Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1–gfp knock-in mice

Jerome Aubert*, Marios P. Stavridis†, Susan Tweedie*, Michelle O’Reilly*, Klemens Vierlinger‡, Meng Li*, Peter Ghazal†, Tom Pratt§, John O. Mason§, Douglas Roy‡, and Austin Smith*¶

*Institute for Stem Cell Research, University of Edinburgh, King’s Buildings, West Mains Road, EH9 3JQ Edinburgh, Scotland; †The Scottish Centre for Genomic Technology and Informatics, University of Edinburgh, Chancellor’s Building, 49 Little France Crescent, EH16 4SB Edinburgh, Scotland; and ‡Division of Biomedical Sciences, University of Edinburgh, Hugh Robson Building, George Square, EH8 9XD Edinburgh, Scotland

The transcription factor Sox1 is the earliest and most specific known marker for mammalian neural progenitors. During fetal development, Sox1 is expressed by proliferating progenitor cells throughout the central nervous system and in no tissue but the lens. We generated a reporter mouse line in which egfp is inserted into the Sox1 locus. Sox1GFP animals faithfully recapitulate the expression of the endogenous gene. We have used the GFP reporter to purify neuroepithelial cells by fluorescence-activated cell sorting from embryonic day 10.5 embryos. RNAs prepared from Sox1GFP+ and Sox1GFP− embryo cells were then used to perform a pilot screen of subtracted cDNAs prepared from differentiating embryonic stem cells and arrayed on a glass chip. Fifteen unique differentially expressed genes were identified, all previously associated with fetal or adult neural tissue. Whole mount in situ hybridization against two genes of previously unknown embryonic expression, Lrrn1 and Musashi2, confirmed the selectivity of this screen for early neuroectodermal markers.

Neural stem cells are promising candidates for the development of cellular and genetic therapies for neurodegenerative disorders such as Parkinson’s disease and Huntington’s disease (1), and for creation of in vitro drug discovery and toxicological screens (2). However, the biomedical application of neural stem cells will require the generation of large homogenous populations of these cells in vitro. One source of neural stem cells is embryonic stem (ES) cells (3). ES cells are derived from the inner cell mass of the preimplantation blastocyst-stage embryo and can be propagated indefinitely in an undifferentiated, pluripotent state (4). The for- preimplantation blastocyst-stage embryo and can be propagated indefinitely in an undifferentiated, pluripotent state (4). The formation of multicellular aggregates called embryoid bodies permits the commitment of pluripotent ES cells into multiple cellular lineages in vitro (5), mimicking aspects of cellular differentiation in early embryos (6). This provides a powerful system for the discovery of genes induced early during development and for functional validation of candidate genes (7). A favored protocol for the commitment of ES into neural lineage is the treatment of embryoid bodies with all-trans retinoic acid (8–10). After induction and outgrowth onto an adhesive substratum, up to 50% of cells express the neural precursor markers Sox1 and Sox2 and can generate neurons and glia (11).

Key advances in defining the optimal conditions for generating and propagating neural stem cells are likely to come from a proper understanding of the molecular mechanisms controlling the fate decisions of pluripotent cells and of fetal and ES cell-derived neural precursors. Here we describe a refined approach to identify genes induced during neural specification and/or maintained in neural progenitor cells in vivo and in vitro. Transgenic mice (Sox1GFP) were generated in which the enhanced GFP (egfp) reporter is inserted into the Sox1 gene via gene targeting (12). Sox1 is the earliest specific marker of neural precursors in the mouse embryo (13). It is present in proliferating neural precursors from the neural plate stage onwards. The only other site of expression in the mid-gestation fetus is the lens (Fig. 1A). Exit from mitosis and neuronal or glial differentiation is accompanied by down-regulation of Sox1 (14). The GFP knock-in allows the visualization of Sox1 expression in these animals by fluorescence microscopy, and the purification of Sox1-positive cells by fluorescence-activated cell sorting (FACS). This purification allows preparation of RNAs for highly selective differential screening of microarrays. We have tested this approach by application to a custom microarray of a subtractive cDNA population prepared from retinoic acid-induced embryoid bodies.

