Hybond-N+ nylon membrane (GE Healthcare)

using Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad) at a constant 0.4 V for 45 minutes. The pre-hybridisation step was carried out in Amersham Rapid-hyb™ Buffer (GE Healthcare) with 100 μg/ml of herring sperm DNA (Promega) at 42°C for 1 hour and was followed by the hybridisation step. The same pre-hybridisation solution was used for hybridisation with addition of 2×10^6 dpm/ml labelled probe prepared using 20 U of T4 Polynucleotide Kinase (Promega) in 1X kinase buffer (Promega) and 50 pmol of [γ-^{32}P]-dATP (GE Healthcare) (3000 Ci/mmol). Hybridisation was carried out for 18 hours and filters were washed in 5 × SSC with 0.1% (w/v) sodium dodecyl sulfate (SDS) (20 minutes at 37°C) followed by 1 × SSC with 0.1% (w/v) SDS and 0.2 × SSC with 0.1% (w/v) SDS (15 minutes each time at 65°C until a clean background signal was obtained). The membrane was exposed to a storage phosphor screen in a cassette at room temperature for 1 day for *miR-3099* blot and 8 days for other blots before scanned using Typhoon™ 9400 (GE Healthcare).

Stemloop RT-qPCR

Reverse transcription of the small RNA was performed based on modified methods [75,76]. cDNA was synthesised from 150 ng-2.5 μg of small RNA enriched total RNA using 0.05 μM of an in-house designed stem loop primer (5'-GTTGGCTCT GGTAGGATG CCGCTC TCA GGGCATCCT ACCAGAGCCA AACTCCCCA-3', GeneWorks), and the Superscript® III Reverse Transcriptase Kit (Invitrogen) with modifications to the manufacturer's protocol. The stem loop primer was added after a denaturation step at 65°C for 5 minutes. The last 6nt at the 3' end of the stem loop primer complements the last 6nt of the 3' end of *miR-3099* small RNA. The stem loop RT primer contains a target site for a universal reverse primer (5'-GTAGGATGCC GCTCTCAGG-3', GeneWorks) and a target site for UniversalProbe Library (UPL) Probe #21 (Roche Diagnostics), which were used in subsequent cDNA amplification processes together with a specific forward primer for *miR-3099* (5'-CGCGTAGGCT AGAGAGAGGT-3', GeneWorks). Briefly, cDNA synthesis was performed at 16°C for 30 minutes followed by 60 cycles of 20°C for 30 seconds, 42°C for 30 seconds and 50°C for 1 second. A final incubation at 75°C for 15 minutes was performed to inactivate the reverse transcriptase enzyme.

Prior to qPCR, pre-PCR of *miR-3099* was performed in a 10 μl reaction volume containing 1X LC480 Probe Master mix (Roche Diagnostics), 50 nM of each forward and universal reverse primers and 0.2X of synthesised cDNA. Pre-PCR was initially carried out at 95°C for 10 minutes, 55°C for 2 minutes and 75°C for 2 minutes and followed by 14 additional cycles of 95°C for 15 seconds and 60°C for 4 minutes. After pre-PCR, 0.01X of amplicons were used for qPCR.

QPCR was carried out in 10 μl reaction volume using 1X LightCycler 480 (LC480) Probe Master mix (Roche Diagnostics), 0.1 μM of a relevant Universal ProbeLibrary probe (Roche Diagnostics), 0.25 μM of each forward and reverse primers and 1 μl of 0.1X of synthesised cDNA. Reactions were prepared in 384-well plates and RT-qPCR was performed using a LightCycler® 480 Real Time PCR System instrument (Roche Diagnostics). QPCR was performed with an initial denaturation at 95°C for 10 minutes followed by 45 cycles at 95°C for 10 seconds, 60°C for 30 seconds and 72°C for 10 seconds, and a final step at 40°C for 1 second.

Real-Time amplification signals were acquired during the elongation step and recorded live using LightCycler® 480 Software version 1.5 (Roche Diagnostics). The cycle threshold or crossing point (Cp) from each signal was calculated based on the Second Derivative Maximum method [77]. A 4-data point standard curve was constructed using serially diluted pooled cDNAs for each primer set used in qPCR in each run. The standard curve was used to determine the PCR efficiency and reproducibility of each PCR system. The *Hmbs* gene was used as reference gene normalisation according to the method as described [18].

Statistical analysis

Two or three independent biological replicates were used for each tissue/organ in each experiment. Two qPCR experiments were performed on the tissue of each biological replicate. The qPCR results were normalized to *Hmbs*, and those that were not outliers, $\log_2$ transformed and then averaged to give the expression data for the biological replicate. One-way ANOVA was used to compare the expression levels among the tissues. A P value of <0.05 was considered statistically significant. Where significant differences were detected among the tissues the least significant difference(s) (LSD) were provided with the analysis (see Additional file 14 for analysis details).

Locked Nucleic Acids - *In situ* hybridisation

Paraffin embedded sections (8 μm) were used for LNA-ISH. Sections were de-paraffinised with washes in xylene (3× for 5 minutes each) and hydrated in a series of ethanol concentrations into RNase-free water. Subsequently, sections were fixed in 4% (w/v) PFA (pH7.0) in 1X PBS (10 minutes) followed by Proteinase K digestion (6.7 μg/ml of Proteinase K, 50 mM of Tris HCl pH7.5, 5 mM of EDTA) for 30 minutes, re-fixed in 4% (w/v) PFA in 1X PBS for 5 minutes and acetylated (0.1 M of triethanolamine, 0.178% (v/v) of concentrated HCl and 0.25% (v/v)

of acetic anhydride) for 10 minutes. Between each step, sections were washed multiple times using 1X PBS.