Materials and Methods

Gene Targeting. A phage genomic DNA library from 129/Ola strain mouse was screened with a 2-kb probe containing the Sox1 ORF (generously provided by Larysa Pevny). From a resulting phage containing 12 kb of Sox1 genomic sequence, 5.5- and 2.5-kb fragments flanking the Sox1 ORF were taken as 5’ and 3’ homology arms to prepare a targeting vector. The gene for egfp was fused in-frame into the second of three consecutive ATGs at the Sox1 translation initiation site (15, 16) via PCR. The fusion product, linked via an internal ribosome entry site (IRES) (17) to the gene encoding puromycin acetyltransferase (pac), was cloned between the homology arms. A cytomegalovirus promoter-driven hygromycinβ-thymidine kinase dual selection cassette (18) flanked by loxP sites was inserted downstream of the GFirepsac cassette (Fig. 1A). After electroporation in E14tg2a ES cells and selection in hygromycin, three targeted clones were identified by Southern analysis with flanking 5’ and 3’ probes, and unique integration was confirmed for two of these by using an egfp probe. Transient transfection with a Cre recombinate expression vector was used to remove the selection cassette. Clones that had undergone excision were selected for in the presence of ganciclovir and screened by Southern analysis (Fig. 1B). One such clone (46C) was injected into blastocystes and...
from the parental E14Tg2a ES cells. (53 are correctly targeted, single integration clones. Clone 46C is a derivative of clone 46 after Cre-mediated deletion of the CMV Hy-TK cassette. Control is DNA populations, Sox1 digested in 0.1% trypsin, and resuspended in cold 10% FCS in embryos were dissected free of extraembryonic membranes, embedding in 2% agarose in PBS and sectioning at 50 µm. Sections were counterstained with propidium iodide and analyzed by confocal microscopy. using a vibratome. Sections were counterstained with propidium iodide and analyzed by confocal microscopy.

Sox1 stained in 0.1% trypsin, and resuspended in cold 10% FCS in PBS. The cells were sorted by flow cytometry to give two cell populations, Sox1-GFP-positive (Sox1GFP+) cells and Sox1-GFP-negative (Sox1GFP−) cells. The sample was kept cold at all times to minimize RNA degradation and cell death during sorting. Viable cells were gated by their forward and side scatter characteristics, and gates were set to sort positive and negative cell populations.

Total RNA was extracted from both cell populations by using the RNeasy Minikit (Qiagen, Valencia, CA), according to the manufacturer’s instructions. RNA yield was determined by measuring absorbance at 260 nm. RNA quality was assessed by electrophoresis of 1 µg of RNA on a standard 1.2% formaldehyde agarose gel.

Passage of the larval line of the Sox1GFP mouse line. Mice were maintained on a mixed 129xMFl background by breeding of heterozygotes to outbred MFl mice.

**Tissue Preparation.** For analysis of Sox1GFP expression, heterozygous Sox1GFP males were crossed with wild-type females. Midday after vaginal plug was considered as embryonic day 0.5 (E0.5). Females were killed by cervical dislocation, and the embryos were dissected free of the uterus, washed in PBS, and observed under a fluorescence microscope. For cryosectioning, embryos were fixed in 4% paraformaldehyde (PFA), cryoprotected in 30% sucrose in PBS, and embedded in OCT compound before cryosectioning at 10 µm. Adult (4–7 weeks old) heterozygous brains were dissected out and fixed in 4% PFA before embedding in 2% agarose in PBS and sectioning at 50 µm by using a vibratome. Sections were counterstained with propidium iodide and analyzed by confocal microscopy.

**FACS Purification and RNA Preparation.** E10.5 Sox1–GFP-positive embryos were dissected free of extraembryonic membranes, digested in 0.1% trypsin, and resuspended in cold 10% FCS in PBS. The cells were sorted by flow cytometry to give two cell populations, Sox1–GFP-positive (Sox1GFP+) and Sox1–GFP-negative (Sox1GFP−) cells. The sample was kept cold at all times to minimize RNA degradation and cell death during sorting. Viable cells were gated by their forward and side scatter characteristics, and gates were set to sort positive and negative cell populations.

Total RNA was extracted from both cell populations by using the RNeasy Minikit (Qiagen, Valencia, CA), according to the manufacturer’s instructions. RNA yield was determined by measuring absorbance at 260 nm. RNA quality was assessed by electrophoresis of 1 µg of RNA on a standard 1.2% formaldehyde agarose gel.

**RT-PCR.** To eliminate contaminating genomic DNA, 1 µg of total RNA was treated with 1 unit of DNase I (GIBCO/BRL) for 15 min at 25°C. DNase I was inactivated with 25 mM EDTA (pH 8.0, GIBCO/BRL) at 65°C for 10 min and chilled on ice. First strand random-primed cDNA was synthesized by using Superscript II Preamplification System (GIBCO/BRL) as described by the supplier. The cDNA was analyzed by PCR amplification using individual primer pairs for specific marker genes. The PCR cycling sequence used was 94.0°C for 3 min, followed by 20–35 cycles of 94.0°C for 30 s, 58.0–60.0°C for 30 s and 72.0°C for 1 min. This was followed by a final extension time of 7 min. All PCR samples were analyzed by electrophoresis on 2% agarose gels stained with ethidium bromide.

**Microarray and Sequence Analysis.** A subtracted library enriched for genes expressed during retinoic acid-induced neural commitment of ES cells (7) was spread out on LB plates containing ampicillin and 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-gal). A total of 384 white bacterial colonies were randomly picked and cultured in 96-well plates. One microliter of each bacterial culture was amplified by using the Advantage cDNA PCR kit (CLONTECH) and PCR primers that are homologous to the flanking regions of the cDNA insert (30–35 cycles: 30 s at 95°C, followed by 3 min at 68°C). PCR fragments were analyzed by electrophoresis on 2% agarose gel. The average insert size was between 200 and 800 bp. PCR products were printed onto poly(llysine)-coated glass slides by using an Affymetrix (Santa Clara, CA) 417 arrayer. Each cDNA insert was spotted in triplicate. A collection of marker genes (see Fig. 2B) whose expression was predetermed by RT-PCR analysis was included to act as control for the specificity of neural expression in the sorted Sox1+/− RNA populations.

Twenty micrograms of Sox1+ or Sox1− total RNA obtained from disaggregated embryos was reverse transcribed with Su-
Scans were performed with an Affymetrix 428 scanner, and the program QUANTARRAY (Perkin-Elmer) was used for image analysis. Multiple scans were taken to define the optimal dynamic range of signal for subsequent analysis. The background was first subtracted for each respective probe element on the array. The median value for the triplicate probe elements representing each gene or insert was then defined. A scaling factor was applied to the arrays representing normalization to the 75th percentile of the global signal distribution. Ratios of expression were then calculated by using these normalized median values. A ratio value (Sox1+/−/Sox1−) of 1.5 or greater was used for the selection of neural specific expression.

Clones encoding 15 up-regulated transcripts (Table 1) were sequenced, and the corresponding genes were identified by BLAST searches of nonredundant (nr), dbEST, and mouse genomic databases. Protein domains were identified by using SMART.

In Situ Hybridization. Subtracted library clones were used to generate antisense RNA probes labeled with digoxigenin-UTP. Automated in situ detection was carried out on E8.5, E10.5, and E11.5 outbred mouse embryos by using an In situPro machine (Abimed, Langenfeld, Germany).