The pre-hybridisation step was carried out in a humidified chamber (50% (v/v) formamide, 5X sodium chloride/sodium citrate, SSC) at 60°C. Amersham Rapidhyb™ Buffer (GE Healthcare) was used for pre-hybridisation with additional *Escherichia coli* tRNA (Sigma Aldrich) and Herring Sperm DNA (Promega) to a final concentration of 100 μg/ml each. After 1-2 hours of pre-hybridisation, custom-made *Sox4_sir3* LNA probes (Cat. no: EQ-70537, Exiqon) were added to the buffer to give a concentration of 0.020 pmol/μl. Hybridisation was carried out in the oven for 16-20 hours.

After the hybridisation step, sections were washed in 5 × SSC (20 minutes at hybridisation temperature) followed by 0.2 × SSC (3 hours at hybridisation temperature). Sections were then rinsed in fresh 0.2 × SSC for 5 minutes and in pre-blocking buffer (0.1 M of Tris HCl pH7.5, 0.15 M of NaCl and 240 μg/ml of levamisole) for a further 5 minutes. In a humidified chamber, sections were blocked in 20% (v/v) foetal calf serum (Sigma Aldrich) and 2% (w/v) blocking powder (Roche Diagnostics) in maleate buffer for 1 hour. After blocking, sections were incubated with 0.0002X (0.00015 U) anti-DIG antibody with alkaline phosphatase, Fab fragments (Roche Diagnostics) in blocking buffer for 16 hours in the dark. Subsequently, sections were washed in NTMT buffer (3× for 10 minutes each: 0.1 M Tris HCl pH9.5, 0.1 M NaCl, 0.05 M $MgCl_2$, 1% (v/v) Tween-20 and 240 μg/ml levamisole) and then with nitro blue tetrazolium chloride (NBT)/5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP) colour reaction (0.375 mg/ml of NBT and 0.188 mg/ml of BCIP in NTMT buffer) for 3 hours to 5 days. After the colour reaction step, sections were washed with Tris EDTA buffer pH8.0 (0.01 M of Tris HCl pH7.5 and 0.001 M EDTA pH8.0) for 10 minutes and were mounted in Entellan® media (ProSciTech).

P19 teratocarcinoma cells

Propagation and differentiation of P19 cells were carried out according to protocols previously described [18,78].

Mouse embryonic stem (mES) cells with Dicer1^{c} conditional allele

Mouse embryonic stem (mES) cells with Dicer1 activity were of a line heterozygous for a conditionally mutant *Dicer1* allele (*Dicer1*^{c}) and a null *Dicer1* allele (*Dicer1*$^{-}$), these genetic modifications have been previously described [79]. mES cells without Dicer1 activity were produced by transient transfection of this *Dicer1*$^{c/-}$ line with Cre recombinase to produce *Dicer1*$^{-/-}$ subclones (JRM and DMM, unpublished data). The mES cells were propagated as previously described [80].

Mouse E3.5 blastocysts

C57BL/6 females of 3-4 weeks of age were superovulated using 5IU of Folligon (PMSG) followed by 5IU of Chorulon (HCG) 47.5 hours later and mated with B6D2F1 entire stud males. Microdrop culture dishes were set up to equilibrate in 37°C, 5% CO_2 incubator 4 hours prior to culture. KSOM (Millipore) media was used in 20 μl droplets in a 35 mm dish, overlaid with Embryo Tested Mineral Oil (Sigma). Superovulated female mice were sacrificed after 2.5 days of superovulation induction and mating, and oviducts were collected into M2 handling media (Millipore). Oviducts were flushed using M2 media, a blunt 30G needle and a1ml syringe. Morulae were collected and cultured in pre-equilibrated KSOM. Blastocysts were collected from culture a day later under a dissecting microscope. These were considered E3.5 blastocysts.

Additional material

Additional file 1: List of 413,494 unique tags.

Additional file 2: List of unique tags mapped to repetitive elements and ncRNAs based on the 3′ end sequences.

Additional file 3: List of unique tags mapped to repetitive elements and ncRNAs based on the 5′ end sequences.

Additional file 4: List of unique tags mapped to miRNAs, pre-miRNAs or miRNA-star based on 3′ end sequences.

Additional file 5: List of unique tags mapped to RefSeq sequences based on 3′ end sequences.

Additional file 6: List of unique tags mapped to RefSeq sequences based on 5′ end sequences.

Additional file 7: List of unique tags mapped to redundant mouse ESTs based on 3′ end sequences.

Additional file 8: List of unique tags mapped to redundant mouse ESTs based on 5′ end sequences.

Additional file 9: List of unique tags mapped to a single locus in the mouse genome based on 3′ end sequences.

Additional file 10: List of unique tags mapped to a single locus in the mouse genome based on 5′ end sequences.

Additional file 11: List of unique tags mapped to multiple loci in the mouse genome based on 3′ end sequences.

Additional file 12: List of unique tags mapped to multiple loci in the mouse genome based on 5′ end sequences.

Additional file 13: List of known miRNAs in the E15.5 developing mouse brain.

Additional file 14: Supplementary figures and data for statistical analysis.

Acknowledgements

This work was supported by National Health and Medical Research Council fellowships (171601 and 461204 to H.S.S); National Health and Medical Research Council Grants 219176, 257501 and 257529 (to H.S.S.) and a fellowship from Pfizer Australia (to P.Q.T.). K.-H.L. was a recipient of the Melbourne International Fee Remission Scholarship (MIFRS) and Universiti Putra Malaysia Staff Training Scholarship (UPMSTS), and an Adelaide Fees Scholarship International equivalent. We thank GeneWorks Pty Ltd for constructing the small RNA (cDNA) library and performing the high-throughput sequencing.

Author details
[1]Department of Molecular Pathology, SA Pathology and Centre for Cancer Biology, P.O. Box 14 Rundle Mall Post Office, Adelaide, SA 5000, Australia. [2]School of Medicine, Faculty of Health Sciences, University of Adelaide, Adelaide, SA 5005, Australia. [3]Department of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia. [4]School of Molecular and Biomedical Science, Faculty of Sciences, University of Adelaide, Adelaide, SA 5005, Australia. [5]Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor DE, Malaysia. [6]eResearchSA, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia. [7]Theme of Laboratory and Community Genetics, Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, VIC 3052, Australia.