Results and Discussion

Expression of Sox1GFP. In undifferentiated ES cells in vitro there is no detectable activity of the Sox1GFP allele but expression is specifically activated on induction of neural differentiation as described elsewhere (12, 14). In vivo after germ-line transmission, the GFP reporter is faithfully expressed in the nervous system and lens, with no apparent ectopic expression. Sox1GFP fluorescence is first detected around E8.5 throughout the neural plate and headfolds (ref. 12 and data not shown). This is slightly

Table 1. Genes identified by microarray analysis

<table>
<thead>
<tr>
<th>Clone</th>
<th>Fold increase</th>
<th>GenBank accession no.</th>
<th>Gene symbol (synonyms), name</th>
<th>Expression</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5C3</td>
<td>2.40</td>
<td>CB968102</td>
<td>Nhlh2 (Henh2, Nscl-2, NSCL2), nescent helix–loop–helix 2</td>
<td>Dev NS and PN cerebellum</td>
<td>HLH</td>
</tr>
<tr>
<td>P2H5</td>
<td>2.40</td>
<td>CB968096</td>
<td>Mm.156164</td>
<td>Dev and mature NS, epididymis</td>
<td></td>
</tr>
<tr>
<td>P2C5</td>
<td>1.98</td>
<td>CB968093</td>
<td>Khdrob3 (SLM-2, Etle, T-STAR), KH domain containing, RNA binding, signal transcription associated 3</td>
<td>Ad B and S muscle/testes</td>
<td>KH</td>
</tr>
<tr>
<td>P387</td>
<td>1.96</td>
<td>CB968098</td>
<td>Slc2a1 (Glut-1), solute carrier family 2 (facilitated glucose transporter), member 1</td>
<td>E10 NT, OV, heart and gut</td>
<td>TM</td>
</tr>
<tr>
<td>P5C11</td>
<td>1.96</td>
<td>CB968101</td>
<td>sFRP2 (Sdf5), secreted frizzled-related sequence protein 2</td>
<td>E neuroepithelium</td>
<td>CRD</td>
</tr>
<tr>
<td>P5H8</td>
<td>1.96</td>
<td>CB968107</td>
<td>Lrn1 (NLRR-1), leucine-rich repeat protein 1, neuronal</td>
<td>Dev CNS, DRG and cartilage</td>
<td>LRR/TM</td>
</tr>
<tr>
<td>P3C1</td>
<td>1.95</td>
<td>CB968099</td>
<td>Sox4, SRY-box-containing gene 4</td>
<td>Dev CNS, thymus, BM</td>
<td>HMG</td>
</tr>
<tr>
<td>P3B1</td>
<td>1.94</td>
<td>CB968097</td>
<td>Zic1, zinc finger protein of the cerebellum 1</td>
<td>E cerebellum and dorsal 1/2 of NT</td>
<td>ZnF</td>
</tr>
<tr>
<td>P2E5</td>
<td>1.81</td>
<td>CB968094</td>
<td>Vim, vimentin</td>
<td>Radial glia in Dev hindbrain</td>
<td>CC</td>
</tr>
<tr>
<td>P5G9</td>
<td>1.76</td>
<td>CB968105</td>
<td>Rtn1, reticulin 1</td>
<td>Dev and mature NS</td>
<td>TM</td>
</tr>
<tr>
<td>P5D11</td>
<td>1.69</td>
<td>CB968103</td>
<td>Msi2h (msi2), musashi homolog 2*</td>
<td>Dev and mature NS, ubiquitous by NB</td>
<td>RRM</td>
</tr>
<tr>
<td>P5D5</td>
<td>1.65</td>
<td>CB968104</td>
<td>Sox11, SRY-box-containing gene 11*</td>
<td>Dev NS, sites of Ep–Mes interactions</td>
<td>HMG</td>
</tr>
<tr>
<td>P5H2</td>
<td>1.63</td>
<td>CB968106</td>
<td>Qk, Quaking</td>
<td>Myelin-forming cells</td>
<td>KH</td>
</tr>
<tr>
<td>P5A9</td>
<td>1.60</td>
<td>CB968100</td>
<td>Hrmt113, HMT1 hnRNP methyltransferase-like 3*</td>
<td>Dev and mature NS</td>
<td>—</td>
</tr>
<tr>
<td>P2E7</td>
<td>1.51</td>
<td>CB968095</td>
<td>Tubal 1 (Talp1), tubulin α1</td>
<td></td>
<td>NPCs</td>
</tr>
</tbody>
</table>

Data are fold increase in expression, accession number of each clone sequence, gene symbols/names, characterized expression pattern, and protein domain information (given where known). Ad, adult; B, brain; BM, bone marrow; CC, coiled coil; CRD, cysteine-rich domain; Dev, developing; DRG, dorsal root ganglia; E, embryo; Ep, epithelial; HLH, helix–loop–helix; HMG, high-mobility group; KH, homology RNA-binding domain; LRR, leucine-rich repeat; M, muscle; Mes, mesenchymal; NB, Northern blot; NPCs, neural precursor cells; NS, nervous system; NT, neural tube; OV, optic vesicle; PN, postnatal; RRM, RNA recognition motif; S, skeletal; TM, transmembrane; ZnF, zinc finger.

*The predicted transcripts from these genes are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession numbers TPA: BK001349, BK001483, and BK001484.

Aubert et al.
later than the reported onset of expression of Sox1 mRNA and protein (13, 14), presumably because of the time required for correct folding and accumulation of GFP to detectable levels. At E9.5 Sox1\textsuperscript{GFP} is expressed along the entire neuraxis but in no other tissue (Fig. 1D). At mid-gestation, Sox1\textsuperscript{GFP} is maintained throughout the brain and the neural tube but is excluded from the roofplate and floorplate (Fig. 1Dii). At this stage, Sox1\textsuperscript{GFP} expression also becomes evident in the lens where Sox1 has been shown to regulate the \( \gamma \text{-crystallin} \) genes and to be necessary for lens fiber cell elongation (21). The distribution of Sox1\textsuperscript{GFP} is in agreement with the published expression of Sox1 mRNA and protein (14). In later stages of embryonic development, Sox1 is excluded from most differentiated neurons and glia but is maintained in the proliferative ventricular zone and in the lens (Fig. 1Diii).

In the brains of adult animals Sox1\textsuperscript{GFP} is prominently expressed in the subgranular layer of the dentate gyrus (Fig. 1Div). Numerous GFP-expressing cells are seen in the-inner subgranular layer, the area where adult neural progenitors have been shown to reside (22). Neurons born in this area subsequently migrate through the granular layer of the dentate gyrus. At higher magnification, smaller numbers of GFP-positive cells can be observed in the granular layer in Sox1\textsuperscript{GFP} mice.

Heterozygous Sox1\textsuperscript{GFP} animals are viable and apparently healthy with no obvious phenotype. Homozygous null Sox1\textsuperscript{GFP} mice have small eyes with opaque lenses and suffer from spontaneous seizures, as described for Sox1 mutants (21).

**FACS Purification and RNA Probe Generation.** E10.5 Sox1\textsuperscript{GFP} embryos were trypsinised, and the resultant pool of cells was sorted by flow cytometry based on GFP expression. Cell sorting yielded a Sox1\textsuperscript{GFP+} neural precursor population (gate R1) and a Sox1\textsuperscript{GFP–} control population (gate R2) (Fig. 2A). RNA was extracted from the Sox1+ and Sox1– cells. To confirm the identity of the two RNA populations, cDNA was produced by reverse transcription and analyzed by RT-PCR. We examined the presence in both populations of a number of known genes whose expression pattern is both spatially and temporally defined during development (Fig. 2B).

As expected, Sox1\textsuperscript{mRNA} was restricted to Sox1\textsuperscript{GFP+} cells. Genes known to be restricted to neural progenitor populations such as Ngn2 and Pax6 (23) displayed a similar restriction to the GFP+ population. Pax7, Nestin, and \( \text{sFRP2} \) all show strong expression in the GFP+ population but are also represented in the GFP– fraction. This is in agreement with a predominant expression of these genes in the developing CNS with additional expression in the somitic mesoderm (Nestin and Pax7) and in the mesonephros (sFRP2) (23–25). RT-PCR analysis revealed weak expression of the ShcC gene in the Sox1+ cell population. ShcC is an adapter protein that is predominantly expressed in mature neurons (26, 27). As neurogenesis in the neural tube begins at E9.5, the expression seen here could be indicative of a small number of early neurons present in the Sox1+ cell population. This could arise from perdurance of GFP after differentiation and Sox1 down-regulation.