Authors' contributions
KHL, JRR, JMR and DLA participated in the small RNA sequence analysis and annotations. KHL, PJB and CNH supervised, designed and carried out the qPCR analysis. KHL and TD performed the small RNA northern analysis. PQT supervised whereas KHL and PSC performed the LNA-ISH analysis. SP performed the procedures for superovulation and mating of mice, cultured the morulae and provided the blastocysts for the expression analysis. DMM and JRM cultured and provided the mES cells with Dicer1 conditional allele. KHL drafted the manuscript. CNH, PQT, DLA and HSS conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

Received: 17 September 2010 Accepted: 5 April 2011
Published: 5 April 2011

References

1. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116**:281-297.
2. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, *et al*: **A uniform system for microRNA annotation.** *RNA* 2003, **9**:277-279.
3. Cullen BR: **Transcription and processing of human microRNA precursors.** *Mol Cell* 2004, **16**:861-865.
4. Hobert O: **Gene regulation by transcription factors and microRNAs.** *Science* 2008, **319**:1785-1786.
5. Alvarez-Garcia I, Miska EA: **MicroRNA functions in animal development and human disease.** *Development* 2005, **132**:4653-4662.
6. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: **The Microprocessor complex mediates the genesis of microRNAs.** *Nature* 2004, **432**:235-240.
7. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ: **Processing of primary microRNAs by the Microprocessor complex.** *Nature* 2004, **432**:231-235.
8. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN: **The Drosha-DGCR8 complex in primary microRNA processing.** *Genes Dev* 2004, **18**:3016-3027.
9. Kim VN: **MicroRNA biogenesis: coordinated cropping and dicing.** *Nat Rev Mol Cell Biol* 2005, **6**:376-385.
10. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U: **Nuclear export of microRNA precursors.** *Science* 2004, **303**:95-98.
11. Lund E, Dahlberg JE: **Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs.** *Cold Spring Harb Symp Quant Biol* 2006, **71**:59-66.
12. Guo L, Lu Z: **The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule?** *PLoS One* **5**:e11387.
13. Zhou X, Duan X, Qian J, Li F: **Abundant conserved microRNA target sites in the 5'-untranslated region and coding sequence.** *Genetica* 2009, **137**:159-164.
14. Lytle JR, Yario TA, Steitz JA: **Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR.** *Proc Natl Acad Sci USA* 2007, **104**:9667-9672.
15. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I: **MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation.** *Nature* 2008, **455**:1124-1128.
16. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R: **Control of translation and mRNA degradation by miRNAs and siRNAs.** *Genes Dev* 2006, **20**:515-524.
17. Baulcombe D: **RNA silencing in plants.** *Nature* 2004, **431**:356-363.
18. Ling KH, Hewitt CA, Beissbarth T, Hyde L, Banerjee K, Cheah PS, Cannon PZ, Hahn CN, Thomas PQ, Smyth GK, *et al*: **Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling.** *Genome Biol* 2009, **10**:R104.
19. Gunnersen JM, Augustine C, Spirkoska V, Kim M, Brown M, Tan SS: **Global analysis of gene expression patterns in developing mouse neocortex using serial analysis of gene expression.** *Mol Cell Neurosci* 2002, **19**:560-573.
20. Gupta A, Tsai LH, Wynshaw-Boris A: **Life is a journey: a genetic look at neocortical development.** *Nat Rev Genet* 2002, **3**:342-355.
21. Smart IH: **Histogenesis of the mesocortical area of the mouse telencephalon.** *J Anat* 1984, **138(Pt 3)**:537-552.
22. Uylings HBM, Van Eden CG, Parnavelas JG, Kalsbeek A: **The prenatal and postnatal development of rat cerebral cortex.** In *The Cerebral Cortex of the Rat*. Edited by: Kolb B, Tees RC. Cambridge, Mass.: MIT Press; 1990:36-76.
23. Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME: **A brain-specific microRNA regulates dendritic spine development.** *Nature* 2006, **439**:283-289.
24. Leucht C, Stigloher C, Wizenmann A, Klafke R, Folchert A, Bally-Cuif L: **MicroRNA-9 directs late organizer activity of the midbrain-hindbrain boundary.** *Nat Neurosci* 2008, **11**:641-648.
25. Makeyev EV, Zhang J, Carrasco MA, Maniatis T: **The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing.** *Mol Cell* 2007, **27**:435-448.
26. Hansen T, Olsen L, Lindow M, Jakobsen KD, Ullum H, Jonsson E, Andreassen OA, Djurovic S, Melle I, Agartz I, *et al*: **Brain expressed microRNAs implicated in schizophrenia etiology.** *PLoS ONE* 2007, **2**:e873.
27. Marti E, Pantano L, Banez-Coronel M, Llorens F, Minones-Moyano E, Porta S, Sumoy L, Ferrer I, Estivill X: **A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing.** *Nucleic Acids Res* 2010, **38(20)**:7219-35.
28. **miRBase.** [http://microrna.sanger.ac.uk/].
29. Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M, Woychik R: **Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms.** *Brief Funct Genomic Proteomic* 2002, **1**:95-104.
30. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, *et al*: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays.** *Nat Biotechnol* 2000, **18**:630-634.
31. Ruan Y, Ooi HS, Choo SW, Chiu KP, Zhao XD, Srinivasan KG, Yao F, Choo CY, Liu J, Ariyaratne P, *et al*: **Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs).** *Genome Res* 2007, **17**:828-838.
32. Lee A, Hansen KD, Bullard J, Dudoit S, Sherlock G: **Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species.** *PLoS Genet* 2008, **4**:e1000299.
33. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, *et al*: **Mammalian microRNAs: experimental evaluation of novel and previously annotated genes.** *Genes Dev* 2010, **24**:992-1009.
34. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, *et al*: **CpG islands influence chromatin structure via the CpG-binding protein Cfp1.** *Nature* 2010, **464**:1082-1086.
35. Linsen SE, de Wit E, de Bruijn E, Cuppen E: **Small RNA expression and strain specificity in the rat.** *BMC Genomics* 2010, **11**:249.
36. Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE, *et al*: **Limitations and possibilities of small RNA digital gene expression profiling.** *Nat Methods* 2009, **6**:474-476.
37. Uziel T, Karginov FV, Xie S, Parker JS, Wang YD, Gajjar A, He L, Ellison D, Gilbertson RJ, Hannon G, Roussel MF: **The miR-17~92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma.** *Proc Natl Acad Sci USA* 2009, **106**:2812-2817.