RNAs for bone morphogenetic protein 4 (BMP4), a marker for early mesodermal differentiation (28), and Sox10, a key regulator in the differentiation of peripheral glial cells, with high expression in neural crest cells and cells of the melanocyte lineage, were preferentially expressed in Sox1\textsuperscript{GFP+} cells. Together, these results confirm the efficient separation of neural and nonneural cell populations by flow cytometry.

\( P311 \) and Riken cDNA 2810027O19 represent two genes that are induced in retinoic acid treated embryoid bodies (7). These genes show similar expression levels in Sox1+ and Sox1– cell populations. This result is consistent with the expression profile revealed by in situ hybridization analyses. Both mRNAs are abundant in the neural tube, but \( P311 \) is also found in the somites and apical ectodermal ridge (7) and Riken cDNA 2810027O19 mRNA is present in migrating neural crest cells, the apical ectodermal ridge, and in condensing mesenchymal cells (data not shown). These two examples illustrate a limitation in the selectivity of screens based on total cell populations from retinoic acid-induced differentiating embryoid bodies (see below).

**Microarray Analysis.** We have previously generated a subtracted library enriched for genes induced during retinoic acid-induced differentiation of ES cells (7). Analysis by differential filter hybridization indicated substantial enrichment for genes of interest. Of 480 clones, 138 (29%) were preferentially expressed during retinoic acid-induced differentiation. These corresponded to 96 unique genes, 40% of which showed enriched expression in the developing or adult CNS (7). However, most of these clones also showed appreciable expression in nonneural tissues of the developing fetus as exemplified by P311 and Riken cDNA 2810027O19. We reasoned that a more stringent screen of the SSH library would be to perform the differential hybridization with Sox1\textsuperscript{GFP+}-purified cell populations from embryos rather than whole cell populations from embryoid bodies.

To test this idea, we generated a cDNA microarray from 384 randomly picked clones from the SSH library and carried out a differential hybridization screen with RNAs generated from the Sox1\textsuperscript{GFP} embryos. Also included on the array were a number of the marker genes used for characterization of the Sox1 GFP RNA isolated from embryos (Fig. 2).

The Sox1+ and Sox1– RNA samples were labeled and hybridized to the array. The majority of the marker genes shown to be preferentially expressed in Sox1+ RNA by RT-PCR analysis (e.g., Sox1, Ngn2, Nestin, and Pax6; Fig. 2B) showed expression fold changes of >1.5 when analyzed by microarray analysis. The expression of Sox10, \( P311 \), and the motor neuron precursor marker Islet1 was not elevated in the Sox1+ population, consistent with additional sites of expression in ventral mesoderm and endoderm (29). The expression of these marker genes determined by the microarray screen is in a broad agreement with the RT-PCR profiles of the sorted RNA populations and known expression within and outwith the embryonic CNS.

**Clone Identification.** Fifteen clones that demonstrated the highest differential Sox1+/– expression ratios (>1.5) were taken for sequencing. Fourteen sequences correspond to known genes, all of which had previously been associated with either developing or adult CNS (Table 1). Clones PSDD1 (Msi2h) and P5D5 (Sox11) do not match known genes directly but are derived from extended 3′ UTR regions that lie downstream of the current gene annotation. In each case, the sequence can be linked to the identified gene via a contiguous assembly of expressed sequences. The discovery of additional 3′ UTR sequences is consistent with the fact that a Musashi2 (Msi2h) 3′ UTR probe detects a 7.1-kb transcript and suggests that the original 2.3-kb Msi2h cDNA sequence (AB056103) is incomplete (30). We also noted that the first 513 base pairs of the original Msi2h cDNA sequence do not align with other Msi2h ESTs or the human Msi2 sequence. In fact, this presumed 5′ UTR sequence has a 100% match with a genomic sequence on mouse chromosome 1, suggesting that the original cDNA is a hybrid of two clones. Msi2h is thought to play a role in the maintenance and proliferation of CNS precursor cells (30, 31).

The relationship between clone PSDD5 and Sox11 was detected by combining sequences from two unigene clusters Mm.41704 and Mm.254253 into a single contiguous sequence of ~7 kb. This relatively long predicted transcript is consistent with human SOX11, which has a 3′ UTR of ~8 kb (32). Interestingly, the screen identified a second SRY-box-containing gene, Sox4, which is expressed in the differentiating subventricular zone progenitors during neurogenesis (33) and has a role during B-cell
differentiation and heart development (34). Sox4 and Sox11 encode class C Sox proteins, are closely related at the sequence level, and have similar expression pattern. It has been proposed that they may be functionally redundant during CNS development (33).

Hrm113 (clone P5A9) has not been described in mouse, but is presumed to be the orthologue of human hnRNP methyltransferase-like 3 (HRMT1L3)-based high identity (>99%) and conserved synteny. The gene encoding P2H5 remains to be identified; this sequence matches a unigene cluster containing ESTs that are derived exclusively from neural tissues and have moderate similarity to Neural Wiskott–Aldrich syndrome protein (N-WASP).

The remaining up-regulated genes were identified as being expressed predominantly in developing or adult nervous system. The identification of Reticulon 1, an endoplasmic reticulum protein of unknown function, in this screen is consistent with the fact that this protein is localized in neuroepithelial progenitors at the pallio-subpallial boundary of the developing telencephalon (35).

The gene for the intermediate filament protein, vimentin, which is expressed in radial glia (36), was up-regulated in the Sox1+ fraction. Radial glia are thought to serve as precursor cells in the developing forebrain (37). The neuronal precursor cell marker tubulin a1 (38) was also enriched in Sox1+ cells.

Nescent helix–loop–helix 2 (Nhlh2) is a basic helix–loop–helix transcription factor that is reported to be transiently expressed in subependymal cells throughout the CNS at mid-gestation and also transiently in the postnatal cerebellum (39, 40). Expression in the Sox1+ fraction could reflect the perdurance of GFP or an earlier onset of Nhlh2 expression than previously described. sFRP2, encoding the Wnt antagonist secreted frizzled-related protein-2, is expressed in the embryonic neuroepithelium (25, 41). This gene was also isolated in the previous screen of this library and demonstrated to promote neural differentiation of ES cells (7). Detection of sFRP2 in the microarray screen demonstrates the potential of this approach for identification of functionally significant players in neural differentiation.

The KH domain containing, RNA binding, signal transduction-associated 3 (Khdrbs3) gene is a predominantly nuclear RNA-binding protein which heterodimerizes with Sam68 (68-kDa Src substrate associated during mitosis). Khdrbs3 expression has been observed in adult brain and also skeletal muscle (42). Its embryonic expression and functional role have yet to be defined.

Expression of Lrrn1 (leucine-rich repeat protein 1, neuronal) has been reported in the CNS at E11.5 by Northern-blot, and whole mount in situ hybridization on E13.5 revealed a predominant expression in the developing nervous system (43). Although its role remains unknown, the LRR domain is proposed to function in cell adhesion and has been implicated in a variety of events in neural development.

In Situ Hybridization. We used whole mount in situ hybridization to examine the embryonic expression of two genes emerging from the microarray screen, Lrrn1 and Msi2h. Both have been suggested to play significant roles in neural development (Fig. 3). Lrrn1 mRNA is detectable along the entire antero-posterior axis of the neuroectoderm, with additional faint expression in somites. At E10.5, Msi2h expression is maintained in the hindbrain and otic vesicle, but also extends along the neural tube. Hybridization is also apparent in dorsal root ganglia and limb bud (Fig. 3D). Lrrn1 mRNA on E10.5 is present in the ventral-most neural tube as well as the hindbrain and the telencephalic vesicle, and is also prominent in the somites.