38. Shao NY, Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Li N, Chen W, Khaitovich P: **Comprehensive survey of human brain microRNA by deep sequencing.** *BMC Genomics* 2010, **11**:409.
39. **National Centre for Biotechnology Information (NCBI): Gene Expression Omnibus.** [http://www.ncbi.nlm.nih.gov/geo/].
40. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, *et al*: **Tiny RNAs associated with transcription start sites in animals.** *Nat Genet* 2009, **41**:572-578.
41. Taft RJ, Simons C, Nahkuri S, Oey H, Korbie DJ, Mercer TR, Holst J, Ritchie W, Wong JJ, Rasko JE, *et al*: **Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans.** *Nat Struct Mol Biol* 2010, **17**:1030-1034.
42. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, *et al*: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells.** *Genome Res* 2008, **18**:610-621.
43. Zhang J, Xu Y, Huan Q, Chong K: **Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response.** *BMC Genomics* 2009, **10**:449.
44. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T: **Identification of tissue-specific microRNAs from mouse.** *Curr Biol* 2002, **12**:735-739.
45. Miska EA, Alvarez-Saavedra E, Townsend M, Yoshii A, Sestan N, Rakic P, Constantine-Paton M, Horvitz HR: **Microarray analysis of microRNA expression in the developing mammalian brain.** *Genome Biol* 2004, **5**:R68.
46. Wulczyn FG, Smirnova L, Rybak A, Brandt C, Kwidzinski E, Ninnemann O, Strehle M, Seiler A, Schumacher S, Nitsch R: **Post-transcriptional regulation of the let-7 microRNA during neural cell specification.** *Faseb J* 2007, **21**:415-426.
47. Smirnova L, Grafe A, Seiler A, Schumacher S, Nitsch R, Wulczyn FG: **Regulation of miRNA expression during neural cell specification.** *Eur J Neurosci* 2005, **21**:1469-1477.
48. Ciafre SA, Galardi S, Mangiola A, Ferracin M, Liu CG, Sabatino G, Negrini M, Maira G, Croce CM, Farace MG: **Extensive modulation of a set of microRNAs in primary glioblastoma.** *Biochem Biophys Res Commun* 2005, **334**:1351-1358.
49. Evangelisti C, Florian MC, Massimi I, Dominici C, Giannini G, Galardi S, Bue MC, Massalini S, McDowell HP, Messi E, *et al*: **MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness.** *FASEB J* 2009, **23**:4276-4287.
50. Lukiw WJ: **Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus.** *Neuroreport* 2007, **18**:297-300.
51. Kuhn DE, Nuovo GJ, Martin MM, Malana GE, Pleister AP, Jiang J, Schmittgen TD, Terry AV Jr, Gardiner K, Head E, *et al*: **Human chromosome 21-derived miRNAs are overexpressed in down syndrome brains and hearts.** *Biochem Biophys Res Commun* 2008, **370**:473-477.
52. Wilfred BR, Wang WX, Nelson PT: **Energizing miRNA research: a review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways.** *Mol Genet Metab* 2007, **91**:209-217.
53. Chen CZ, Li L, Lodish HF, Bartel DP: **MicroRNAs modulate hematopoietic lineage differentiation.** *Science* 2004, **303**:83-86.
54. Li JY, Yong TY, Michael MZ, Gleadle JM: **Review: The role of microRNAs in kidney disease.** *Nephrology (Carlton)* 2010, **15**:599-608.
55. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I: **A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes.** *Cell* 2006, **126**:1203-1217.
56. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**:W70-74.
57. O'Carroll D, Mecklenbrauker I, Das PP, Santana A, Koenig U, Enright AJ, Miska EA, Tarakhovsky A: **A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway.** *Genes Dev* 2007, **21**:1999-2004.
58. Bussing I, Yang JS, Lai EC, Grosshans H: **The nuclear export receptor XPO-1 supports primary miRNA processing in C. elegans and Drosophila.** *EMBO J* 2010, **29**:1830-1839.
59. Saito K, Ishizuka A, Siomi H, Siomi MC: **Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells.** *PLoS Biol* 2005, **3**: e235.
60. **UCSC Genome Bioinformatics.** [http://genome.ucsc.edu/].
61. Pierce ML, Weston MD, Fritzsch B, Gabel HW, Ruvkun G, Soukup GA: **MicroRNA-183 family conservation and ciliated neurosensory organ expression.** *Evol Dev* 2008, **10**:106-113.
62. Hilgers V, Bushati N, Cohen SM: **Drosophila microRNAs 263a/b confer robustness during development by protecting nascent sense organs from apoptosis.** *PLoS Biol* 2010, **8**:e1000396.
63. Lin SY, Johnson SM, Abraham M, Vella MC, Pasquinelli A, Gamberi C, Gottlieb E, Slack FJ: **The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target.** *Dev Cell* 2003, **4**:639-650.
64. Lai EC: **Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation.** *Nat Genet* 2002, **30**:363-364.
65. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, *et al*: **A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity.** *Science* 2010, **328**:1694-1698.
66. Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D: **MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster.** *J Biol Chem* 2007, **282**:25053-25066.
67. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, *et al*: **A mammalian microRNA expression atlas based on small RNA library sequencing.** *Cell* 2007, **129**:1401-1414.
68. Lim LP, Linsley PS: **Mustering the micromanagers.** *Nat Biotechnol* 2007, **25**:996-997.
69. McBurney MW, Reuhl KR, Ally AI, Nasipuri S, Bell JC, Craig J: **Differentiation and maturation of embryonal carcinoma-derived neurons in cell culture.** *J Neurosci* 1988, **8**:1063-1073.
70. Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI: **Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells.** *J Cell Biol* 1982, **94**:253-262.
71. Jones-Villeneuve EM, Rudnicki MA, Harris JF, McBurney MW: **Retinoic acid-induced neural differentiation of embryonal carcinoma cells.** *Mol Cell Biol* 1983, **3**:2271-2279.
72. MacPherson PA, Jones S, Pawson PA, Marshall KC, McBurney MW: **P19 cells differentiate into glutamatergic and glutamate-responsive neurons in vitro.** *Neuroscience* 1997, **80**:487-499.
73. Huang B, Li W, Zhao B, Xia C, Liang R, Ruan K, Jing N, Jin Y: **MicroRNA expression profiling during neural differentiation of mouse embryonic carcinoma P19 cells.** *Acta Biochim Biophys Sin (Shanghai)* 2009, **41**:231-236.
74. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
75. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, *et al*: **Real-time quantification of microRNAs by stem-loop RT-PCR.** *Nucleic Acids Res* 2005, **33**:e179.
76. Tang F, Hajkova P, Barton SC, Lao K, Surani MA: **MicroRNA expression profiling of single whole embryonic stem cells.** *Nucleic Acids Res* 2006, **34**: e9.
77. Luu-The V, Paquet N, Calvo E, Cumps J: **Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction.** *Biotechniques* 2005, **38**:287-293.
78. Ling KH, Hewitt CA, Beissbarth T, Hyde L, Cheah PS, Smyth GK, Tan SS, Hahn CN, Thomas T, Thomas PQ, Scott HS: **Spatiotemporal Regulation of Multiple Overlapping Sense and Novel Natural Antisense Transcripts at the Nrgn and Camk2n1 Gene Loci during Mouse Cerebral Corticogenesis.** *Cereb Cortex* 2010.
79. Mattiske DM, Han L, Mann JR: **Meiotic maturation failure induced by DICER1 deficiency is derived from primary oocyte ooplasm.** *Reproduction* 2009, **137**:625-632.
80. Mann JR: **Deriving and propagating mouse embryonic stem cell lines for studying genomic imprinting.** *Methods Mol Biol* 2001, **181**:21-39.

doi:10.1186/1471-2164-12-176
Cite this article as: Ling *et al.*: Deep sequencing analysis of the developing mouse brain reveals a novel microRNA. *BMC Genomics* 2011 **12**:176.

CHAPTER 7

General discussion and conclusion

7.1 Molecular networks involved in mammalian cerebral corticogenesis

Global gene expression profiling of mouse cerebral cortices procured from different developmental stages has the potential to provide insights on the underlying mechanistic events, which govern the embryonic development as well as postnatal function of the cerebral cortex. With as much as 85% similarities in genes shared between the mouse and human, outcomes of this study can be extrapolated to explain the regulatory processes that govern human cerebral corticogenesis. To date, information on causative genes that predispose an individual with a higher risk to acquire neurological and neuropsychiatric disorders such as Alzheimer's disease, Parkinson's disease, motor-neurone disorders, intellectual disabilities and schizophrenia remains elusive. Additionally, the onset as well as progression of these disorders is often caused by the dysregulation of expression profiles involving multiple genes that interact with each other and the environment, thus forming a 'mesh of interconnected molecules'. It is essential to elucidate the fundamental molecular networks in normal cerebral corticogenesis before comprehending the underlying causative molecular networks that lead to various neurological or neuropsychiatric disorders. Therefore, this study attempts to decipher the fundamental molecular networks in cerebral corticogenesis and will serve as benchmark information for future comparative analysis involving samples with pathological conditions.

SAGE technique was employed in this study to characterise the different fundamental molecular networks in cerebral cortices obtained from mouse aged E15.5, E17.5, P1.5 and P150. These timepoints resemble all the major developmental events throughout cerebral corticogenesis ranging from the neuronal cells migration to neuronal maturation, differentiation, axonogenesis, gliogenesis and synaptogenesis. Network analysis involving all the 70 validated differentially expressed tags (DETs) showed a massive interconnection among these DETs throughout the embryonic, postnatal and adult stages of development. These networks have been associated with disease-causative DETs such as *Actb*, *App*, *Atp7a*, *Cdkn1c*, *Clcn2*, *Dcx*, *Hprt1*, *Mapt*, *Mbp*, *Plp1*, *Sncb* and *Tspan7*. More importantly, these DETs are also interconnected with 150 molecules, which

have never been associated with any neurological conditions in human. These molecules provide immense insights on how these molecular networks (contain disease-causative DETs) play a role in the onset and progression of neurological disorders. These molecules not only provide information to understand the role of the known disease causative DETs, but also serve as therapeutic candidates in the future.

A different analytical approach such as genomic clustering of DETs revealed 4 loci within the genome with a significantly higher number of mapped DETs as compared to other loci. These DETs were mapped to loci around *Sox4*, *Sox11*, *Nrgn* and *Camk2n1* genes. DETs mapped to *Sox4* and *Sox11* genomic loci were embryonic specific tags whereas those mapped to *Nrgn* and *Camk2n1* genomic loci were specific to the adult cerebral cortex. All genomic clusters have identical features where multiple transcripts were expressed from both sense and antisense orientations, overlapping each other and have different 3' UTR lengths due to different polyadenylation sites. These loci are considered actively transcribed and only specific to certain stages of cerebral corticogenesis. This is the first report of such peculiar characteristics at these genomic loci and is the landmark finding of the entire study. Genomic clustering of different overlapping transcripts in both directions demonstrate the capability of the mammalian genome to store information in a multilayer format, hence renewing our faith on existing transcriptional regulation mechanisms.