Conclusion

In this study, we have shown that Sox1–gfp knock-in mice allow reliable visualization and purification of pan-neural progenitor cells from mid-gestation mouse embryos. Importantly, Sox1 is expressed in neuroepithelial cells throughout the entire neuraxis, labeling all categories of regionally specified neural precursor. The particular advantage of Sox1 over the other well established pan-neural marker nestin is that there is no detectable expression outside the CNS during early to mid-fetal development apart from in the well defined structure of the lens. Examination in whole mount embryos shows that the Sox1GFP reporter reproduces faithfully the expression of Sox1. Interestingly, preliminary analyses of adult brains have highlighted expression in the subgranular layer of the dentate gyrus (Fig. 1), a region for which there is now overwhelming evidence of the persistence of neural precursor cells (22, 44). Thus, Sox1GFP may be a useful marker of adult neural precursors. Further studies are required to test this directly.

In the present study, we used FACS to separate Sox1-GFP-positive and -negative populations from whole E10.5 mouse embryos. Analysis of a panel of markers by RT-PCR yielded expression data consistent with the substantial elimination of nonneural cells from the GFP-positive population and conversely the absence of neural precursors from the GFP-negative population.

We then carried out a pilot microarray screen with the aim of identifying genes specifically expressed both during neural commitment of ES cells and in neural progenitor cells in vivo. From 384 arrayed SSH clones, we identified 15 unique clones showing preferential expression in the GFP-positive cell population. Of these, 11 represent known genes previously reported as expressed in embryonic and/or adult neural tissues, particularly in the brain. The remaining genes were ESTs, each of which has originated from libraries derived from neural tissues. We have been able to identify three of the ESTs as corresponding to Musashi2, Sox11, and Hrm113. Musashi2 has previously been described as ubiquitously expressed based on Northern analyses.
of adult tissues (30). However, our in situ hybridization data show that this gene is preferentially expressed in neural tissue in the fetus. It is noteworthy that marker genes with expression in the developing nervous system but substantial additional nonneural expression (i.e., Sox10, Islet1, P311, and Riken cDNA 2810027019) were not significantly enriched in the Sox1+ population. Furthermore, several clones were found to be present at higher levels in the Sox1– RNA population (data not shown). These clones most likely correspond to transcripts expressed in nonneural tissues induced by retinoic acid treatment of embryoid bodies.

The fold enrichment value is not an absolute measure of differential expression, and in many cases may be a considerable underestimate of the selectivity of expression caused by the heterogeneity of the Sox1+ population. Sox1 marks the entire pool of proliferating precursors in the neural tube, whereas all of the genes identified have a regionally restricted expression. A previous study has indicated that complex tissues such as the brain are prone to a “dilution effect” when analyzed by microarray, yielding lower levels of fold change and smaller numbers of differentially expressed genes compared with studies using cell lines (45). Nonetheless, larger-scale screening may identify genes with broader neural expression and consequent higher fold enrichment values.

Overall, this study demonstrates the potential of combining in vitro ES cell differentiation and in vivo lineage purification with microarray technology to achieve rapid, efficient identification of genes expressed selectively in tissues and stages of interest. Previously described expression profiles and our in situ hybridization data of the differentially regulated clones examined confirm the underlying principle of using RNA prepared from Sox1-selected cells to screen custom-built microarrays enriched for neural genes. This pilot scale screen has been sufficient to highlight several genes, notably Nhlh2, Lmn1, H11001, Rtn1, and the unknown gene corresponding to the P2H5 EST, for further investigation as potential regulators of neural development. A rapid means of assessing the significance of these genes would be via episomal gain-of-function analyses in ES cells (7).

We thank Hannah Dunstan for in situ hybridization, Steve le Moenec for MoFlo operation, Jenny Nichols for chimacra production, and Carol Manson and staff for mouse husbandry. This research was supported by the Biotechnology and Biological Sciences Research Council, the Medical Research Council of the United Kingdom, and the Human Frontiers Science Program Organisation. J.A. was the recipient of an Institut National de la Sante et de la Recherche Médicale fellowship and a Marie Curie individual fellowship. M.O.R. is supported by the Wellcome Trust Ph.D. Program. The Scottish Centre for Genomic Technology and Informatics was established with support from the Scottish Higher Education Funding Council.