7.2 Discovery of novel transcripts

The aim of this study is to identify genes involved in the development of the mouse cerebral cortex as well as to discover any novel genes that could play a crucial role during cerebral corticogenesis. One of the many benefits of SAGE technique is its ability to discover new transcripts without having to know the sequences being analysed. Besides the validation of 10 ESTs with differential expression profiles throughout cerebral corticogenesis, further characterisation of *Sox4*, *Sox11*, *Nrgn* and *Camk2n1* genomic clusters lead to the discovery of novel long noncoding RNAs (lncRNAs) or to be more specific, *cis* natural antisense transcripts (NATs).

Since then, these NATs have become the focus of the study. Initial characterisation of these loci showed full overlapping between the sense and NATs. The presence of an NAT within a protein-coding gene locus is a widely accepted phenomenon but the existence of multiple NAT variants with different transcript features is relatively novel and has not been reported elsewhere. Expression profiling of these NATs in whole brains of different developmental stages, different adult brain regions, adult organs and cell lines of mouse origins revealed differential expression profile between the sense and NATs from *Sox11* and *Nrgn* but not the *Sox4* and *Camk2n1* genomic loci suggesting that these NATs may be regulated by different factors or transcriptional elements. Interestingly, some of these NAT variants with different 3' UTR lengths have different spatiotemporal expression profiles throughout cerebral corticogenesis. Since all the transcript variants are expressed from the same gene locus at either strand, it is unclear what determines early polyadenylation of transcripts and how could different 3' UTR lengths dominate when and where a transcript should be preserved intracellularly. Additionally, these transcripts do not encode for any peptides/proteins of known function, hence their presence during cerebral corticogenesis is questionable.

7.3 Novel role of long noncoding RNAs

Neuronal migration, morphogenesis, maturation and synaptogenesis are the fundamental processes leading to neuronal network formation or rewiring thus underlies the neuroplasticity of the cerebral cortex [reviewed in 1, reviewed in 2, reviewed in 3]. The discovery of such an unusual NATs clustering at *Sox4*, *Sox11*, *Nrgn* and *Camk2n1* genomic loci warrant a further study on the role of these NATs especially in the development and plasticity of the cerebral cortex. Interestingly, both the embryonic brain specific Sox4 and Sox11 proteins encoded within these loci are closely related to each other [4-6]. Both Sox4 and Sox11 are categorised under the same SoxC group featuring high structural similarity with each other and both are needed for pan-neuronal gene expression that directs neuronal-fated cells through migration, differentiation and maturation within the central nervous system. In addition, both the adult brain specific Nrgn and Camk2n1 proteins encoded in these loci have a closely related function [7-10].

Both of them regulate CaMKII protein, which are involved in long-term potentiation processes leading to long-term memory formation. However, the role of the NATs transcribed from the opposite strand of these loci remains uncharacterised and controversial. Although the unusual clustering of sense and NATs at *Sox4*, *Sox11*, *Nrgn* and *Camk2n1* genomic loci are yet to be proven as the factor responsible for complex developmental events, these features suggest a peculiar molecular organisation of transcriptional units within the mammalian genome.

Both sense and NATs have perfect sequence complementarities. Based on this feature alone, one can predict that the NATs may bind to the sense transcripts or DNA sequence and exert different mechanistic outcomes such as transcriptional interference, chromatin remodelling, interference of sense transcript trafficking, translational repression and transcript stabilisation or degradation. These mechanisms have been demonstrated in isolated cases involving the study of different organisms [11-16]. RNA FISH analysis of *Sox4*, *Nrgn* and *Camk2n1* sense and NATs confirmed the interactions between the two types of transcripts leading to the formation of dsRNA pairs. These dsRNA pairs are cytoplasmic-localised in the cells prepared from the cerebral cortex, hippocampus, cerebellum and olfactory bulbs. When *Sox4* NAT-overexpression analysis was performed on NIH/3T3 cells, the *Sox4* NAT did not cause transcriptional interference within the nuclear compartment as expected. Surprisingly, the *Sox4* NAT also did not cause any translational repression or sense transcripts degradation within the cytoplasm but they do cooperate with the sense transcript to produce a novel endogenous small interfering RNA (endo-siRNA), *Sox4_sir3*. However, we may only capture the downstream effect of the *Sox4* NAT based on the artificially induced ectopic expression. The system may somehow miss any potential *cis* regulatory role of *Sox4* NAT within the nucleus during the interference of transcription or chromatin structure beyond *Sox4* gene locus.

The biogenesis of an endo-siRNA from dsRNAs is commonly seen in plants or viruses due to their ability to replicate RNA via RNA-dependent RNA polymerase (RdRP) activity to form dsRNAs [17, 18]. However, the sources of endo-siRNAs in the mammalian system are mainly retrotransposons, pseudogenes, viral RNAs, short-hairpin RNAs or sense-NATs dsRNA pairs. To date, only limited numbers

of endo-siRNAs have been discovered in the mammalian system and they were mainly originated from naturally formed dsRNA portion of mRNAs or retrotransposons [19, 20]. The *Sox4* sense-NATs scenario is the first of its kind reported in the mammalian system and bears great consequences in the mammalian genome regulation mechanism. *Sox4_sir3* originated from *Sox4* sense transcript in the presence of the lowly expressed *Sox4* NAT, which the biogenesis did not disrupt the overall expression of the Sox4 proteins. This unique mechanism provides a win-win situation to preserve the original Sox4-mediated cellular regulatory processes and at the same time diversifies Sox4 regulomes via *Sox4_sir3*-mediated RNA interference (RNAi) mechanism against other transcripts. The finding may explain the origin of small RNAs abundantly found in the cytoplasm, which could serve as another hidden mode of gene expression regulation diversity within the cell.

7.4 Discovery of novel miRNAs in the brain

The revolution of sequencing technologies has enabled the discovery of lowly expressed small RNAs in an ultra high-throughput manner [21, 22]. Of the many different small RNA classes, microRNA (miRNA) is the most studied small RNA which has been demonstrated to play a diverse regulatory role during organismal development [reviewed in 23, reviewed in 24]. At molecular level, miRNA causes either translational repression or transcript degradation via RNAi pathway. Only very few miRNAs were reported and even fewer were described to have brain-specific expression or function. The next-generation sequencing platform that utilises new sequencing chemistry was used to discover novel miRNAs in an E15.5 mouse whole brain. Analysis of the small RNA sequences revealed 4 novel miRNAs with one of them known as *miR-3099*, which was later found to have expression in early mouse embryogenesis and then expressed specifically in the brain after E11.5. The study provides not only the first small RNA transcriptome generated using the deep sequencing technology for discovering novel miRNA but also as a reference dataset describing known miRNAs found in the dynamically developing mouse brain.

7.5 Limitation to the study and recommendations for future work

The study has several limitations. The amount of SAGE tags generated from the cerebral cortex using the conventional sequencing technology is considered shallow by today's standard. Employing the next-generation sequencing platform to generate SAGE tags will improve the transcriptome coverage tremendously hence provide a full sampling of all the expressed transcripts in the cell, especially those with low counts or abundance. Subsequent analysis of DETs depends profoundly on RT-qPCR validation and pathway analysis approach was adopted to implicate the functional properties of DETs. Interactions of novel molecules with DETs associated with human neurological disorders can be studied using the co-immunoprecipitation technique, electrophoretic mobility shift assay (EMSA) and western blotting analysis on various mouse cerebral cortices of different developmental stages. Loss-of-function approach such as systematic deletion or depletion of a candidate DET using either RNAi or relevant transgenic animal models would improve our understanding on the role of upregulated DETs in the predicted molecular networks. Alternatively, gain-of-function study such as transient or stable expression of candidate DETs may be effectively performed on downregulated candidates reported in the study. Correlation between transcript and protein expression would provide a more complete picture of how the cerebral corticogenesis is regulated by differentially expressed molecules.

The study of naturally occurring antisense transcripts (NATs) for all the genomic clusters, *Sox4*, *Sox11*, *Nrgn* and *Camk2n1*, has been very challenging. Multiple attempts to fully sequence the NATs using the RACE method have been unsuccessful due to the overlapping nature of NATs at these loci. The employed method failed to differentiate sequences among overlapping NATs especially when primer-walking approach was adopted for lengthy NATs. The approach of incorporating PET sequences in the study only confirmed the TSSs and polyadenylation sites for NATs but did not provide internal structural information such as intron-exon splicing sites. To completely characterise these NATs, a full-length cDNA library screening should be carried out.

When it comes to the preliminary functional characterisation of sense and NATs using *Sox4* genomic locus as a model, the ectopic expression study of PET2, PET3, PET5 and PET6 *Sox4* NATs in NIH/3T3 cells are only effectively captured in the downstream function of these NATs such as the production of *Sox4-sir3*. To evaluate their potential roles in regulating transcription or chromatin structure beyond the *Sox4* gene locus, an induction of *Sox4* NATs expression in a dosage-dependent manner should be carried out. This method would unmask any mechanism mediated by the transcript in the dosage-related manner within the cell. Additional analysis such as RNA protection assay could be carried out using independent nuclear and cytoplasmic extracted total RNAs to assess the potential role of *Sox4* NATs in different cellular compartments. As we demonstrated the formation of dsRNA aggregates within the cytoplasm, an additional RNase III or VI (digest dsRNAs) treated cells should be included as a negative control for the experiment. Performing the RNA ISH experiment on specialised cells such as neurones may provide interesting findings on RNA trafficking to synapses and subcellular localisation of these aggregates at dendritic or axonal extensions in response to long-term potentiation. To date, the study has demonstrated dsRNA aggregations of transcripts from four genomic loci in brain cells. Given these structures are RNase A resistant, a deep sequencing of transcripts from RNase A treated cells may be able to distil more dsRNAs from the transcriptome and provide a genomic view of sense and NATs that form dsRNAs.

The study also reported the discovery of novel miRNAs in the developing mouse brain. The employment of very stringent screening criteria to filter the deep sequencing dataset may have excluded potentially useful small regulatory RNAs. Small regulatory RNAs and miRNA-like small RNAs have been previously reported to derive from 'housekeeping' RNAs such as tRNA [25, 26]. Although the main aim of this study was to identify novel miRNA, *in silico* analysis of filtered small RNA sequences for abundantly expressed and precisely 5'- and 3'-processed sequences, followed by validation using northern blotting analysis may lead to the discovery of new classes of potentially functional small RNAs. However, to define a small RNA such as *miR-3099* as functional, more functional analysis has to be carried out. Simple mimicking and inhibitory studies using Locked Nucleic Acid (LNA) probe for *miR-3099* can be performed using an *in vitro* system such as P19 teratocarcinoma cell line. Functional assays can then be

carried out to characterise the proliferation, differentiation and migration capability of the cell in the excess or absence of *miR-3099* mimics or inhibitors. Core component of RISC complexes include Argonaute2 (Ago2) and GW182 proteins [reviewed in 27, 28]. To investigate downstream targets for *miR-3099*, ribonucleoprotein immunoprecipitation (RNP-IP) of the cells transfected with *miR-3099* mimics or inhibitors can be performed to isolate miRNA-RISC-mRNA complexes [29]. These complexes can be obtained by using Ago2 and GW182 antibodies followed by reverse-transcription of bound mRNAs into cDNAs using the *miR-3099* seed sequence and 3'UTR derived primers. Microarray analysis can then be performed on these cDNAs to determine DETs between *miR-3099* depleted/overexpressed cells and controls followed by validation at both mRNA and protein levels using RT-qPCR and western blotting, respectively. Target sites within the validated DETs can be predicted using *in silico* programs such as DIANA micro-T 3.0 [30] and TargetScan [31] followed by laboratory confirmation using a gene reporter assay in an *in vitro* model. Similarly to *miR-3099*, the functional properties of the endo-siRNA, *Sox4_sir3*, produced as a result of dsRNA formation between *Sox4* sense and NATs can be evaluated using a similar approach.

7.6 Conclusion

The information obtained from the above studies has provided an overview of complicated molecular networks in governing the development of the sophisticated mouse cerebral cortex. More importantly, these studies have led to the discovery of various novel NATs with potential regulatory properties that are involved in novel genetic mechanisms during cerebral corticogenesis. These studies have uncovered various novel transcripts that could lead to novel genetic mechanisms and also leave many questions unanswered. Additional studies are required to further understand the relationships between various molecular networks or differentially regulated genomic loci characterised and their encoded or regulated proteins, which determine the phenotypic outcomes of an organism. In summary, this information will shed light on known/novel candidate molecular targets/networks involved in the central nervous system development as well as potential causative/therapeutic targets for neurological or neuropsychiatric disorders.

7.7 References

1. Feng Y, Walsh CA: **Protein-protein interactions, cytoskeletal regulation and neuronal migration**. *Nat Rev Neurosci* 2001, **2**(6):408-416.
2. Marin O, Rubenstein JL: **A long, remarkable journey: tangential migration in the telencephalon**. *Nat Rev Neurosci* 2001, **2**(11):780-790.
3. Nadarajah B, Parnavelas JG: **Modes of neuronal migration in the developing cerebral cortex**. *Nat Rev Neurosci* 2002, **3**(6):423-432.
4. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J: **The establishment of neuronal properties is controlled by Sox4 and Sox11**. *Genes Dev* 2006, **20**(24):3475-3486.
5. Dy P, Penzo-Mendez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V: **The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties**. *Nucleic Acids Res* 2008, **36**(9):3101-3117.
6. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ: **Roles of Sox4 in central nervous system development**. *Brain Res Mol Brain Res* 2000, **79**(1-2):180-191.
7. Pak JH, Huang FL, Li J, Balschun D, Reymann KG, Chiang C, Westphal H, Huang KP: **Involvement of neurogranin in the modulation of calcium/calmodulin-dependent protein kinase II, synaptic plasticity, and spatial learning: a study with knockout mice**. *Proc Natl Acad Sci U S A* 2000, **97**(21):11232-11237.
8. Chang BH, Mukherji S, Soderling TR: **Calcium/calmodulin-dependent protein kinase II inhibitor protein: localization of isoforms in rat brain**. *Neuroscience* 2001, **102**(4):767-777.
9. Prichard L, Deloulme JC, Storm DR: **Interactions between neurogranin and calmodulin in vivo**. *J Biol Chem* 1999, **274**(12):7689-7694.
10. Saha S, Datta K, Rangarajan P: **Characterization of mouse neuronal Ca2+/calmodulin kinase II inhibitor alpha**. *Brain Res* 2007, **1148**:38-42.
11. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P: **The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin**. *Science* 2008, **322**(5908):1717-1720.
12. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes**. *Nature* 2002, **415**(6873):810-813.
13. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C: **Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation**. *Mol Cell* 2008, **32**(2):232-246.
14. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al*: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs**. *Cell* 2007, **129**(7):1311-1323.
15. Kalantry S, Purushothaman S, Bowen RB, Starmer J, Magnuson T: **Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation**. *Nature* 2009.

16. Panning B, Dausman J, Jaenisch R: **X chromosome inactivation is mediated by Xist RNA stabilization**. *Cell* 1997, **90**(5):907-916.
17. Ahlquist P: **RNA-dependent RNA polymerases, viruses, and RNA silencing**. *Science* 2002, **296**(5571):1270-1273.
18. Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SI: **RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production**. *Proc Natl Acad Sci U S A* 2005, **102**(1):152-157.
19. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T *et al*: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes**. *Nature* 2008.
20. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: **Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs**. *Genes Dev* 2008, **22**(20):2773-2785.
21. Shao NY, Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Li N, Chen W, Khaitovich P: **Comprehensive survey of human brain microRNA by deep sequencing**. *BMC Genomics* 2010, **11**:409.
22. Zhang J, Xu Y, Huan Q, Chong K: **Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response**. *BMC Genomics* 2009, **10**:449.
23. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.
24. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.
25. Reese TA, Xia J, Johnson LS, Zhou X, Zhang W, Virgin HW: **Identification of novel microRNA-like molecules generated from herpesvirus and host tRNA transcripts**. *J Virol* 2010, **84**(19):10344-10353.
26. Lee YS, Shibata Y, Malhotra A, Dutta A: **A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)**. *Genes Dev* 2009, **23**(22):2639-2649.
27. Ding L, Han M: **GW182 family proteins are crucial for microRNA-mediated gene silencing**. *Trends Cell Biol* 2007, **17**(8):411-416.
28. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD: **Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes**. *Cell* 2005, **123**(4):607-620.
29. Hassan MQ, Gordon JA, Lian JB, van Wijnen AJ, Stein JL, Stein GS: **Ribonucleoprotein immunoprecipitation (RNP-IP): a direct in vivo analysis of microRNA-targets**. *J Cell Biochem* 2010, **110**(4):817-822.
30. Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K *et al*: **DIANA-microT web server: elucidating microRNA functions through target prediction**. *Nucleic Acids Res* 2009, **37**(Web Server issue):W273-276.
31. **TargetScan** [http://www.targetscan.org/]